DAGStat 2019

Statistics under one umbrella

5TH JOINT STATISTICAL MEETING
March 18 – 22, LMU Munich

Book of Abstracts

List of Contents

PROGRAM

Plenary Sessions

Plenary Session 1

Tuesday, March 19, 11:10 am – 12:20 pm, Room: Audimax

Sara van de Geer. *"Adaptivity of signal priors"*.
(Page 44, abstract 1)

Plenary Session 2

Wednesday, March 20, 11:10 am – 12:20 pm, Room: Audimax

Per Kragh Andersen. *"Measuring Expected Years of Life Lost"*.
(Page 44, abstract 2)

Plenary Session 3

Thursday, March 21, 11:25 am – 12:30 pm, Room: Audimax

Michael Jordan. *"Statistical Machine Learning: Dynamical, Economic and Stochastic Perspectives"*.
(Page 45, abstract 3)

Plenary Session 4

Friday, March 22, 11:10 am – 12:20 pm, Room: Audimax

Donald B. Rubin. *"Modern Computing Implementing Classical, But Heretofore Unnurtured Statistical Ideas"*.
(Page 45, abstract 4)

Tuesday, March 19, 9:00 am – 10:40 am

Advanced Regression Modeling I (Model Selection and Model Choice in Flexible Regression Models)

Chair: Andreas Groll, Room: D 209

Stanislav Anatolyev. *"A ridge to homogeneity"*.
09:00 am – 09:20 am (page 46, abstract 5)

Konstantinos Perrakis, Sach Mukherjee. *"Scalable Bayesian regression in high dimensions with multiple data sources"*.
09:20 am – 09:40 am (page 46, abstract 6)

Marinela Capanu, Colin Begg, Mithat Gonen. *"Optimized variable selection via repeated data splitting"*.
09:40 am – 10:00 am (page 47, abstract 7)

Lisa Schlosser, Torsten Hothorn, Achim Zeileis. *"The Power of Unbiased Recursive Partitioning: A Unifying View of CTree, MOB, and GUIDE"*.
10:00 am – 10:20 am (page 47, abstract 8)

Paul Wiemann, Thomas Kneib. *"Should positivity imply a multiplicative model? Introducing the Softplus function as an alternative to the common log link"*.
10:20 am – 10:40 am (page 48, abstract 9)

Computational Statistics and Statistical Software I (Temporal Data)

Chair: Roland Fried, Room: A 240

Alexander Meier, Haeran Cho, **Claudia Kirch**. *"Estimating multiple changes in the mean using moving sum statistics"*.
09:00 am – 09:20 am (page 49, abstract 10)

Katharina Hees, Roland Fried. *"How to model extreme events that occur in clusters?"*.
09:20 am – 09:40 am (page 49, abstract 11)

Charles Driver. *"Hierarchical continuous time state space modelling with ctsem"*.
09:40 am – 10:00 am (page 50, abstract 12)

Carolin Malsch, Stefan Störk, Peter U. Heuschmann. *"Confidence intervals for average sequential attributable fraction – A simulation study"*.
10:00 am – 10:20 am (page 51, abstract 13)

Design of Experiments and Clinical Trials I (Optimal Design I)

Chair: Norbert Gaffke, Room: E 004

Luzia Trinca. *"Design of experiments for fitting flexible curves"*.
09:00 am – 09:20 am (page 52, abstract 14)

Heiko Großmann. *"A practical approach to designing partial-profile choice experiments for estimating main effects and interactions"*.
09:20 am – 09:40 am (page 52, abstract 15)

Eric Nyarko, Rainer Schwabe. *"Designs for Second-Order Interactions in Paired Comparison Experiments of Full and Partial Profiles"*.
09:40 am – 10:00 am (page 53, abstract 16)

Frank Miller, Mahmood Ul-Hassan. *"Optimal item calibration designs for computerized achievement tests"*.
10:00 am – 10:20 am (page 53, abstract 17)

Fritjof Freise, Rainer Schwabe. *"On optimal designs for multi-factor two-level models on a design region restricted by the number of active factors"*.
10:20 am – 10:40 am (page 54, abstract 18)

Empirical Economics and Applied Econometrics I

Chair: Robert Jung, Room: A 214

Fabian Krüger, Johanna F. Ziegel. *"Generic Conditions for Forecast Dominance".*
09:00 am – 09:20 am (page 55, abstract 19)

C. S. Bos, S. J. Koopman, **M. Massmann**. *"Maximum likelihood analysis of high-dimensional reduced-rank regressions".*
09:20 am – 09:40 am (page 55, abstract 20)

Maurizio Daniele. *"Selecting the Number of Factors in Approximate Factor Models using Group Variable regularization".*
09:40 am – 10:00 am (page 56, abstract 21)

Magnus Sass. *"Detecting periods of excessive credit in the EU - A structural counterfactual approach".*
10:00 am – 10:20 am (page 56, abstract 22)

Tobias Hartl, **Rolf Tschernig**, Enzo Weber. *"Identification of structural shocks via common fractional components".*
10:20 am – 10:40 am (page 57, abstract 23)

Preclinical and Pharmaceutical Statistics I (Clinical Drug Development)

Chair: Antonia Zapf and Vivian Lanius, Room: A 021

Benjamin Hofner. *"Statistical issues in drug development and the role of statisticians in regulatory agencies".*
09:00 am – 09:40 am (page 58, abstract 24)

Ekkehard Glimm. *"Adjusting for selection bias in assessing treatment effect estimates from multiple subgroups".*
09:40 am – 10:00 am (page 59, abstract 25)

Heiko Götte, Junyuan Xiong, Marietta Kirchner, Meinhard Kieser. *"An efficient phase II/III development program utilizing information on short-term response and long-term survival".*
10:00 am – 10:20 am (page 59, abstract 26)

Manuel Wiesenfarth, Silvia Calderazzo. *"Quantification of prior impact in terms of prior effective historical and current sample size".*
10:20 am – 10:40 am (page 60, abstract 27)

Small Area Analysis and Spatial Statistics I

Chair: Ralf Münnich, Room: A 125

Timo Schmid, Fabian Bruckschen, Sandra Hadam, Nicola Salvati, Till Zbiranski. *"Estimating socio-demographic indicators using mobile phone data with applications in Germany and Senegal".*
09:00 am – 09:40 am (page 61, abstract 28)

Sören Pannier, Timo Schmid. *"Estimating proportions of multidimensional poverty in small areas"*.
09:40 am – 10:00 am (page 62, abstract 29)

Anne-Sophie Stelzer. *"Small area estimation in forest inventories: Overview of methods and challenges in practical applications"*.
10:00 am – 10:20 am (page 63, abstract 30)

Marcel Neunhoeffer, Richard Traunmüller. *"Generative Adversarial Imputation Nets for Small Area Estimation"*.
10:20 am – 10:40 am (page 64, abstract 31)

Statistical Literacy and Statistical Education I

Chair: Katharina Schüller, Room: A 140

Rolf Biehler. *"Data science education at school level – Conceptions, examples and experience from a pilot project"*.
09:00 am – 09:20 am (page 65, abstract 32)

Karsten Lübke, Matthias Gehrke, Jörg Horst, Sebastian Sauer. *"Causal Modelling in Introductory Statistics?"*.
09:20 am – 09:40 am (page 66, abstract 33)

Constantin Weiser, Manuel Förster, Florian Heiss, Sigbert Klinke, Andreas Maur, Thorsten Schank, Kirsten Winkel. *"Flipped Classroom Implementation in Large Statistics Lectures"*.
09:40 am – 10:00 am (page 67, abstract 34)

Sigbert Klinke. *"shinyExample - ein R Paket zur Unterstützung der Entwicklung einfacher Shiny-Apps"*.
10:00 am – 10:20 am (page 67, abstract 35)

Ulrich Rendtel, Sören Pannier. *"Die Prognose des Studienerfolgs auf Basis individueller Studienveräufe im Fach Wirtschaftswissenschaft"*.
10:20 am – 10:40 am (page 68, abstract 36)

Tuesday, March 19, 1:30 pm – 2:50 pm

Advanced Regression Modeling II (Nonparametric Regression Beyond the Mean)

Chair: Nikolaus Umlauf, Room: D 209

Paul Eilers. *"Goodbye moments, hello expectiles"*.
01:30 pm – 01:50 pm (page 69, abstract 37)

Elmar Spiegel, Thomas Kneib, Petra von Gablenz, Inga Holube, Fabian Otto-Sobotka. *"Generalized Expectile Regression with Flexible Response Function for patient reported outcomes"*.
01:50 pm – 02:10 pm (page 70, abstract 38)

Manuel Carlan, Thomas Kneib, Nadja Klein. *"Bayesian Conditional Transformation Models"*.

02:10 pm – 02:30 pm (page 71, abstract 39)

Almond Stöcker, Eva-Maria Maier, Bernd Fitzenberger, Sonja Greven. *"Flexible regression for probability densities in Bayes spaces"*.
02:30 pm – 02:50 pm (page 72, abstract 40)

Causal Inference I (Modelling Causal Structures)

Chair: Heinz Leitgöb and Martin Elff, Room: A 240

Nanny Wermuth. *"How can graphical Markov models aid causal inference?"*.
01:30 pm – 02:10 pm (page 73, abstract 41)

Tobias Wolbring, Lars Leszczensky. *"How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study"*.
02:10 pm – 02:30 pm (page 73, abstract 42)

Marco Steenbergen, Lukas F. Stoetzer. *"Measurement, Causal Models and Treatment Effects"*.
02:30 pm – 02:50 pm (page 74, abstract 43)

Design of Experiments and Clinical Trials II (Adaptive Designs I)

Chair: Gernot Wassmer, Room: E 004

Maximilian Pilz, Kevin Kunzmann, Carolin Herrmann, Geraldine Rauch, Meinhard Kieser. *"Optimal adaptive two-stage designs for normally distributed outcomes"*.
01:30 pm – 01:50 pm (page 75, abstract 44)

Manuel Feißt, Meinhard Kieser. *"Incorporating historical two-arm data in clinical trials with binary outcome"*.
01:50 pm – 02:10 pm (page 76, abstract 45)

Geraldine Rauch, Carolin Herrmann, Maximilian Pilz, Meinhard Kieser. *"A new rule for sample size recalculation based on resampling in an adaptive design setting"*.
02:10 pm – 02:30 pm (page 77, abstract 46)

Maria Stark, Antonia Zapf. *"Reestimation of the prevalence in a confirmatory diagnostic accuracy study"*.
02:30 pm – 02:50 pm (page 78, abstract 47)

Empirical Economics and Applied Econometrics II

Chair: Ralf Brüggemann, Room: A 214

Christiane Baumeister, James D. Hamilton. *"Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks"*.
01:30 pm – 02:10 pm (page 79, abstract 48)

Dominik Bertsche, **Robin Braun**. *"Identification of Structural Vector Autoregressions by Stochastic Volatility"*.

02:10 pm – 02:30 pm (page 79, abstract 49)

Simone Maxand. *"Identification of independent structural shocks in the presence of multiple Gaussian components"*.
02:30 pm – 02:50 pm (page 80, abstract 50)

Latent Variable Modelling I

Chair: Daniel Seddig, Room: A 021

David Kaplan. *"An Approach to Addressing Multiple Imputation Model Uncertainty Using Bayesian Model Averaging"*.
01:30 pm – 02:10 pm (page 81, abstract 51)

Felix Naumann. *"Estimation of a Nonparametric Multidimensional Item Response Model Using Dirichlet Process Mixtures"*.
02:10 pm – 02:30 pm (page 81, abstract 52)

Timo von Oertzen. *"Dirichlet Clustering in Onyx"*.
02:30 pm – 02:50 pm (page 82, abstract 53)

Mathematical Statistics I

Chair: Alexander Meister, Room: A 119

Oliver Feng, Aditya Guntuboyina, Arlene Kim, **Richard Samworth**. *"Log-concave density estimation"*.
01:30 pm – 02:10 pm (page 83, abstract 54)

Matthias Loeffler, Antoine Picard. *"Spectral thresholding for the estimation of Markov chain transition operators"*.
02:10 pm – 02:30 pm (page 83, abstract 55)

Viktor Bengs, Hajo Holzmann. *"Adaptive confidence sets for kink-location and kink-size in nonparametric regression"*.
02:30 pm – 02:50 pm (page 84, abstract 56)

Preclinical and Pharmaceutical Statistics II (Estimands)

Chair: Vivian Lanius and Antonia Zapf, Room: A 213

Kaspar Rufibach, Evgeny Degtyarev, Jonathan Siegel, Viktoriya Stalbovskaya, Steven Sun. *"Estimand framework in Oncology drug development – impact and opportunities"*.
01:30 pm – 01:50 pm (page 85, abstract 57)

Bjoern Bornkamp, Audrey Boruvka, Evgeny Degtyarev, Vera Kuehnl, Feng Liu, Yi Liu, Emily Martin, Devan Mehrotra, Satrajit Roychoudhury, Kaspar Rufibach, An Vandebosch. *"Estimation of Principal stratum effects, an overview and potential applications in oncology"*.
01:50 pm – 02:10 pm (page 86, abstract 58)

Viktoriya Stalbovskaya, Juliane Manitz, Marie-Laure Casadebaig, Emily Martin, Rui Sammi Tang, Godwin Yung, Vincent Haddad, Fei Jie, Christelle Lorenzato, Jiangxiu

Zhou, Evgeny Degtyarev. *"Estimands in the presence of treatment switching"*.
02:10 pm – 02:30 pm (page 87, abstract 59)

Hans-Jochen Weber, Marie-Laure Casadebaig, Emily Butler, Satrajit Roychoudhury, Kaspar Rufibach, Viktoriya Stalbovskaya, Steven Sun. *"Implementation of the ICH E9 addendum: A case study in hematology"*.
02:30 pm – 02:50 pm (page 88, abstract 60)

Small Area Analysis and Spatial Statistics II

Chair: Timo Schmid, Room: A 125

Helmut Waldl. *"Comparing designs for prediction based on stationary vs. non-stationary space-time covariance functions"*.
01:30 pm – 01:50 pm (page 89, abstract 61)

Marc Hüsch, Bruno U. Schyska, Lueder von Bremen. *"Identifying spatial dependence structures with copulas and generalized additive models"*.
01:50 pm – 02:10 pm (page 90, abstract 62)

Jonathan Rathjens, Eva Becker, Arthur Kolbe, Katharina Olthoff, Sabine Bergmann, Jürgen Hölzer, Katja Ickstadt. *"Spatio-Temporal Smoothing of Drinking Water Contamination Data"*.
02:10 pm – 02:30 pm (page 91, abstract 63)

Dany Djeudeu, Susanne Moebus, Katja Ickstadt. *"Multilevel Conditional Autoregressive models for longitudinal data nested in geographical units with dynamic characteristics"*.
02:30 pm – 02:50 pm (page 92, abstract 64)

Statistical Literacy and Statistical Education II

Chair: Rolf Biehler, Room: A 140

Laura Martignon. *"Statistical Literacy and Statistical Education"*.
01:30 pm – 02:10 pm (page 93, abstract 65)

Joachim Engel. *"Civic Statistics: Big Ideas, Needs and Challenges. Why we need a new subdiscipline"*.
02:10 pm – 02:30 pm (page 94, abstract 66)

Ana Kolar. *"A New Approach for Developing Statistical Thinking"*.
02:30 pm – 02:50 pm (page 95, abstract 67)

Tuesday, March 19, 3:20 pm – 4:40 pm

Computational Statistics and Statistical Software II (Omics)

Chair: Jörg Rahnenführer, Room: A 240

Janine Wiebach, **Miriam Sieg**, Jochen Kruppa. *"Comparison of different preprocessing methods for the analysis of metabolite data"*.
03:20 pm – 03:40 pm (page 96, abstract 68)

Philipp Probst, Moritz Herrmann, Roman Hornung, Vindi Jurinovic, Anne-Laure Boulesteix. *"Benchmarking survival prediction methods using 18 multi-omics datasets from the „The cancer genome atlas" (TCGA)".*
03:40 pm – 04:00 pm (page 97, abstract 69)

Roman Hornung, Marvin N Wright. *"Block Forests: random forests for blocks of clinical and omics covariate data".*
04:00 pm – 04:20 pm (page 98, abstract 70)

Jan Klosa, Noah R. Simon, Volkmar Liebscher, Dörte Wittenburg. *"Sparse-group lasso variants for whole-genome regression models in livestock".*
04:20 pm – 04:40 pm (page 99, abstract 71)

Design of Experiments and Clinical Trials III (Adaptive Designs II)

Chair: Werner Brannath, Room: E 004

Tobias Mütze, Susanna Salem, Norbert Benda, Heinz Schmidli, **Tim Friede**. *"Blinded continuous information monitoring of recurrent events endpoints with time trends".*
03:20 pm – 03:40 pm (page 100, abstract 72)

Tobias Mielke, Vladimir Dragalin. *"Adaptive designs for drug combination informed by longitudinal model for the response".*
03:40 pm – 04:00 pm (page 101, abstract 73)

Laura Kerschke, Andreas Faldum, Rene Schmidt. *"An Alternative Log-Rank Test for Adaptive Survival Trials".*
04:00 pm – 04:20 pm (page 102, abstract 74)

Cornelia Ursula Kunz, Nigel Stallard. *"Combining Parallel Adaptive Seamless Phase 2/3 Trials".*
04:20 pm – 04:40 pm (page 103, abstract 75)

Latent Variable Modelling II

Chair: Timo von Oertzen, Room: A 021

Carsten Szardenings, Anna Doebler, Philipp Doebler. *"A Recent Perspective on Differential Item Functioning and its Implications in the Rasch model".*
03:20 pm – 03:40 pm (page 104, abstract 76)

Malgorzata Bogdan, **Wei Jiang**, Julie Josse, Blazej Miasojedow, Veronika Rockova. *"Adaptive Bayesian SLOPE – High-dimensional Model Selection with Missing Values".*
03:40 pm – 04:00 pm (page 104, abstract 77)

Omololu Stephen Aluko, Birhanu Ayele. *"Statistical methodologies for handling ordinal longitudinal responses with intermittent missingness".*
04:00 pm – 04:20 pm (page 105, abstract 78)

Mathematical Statistics II

Chair: Hajo Holzmann, Room: A 119

Dominik Poß, **Dominik Liebl**, Alois Kneip, Hedwig Eisenbarth, Tor Wager, Lisa Feldman Barrett. *"Super-Consistent Estimation of Points of Impact in Nonparametric Regression with Functional Predictors"*.
03:20 pm – 03:40 pm (page 106, abstract 79)

Melanie Birke, Christoph Reihl, Hajo Holzmann. *"Simultaneous confidence bands for the covariance kernel of Banach space valued functional data"*.
03:40 pm – 04:00 pm (page 107, abstract 80)

Robert Bassett, James Sharpnack. *"Fused Density Estimation on Infrastructure Networks"*.
04:00 pm – 04:20 pm (page 107, abstract 81)

Marcel Klatt, Carla Tameling, Axel Munk. *"Empirical Regularized Optimal Transport: Statistical Theory and Applications"*.
04:20 pm – 04:40 pm (page 108, abstract 82)

Small Area Analysis and Spatial Statistics III

Chair: Hanna Brenzel, Room: A 125

Anna Schritz, Andrew Lawson, Gloria Aguayo. *"Joint spatial modelling of disease outcomes of Chilean survey data"*.
03:20 pm – 03:40 pm (page 109, abstract 83)

Britta Stöver. *"Local economic impact of universities"*.
03:40 pm – 04:00 pm (page 110, abstract 84)

Markus Zwick, **Ralf Münnich**, Johannes Kopp, Petra Stein, Rainer Schnell. *"MikroSim – Sektorenübergreifendes kleinräumiges Mikrosimulationsmodell"*.
04:00 pm – 04:20 pm (page 111, abstract 85)

Nora Würz, Timo Schmid, Nikos Tzavidis. *"Data-driven Transformations for the Estimation of Small Area Means"*.
04:20 pm – 04:40 pm (page 112, abstract 86)

Statistics in Agriculture and Ecology I

Chair: Roland Langrock, Room: D 209

Janine Baerbel Illian. *"Point processes — abstraction and practical relevance in ecology"*.
03:20 pm – 04:00 pm (page 113, abstract 87)

Sina Mews, Roland Langrock, Nicola Quick, Ruth King. *"A continuous-time multi-state capture-recapture model for the annual movement of bottlenose dolphins on the east coast of Scotland"*.
04:00 pm – 04:20 pm (page 114, abstract 88)

Hans-Peter Piepho. *"A Coefficient of Determination (R2) for Generalized Linear Mixed Models"*.

04:20 pm – 04:40 pm (page 115, abstract 89)

Survival and Event History Analysis I (Non-standard Sampling)

Chair: Tobias Bluhmki, Room: A 213

Achim Dörre. *"Semiparametric Modeling of Doubly Truncated Lifetimes in Registry Data"*.
03:20 pm – 03:40 pm (page 116, abstract 90)

Rafael Weissbach, Achim Doerre, Anne Fink, Gabriele Doblhammer. *"Left-censoring in survival analysis: An application to dementia incidence"*.
03:40 pm – 04:00 pm (page 116, abstract 91)

Jan Feifel, Dennis Dobler. *"Time-simultaneous inference in general nested case-control designs"*.
04:00 pm – 04:20 pm (page 117, abstract 92)

Kwun Chuen Gary Chan. *"On modeling complex longitudinal and survival data with a terminal trend"*.
04:20 pm – 04:40 pm (page 117, abstract 93)

Statistical Literacy and Statistical Education III

Chair: Joachim Engel, Room: A 140

Leah Braun, Karin Binder. *"Increase in the speed of medical decisions due to natural frequencies"*.
03:20 pm – 03:40 pm (page 118, abstract 94)

Ursula Berger, Cornelia Oberhasuer, Michaela Coenen. *"The development of epiLEARNER: an innovative e-learning project by and for medical students"*.
03:40 pm – 04:00 pm (page 119, abstract 95)

Michael Gabel, Marietta Kirchner, Lorenz Uhlmann, Maximilian Pilz, Dorothea Weber, Meinhard Kieser. *"Challenges in teaching Medical Data Science"*.
04:00 pm – 04:20 pm (page 120, abstract 96)

Paul Bach, Lorena Hafermann, Geraldine Rauch, Nadja Klein. *"Regression model building in medical statistics"*.
04:20 pm – 04:40 pm (page 121, abstract 97)

Time Series Analysis I (Time Series Econometrics)

Chair: Paulo M. Rodrigues, Room: A 214

Brian D O Anderson, **Manfred Deistler**, Jean-Marie Dufour. *"On the Sensitivity of Granger Causality to Errors-in-Variables, Linear Transformations and Subsampling"*.
03:20 pm – 03:40 pm (page 122, abstract 98)

Tobias Hartl, Rolf Tschernig, Enzo Weber. *"Fractional trends in unobserved components models"*.

03:40 pm – 04:00 pm (page 122, abstract 99)

Oliver Stypka, **Martin Wagner**, Peter Grabarczyk, Rafael Kawka. *"The Asymptotic Validity of "Standard" Fully Modified OLS Estimation and Inference in Cointegrating Polynomial Regressions".*
04:00 pm – 04:20 pm (page 123, abstract 100)

Marina Balboa, **Paulo M. M. Rodrigues**, Antonio Rubia, Robert Taylor. *"Multivariate Testing for Fractional Integration".*
04:20 pm – 04:40 pm (page 123, abstract 101)

Tuesday, March 19, 5:00 pm – 6:20 pm

Young Statisticans

Chair: Tobias Bluhmki and Julia Krzykalla, Room: A 140

Martina Schlenker. *"Selection Effects in Bayesian Hierarchical Models Bachelor Thesis in Cooperation with Boehringer Ingelheim Pharma GmbH & Co. KG".*
05:00 pm – 05:20 pm (page 124, abstract 102)

Kaya Miah. *"Risks and benets of autologous stem cell transplantations in treating elderly patients with multiple myeloma: Competing risks analyses".*
05:20 pm – 05:40 pm (page 125, abstract 103)

Aisouda Hoshiyar. *"Challenging the commonly used log-link in statistical models for count data with an application to infection disease data".*
05:40 pm – 06:00 pm (page 126, abstract 104)

Guillermo Briseño Sanchez, Maike Hohberg, Andreas Groll, Thomas Kneib. *"Flexible instrumental variable distributional regression".*
06:00 pm – 06:20 pm (page 127, abstract 105)

Data Fusion and Meta-Analysis I

Chair: Tim Friede and Masayuki Henmi, Room: E 004

Thomas P.A. Debray. *"Clinical Prediction Models and the role of Evidence Synthesis".*
05:00 pm – 05:40 pm (page 128, abstract 106)

Satoshi Hattori. *"Summray concordance index for meta-analysis of prognostic studies with survival outcome".*
05:40 pm – 06:00 pm (page 128, abstract 107)

Cornelia Frömke, Mathia Kirstein, Antonia Zapf. *"A nonparametric approach for meta-analysis of diagnostic accuracy studies with multiple cut-offs".*
06:00 pm – 06:20 pm (page 129, abstract 108)

Latent Variable Modelling III

Chair: Augustin Kelava, Room: A 021

Marius Ötting, Roland Langrock, Antonello Maruotti. *"A copula-based multivariate hidden Markov model for modelling momentum in football".*

05:00 pm – 05:20 pm (page 130, abstract 109)

Augustin Kelava, Holger Brandt. *"A Nonlinear Dynamic Latent Class Structural Equation Model"*.
05:20 pm – 05:40 pm (page 131, abstract 110)

Mirka Henninger. *"A new varying threshold approach to model response styles in the IRT framework"*.
05:40 pm – 06:00 pm (page 132, abstract 111)

Holger Brandt, Zachary Roman, Mark Anderson, Augustin Kelava. *"Identifying inattentive responses using dynamic latent class modeling"*.
06:00 pm – 06:20 pm (page 133, abstract 112)

Mathematical Statistics III

Chair: Alexander Meister, Room: A 119

Lutz Mattner. *"The sufficiency principle: How to teach it, and what does it entail?"*.
05:00 pm – 05:20 pm (page 134, abstract 113)

Paul Jacobus van Staden. *"A bounded quantile-based measure of kurtosis"*.
05:20 pm – 05:40 pm (page 134, abstract 114)

Fatemeh Ghaderinezhad, Christophe Ley. *"To choose or not to choose a prior. That's the question!"*.
05:40 pm – 06:00 pm (page 134, abstract 115)

Houda Yaqine, Hamid El Maroufy, Christiane Fuchs. *"Parameter Estimation for Lotka-Volterra Switching model"*.
06:00 pm – 06:20 pm (page 135, abstract 116)

Official Statistics and Survey Statistics I

Chair: Markus Zwick, Room: A 125

Piet J.H. Daas. *"Using Big Data in Official Statistics"*.
05:00 pm – 05:40 pm (page 136, abstract 117)

Markus Zwick, Clara Schartner. *"Smart Business Cycle Statistics"*.
05:40 pm – 06:00 pm (page 136, abstract 118)

Florian Keusch, Mark Trappmann, Georg-Christoph Haas, Sebastian Bähr, Frauke Kreuter. *"Enriching an Ongoing Panel Survey with Mobile Phone Measures: The IAB-SMART Study"*.
06:00 pm – 06:20 pm (page 137, abstract 119)

Preclinical and Pharmaceutical Statistics III (Preclinical studies)

Chair: Tina Lang, Room: A 213

Hannes Buchner, Robert Reidy, Michael Matiu, Johannes Solzin, Alexander Berger, Armin Boehrer, Erich Bluhmki. *"A Novel Approach to Outlier Identification in Bioas-*

says".
05:00 pm – 05:20 pm (page 138, abstract 120)

Franziska Kappenberg, Jan Hengstler, Jörg Rahnenführer. *"How to handle deviating control values in dose-response curves".*
05:20 pm – 05:40 pm (page 139, abstract 121)

Reinhard Meister. *"New Approaches for Bivariate Quantitative Dose-Response — A Screening Study from Hormone- Research and Development".*
05:40 pm – 06:00 pm (page 140, abstract 122)

Konrad Neumann, Samuel Knauss, Ulrike Grittner. *"Adaptive designs in preclinical dose finding studies".*
06:00 pm – 06:20 pm (page 141, abstract 123)

Robust and Nonparametric Statistics & Computational Statistics and Statistical Software

Chair: Claudia Becker, Room: A 240

Ahmed R.M. Alsayed, Giancarlo Manzi, Sek Siok Kun. *"Comparison of Dependence Coefficients in Presence of Outliers, A Simulation Study".*
05:00 pm – 05:20 pm (page 142, abstract 124)

Mustafa Çavus, Berna Yazici, Ahmet Sezer. *"Comparison of some normality tests in the presence of outliers".*
05:20 pm – 05:40 pm (page 142, abstract 125)

Patrick Schenk. *"Removing Outliers: Effects on Statistical Inference and Suggestions for Choosing Exclusion Boundaries".*
05:40 pm – 06:00 pm (page 143, abstract 126)

Statistics in Agriculture and Ecology II

Chair: Hans-Peter Piepho, Room: D 209

Waqas Malik, Hans-Peter Piepho. *"Testing Multiplicative Terms in AMMI and GGE Models for Multienvironment Trials".*
05:00 pm – 05:20 pm (page 144, abstract 127)

Inga Blunk, Manfred Mayer, Henning Haman, Norbert Reinsch. *"How to detect imprinted loci using estimated parent-of-origin effects and simple gene counts only".*
05:20 pm – 05:40 pm (page 145, abstract 128)

Hayal Boyacioglu, **Hulya Boyacioglu**. *"Application of Multivariate Statistical Methods in Water Pollution Footprinting".*
05:40 pm – 06:00 pm (page 146, abstract 129)

Time Series Analysis II

Chair: Alexander Mayer, Room: A 214

Mustafa Kilinc. *"Detecting multiple location shifts under long-memory stationary er-*

rors".
05:00 pm – 05:20 pm (page 147, abstract 130)

Mawuli Segnon, **Manuel Stapper**. *"Long Memmory Conditional Heteroscedasticity in Count Data".*
05:20 pm – 05:40 pm (page 147, abstract 131)

Alexander Mayer. *"Estimation and Inference in Adaptive Learning Models with Slowly Decreasing Gains".*
05:40 pm – 06:00 pm (page 148, abstract 132)

Thiago do Rego Sousa, Richard Davis, Claudia Klüppelberg. *"Parameter estimation for time series models based on the simulated characteristic function".*
06:00 pm – 06:20 pm (page 149, abstract 133)

Wednesday, March 20, 9:00 am – 10:40 am

Verleihung der IBS-DR Nachwuchspreise

Chair: Andreas Faldum, Room: D 209

Jasmin Rühl . *"Sample Size Calculation in Time-To-Event Trials with Non-Proportional Hazards Using GESTATE".*
09:00 am – 09:15 am (page 150, abstract 134)

Stefanie Krügel, Prof. Dr. Anne-Laure Boulesteix, Dr. Martin Depner . *"Statistical Approaches to Characterize and Compare Networks of Microbiome Data".*
09:15 am – 09:30 am (page 151, abstract 135)

Tobias Mütze, Ekkehard Glimm, Heinz Schmidli, Tim Friede . *"Group sequential designs with robust semiparametric recurrent event models".*
09:30 am – 09:50 am (page 152, abstract 136)

Philipp Wittenberg, Fah Fatt Gan, Sven Knoth . *"A simple signaling rule for variable life-adjusted display derived from an equivalent risk-adjusted CUSUM chart".*
09:50 am – 10:10 am (page 153, abstract 137)

Advanced Regression Modeling III (Distributional Regression/ GAMLSS)

Chair: Helga Wagner, Room: A 140

Nikolaus Umlauf, Nadja Klein, Thorsten Simon, Achim Zeileis. *"Neural Network Distributional Regression".*
09:00 am – 09:20 am (page 154, abstract 138)

Hannes Riebl, Nadja Klein, Thomas Kneib. *"Random Function Responses in Distributional Regression".*
09:20 am – 09:40 am (page 155, abstract 139)

Nadja Klein, Manuel Carlan, Thomas Kneib, Stefan Lang, Helga Wagner. *"Bayesian Effect Selection in Structured Additive Distributional Regression Models".*
09:40 am – 10:00 am (page 156, abstract 140)

Tobias Hepp, Angelina Hammon. *"Proper imputation for GAMLSS inference"*.
10:00 am – 10:20 am (page 157, abstract 141)

Alexander Volkmann, Almond Stöcker, Fabian Scheipl, Sonja Greven. *"Multivariate Functional Additive Mixed Models"*.
10:20 am – 10:40 am (page 158, abstract 142)

Computational Statistics and Statistical Software III(Open science and reproducibility)

Chair: Gero Szepannek, Room: A 240

Anne-Laure Boulesteix, Felix Schönbrodt. *"Open Science and statistics"*.
09:00 am – 09:20 am (page 159, abstract 143)

Heidi Seibold. *"Research Software Engineers"*.
09:20 am – 09:40 am (page 159, abstract 144)

Bernd Bischl, Benjamin Hofner, **Fabian Scheipl**. *"Reproducible Methodological Research and Scientific Publishing"*.
09:40 am – 10:00 am (page 160, abstract 145)

Felix D. Schönbrodt. *"Correcting for bias in the literature: A comparison of meta-analytic methods for bias-correction"*.
10:00 am – 10:20 am (page 161, abstract 146)

Sabine Hoffmann, Felix Schönbrodt, Ralf Elsas, Simon Klau, Anne-Laure Boulesteix. *"The multiplicity of possible analysis strategies and how it is handled across scientific disciplines"*.
10:20 am – 10:40 am (page 162, abstract 147)

Design of Experiments and Clinical Trials IV (Adaptive Designs III)

Chair: Geraldine Rauch, Room: E 004

James Wason. *"Novel designs for trials with multiple treatments and subgroups"*.
09:00 am – 09:40 am (page 163, abstract 148)

Stella Preussler, Marietta Kirchner, Heiko Götte, Meinhard Kieser. *"Optimal designs for multi-arm phase II/III drug development programs"*.
09:40 am – 10:00 am (page 164, abstract 149)

Maja Krajewska, Geraldine Rauch. *"Comparison of the efficacy of Bayesian and frequentist designs for oncological phase II basket trials."*.
10:00 am – 10:20 am (page 165, abstract 150)

Deepak Parashar. *"Geometric representation of master protocols"*.
10:20 am – 10:40 am (page 166, abstract 151)

Data Fusion and Meta-Analysis II

Chair: Dankmar Boehning, Room: A 119

Annika Hoyer, Oliver Kuss. *"Meta-analysis of full ROC curves: A parametric model based on flexible distributions of diagnostic test values".*
09:00 am – 09:20 am (page 167, abstract 152)

Orestis Efthimiou, Ian White. *"The dark side of the force: multiple testing issues in network meta-analysis and how to address them".*
09:20 am – 09:40 am (page 167, abstract 153)

Gerta Rücker, Adriani Nikolakopoulou, Theodoros Papakonstantinou, Guido Schwarzer. *"The importance of a study for treatment estimates in network meta-analysis".*
09:40 am – 10:00 am (page 168, abstract 154)

Georgia Salanti, Adiani Nikolakopoulou, Dimitris Mavridis. *"On ranking multiple health interventions".*
10:00 am – 10:20 am (page 169, abstract 155)

Theodoros Papakonstantinou, Adriani Nikolakopoulou, Gerta Rücker, Anna Chaimani, Guido Schwarzer, Matthias Egger, Georgia Salanti. *"Using flow decomposition to estimate the contribution of studies in network meta-analysis".*
10:20 am – 10:40 am (page 170, abstract 156)

Measurement and Measurement Error I (Measurement Error)

Chair: Helmut Kuechenhoff, Room: A 213

Laurence S. Freedman, Paul Gustafson, Pamela Shaw, Raymond J. Carroll, **Veronika Deffner**, Kevin Dodd, Victor Kipnis, Ruth Keogh, Helmut Küchenhoff, Janet Tooze. *"Measurement error and misclassification of variables in observational epidemiology: basic knowledge and practical guidance".*
09:00 am – 09:20 am (page 171, abstract 157)

Pamela Ann Shaw, Eric Oh, Bryan Shepherd, Thomas Lumley. *"Estimation methods to address correlated covariate and time-to-event error".*
09:20 am – 10:00 am (page 172, abstract 158)

Timm Intemann, Kirsten Mehlig, Stefaan De Henauw, Alfonso Siani, Tassos Constantinou, Luis A. Moreno, Dénes Molnár, Thomas Veidebaum, Iris Pigeot. *"SIMEX for Box-Cox transformed measurements".*
10:00 am – 10:20 am (page 173, abstract 159)

Felix Günther, Caroline Brandl, Iris M. Heid, Helmut Küchenhoff. *"Accounting for misclassification in automated disease diagnosis based on medical image data".*
10:20 am – 10:40 am (page 175, abstract 160)

Official Statistics and Survey Statistics II

Chair: Andreas Quatember, Room: A 125

Simon Kühne, Jannes Jacobsen, Martin Kroh. *"Sampling in Times of High Immigrati-*

on: The IAB-BAMF-SOEP Survey of Refugees".
09:20 am – 09:40 am (page 176, abstract 161)

Rossalina Latcheva, David Reichel, **Ursula Till-Tentschert**. *"European Union Minorities and Discrimination Survey (EU-MIDIS II) - Surveying immigrants and ethnic minorities in the 28 EU Member States".*
09:40 am – 10:00 am (page 177, abstract 162)

Dimitri Prandner, Martin Weichbold. *"Building a Sampling Frame for Migrant Populations via an Onomastic Approach – One or more lessons learned from the Austrian Immigrant Survey 2016".*
10:00 am – 10:20 am (page 178, abstract 163)

Marie-Luise Zeisler, **Johannes Lemcke**, Leman Bilgic, Claudia Santos-Hövener, Patrick Schmich. *"Integration of migrant populations into health monitoring in Germany - Results from a feasibility study".*
10:20 am – 10:40 am (page 179, abstract 164)

Statistics in Behavorial and Educational Sciences I (Educational Sciences)

Chair: Timo Gnambs, Room: A 021

Steffi Pohl. *"Using timing information to model missing values in test data".*
09:00 am – 09:40 am (page 180, abstract 165)

Pascal Jordan. *"Paradoxical properties of parameter estimates in multidimensional models".*
09:40 am – 10:00 am (page 180, abstract 166)

Esther Ulitzsch, Matthias von Davier, Steffi Pohl. *"A Hierarchical Latent Response Model for Inferences about Examinee Engagement in Terms of Guessing and Item-Level Nonresponse".*
10:00 am – 10:20 am (page 181, abstract 167)

Boris Forthmann, **Daniela Gühne**, Philipp Doebler. *"Revisiting Dispersion in Count Data Item Response Theory Models: The Conway-Maxwell-Poisson Counts Model".*
10:20 am – 10:40 am (page 182, abstract 168)

Time Series Analysis III (Change Points)

Chair: Annika Betken, Room: A 214

Kai Rouven Wenger, Christian Leschinski. *"Fixed-Bandwidth CUSUM Tests Under Long Memory".*
09:00 am – 09:20 am (page 183, abstract 169)

Sven Otto, Jörg Breitung. *"Backward CUSUM for Testing and Monitoring Structural Change".*
09:20 am – 09:40 am (page 183, abstract 170)

Maria Mohr, Natalie Neumeyer. *"Consistent change point detection in a nonparametric time series regression model".*
09:40 am – 10:00 am (page 184, abstract 171)

Annika Betken, Martin Wendler. *"Change-point tests based on self-normalization and subsampling for LRD data"*.
10:00 am – 10:20 am (page 184, abstract 172)

Alexander Dürre, Roland Fried. *"Robust change point tests using bounded transformations"*.
10:20 am – 10:40 am (page 185, abstract 173)

Wednesday, March 20, 1:30 pm – 2:50 pm

Data Fusion and Meta-Analysis III

Chair: Christian Röver, Room: A 119

Dankmar Boehning, Patarawan Sangnawakij, Holling Heinz. *"Count Outcome Meta-Analysis with Mixed Arm Information"*.
01:30 pm – 01:50 pm (page 186, abstract 174)

Thilo Welz, Markus Pauly. *"Robust covariance estimation in mixed-effects meta-regression models - A simulation study"*.
01:50 pm – 02:10 pm (page 187, abstract 175)

Malgorzata Roos, Sona Hunanyan, Leonhard Held. *"Classification of tail-adjusted heterogeneity priors in the Bayesian meta-analysis estimated by bayesmeta"*.
02:10 pm – 02:30 pm (page 188, abstract 176)

Federico Bonofiglio, Harald Binder, Martin Schumacher. *"Recovery of IPD inferences from key IPD summaries only: application to distributed computing under privacy constraints"*.
02:30 pm – 02:50 pm (page 189, abstract 177)

Measurement and Measurement Error II (Measurement Error and missing data)

Chair: Moritz Heene, Room: A 213

Dominikus Stelzer, Julia Ortner, Louis Velthuis, Reyn van Ewijk, Anita Arslanow, Michael Nagel, Marc Nguyen-Tat, Peter R. Galle, Frank Lammert, Erik Farin-Glattacker, Harald Binder, Erika Graf. *"Comparing cohorts from distinct sources: The issue of differently operationalized predictor variables"*.
01:30 pm – 01:50 pm (page 190, abstract 178)

Xavier Fernández-i-Marín. *"Measurement for better public administration research (and better theory, too)"*.
01:50 pm – 02:10 pm (page 191, abstract 179)

Dominik de Sordi, Fabian Otto-Sobotka, Antje Timmer. *"Systematic Review on handling missing participant data in longitudinal studies"*.
02:10 pm – 02:30 pm (page 192, abstract 180)

Official Statistics and Survey Statistics III

Chair: Florian Keusch, Room: A 125

Murray Aitkin. *"An alternative measure of income inequality over successive surveys"*.
01:30 pm – 01:50 pm (page 193, abstract 181)

Alexander Bauer, Andreas Bender, André Klima, Helmut Küchenhoff. *"KOALA: A new paradigm for election coverage - An opinion poll based "now-cast" of probabilities of events in multi-party electoral systems"*.
01:50 pm – 02:10 pm (page 194, abstract 182)

Jörg Drechsler, Birgit Pech. *"Nonparametric Multiple Imputation for Bridging Between Different Industry Coding Systems"*.
02:10 pm – 02:30 pm (page 194, abstract 183)

André Klima. *"Estimation of voter transitions in the immediate post-election period"*.
02:30 pm – 02:50 pm (page 195, abstract 184)

Robust and Nonparametric Statistics I

Chair: Natalie Neumeyer, Room: A 214

Oliver Thunich, Sebastian Schoneberg, Bertram Schäfer. *"An extension for smoothed empirical likelihood confidence intervals for extreme quantiles and small sample sizes"*.
01:30 pm – 01:50 pm (page 196, abstract 185)

Markus Sebastian Doktor, Wolfgang Kurz, Peter Ruckdeschel, Jean-Pierre Stockis. *"Stochastic models for non-destructive testing in civil engineering"*.
01:50 pm – 02:10 pm (page 197, abstract 186)

Mirko Alexander Jakubzik. *"Applications of a minimum distance estimator for self-exciting counting processes"*.
02:10 pm – 02:30 pm (page 198, abstract 187)

Beat Hulliger. *"Regression based on medians with application to survey data"*.
02:30 pm – 02:50 pm (page 199, abstract 188)

Survival and Event History Analysis II (Complex Modeling)

Chair: Arthur Allignol, Room: B106

Marie Böhnstedt, Jutta Gampe, Hein Putter. *"Detecting Deceleration in Old-Age Mortality Rates Using Focused Model Selection"*.
01:30 pm – 01:50 pm (page 200, abstract 189)

Moritz Berger, Marie-Therese Puth, Gerhard Tutz, Nils Heim, Matthias Schmid. *"Tree-Structured Modeling of Time-Varying Coefficients for Discrete Time-to-Event Data"*.
01:50 pm – 02:10 pm (page 201, abstract 190)

Steffen Unkel, Steven Abrams, Andreas Wienke, Niel Hens. *"The genesis and use of time-varying frailty models for representing heterogeneities in the transmission of infectious diseases"*.
02:10 pm – 02:30 pm (page 202, abstract 191)

Alexander Seipp, Verena Jürgens, Antje Timmer, Fabian Otto-Sobotka. *"Weighting Expectile Regression for Survival Analysis with Right-Censoring"*.
02:30 pm – 02:50 pm (page 203, abstract 192)

Statistics in Finance I

Chair: Markus Bibinger, Room: E 004

Mathieu Rosenbaum. *"How do market participants contribute to market quality? A statistical approach"*.
01:30 pm – 02:10 pm (page 204, abstract 193)

Claudia Klüppelberg, **Miriam Isabel Seifert**. *"Measuring risks in a network of light-tailed financial objects"*.
02:10 pm – 02:30 pm (page 204, abstract 194)

Bastian Gribisch, **Jan Patrick Hartkopf**, Roman Liesenfeld. *"Factor State-Space Models for High-Dimensional Realized Covariance Matrices of Asset Returns"*.
02:30 pm – 02:50 pm (page 205, abstract 195)

Statistics of High Dimensional Data I

Chair: Arne Bathke, Room: D 209

Patrick Laurie Davies, Lutz Dümbgen. *"A Model-free Approach to Linear Least Squares Regression with Exact Probabilities and Applications to Covariate Selection"*.
01:30 pm – 01:50 pm (page 206, abstract 196)

Philipp Hermann, Hajo Holzmann. *"Random coefficient model - model selection and estimation of first and second moments"*.
01:50 pm – 02:10 pm (page 206, abstract 197)

Vladimir Pyrlik. *"Shrinkage in Estimating High Dimensional Copulas"*.
02:10 pm – 02:30 pm (page 207, abstract 198)

Dominik Liebl, Stefan Rameseder, **Christoph Rust**. *"Improving Estimation in Functional Linear Regression with Points of Impact: Insights into Google AdWords"*.
02:30 pm – 02:50 pm (page 207, abstract 199)

Statistics in Science, Technology and Industry I

Chair: Ansgar Steland, Room: A 240

Axel Gandy. *"Some examples of handling uncertainty in industrial applications"*.
01:30 pm – 02:10 pm (page 208, abstract 200)

Jürgen Franke, Pak Hang Lo. *"Fully automatic nonparametric intensity estimates for studying the microstructure of composite materials from 2d and 3d images"*.
02:10 pm – 02:30 pm (page 208, abstract 201)

Jürgen Pilz, Natalie Vollert, Konstantin Posch. *"Statistical Modelling and Design for Quality Control and Reliability Analysis in Power Semiconductor Manufacturing Processes"*.
02:30 pm – 02:50 pm (page 209, abstract 202)

Time Series Analysis IV (Discrete and Functional Time Series)

Chair: Christian Weiß, Room: A 140

Christian Weiß. *"Distance-based Analysis of Ordinal Time Series"*.
01:30 pm – 01:50 pm (page 210, abstract 203)

Lena Reichmann, Carsten Jentsch. *"Autoregressive-type time series models with bounded support"*.
01:50 pm – 02:10 pm (page 210, abstract 204)

Daniel Constantin Rademacher, Jens-Peter Kreiß, Efstathios Paparoditis. *"Asymptotic Normality of Integrated Periodogram Operators"*.
02:10 pm – 02:30 pm (page 211, abstract 205)

Timo Adam, Roland Langrock, Christian H. Weiß. *"Semi-parametric hidden Markov models for time series of counts"*.
02:30 pm – 02:50 pm (page 212, abstract 206)

Wednesday, March 20, 3:20 pm – 4:40 pm

Clustering I (Copula and Genetics)

Chair: Ingo Steinwart, Room: A 119

Marta Nai Ruscone. *"Model-based Clustering with R-vine copulas"*.
03:20 pm – 03:40 pm (page 213, abstract 207)

F. Marta L. Di Lascio, Fabrizio Durante, **Sebastian Fuchs**. *"Dissimilarity functions for copula-based hierarchical clustering of continuous variables"*.
03:40 pm – 04:00 pm (page 213, abstract 208)

Cesaire Joris Kuete Fouodo, Inke R. König. *"Detection of Genetic Similarities using Unsupervised Random Forest"*.
04:00 pm – 04:20 pm (page 214, abstract 209)

Claus-Dieter Mayer. *"Alternative splicing based clustering of genes"*.
04:20 pm – 04:40 pm (page 215, abstract 210)

Official Statistics and Survey Statistics IV

Chair: Florian Keusch, Room: A 125

Annelies Blom, Carina Cornesse, **Barbara Felderer**, Marina Fikel, Ulrich Krieger. *"Push-to-web recruitment of a probability-based online panel: Experimental evidence"*.
03:20 pm – 03:40 pm (page 216, abstract 211)

Patrick Schenk. *"Are Paradata Worth the Effort? Using Adjusted Response Times and Other Paradata To Predict Data Quality in a Survey."*.
03:40 pm – 04:00 pm (page 217, abstract 212)

Andrei Veikher. *"Multi factor modelling of survey external validity by using statistic and administrative data"*.
04:00 pm – 04:20 pm (page 218, abstract 213)

Sandra Hadam. *"Mobilfunkdaten in der amtlichen Statistik"*.
04:20 pm – 04:40 pm (page 219, abstract 214)

Robust and Nonparametric Statistics II

Chair: Natalie Neumeyer, Room: A 214

Andreas Christmann. *"Robustness and Stability of Kernel-Based Machine Learning"*.
03:20 pm – 03:40 pm (page 220, abstract 215)

Robert Hable. *"Fitting additive models with regularized kernel methods: methodology, robustness properties, and business applications"*.
03:40 pm – 04:00 pm (page 220, abstract 216)

Heike Deutelmoser, Dominique Scherer, Justo Lorenzo Bermejo. *"Empirical examination of the potential of robust regularized regression to examine genetic associations with circulating metabolite levels"*.
04:00 pm – 04:20 pm (page 221, abstract 217)

Benedikt Funke, **Masayuki Hirukawa**. *"Bias Correction for Local Linear Regression Estimation Using Asymmetric Kernels via the Skewing Method"*.
04:20 pm – 04:40 pm (page 222, abstract 218)

Statistics in Behavorial and Educational Sciences II (Behavioural Sciences)

Chair: Steffi Pohl, Room: A 213

Christoph Frohn, Monika Obersneider. *"Dynamic Microsimulation Modelling of Care Needs in Germany"*.
03:20 pm – 03:40 pm (page 223, abstract 219)

Christoph Herrmann, Susanne Kirschstein-Barczewski. *"Elicited preferences of potential spontaneous unaffiliated on-site volunteers in the context of natural disasters"*.
03:40 pm – 04:00 pm (page 224, abstract 220)

Martin Elff. *"Tactical Voting and Ticket-Splitting in Mixed Electoral Systems: A Finite-Mixture Approach Applied to the Case of Germany"*.
04:00 pm – 04:20 pm (page 224, abstract 221)

Patrick Michael Schwaferts. *"Bayes Factor: Inconsistency in Sequential Updating."*.
04:20 pm – 04:40 pm (page 225, abstract 222)

Survival and Event History Analysis III (High-Dimensional Analysis)

Chair: Moritz Maximilian Berger, Room: B106

Thomas Welchowski, Verena Zuber, Matthias Schmid. *"Correlation-Adjusted Regression Survival Scores for High-Dimensional Variable Selection"*.
03:20 pm – 03:40 pm (page 226, abstract 223)

Katrin Madjar, Manuela Zucknick, Katja Ickstadt, Jörg Rahnenführer. *"Bayesian variable selection for Cox models with network-structured covariates"*.
03:40 pm – 04:00 pm (page 227, abstract 224)

Maike Hohberg, Andreas Groll. *"Adaptive LASSO Cox frailty models based on the full likelihood"*.
04:00 pm – 04:20 pm (page 227, abstract 225)

Stephan M. Bischofberger, Munir Hiabu, Enno Mammen, Jens P. Nielsen. *"Smooth backfitting of additively structured hazard rates for in-sample forecasting"*.
04:20 pm – 04:40 pm (page 228, abstract 226)

Statistics in Finance II

Chair: Yarema Okhrin, Room: E 004

V Golosnoy, J Kellermann. *"Testing for Daily Jumps in Risky Asset Returns: a novel approach based on Gini concentration measure"*.
03:20 pm – 03:40 pm (page 229, abstract 227)

Alexander Kreuzer, Claudia Czado. *"Dynamic regular vine copulas with an application to exchange rates dependence"*.
03:40 pm – 04:00 pm (page 229, abstract 228)

Ekaterina Kazak, Winfried Pohlmeier. *"Portfolio Pretesting with Machine Learning"*.
04:00 pm – 04:20 pm (page 230, abstract 229)

Ingo Hoffmann, Christoph J. Börner. *"The risk function of the goodness-of-fit tests for tail models."*.
04:20 pm – 04:40 pm (page 230, abstract 230)

Statistics of High Dimensional Data II

Chair: Markus Pauly, Room: D 209

Eva-Maria Huessler, Martin Schäfer, Pablo Landgraf, Holger Schwender. *"Detecting binding sites in PAR-CLIP data using a Bayesian hierarchical model"*.
03:20 pm – 03:40 pm (page 231, abstract 231)

Miriam Kesselmeier, Anke Hinney, André Scherag. *"High-throughput DNA methylation analysis with reference-free cell type adjustment: method comparison in a real data application"*.
03:40 pm – 04:00 pm (page 232, abstract 232)

Simon Lineu Umbach. *"Forecasting with Supervised Factor Models"*.
04:00 pm – 04:20 pm (page 233, abstract 233)

Anke Huels, Michael P. Epstein. *"Impact of Population Stratification on Polygenic Risk Score Approaches"*.
04:20 pm – 04:40 pm (page 234, abstract 234)

Statistics in Science, Technology and Industry II

Chair: Axel Gandy, Room: A 240

Ansgar Steland. *"Inference and Change Detection for LSHD Time Series and Applications to Ozone Monitoring"*.
03:20 pm – 03:40 pm (page 235, abstract 235)

Philipp Otto, Rick Steinert. *"Estimation of the Spatial Weighting Matrix for Spatiotemporal Data with Structural Breaks"*.
03:40 pm – 04:00 pm (page 236, abstract 236)

Matthias Gärtner, Solveig Plomer, Sevil Duvarci, Jochen Roeper, Michael Messer, **Gaby Schneider**. *"Statistical analysis of joint pausing in parallel spike trains"*.
04:00 pm – 04:20 pm (page 237, abstract 237)

Eva-Christina Becker-Emden, Sonja Kuhnt. *"Simultaneous optimization of several correlated response variables"*.
04:20 pm – 04:40 pm (page 237, abstract 238)

Wednesday, March 20, 5:00 pm – 6:20 pm

Advanced Regression Modeling IV (Longitudinal Data and Mixed Models)

Chair: Nadja Klein, Room: D 209

Helga Wagner. *"Flexible Bayesian modelling of treatment effects on panel outcomes"*.
05:00 pm – 05:20 pm (page 238, abstract 239)

Claudia Czado, Matthias Killiches. *"A D-Vine Copula-Based Model for Repeated Measurements Extending Linear Mixed Models with Homogeneous Correlation Structure"*.
05:20 pm – 05:40 pm (page 238, abstract 240)

Georg Heinze, Rok Blagus. *"Solving separation in the mixed effects logistic regression model"*.
05:40 pm – 06:00 pm (page 239, abstract 241)

Stanislav Anatolyev. *"Second order asymptotic biases of consistent estimators under many instruments"*.
06:00 pm – 06:20 pm (page 239, abstract 242)

Computational Statistics and Statistical Software IV(Software)

Chair: Gero Szepannek, Room: A 214

Fabian Scheipl, Jeff Goldsmith. *"tidyfun: a new framework for representing and working with function-valued data"*.
05:00 pm – 05:20 pm (page 240, abstract 243)

Sebastian Doehler, Guillermo Durand, Etienne Roquain. *"Controlling the false discovery rate for discrete data: New results and software"*.
05:20 pm – 05:40 pm (page 240, abstract 244)

Xudong Sun. *"recent advances in deep reinforcement learning and the R implementation rlR package"*.
05:40 pm – 06:00 pm (page 241, abstract 245)

Ludwig A. Hothorn, Torsten Hothorn, **Paola G. Ferrario**. *"A workflow for metabolomics using CRAN packages to demonstrate association between a covariate and multiple analytes (some with detection limit)"*.
06:00 pm – 06:20 pm (page 242, abstract 246)

Design of Experiments and Clinical Trials V (Optimal Design II)

Chair: Maryna Prus, Room: A 125

Kira Alhorn, Kirsten Schorning, Holger Dette. *"Optimal designs for frequentist model averaging"*.
05:00 pm – 05:20 pm (page 243, abstract 247)

Markus Hainy, David Price, Olivier Restif, Christopher Drovandi. *"Optimal Bayesian design for model discrimination via classification"*.
05:20 pm – 05:40 pm (page 243, abstract 248)

Norbert Gaffke. *"The adaptive Wynn-algorithm in generalized linear models with univariate response"*.
05:40 pm – 06:00 pm (page 244, abstract 249)

Osama Idais. *"Locally optimal designs for gamma models"*.
06:00 pm – 06:20 pm (page 244, abstract 250)

Survival and Event History Analysis IV (Competing Risks and Multistate Models I)

Chair: Jan Beyersmann and Matthias Schmid, Room: B106

Lucas Radloff, Rafael Weißbach. *"Consistent estimation in non-Markov multi-state models"*.
05:00 pm – 05:20 pm (page 245, abstract 251)

Ursula U. Müller, Ingrid Van Keilegom. *"Goodness-of-fit tests for the cure rate in a mixture cure model"*.
05:20 pm – 05:40 pm (page 246, abstract 252)

Fabian Otto-Sobotka, Alexander Seipp, Verena Jürgens, Antje Timmer. *"Semiparametric Accelerated Failure Times Quantile and Expectile Regression using Auxiliary Likelihoods"*.
05:40 pm – 06:00 pm (page 247, abstract 253)

Statistics in Finance III

Chair: Markus Bibinger, Room: E 004

Gustav Alfelt, Taras Bodnar, Joanna Tyrcha. *"Goodness-of-fit tests for centralized Wishart processes"*.

05:00 pm – 05:20 pm (page 248, abstract 254)

Eugen Ivanov, Yarema Okhrin. *"Matrixvariate Factor Model for Realized Covariances"*.
05:20 pm – 05:40 pm (page 248, abstract 255)

Fabian Krüger, Roman Liesenfeld, **Laura Reh**. *"Dynamic Modeling of the Global Minimum Variance Portfolio weights"*.
05:40 pm – 06:00 pm (page 249, abstract 256)

Lars Winkelmann. *"Detecting a hidden component in high-frequency yield curves using rank tests for the covolatility process"*.
06:00 pm – 06:20 pm (page 249, abstract 257)

Statistics in Science, Technology and Industry III

Chair: Sonja Kuhnt, Room: A 240

Nils Mause, Ansgar Steland. *"Inference on the Second Moment Structure of High-Dimensional Sensor-Type Data in a K - Sample Setting"*.
05:00 pm – 05:20 pm (page 250, abstract 258)

Lena Hubig, Nicholas Lack, Ulrich Mansmann. *"Statistical Process Monitoring to Improve Quality Assurance of Inpatient Care"*.
05:20 pm – 05:40 pm (page 251, abstract 259)

Sven Knoth. *"On Steady-state Performance Characteristics of Control Charts – Meaning and Numerics"*.
05:40 pm – 06:00 pm (page 252, abstract 260)

Thursday, March 21, 9:00 am – 10:40 am

Clustering II (Mixture Models)

Chair: Christian Hennig, Room: A 213

Jean-Patrick Baudry, Gilles Celeux. *"Model-based clustering for cytometry"*.
09:20 am – 09:40 am (page 253, abstract 261)

Bernd Taschler, Frank Dondelinger, Sach Mukherjee. *"Model-based clustering in very high dimensions via adaptive projections"*.
09:40 am – 10:00 am (page 254, abstract 262)

Malte Jastrow, Claus Weihs. *"Highly Multimodal Likelihood Functions of Mixture Distributions"*.
10:00 am – 10:20 am (page 255, abstract 263)

Carlos Améndola, Christian Haase, Alexander Engström. *"Maximum Number of Modes of Gaussian Mixtures"*.
10:20 am – 10:40 am (page 256, abstract 264)

Classification and Pattern Recognition I

Chair: Hans Kestler, Room: E 004

Eyke Hüllermeier. *"Analyzing and Learning from Ranking Data: New Problems and Challenges"*.
09:00 am – 09:40 am (page 257, abstract 265)

Georg Schollmeyer. *"Classification with stylized betweenness-relations allowing for regularization with uniform Vapnik-Chervonenkis-guarantees"*.
09:40 am – 10:00 am (page 258, abstract 266)

Lu-Hung Chen, Ci-Ren Jiang. *"Sensible functional linear discriminant analysis"*.
10:00 am – 10:20 am (page 259, abstract 267)

Dennis Hammerschmidt. *"Talk and Action in the United Nations General Assembly - Vote-buying and the power to induce states to vote against their own preferences"*.
10:20 am – 10:40 am (page 259, abstract 268)

Design of Experiments and Clinical Trials VI (Optimal Design III)

Chair: Heiko Großmann, Room: A 214

Eva Benkova, Radoslav Harman, **Werner G. Müller**. *"Privacy sets for distance constraints"*.
09:00 am – 09:20 am (page 260, abstract 269)

Dennis Schmidt. *"c- and ϕ_k-optimal designs for a class of nonlinear multiple regression models"*.
09:20 am – 09:40 am (page 260, abstract 270)

Nadja Malevich, Christine Müller. *"Optimal inspection times for lifetime estimation based on interval-censored samples"*.
09:40 am – 10:00 am (page 261, abstract 271)

Martin Radloff, Rainer Schwabe. *"Locally D-optimal Designs for Non-linear Models on the k-dimensional Ball"*.
10:00 am – 10:20 am (page 261, abstract 272)

Eirini Renata Tsirpitzi, Frank Miller. *"Optimal dose-finding for efficacy-safety-models"*.
10:20 am – 10:40 am (page 262, abstract 273)

Data Science

Chair: Thomas Seidl, Room: D 209

Joachim M. Buhmann. *"Robust algorithmics: a foundation for science?!"*.
09:00 am – 09:40 am (page 263, abstract 274)

Ulrich Pötter, Ingrid Schockaert. *"Dynamic topological analysis of residential mobility"*.
09:40 am – 10:00 am (page 264, abstract 275)

Daniela Zöller, Marlon Claaßen, Stefan Lenz, Martin Treppner, Harald Binder. *"Dealing with complex patterns in mobile app and wearable device data"*.

10:00 am – 10:20 am (page 265, abstract 276)

Jochen Kruppa. *"Risk factors with a spike at zero in epigenome-wide association studies"*.
10:20 am – 10:40 am (page 266, abstract 277)

Epidemiology I (Causal inference methods)

Chair: Vanessa Didelez, Room: A 240

Miguel Hernan. *"Estimating per-protocol effects. Randomized trials analyzed like observational studies"*.
09:00 am – 09:40 am (page 267, abstract 278)

Dirk Enders. *"A comparison of sequential and simultaneous Propensity Score matching in a study with three treatment groups"*.
09:40 am – 10:00 am (page 268, abstract 279)

Christoph Kurz, Michael Laxy. *"Association of Obesity with Health Care Costs: Strengthening the Instrument in Mendelian Randomization Studies"*.
10:00 am – 10:20 am (page 269, abstract 280)

Roland Weigand. *"Propensity Scores aus hochdimensionalen Routinedaten und das DMP Koronare Herzkrankheit"*.
10:20 am – 10:40 am (page 270, abstract 281)

Marketing and E-Commerce

Chair: Winfried Steiner, Room: A 119

Friederike Paetz. *"Latent Class Analysis in Marketing: Drawing Inferences for Social Brand Personalities"*.
09:00 am – 09:40 am (page 271, abstract 282)

Toshifumi Sugitani. *"Statistical considerations on assessment of responsiveness of sales to salesforce effort: A Japanese pharmaceutical company's example"*.
09:40 am – 10:00 am (page 272, abstract 283)

Felix Meyer. *"Modeling price-sensitive demand: An application to continuous pricing"*.
10:00 am – 10:20 am (page 272, abstract 284)

Shuai Shao. *"Whether, when and which: modelling advanced seat reservations by airline passengers"*.
10:20 am – 10:40 am (page 273, abstract 285)

Machine Learning I

Chair: Tim Beißbarth and Hans Kestler, Room: A 140

Michel Philipp, Thomas Rusch, Kurt Hornik, Achim Zeileis, Carolin Strobl. *"Stability Assessment for Trees and other Supervised Statistical Learning Results"*.
09:00 am – 09:20 am (page 274, abstract 286)

Stephan Seifert, Sven Gundlach, Silke Szymczak. *"Surrogate minimal depth as an importance measure for variables in random forests".*
09:20 am – 09:40 am (page 274, abstract 287)

Annette Möller, Jan Gertheiss. *"A classification tree for functional data".*
09:40 am – 10:00 am (page 275, abstract 288)

Christoph Molnar, Giuseppe Casalicchio, Bernd Bischl. *"Measuring and optimizing machine learning interpretability".*
10:00 am – 10:20 am (page 276, abstract 289)

Max Westphal, Werner Brannath. *"A multiple testing framework for the efficient statistical evaluation of (machine-learned) prediction models".*
10:20 am – 10:40 am (page 277, abstract 290)

Small Sample Statistics

Chair: Jörg Rahnenführer and Robert Kwiecien, Room: A 125

Malgorzata Bogdan. *"Convex optimization methods for identifying predictors when n<p".*
09:20 am – 10:00 am (page 278, abstract 291)

Dario Zocholl, Manuel Wiesenfarth, Geraldine Rauch, Annette Kopp-Schneider. *"Designing pediatric phase I clinical trials in oncology by borrowing information from trials with adult patients".*
10:00 am – 10:20 am (page 279, abstract 292)

Annette Kopp-Schneider, Silvia Calderazzo, Manuel Wiesenfarth. *"Use of external information in clinical trials: What can be gained in terms of frequentist power?".*
10:20 am – 10:40 am (page 280, abstract 293)

Thursday, March 21, 1:30 pm – 2:50 pm

Clustering III (General clustering and classification)

Chair: Jean-Patrick Baudry, Room: A 213

Ingo Steinwart. *"Aspects of adaptive density-based cluster analysis".*
01:30 pm – 02:10 pm (page 281, abstract 294)

Christian Hennig, Cinzia Viroli, Laura Anderlucci. *"K-quantiles clustering".*
02:10 pm – 02:30 pm (page 281, abstract 295)

Adalbert F.X. Wilhelm. *"Hybrid Image Classification using Captions and Image Features".*
02:30 pm – 02:50 pm (page 282, abstract 296)

Design of Experiments and Clinical Trials VII (Optimal Design IV)

Chair: Werner G. Müller, Room: A 214

Florin Vaida, Kristen Hansen, Ming Tai-Seale. *"Efficient design for longitudinal, cluster-*

randomized clinical trials with repeated measures".
01:30 pm – 01:50 pm (page 283, abstract 297)

Maryna Prus. *"Optimal Designs in Multiple Group Random Coeffcient Regression Models".*
01:50 pm – 02:10 pm (page 284, abstract 298)

Joachim Kunert, Johanna Mielke. *"Efficient Designs for the estimation of mixed and self carryover effects".*
02:10 pm – 02:30 pm (page 284, abstract 299)

Marius Schmidt, Rainer Schwabe. *"Standardized Maximin D- and c-optimal Designs for Poisson Count Data with Gamma Block Effects".*
02:30 pm – 02:50 pm (page 285, abstract 300)

Epidemiology II (Enviromental risks)

Chair: Dirk Enders, Room: E 004

Matthias Aßenmacher, Jan Christian Kaiser, Ignacio Zaballa, Antonio Gasparrini, Helmut Küchenhoff. *"Exposure-lag response associations between lung cancer mortality and radon exposure in German uranium miners".*
01:30 pm – 01:50 pm (page 286, abstract 301)

Noemi Castelletti, Kyoji Furukawa, Cristoforo Simonetto, Helmut Küchenhoff, Georgios T. Stathopoulos, Jan Christian Kaiser. *"Independent estimation of risk from smoking and radiation for different histologic lung cancer types using generalized additive models and biologically-based models of carcinogenesis".*
01:50 pm – 02:10 pm (page 287, abstract 302)

Claudia Wigmann, Anke Hüls, Jean Krutmann, Tamara Schikowski. *"Application of weighted risk scores to estimate the relative contribution of environmental and genetic factors to skin aging".*
02:10 pm – 02:30 pm (page 288, abstract 303)

Machine Learning II

Chair: Tim Beißbarth and Hans Kestler, Room: A 140

Aude Sportisse, Claire Boyer, Julie Josse. *"Low-rank estimation with Missing Non At Random data".*
01:30 pm – 01:50 pm (page 289, abstract 304)

Umberto Noè, Dirk Husmeier. *"Scaled Expected Improvement for Bayesian Optimization".*
01:50 pm – 02:10 pm (page 290, abstract 305)

Dirk Surmann, Claus Weihs, Uwe Ligges. *"Infill Criterion for Multimodal Model-Based Optimisation".*
02:10 pm – 02:30 pm (page 290, abstract 306)

Andrea Bommert, Jörg Rahnenführer, Michel Lang. *"Stable Feature Selection".*
02:30 pm – 02:50 pm (page 291, abstract 307)

Network Analysis I

Chair: Goeran Kauermann, Room: A 119

Pavel N. Krivitsky, Martina Morris. *"Inference for Social Network Models from Egocentrically-Sampled Data"*.
01:30 pm – 02:10 pm (page 292, abstract 308)

Ann-Kristin Becker, **Hajo Holzmann**. *"Nonparametric inference in the dynamic stochastic block model"*.
02:10 pm – 02:30 pm (page 292, abstract 309)

Goeran Kauermann, **Benjamin Sischka**. *"Bayesian and Spline based Approaches for (EM based) Graphon Estimation"*.
02:30 pm – 02:50 pm (page 293, abstract 310)

Robust and Nonparametric Statistics III

Chair: Peter Ruckdeschel, Room: D 209

Sermad Abbas, Roland Fried. *"Real-time detection of sudden location changes in time series with a time-varying trend"*.
01:30 pm – 01:50 pm (page 294, abstract 311)

Edgar Brunner. *"Paradoxical Results with Ranks for Unequal Sample Sizes"*.
01:50 pm – 02:10 pm (page 295, abstract 312)

Bernhard Spangl. *"On robust two-way MANOVA tests with applications"*.
02:10 pm – 02:30 pm (page 296, abstract 313)

Kevin Leckey, Christine H. Müller, Dennis Malcherczyk. *"Generalized sign tests: From asymptotics to efficient computation"*.
02:30 pm – 02:50 pm (page 296, abstract 314)

Survival and Event History Analysis V (Competing Risks and Multistate Models II)

Chair: Jan Beyersmann, Room: A 021

Jörn Schulz. *"State transition modeling of complex monitored health data"*.
01:30 pm – 01:50 pm (page 297, abstract 315)

Tobias Bluhmki, Hein Putter, Arthur Allignol, Jan Beyersmann on behalf of the COMBACTE-MAGNET consortium. *"Resampling complex time-to-event data without individual patient data, with a view towards time-dependent exposures"*.
01:50 pm – 02:10 pm (page 298, abstract 316)

Marvin N Wright, Laust H. Mortensen, Sasmita Kusumastuti, Rudi G.J. Westendorp, Thomas A. Gerds. *"Recurrent neural networks for time to event predictions with competing risks"*.
02:10 pm – 02:30 pm (page 299, abstract 317)

Regina Stegherr, Jan Beyersmann, Claudia Schmoor, Michael Luebbert, Tim Friede. *"Methodological aspects in the analysis of adverse events in time-to-event data"*.
02:30 pm – 02:50 pm (page 300, abstract 318)

Time Series Analysis V

Chair: Marc-Oliver Pohle, Room: A 240

Marc-Oliver Pohle. *"Analyzing Different Facets of Forecast Quality through Decompositions of Loss Functions"*.
01:30 pm – 01:50 pm (page 301, abstract 319)

Jennifer Pohle, Roland Langrock, Ruth King, Mihaela van der Schaar. *"Coupled state-switching models with applications in ecology and medicine"*.
01:50 pm – 02:10 pm (page 302, abstract 320)

Paul Bürkner, Jonah Gabry, Aki Vehtari. *"Approximate leave-future-out cross-validation for time series models"*.
02:10 pm – 02:30 pm (page 303, abstract 321)

Visualisation and Exploratory Data Analysis

Chair: Heike Hofmann, Room: A 125

Heike Hofmann. *"Visual Inference: leveraging the power of our eyes"*.
01:30 pm – 02:10 pm (page 304, abstract 322)

Asbel Bohigues, Cristina Rivas. *"HJ-Biplot as a data visualization tool in Social Sciences"*.
02:10 pm – 02:30 pm (page 304, abstract 323)

Achim Zeileis, Jason C. Fisher, Kurt Hornik, Ross Ihaka, Claire D. McWhite, Paul Murrell, Reto Stauffer, Claus O. Wilke. *"A Toolbox for Manipulating and Assessing Color Palettes for Statistical Graphics"*.
02:30 pm – 02:50 pm (page 305, abstract 324)

Thursday, March 21, 3:20 pm – 4:40 pm

Computational Statistics and Statistical Software V (invited)

Chair: Roland Fried, Room: A 240

Gertraud Malsiner-Wallli, Sylvia Frühwirth-Schnatter, **Bettina Grün**. *"Identifying Mixtures of Mixtures Using Bayesian Estimation"*.
03:20 pm – 04:00 pm (page 306, abstract 325)

Nicole Schüller, Anne-Laure Boulesteix, Bernd Bischl, Roman Hornung. *"Robust outcome prediction across data sources through altered tuning parameter value selection"*.
04:00 pm – 04:20 pm (page 307, abstract 326)

Jonas Rieger, Lars Koppers, Carsten Jentsch, Jörg Rahnenführer. *"Measuring Stability of Replicated LDA Runs"*.
04:20 pm – 04:40 pm (page 308, abstract 327)

Design of Experiments and Clinical Trials VIII (Clinical Trials I)

Chair: Tim Friede, Room: A 214

Chul Ahn. *"Sample size considerations for paired experimental design with incomplete outcomes"*.
03:20 pm – 03:40 pm (page 309, abstract 328)

Benjamin Mayer. *"A two-level matching algorithm for a multi-center case-control study using registry data"*.
03:40 pm – 04:00 pm (page 310, abstract 329)

Christina Loley, Beate Krüger, Michael Matiu, Jürgen von Frese, Hannes Buchner, Armin Boehrer, Erich Bluhmki. *"Equivalence testing with dependent data and unequal variances: Simulation of power and type 1 error for modifications of the TOST procedure"*.
04:00 pm – 04:20 pm (page 311, abstract 330)

Robert Richmond Peck. *"Optimal Decisions in the Portfolio Problem"*.
04:20 pm – 04:40 pm (page 312, abstract 331)

Epidemiology III (Chronic and infectious disease methodology)

Chair: André Karch, Room: E 004

Tim Filla, Annika Hoyer, Thaddäus Tönnies, Ralph Brinks. *"Prevalence of chronic diseases: Comparison between an analytical relationship and a micro-simulation."*.
03:20 pm – 03:40 pm (page 313, abstract 332)

Ralph Brinks, Annika Hoyer. *"Compression of morbidity due to chronic diseases in Germany? Results from the Survey of Health, Ageing and Retirement in Europe (SHARE) 2004-2015"*.
03:40 pm – 04:00 pm (page 314, abstract 333)

Sebastian Meyer. *"Evaluating forecasts of infectious disease spread"*.
04:00 pm – 04:20 pm (page 315, abstract 334)

Birgit Debrabant. *"Estimation of multivariate hidden population sizes from register data"*.
04:20 pm – 04:40 pm (page 315, abstract 335)

Machine Learning III

Chair: Tim Beißbarth and Hans Kestler, Room: A 140

Jakob Richter, Katrin Madjar, Jörg Rahnenführer. *"Calculating Optimal Subgroup Weights for Survival Analysis using Model-Based Optimization"*.
03:20 pm – 03:40 pm (page 316, abstract 336)

David Rügamer, Sonja Greven. *"Inference for L2-Boosting"*.
03:40 pm – 04:00 pm (page 316, abstract 337)

Eliud Silva, Miguel A. Villalobos. *"An application of Statistical learning to the analysis*

of mortality by homicide in Mexico, 2014-2017".
04:00 pm – 04:20 pm (page 317, abstract 338)

Daniel Horn, Nils Jannik Schüßler. *"Statistical Analysis of Benchmark Results".*
04:20 pm – 04:40 pm (page 318, abstract 339)

Network Analysis II

Chair: Alexander Günther Kreiß, Room: A 119

Sevag Kevork, Goeran Kauermann. *"Iterative Estimation for Exponential Random Graph Models with Nodal Random Effects".*
03:20 pm – 03:40 pm (page 319, abstract 340)

Alexander Günther Kreiß. *"Modelling Time-Varying Dependence in Dynamic Networks with Applications to Regression and Model-Checking in Survival Analysis".*
03:40 pm – 04:00 pm (page 319, abstract 341)

Andreas Geyer-Schulz, Fabian Ball. *"On the Construction of Invariant Measures for Graph Partition Comparison".*
04:00 pm – 04:20 pm (page 320, abstract 342)

Robust and Nonparametric Statistics IV

Chair: Peter Ruckdeschel, Room: D 209

Davy Paindaveine, Germain Van Bever. *"Halfspace depth for scatter matrices".*
03:20 pm – 04:00 pm (page 321, abstract 343)

Karl Mosler, Pavlo Mozharovskyi. *"Choosing among notions of depth for multivariate data".*
04:00 pm – 04:20 pm (page 321, abstract 344)

Melanie Horn, Christine Müller. *"Test based on sign depth for multiple regression".*
04:20 pm – 04:40 pm (page 322, abstract 345)

Survival and Event History Analysis VI (Prediction)

Chair: Matthias Schmid, Room: A 021

Colin Griesbach, Andreas Groll, Elisabeth Waldmann. *"Joint Modelling approaches to survival analysis via likelihood-based boosting techniques.".*
03:20 pm – 03:40 pm (page 323, abstract 346)

Matthias Brueckner. *"Bayesian joint latent class models of longitudinal and time-to-event outcomes".*
03:40 pm – 04:00 pm (page 324, abstract 347)

Bernhard Haller. *"A simulation study comparing different approaches for detection of covariate-by-treatment interactions".*
04:00 pm – 04:20 pm (page 325, abstract 348)

Grigoriy Volovskiy, Udo Kamps. *"Maximum Likelihood Prediction of Record Values".*
04:20 pm – 04:40 pm (page 325, abstract 349)

Statistics in Science, Technology and Industry IV

Chair: Jürgen Pilz, Room: A 213

Ben Moews, Rafael S. de Souza, Emille E. O. Ishida, Alex I. Malz, Caroline Heneka, Ricardo Vilalta, Joe Zuntz. *"What we might miss: Stress-testing measurements of dark energy"*.
03:20 pm – 03:40 pm (page 326, abstract 350)

F. Marta L. Di Lascio, Andrea Menapace, Maurizio Righetti. *"Joint and conditional dependence modelling of district heating demand and weather conditions: a copula-based approach"*.
03:40 pm – 04:00 pm (page 327, abstract 351)

Maximilian Coblenz, Oliver Grothe, Simon Holz, Rainer Koch. *"Modeling Fuel Injector Spray Characteristics of Jet Engines Using Vine Copulas"*.
04:00 pm – 04:20 pm (page 328, abstract 352)

Sophie Tchanyou Ganme, Katja Ickstadt. *"Bayesian Prediction for failure times in Fatigue Behavior of Prestressed Concrete"*.
04:20 pm – 04:40 pm (page 328, abstract 353)

Visualisation and Exploratory Data Analysis

Chair: Hans Kestler, Room: A 125

Antony Unwin. *"Graphics in Research and Teaching illustrated in Forschung und Lehre"*.
03:20 pm – 03:40 pm (page 329, abstract 354)

Bernhard Hrobath, Friedrich Leisch. *"A new approach to model and visualize Airbnb listing prices by the use of a smoothing surface on spatial information"*.
03:40 pm – 04:00 pm (page 329, abstract 355)

Federico Marini, Aaron Lun, Charlotte Soneson, Kevin Rue-Albrecht. *"iSEE: RNA-sequencing data exploration made easy and reproducible"*.
04:00 pm – 04:20 pm (page 330, abstract 356)

Karsten Keller. *"Short ordinal patterns in time series analysis"*.
04:20 pm – 04:40 pm (page 331, abstract 357)

Thursday, March 21, 5:00 pm – 6:20 pm

Advanced Regression Modeling V

Chair: Elmar Spiegel, Room: A 021

Manuel Batram, Sebastian Büscher, Dietmar Bauer. *"Using mixed multinomial probit models to explain daily mobility behavior in a large panel data set"*.
05:00 pm – 05:20 pm (page 332, abstract 358)

Muriel Buri, Armin Curt, John Steeves, Torsten Hothorn. *"Baseline-adjusted Proportional Odds Models for Quantification of Treatment Effects in Neurological Trials with Ordinal Outcomes"*.
05:20 pm – 05:40 pm (page 333, abstract 359)

Causal Inference II (Aspects of Propensity Score Methods)

Chair: Heinz Leitgöb and Martin Elff, Room: B106

Lina Glaubitz, Tim Filla, Oliver Kuß. *"Measuring global covariate balance in matched propensity score analysis"*.
05:00 pm – 05:20 pm (page 334, abstract 360)

Christiana Drake, Julie Smith-Gagen. *"Propensity Weighting in the Estimation of Direct Effects."*.
05:20 pm – 05:40 pm (page 334, abstract 361)

Classification and Pattern Recognition II

Chair: Andreas Geyer-Schulz, Room: A 125

Cuiling Wang, Richard Lipton, Ellen Grober. *"Evaluate the diagnostic accuracy for disease of longitudinal markers with missing data"*.
05:00 pm – 05:20 pm (page 335, abstract 362)

Alexander Hapfelmeier. *"Measuring conditional agreement in method comparison studies by mixed-effects model trees"*.
05:20 pm – 05:40 pm (page 336, abstract 363)

Design of Experiments and Clinical Trials IX (Clinical Trials II)

Chair: Ekkehard Glimm, Room: A 214

Dorothea Weber, Lorenz Uhlmann, Meinhard Kieser. *"Adaptive Propensity Score Procedure Improves Matching in Prospective Observational Trials"*.
05:00 pm – 05:20 pm (page 337, abstract 364)

Ralf Bender. *"Diskussion der Estimand-Strategien aus Sicht der Nutzenbewertung"*.
05:20 pm – 05:40 pm (page 338, abstract 365)

Data Fusion and Meta-Analysis IV

Chair: Georgia Salanti and Sibylle Sturtz, Room: A 119

Christian Röver, Tim Friede. *"Dynamically borrowing strength from another study"*.
05:00 pm – 05:20 pm (page 339, abstract 366)

Burak Kürsad Günhan, Christian Röver, Tim Friede. *"Meta-analysis of few studies involving rare events"*.
05:20 pm – 05:40 pm (page 340, abstract 367)

Epidemiology IV

Chair: Ralph Brinks, Room: E 004

Mercè Garí. *"Statistical Tools for Assessing the Exposome"*.
05:00 pm – 05:20 pm (page 341, abstract 368)

Andreas Beyerlein. *"Quantile regression for the applied user – opportunities, challenges, examples"*.
05:20 pm – 05:40 pm (page 342, abstract 369)

Machine Learning IV

Chair: Tim Beißbarth and Hans Kestler, Room: A 140

Lyudmila Grigoryeva, Oleksandra Kukharenko, Juan-Pablo Ortega. *"Forecasting of high-dimensional realized covariances with reservoir computing"*.
05:00 pm – 05:20 pm (page 343, abstract 370)

Matthias Katzfuss. *"Gaussian-Process Approximations for Big Data"*.
05:20 pm – 05:40 pm (page 343, abstract 371)

Statistics of High Dimensional Data III

Chair: Taras Bodnar, Room: D 209

Dietrich von Rosen. *"The Growth Curve model under high dimensions with applications to profile analysis"*.
05:00 pm – 05:40 pm (page 344, abstract 372)

Statistics in Science, Technology and Industry V

Chair: Sven Knoth, Room: A 213

Undine Falkenhagen, Wolfgang Kössler, Hans-Joachim Lenz. *"Inlier Detection"*.
05:00 pm – 05:20 pm (page 345, abstract 373)

Gloria Gheno. *"A new statistical index to evaluate sleep quality using sensors"*.
05:20 pm – 05:40 pm (page 345, abstract 374)

Friday, March 22, 9:00 am – 10:40 am

Advanced Regression Modeling VI (Modeling Multivariate Dependence)

Chair: Thomas Kneib, Room: A 021

Giampiero Marra, Rosalba Radice. *"Generalised Joint Regression Modelling"*.
09:00 am – 09:40 am (page 346, abstract 375)

Hendrik van der Wurp, Andreas Groll. *"Predicting matches in international football tournaments via generalised joint regression modelling"*.
09:40 am – 10:00 am (page 346, abstract 376)

Bruno Santos, Thomas Kneib. *"Structured additive multiple-output noncrossing Bayesian quantile regression models"*.
10:00 am – 10:20 am (page 347, abstract 377)

Isa Marques, Thomas Kneib, Nadja Klein. *"Non-stationary spatial regression for modelling monthly precipitation in Germany"*.
10:20 am – 10:40 am (page 347, abstract 378)

Causal Inference III (Neyman-Rubin Model and Observational Studies)

Chair: Heinz Leitgöb and Martin Elff, Room: A 125

Ming-Yueh Huang. *"Estimating continuous treatment effect functions with joint sufficient dimension reduction"*.
09:20 am – 09:40 am (page 348, abstract 379)

Maja von Cube, Martin Schumacher, Martin Wolkewitz. *"Causal inference in multistate models - estimands and estimators of the population-attributable fraction"*.
09:40 am – 10:00 am (page 349, abstract 380)

Michael Lechner. *"Practical and Effective Estimation of Effect Heterogeneity by Modified Causal Forests"*.
10:00 am – 10:20 am (page 350, abstract 381)

Julia Krzykalla, Axel Benner, Annette Kopp-Schneider. *"Search for predictive factors based on observational studies"*.
10:20 am – 10:40 am (page 351, abstract 382)

Design of Experiments and Clinical Trials X (Clinical Trials III)

Chair: Joachim Gerss, Room: A 214

Markus Harden, Tim Friede. *"Blinded sample size reestimation in multi-centre randomized controlled clinical trials"*.
09:20 am – 09:40 am (page 352, abstract 383)

Silvia Calderazzo, Manuel Wiesenfarth, Annette Kopp-Schneider. *"A Bayesian decision-theoretic framework for evaluation of Bayesian clinical trials performance and robustness to prior-data conflict"*.
09:40 am – 10:00 am (page 353, abstract 384)

Angelika M. Stefan, Felix Schönbrodt. *"Planning sequential Bayesian designs: Sample size prediction and stopping boundary specification"*.
10:00 am – 10:20 am (page 354, abstract 385)

Carolin Herrmann, Kevin Kunzmann, Maximilian Pilz, Meinhard Kieser, Geraldine Rauch. *"A new conditional performance score for evaluating sample size recalculation rules in adaptive designs"*.
10:20 am – 10:40 am (page 355, abstract 386)

Machine Learning V

Chair: Tim Beißbarth and Hans Kestler, Room: A 140

Günther Palm. *"Learning in artificial and real neural networks"*.

09:00 am – 09:40 am (page 356, abstract 387)

Moritz Hess, Stefan Lenz, Harald Binder. *"Conditional sampling for exploring biological connections in single cell RNA-Seq data with Deep Boltzmann Machines".*
09:40 am – 10:00 am (page 356, abstract 388)

Göran Köber, Stefan Lenz, Haakon Engen, Kenneth S.L. Yuen, Anita Schick, Raffael Kalisch, Harald Binder. *"Pattern Detection of Life Events and Daily Hassles Using Longitudinal Deep Boltzmann Machines".*
10:00 am – 10:20 am (page 357, abstract 389)

Christian Arnold, Marcel Neunhoeffer, Sebastian Sternberg. *"Releasing Differentially Private Synthetic Micro-Data with Bayesian GANs".*
10:20 am – 10:40 am (page 358, abstract 390)

Network Analysis III

Chair: Pavel N: Krivitsky, Room: A 213

Pascal Schlosser, Anna Köttgen, Martin Schumacher. *"Preservation of multivariate correlation-based networks constructed from high-dimensional data".*
09:20 am – 09:40 am (page 359, abstract 391)

Angelika Schmid, Sven Apel. *"Tensor decomposition for dynamic clustering in multimodal social networks".*
09:40 am – 10:00 am (page 360, abstract 392)

Michael Lebacher, Göran Kauermann. *"Regression-based Network Reconstruction with Nodal and Dyadic Covariate Effects".*
10:00 am – 10:20 am (page 360, abstract 393)

Verena Bauer, Göran Kauermann. *"A smooth dynamic network model for patent collaboration data".*
10:20 am – 10:40 am (page 361, abstract 394)

Robust and Nonparametric Statistics V

Chair: Peter Ruckdeschel, Room: E 004

Gabriel Frahm, Klaus Nordhausen, Hannu Oja. *"M-Estimation with Incomplete and Dependent Multivariate Data".*
09:40 am – 10:00 am (page 362, abstract 395)

Daniel Vogel. *"Robust estimators and tests for Gaussian graphical models".*
10:00 am – 10:20 am (page 362, abstract 396)

Miriam Jaser, **Stephan Haug**, Aleksey Min. *"A simple non-parametric goodness-of-ft test for elliptical copulas".*
10:20 am – 10:40 am (page 363, abstract 397)

Statistics of High Dimensional Data IV

Chair: Erik Thorsén, Room: D 209

Taras Bodnar, Arjun K. Gupta, Nestor Parolya, **Erik Thorsén**. *"Sampling Distributions of Optimal Portfolio Weights and Characteristics"*.
09:20 am – 09:40 am (page 364, abstract 398)

Louis Dijkstra*, Moritz Hanke*, Ronja Foraita, Iris Pigeot, Vanessa Didelez. *"Best Subset Selection: The Holy Grail for Variable Selection?"*.
09:40 am – 10:00 am (page 365, abstract 399)

Humera Razzak, Christian Heumann. *"Nonparametric Bayesian dependent Chained Equation Multiple Imputation for Incomplete Surveys"*.
10:00 am – 10:20 am (page 366, abstract 400)

Martin Spindler, Xi Chen, Victor Chernozhukov, Ye Luo. *"Adaptive Discrete Smoothing for (High-Dimensional and Nonlinear) Panel Data"*.
10:20 am – 10:40 am (page 366, abstract 401)

Time Series Analysis VI (Time Series Resampling)

Chair: Matei Demetrescu, Room: A 240

A.M. Robert Taylor. *"Detecting Regimes of Predictability in the U.S. Equity Premium"*.
09:00 am – 09:40 am (page 367, abstract 402)

Marco Meyer, Efstathios Paparoditis, Jens-Peter Kreiß. *"Extending the validity of frequency domain bootstrap methods to general stationary processes"*.
09:40 am – 10:00 am (page 367, abstract 403)

Carina Beering, Carsten Jentsch, Anne Leucht, Marco Meyer. *"Bootstrapping characteristic functions under local stationarity"*.
10:00 am – 10:20 am (page 368, abstract 404)

Rainer Buschmeier. *"The impact of selecting the truncation indices on the order estimation of subspace methods—a simulation study with seasonally integrated processes."*.
10:20 am – 10:40 am (page 369, abstract 405)

ABSTRACTS

Plenary Sessions

1 Adaptivity of signal priors

Sara van de Geer
ETH Zürich, Switzerland

In recent years we have seen continuing novel developments concerning sparsity inducing estimation methodology with the Lasso as prime example.
We will extend this to total variation type penalties on graphs. The estimator is there

$$\hat{f} := \arg\min_{f \in \boldsymbol{R}^n} \left\{ \|Y - f\|_2^2 + 2\lambda \|Df\|_1 \right\}$$

where $Y \in \boldsymbol{R}^n$ is a vector of observations with mean vector f^0 and $D \in \boldsymbol{R}^{m \times n}$ is a given matrix, with possibly $m \geq n$ and typically with $\text{rank}(D) < n$. Moreover, $\lambda > 0$ is a given tuning parameter.
We will provide high probability bounds for $\|\hat{f} - f^0\|_2$ that adapt to the sparsity of Df^0. This is joint work with Francesco Ortelli.

2 Measuring Expected Years of Life Lost

Per Kragh Andersen
University of Copenhagen, Denmark

We will study two classes of models for estimating the expected number of life years lost for a given group of subjects. One is phrased in the framework of a (Markov or non-Markov) illness-death model in combination with a population life-table. It is compared to a number of previously suggested measures of life years lost among patients with a given disease. The other is based on the competing risks model and we show that the cause j cumulative incidence function integrated from 0 to t has a natural interpretation as the expected number of life years lost due to cause j before time t. The large sample properties of a non-parametric estimator are outlined, and it is discussed how the expected number of years lost may be related to explanatory variables in a regression model based on pseudo-observations. The methods are illustrated using data on Danish patients with cancer or psychiatric disorder.

3 Statistical Machine Learning: Dynamical, Economic and Stochastic Perspectives

Michael Jordan

University of California, Berkeley, USA

Statistical Machine Learning: Dynamical, Economic and Stochastic Perspectives
While there has been significant progress in the theory and practice in statistical machine learning in recent years, many fundamental challenges remain. Some are mathematical in nature, such as the challenges associated with optimization and sampling in high-dimensional spaces. Some are statistical in nature, including the challenges associated with multiple decision-making. Others are economic in nature, including the need to price services and provide incentives in learning-based two-way markets. I will present recent progress on each of these fronts.

4 Modern Computing Implementing Classical, But Heretofore Unnurtured Statistical Ideas

Donald B. Rubin

Tsinghua University, Beijing, China
Temple University, Philadelphia, USA

Many seeds of important statistical ideas were never nurtured because of the absence of modern computing. These range from obvious ideas that rely on speed of computation to implement to more recondite ones that rely on visual displays to trigger aspects of human cognition used to make decisions. A few idiosyncratic examples will be presented, including re-randomization in the design of studies, enhanced tipping point displays to reveal sensitivity of inferential conclusions to assumptions, and conditional calibration plots used to select sage statistical procedures.

Advanced Regression Modeling I (Model Selection and Model Choice in Flexible Regression Models)

5 A ridge to homogeneity

Stanislav Anatolyev
CERGE-EI, Czech Republic

In some heavily parameterized models, one may benefit from shrinking a subset of parameters towards a common target. We consider L2 shrinkage towards an equal parameter value that balances between unrestricted estimation (i.e., allowing full heterogeneity) and estimation under equality restriction (i.e. imposing full homogeneity). The penalty parameter of such ridge regression estimator is tuned using one-leave-out cross-validation. The reduction in predictive mean squared error tends to increase with the dimensionality of the parameter set. We illustrate the benefit of such shrinkage with a few stylized examples. We also work out, both theoretically and empirically, a heterogenous linear panel data setup and compare several estimators and corresponding confidence intervals.

6 Scalable Bayesian regression in high dimensions with multiple data sources

Konstantinos Perrakis, Sach Mukherjee
German Center for Neurodegenerative Diseases, Germany

Applications of high-dimensional regression often involve multiple sources or types of covariates. We propose methodology for this setting, emphasizing the "wide dataregime with large total dimensionality p and sample size $n \ll p$. We focus on flexible ridge-type priors with shrinkage levels that are specific to each data type or source and that are set automatically by empirical Bayes. All estimation, including setting of shrinkage levels, is formulated mainly in terms of inner product matrices of size $n \times n$. This renders computation efficient in the wide data regime and allows scaling to problems with millions of features. Furthermore, the proposed procedures are free of user-set tuning parameters. We show how sparsity can be achieved by post-processing of the Bayesian output via constrained minimization of a certain Kullback-Leibler divergence. This yields sparse solutions with adaptive, source-specific shrinkage, including a closed-form variant that scales to very large p. We present empirical results from a simulation study based on real data and a case study in Alzheimer's disease involving millions of features and multiple data sources.

7 Optimized variable selection via repeated data splitting

Marinela Capanu, Colin Begg, Mithat Gonen

Memorial Sloan Kettering Cancer Center, United States of America

We introduce a new variable selection procedure that repeatedly splits the data into two sets, one for estimation and one for validation, to obtain an empirically optimized threshold which is then used to screen for variables to include in the final regression model. In an extensive simulation study we show that the proposed variable selection technique enjoys superior performance compared to candidate methods, being amongst those with the lowest inclusion of noisy predictors while having the highest power to detect the correct model and being unaffected by correlations among the predictors. We illustrate the methods by applying them to a cohort of patients undergoing hepatectomy at our institution.

8 The Power of Unbiased Recursive Partitioning: A Unifying View of CTree, MOB, and GUIDE

Lisa Schlosser[1], Torsten Hothorn[2], Achim Zeileis[1]

[1]Universität Innsbruck, Austria
[2]Universität Zürich, Switzerland

A core step of every algorithm for learning regression trees is the decision if, how, and where to split the underlying data - or, in other words, the selection of the best splitting variable from the available covariates and the corresponding split point. Early tree algorithms (e.g., AID, CART) employ greedy search strategies that directly compare all possible split points in all available covariates. However, subsequent research showed that this is biased towards selection of covariates with more potential split points. Therefore, unbiased recursive partitioning algorithms have been suggested (e.g., QUEST, GUIDE, CTree, MOB) that first select the covariate based on statistical inference using p-values that are appropriately adjusted for the possible split points. In a second step a split point optimizing some objective function is selected in the chosen split variable. However, different unbiased tree algorithms employ different inference frameworks for computing these p-values and their relative advantages or disadvantages are not well understood, yet.
Therefore, three different approaches are considered here and embedded into a common modeling framework with special emphasis to linear model trees: classical categorical association tests (GUIDE), conditional inference (CTree), parameter instability tests (MOB). It is assessed how different building blocks affect the power of the tree algorithms to select the appropriate covariates for splitting: residuals vs. full model scores, binarization of residuals/scores at zero, binning of covariates, conditional vs. unconditional approximations of the null distribution.

9 Should positivity imply a multiplicative model? Introducing the Softplus function as an alternative to the common log link

Paul Wiemann, Thomas Kneib
University of Göttingen, Germany

The response function is an integral component of every generalized linear model. Its choice can have a great influence on estimation dynamics, interpretation and interrelation of the explanatory variables. The same is true in distributional regression models, in which all distributional parameters are linked to semiparametric additive predictors. The choice of the link function often depends only on the domain of the distribution parameter and is not further scrutinized, i.e. if the parameter is restricted to the positive domain the exponential function is usually used, if the parameter lives in the unit interval, the expit function is selected by default. A consequence of this is a multiplicative model of effects when the parameter domain is positive, usually without substantive reasoning. Statistical modeling should reflect the data generating process, but aspects such as interpretability or numerical stability may be additional concerns. We propose to consider the softplus function $\text{softplus}(x, b) = \log(\exp(bx) + 1)/b > 0$ as a response function. It is a smooth approximation of arbitrary accuracy of the ramp function $\max(0, x)$. On the positive domain it corresponds for the most part to the identity function, which implies in this domain an additive model of effects and allows for their straightforward interpretation. Furthermore, the calculation of the softplus function does not show numerical difficulties. We demonstrate the applicability of the softplus function as a response function in the context of Bayesian distribution regression using simulations and real data. In addition, we compare the performances of the softplus response function and the log link w.r.t. predictive performance.

Computational Statistics and Statistical Software I (Temporal Data)

10 Estimating multiple changes in the mean using moving sum statistics

Alexander Meier[1], Haeran Cho[3], **Claudia Kirch**[1,2]

[1]Otto-von-Guericke University Magdeburg, Germany
[2]Center for Behavioral Brain Sciences (CBBS), Magdeburg, Germany
[3]University of Bristol, UK

Extracting a piecewise constant signal plays an important role in many fields of science, economy, technology and medicine. Change point analysis provides the mathematical tools for this as well as related problems. In this talk we discuss how to use moving sum statistics in order to solve this problem, where we use an information criterion based approach to merge information obtained from a variety of bandwidths. We discuss the implementation of this new approach in the R-package `mosum` and shed some light on possible future extensions.

11 How to model extreme events that occur in clusters?

Katharina Hees, Roland Fried

TU Dortmund, Germany

In the analysis of extreme events, one is often concerned with observations above a high threshold, called exceedances. Classical extreme value theory is based on the assumption that events occur regularly. It is a wellknown result that in this case the times between the exceedances have an exponential distribution for high thresholds. However, there are also applications, for example earthquakes, storms and solar fares, where the interexceedance times are not exponentially distributed. More precisely, the extreme events occur in clusters: there are periods with high activity and periods where no extreme events happen. One possible model for such a behavior is to assume shortterm dependence. The classical approach to analyze extremes of shortrange dependent data is to first identify the clusters, a procedure that is known as declustering, and then to extract the maximum from each cluster. The clusters of exceedances above a high threshold are independent of each other and the times between such clusters are again exponentially distributed. Hence, one can then proceed by applying classical extreme value tools for iid data to the cluster maxima. Another model that leads to clustering of extreme events is to assume heavy tailed waiting times between the events. The return times of the extremes, or more precisely of the exceedances above a high threshold, are then no longer exponentially distributed, more precisely have a Mittag-Leffler distribution, and this results in a serial clustering of the extreme events as well. The two above described models are completely different: while in the first model, we have clustering due to dependencies, the second model leads to a serial clustering due to heavy tailed interarrival times. Nevertheless, by just looking at a time series with clustering behavior it is not obvious which model one should choose. In this talk, we will discuss how to choose the appropriate model for analyzing extremes of clustered data.

12 Hierarchical continuous time state space modelling with ctsem

Charles Driver

Max Planck Institute for Human Development, Germany

In this talk I will discuss ctsem, an R package for estimating continuous time state space models over multiple subjects. Continuous time systems are of interest in psychology and related fields, in part because they offer a direct specification of a causal system. While the field has seen substantial development in the single subject case, less emphasis has been seen in the multiple subjects scenario. ctsem allows relatively straightforward specification of linear and nonlinear dynamic systems with exogenous inputs, allows for the inclusion of priors, as well as random effects and covariates across subjects. ctsem uses the Stan software as a backend for model construction and estimation purposes, and offers estimation options including optimization, importance sampling, and Hamiltonian Monte Carlo. For the most part, latent states are integrated out using variants of the Kalman filter, and parameters for individual subjects' systems may be either directly sampled or in some circumstances integrated out via extension of the state matrices. I will present some empirical examples using data on wellbeing from the German Socio-Economic Panel survey to demonstrate various aspects of the available output.

13 Confidence intervals for average sequential attributable fraction – A simulation study

Carolin Malsch[1], Stefan Störk[2,3], Peter U. Heuschmann[1,2,4]

[1]Institute of Clinical Epidemiology and Biometry, University of Würzburg, Germany
[2]Comprehensive Heart Failure Center, University Hospital of Würzburg, Germany
[3]Department of Internal Medicine I, Cardiology, University Hospital Würzburg, Germany
[4]Clinical Trial Centre Würzburg, University Hospital Würzburg, Germany

Attributable fractions (AF) estimate the amount of an outcome associated with a risk factor. The average sequential AF is frequently used to estimate the AF in a multifactorial situation. Multiple exposures are considered simultaneously by calculating the sequential AF in every permutation of risk factor removal from the population and subsequent averaging of results. Another description of the very same approach is possible using partial AF: the population is decomposed disjointly with regard to these factors, i.e. the resulting strata are associated with the appearance of exactly one subset of risk factors, respectively. Partial AF are calculated for each stratum individually. AF for each risk factor is subsequently calculated from strata where the risk factor occurs alone or in combination with other risk factors. In strata with multiple risk factors, the partial AF are divided into equal parts and assigned to the risk factors of interest. This approach ensures additivity, i.e. average AF of individual risk factors add up to the overall AF observed in the sample.
Confidence intervals for average AF in the univariate situation have been studied in the past. The asymptotic confidence interval of the average AF obtained via the delta method cannot be recommended due to underestimation of the standard error, as was shown by Lehnert-Batar et al. (2006) by use of simulation studies. Confidence intervals for average AF in multifactorial situations have not been studied so far. The aim of this study is to compare Monte Carlo approaches for the calculation of confidence intervals for average AF in the multifactorial situation including model-based estimation, bootstrap and jackknife. For this purpose, confidence intervals are investigated and characterized using computer simulation and a real life example is provided for data from a cohort study on risk factors for recurrent cardiovascular event in patients with established coronary heart disease. The results of the study will improve the understanding on best approaches to estimate confidence intervals for average AF in multifactorial situations.

Design of Experiments and Clinical Trials I (Optimal Design I)

14 Design of experiments for fitting flexible curves

Luzia Trinca
Unesp, Brazil

Low order polynomials are widely used to approximate the relation between a continuous response and quantitative factors from experimental data. However limitations are frequent due to lack-of-fit. Increasing the order of the model may solve the lack-of-fit problem but introduces implausible relations and lack of interpretation. Quite often we are presented to data showing asymmetric curves or surfaces, sometimes with asymptotes but without any mechanistic model as justification. Royston and Altman (1994), inspired in the Box-Tidwell transformation family, proposed fractional polynomial models as flexible alternatives to model observational data empirically. The fractional polynomial of low order is also quite interesting to model experimental data mainly for response surface studies. However usual experiments with few equally spaced levels do not have enough information to estimate the extra parameters needed. As defined by Royston and co-authors, for each degree and factor, there is a power parameter to estimate, besides the regression coefficients.
In this work we define a more parsimonious second order fractional polynomial (as function of just one power parameter) and obtain optimum designs for fitting such type of models. As the model is non-linear, prior information on the parameters is required to construct the design. We then investigate the behaviour and sensitivity of optimum designs for several models.

15 A practical approach to designing partial-profile choice experiments for estimating main effects and interactions

Heiko Großmann
Otto-von-Guericke-Universität Magdeburg, Germany

A method is presented which facilitates the practical construction of designs for stated choice experiments in which the choice sets are pairs of partial profiles and where, for a potentially large number of two-level attributes, the main effects and two-factor interactions are to be estimated. Although partly heuristic, the approach has a sound theoretical basis and can be used to generate utility-neutral designs for the multinomial logit model which possess a high statistical efficiency. Applying the method neither requires expert knowledge of design theory nor specialized software and is illustrated with an example.

16 Designs for Second-Order Interactions in Paired Comparison Experiments of Full and Partial Profiles

Eric Nyarko, Rainer Schwabe
Otto-von-Guericke-University Magdeburg, Germany

Paired comparison experiments have received considerable attention over the last few years in many fields of application like psychology, transport economics, health economics and marketing for learning consumer preferences towards new products or services. Typically, in such experiments competing options (alternatives) which are described by the levels of various attributes known as full-profiles or a predefined
subset of the attributes known as partial-profiles to mitigate cognitive burden are presented to respondents to evaluate and indicate their preferences.
By the concept of utility maximization, the choice behavior can then be described by a logit or probit model. In this setting the aim of optimal design is to identify the best choice sets by manipulating attribute levels to be presented together. While much work has been done on optimal design for additive model in the literature, there are few results taking into consideration interactions between the levels of the attributes.
For this situation, and under the indifference assumption of equal choice probabilities where the information matrix of the multinomial logit model for pairs and the corresponding linear paired comparison model coincides, we derive optimal designs in the presence of second-order interactions, and where the alternatives in the choice sets are described by either full or partial profiles.

17 Optimal item calibration designs for computerized achievement tests

Frank Miller, Mahmood Ul-Hassan
Stockholm University, Sweden

The importance of large-scale achievement tests like national tests in school, eligibility tests for university, or international assessments for evaluation of students is increasing. Pretesting of questions which is called "item calibration" for such tests is done to determine characteristic properties of the questions. Usually, the calibration part of a test should be small and should have a negligible burden on examinees.
We want to choose examinees which are specifically suitable for calibration of a specific question based on their estimated ability. Statistically, we use item-response-models for the probability to answer a question correctly and want to improve precision of parameter-estimators in the models. This leads us to a restricted optimal design problem where the restriction is given by the available population of examinees. We present an equivalence theorem which characterizes conditions for an optimal design in this situation. After outlining an algorithm for computation of optimal calibration designs, we illustrate the method with some examples. Computed design efficiencies for the examples show us that the optimal designs are considerably better than the standard way of calibrating.

18 On optimal designs for multi-factor two-level models on a design region restricted by the number of active factors

Fritjof Freise[1], Rainer Schwabe[2]

[1]Technische Universität Dortmund, Germany
[2]Otto von Guericke University Magdeburg, Germany

There has been a vast amount of research over the last decades on optimal design for analysis of variance models with binary predictors when all factors can be adjusted independently. However, in some applications natural constraints arise on the number of factors allowed to be active. This can be accomplished by imposing restrictions on the full factorial design region by bounds on the number of factors set to a high and those set to a low level. For this situations optimal designs are presented for models containing only main effects as well as for models with first-order interactions. In particular, for the main effects model moderately restrictive bounds allow for designs which are optimal even for the unrestricted full factorial design region and are, hence, as efficient as the full factorial design. However, for narrow or very asymmetric margins the settings of the optimal design are located on the upper and lower boundary, where equality in the restrictions is attained. Analogous results are obtained in the case of first-order interactions under symmetric constraints. There, for narrow bounds, additional central design points are required, for which half of the factors are active.

Empirical Economics and Applied Econometrics I

19 Generic Conditions for Forecast Dominance

Fabian Krüger[1], Johanna F. Ziegel[2]
[1]Heidelberg University, Germany
[2]University of Bern, Switzerland

Recent studies have analyzed whether one forecast method dominates another under a class of consistent scoring functions. While the existing literature focuses on empirical tests of forecast dominance, little is known about the theoretical conditions under which one forecast dominates another. To address this question, we first derive a new characterization of dominance among forecasts of the mean functional. We then present various scenarios under which dominance occurs. Unlike existing results, our results allow for the case that the forecasts' underlying information sets are not nested, and allow for uncalibrated forecasts that suffer, e.g., from model misspecification or parameter estimation error.

20 Maximum likelihood analysis of high-dimensional reduced-rank regressions

C. S. Bos[1], S. J. Koopman[1], **M. Massmann**[2]
[1]Vrije Universiteit Amsterdam
[2]WHU – Otto Beisheim School of Management, Germany

This paper considers the simultaneous estimation by exact maximum likelihood of high-dimensional reduced-rank regressions with vector autoregressive error terms. In particular, a model is entertained that has a few variables of interest, i.e. three to five, and a large number of explanatory variables, viz. one to two hundred. A novel linear state-space representation of the model is suggested, decomposing the set of parameters into one large subset that is put into the state while the remaining few paramaters enter the system matrices. This setup not only accomodates a large number of covariates but also allows to employ standard and computationally efficient likelihood-based inference. For instance, likelihood ratio tests can be used to examine the rank of the regression coefficient and thus to construct 'factors', or to impose restrictions on the weights given to covariates in the constructions of the factors. The model can hence be seen as using approximate canonical correlations for purposes of dimension reduction. To estimate the hyperparameters numerically, the algorithm by Morf, Vieira & Kailath (1978, Annals of Statistics) is implemented. In an empirical illustration, the FRED-MD database is used to estimate a joint model of industrial production, employment as well as the CPI, and to produce pseudo out-of-sample forecast. For the most part, the forecasts outperform Stock & Watson type diffusion index forecasts.

21 Selecting the Number of Factors in Approximate Factor Models using Group Variable regularization

Maurizio Daniele

Universität Konstanz, Germany

We propose a novel method for the estimation of the correct number of factors in approximate factor models. The model is based on a penalized maximum likelihood approach incorporating an adaptive hierarchical lasso penalty function that enables setting entire columns of the factor loadings matrix to zero, which corresponds to removing irrelevant factors. Additionally, it is capable to estimate weak factors by allowing for sparsity in the non-zero columns. We show that the proposed estimator consistently estimates the correct number of factors.
Simulation experiments reveal superior selection accuracies for our method in finite samples over existing approaches, especially for data with cross-sectional and serial correlations.
In an empirical application on a large macroeconomic dataset, we show that the average mean squared forecast errors of an approximate factor model are lower if the number of included factors is selected by our method.

22 Detecting periods of excessive credit in the EU - A structural counterfactual approach

Magnus Sass

FU Berlin, Germany

Periods of excessive credit growth are often followed by financial crises. A reliable and timely measure of cyclical systemic risk emanating from excessive developments of credit, is therefore required in order to activate countercyclical macroprudential instruments appropriately. The Basel III HP-filtered credit-to-GDP gap, however, has been questioned on methodological and economic ground recently. This paper, therefore, proposes an alternative measure based on economic fundamentals. As a structural counterfactual gap, the paper extracts the difference between the observed level and a "fundamentals-justified" level of credit. Fundamental justification is ensured by means of a counterfactual exercise, where only fundamental drivers of credit are allowed to materialise as estimated from a small monetary SVAR. Fundamentally driving factors are assumed to be shocks to aggregate supply and aggregate demand as commonly acknowledged in the literature. In the empirical application the structural counterfactual gap is constructed for 13 EU countries from 1975 to 2017. The paper provides evidence that the proposed measure outperforms the Basel III credit-to-GDP gap in an in-sample early warning exercise, while performing equally well as the standard measure within real time settings. Besides the very good early warning properties, the structural counterfactual gap contributes to the literature by providing an economic interpretation to the developments within credit.

23 Identification of structural shocks via common fractional components

Tobias Hartl[1,2], **Rolf Tschernig**[1], Enzo Weber[1,2]

[1]Universität Regensburg, Germany

[2]Institut für Arbeitsmarkt- und Berufsforschung, Nürnberg

We derive a structural factor model framework that models fractional cointegration relations via common fractional components, exhibits an error-correction representation and permits an economic interpretation. In contrast to classical SVARs, our model allows for multicointegration relations, different integration orders and mixed frequencies in the data. By taking into account the long-run dynamic characteristics via the fractional factors, our framework generalizes dynamic factor models that typically require stationary data. In addition, identifying restrictions for structural shocks of different persistence emerge from the fractional factors. For an underlying macroeconomic dataset for the US we study the propagation of structural shocks using impulse responses and historical decompositions based on the structural fractional components model.

Preclinical and Pharmaceutical Statistics I (Clinical Drug Development)

24 Statistical issues in drug development and the role of statisticians in regulatory agencies

Benjamin Hofner

Paul-Ehrlich-Institut (PEI), Germany

Drug development is a difficult, time-consuming and expensive task with many challenges and pitfalls. One starts with early preclinical trials to assess toxicity, pharmacokinetics and first signs of efficacy in non-human subjects. After a successful preclinical development program, pharmaceutical companies or research consortia conduct a series of clinical trials in humans from Phase I (safety, dose finding) to Phase III (confirmatory proof of efficacy). These data are submitted to regulatory agencies (e.g. the EMA or national agencies in Europe) to be granted market access. The talk will show that drug development is also a challenging task for regulators who need to keep pace with the pharmaceutical industry in order to be able to talk to companies at eye level and to appropriately and timely assess submitted documents.

To set the current practice into perspective, this talk will begin with an historical overview on drug development and drug regulation with a special focus on statistics: With randomized controlled clinical trials evolving since the late 1940s many of the current standards are quite modern. Only since 1976, a proof of efficacy is needed to gain market access in Germany.

The different roles of the European Medicines Agency (EMA) and national agencies (here the Paul-Ehrlich-Institut; PEI), will be illustrated and regulatory procedures from clinical trial applications, over scientific advice to marketing authorization applications will be briefly explained. The special role of statisticians in the regulatory system will be of particular interest.

The talk will end with a brief overview of current hot topics in regulatory statistics.

25 Adjusting for selection bias in assessing treatment effect estimates from multiple subgroups

Ekkehard Glimm

Novartis Pharma, Switzerland

In this talk, we will review several methods for adjusting treatment effect estimates in clinical trials where it is suspected that the effects of one or more treatments differ in several subpopulations. In such situations, the most extreme subpopulation estimates usually get more attention than the other results. Few clinical teams resist the temptation to read too much into these results.
The talk focusses on the construction of simultaneous confidence intervals to quantify the uncertainty of the treatment effects more realistically than usual marginal confidence intervals. A method for adjusting the point estimates is suggested as well. These methods are compared with shrinkage estimates and credibility intervals arising from Bayesian hierarchical models. Important features of the two approaches are compared in a typical application in clinical trials. Finally, we present simulations to illustrate how the presented approaches behave across a range of scenarios that may occur in practice. The focus is on situations with relatively few subpopulations (roughly 4 to 20).

26 An efficient phase II/III development program utilizing information on short-term response and long-term survival

Heiko Götte[1], Junyuan Xiong[2], Marietta Kirchner[3], Meinhard Kieser[3]

[1]Merck KGaA, Darmstadt, Germany
[2]EMD Serono, Billerica, MA, USA
[3]Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

In the highly competitive oncology area, efficiency in drug development is critical. In phase II, there is a trade-off between reduced investment and high confidence in the potential of the drug before going into phase III. Reduced investment requires use of short-term endpoints. The probability of success for a future phase III trial based on long-term endpoints is a measure of confidence. Both aspects can be investigated under the framework of the probability of success using short-term endpoints data. We present a Bayesian and frequentist mixed approach to utilize information from a multi-categorical short-term endpoint (response status) and a long-term endpoint (overall survival) to determine the probability of success. The Bayesian approach is used to obtain predicted overall survival of censored subjects while the analyses for estimating the treatment effects is based on frequentist methods.
The Bayesian approach requires prior information about the association between response and overall survival. When developing a new drug, there is limited reliable prior information. We follow the concept of a "phase 2+" design where after a go-to phase-III-decision, further follow-up data from phase II is collected and used to make interim decisions on phase III. The phase III will be stopped early when updated phase II and partial phase III data lead to a low probability of success. A simulation study is used to determine optimal combinations of decision boundaries and time points of analysis.

27 Quantification of prior impact in terms of prior effective historical and current sample size

Manuel Wiesenfarth, Silvia Calderazzo

German Cancer Research Center, Germany

Bayesian methods can be advantageous due to their capability of borrowing historical information through elicitation of prior distributions. To quantify and communicate prior informativeness, the concept of prior effective sample sizes (ESS) has proven to be convenient, equating information provided by a prior to a sample size. As prior information can - practically or hypothetically - arise from a certain number of historical samples, the common approach equates the prior ESS to the number of samples in such historical data set. This measure is however independent from the newly observed data, and thus would not capture an actual "loss of information" induced by the prior in case of prior-data conflict. We build on Reimherr et al (2104) to relate prior information to a number of (virtual) samples from the current data model and introduce the prior effective current sample size (ECSS), tailored to the application in a Bayesian clinical trial design setting. The approach is applied to an adaptive two-arm trial design which has been the focus of several recent publications. Thereby, the number of control patients recruited is adjusted depending on the prior effective sample size at an interim analysis. In this setting, we argue that the ECSS is the appropriate data-dependent measure, as the final aim is to save current patients (as opposed to historical patients used to elicit a prior) from recruitment. Special emphasis is put on robust mixture priors, power priors and hierarchical models. We argue that the ECSS can help overcoming lack of consensus in the assessment of a data-dependent ESS for mixture priors in particular, and that it can provide further insights into the impact of priors in general. An efficient R implementation is provided.
Reimherr, M., Meng, X.-L., and Nicolae, D. L. (2014). Being an informed bayesian: Assessing prior informativeness and prior likelihood conflict. arXiv preprint arXiv:1406.5958

Small Area Analysis and Spatial Statistics I

28 Estimating socio-demographic indicators using mobile phone data with applications in Germany and Senegal

Timo Schmid[1], Fabian Bruckschen[1], Sandra Hadam[2], Nicola Salvati[3], Till Zbiranski[1]
[1]Freie Universität Berlin, Germany
[2]Statistisches Bundesamt, Germany
[3]University of Pisa, Italy

Modern systems of official statistics require the accurate and timely estimation of socio-demo-graphic indicators for disaggregated geographical regions. Traditional data collection methods such as censuses or household surveys impose great financial and organizational burdens for National Statistical Institutes. The rise of new information and communication technologies offers promising sources to mitigate these shortcomings. In this paper, we propose a unified approach for Governmental Institutions based on small area estimation that allows for the estimation of socio-demographic indicators by using mobile phone data.
In particular, the methodology is applied to mobile phone data from a) Germany for estimating unemployment rates at spatially disaggregated levels (functional urban areas) and b) Senegal for deriving sub-national estimates of the share of illiterates disaggregated by gender. The estimates are used to identify hot spots of illiterates/ unemployed persons with a need for additional infrastructure or policy adjustments.
From a methodological point of view, the presented paper uses area-level small area models in combination with covariates from alternative data sources. The resulting estimates are benchmarked such that the aggregated small area estimates produce the official national estimate for the country. We also apply transformation to restrict the indicator of interest, for instance the literacy rate, to particular intervals when necessary.
Although the paper focuses on literacy and unemployment as particular socio-demographic indicators, the proposed approach is applicable to indicators from national statistics in general.

29 Estimating proportions of multidimensional poverty in small areas

Sören Pannier, Timo Schmid
Freie Universität Berlin, Germany

The United Nations have declared the eradication of extreme poverty and hunger their first millennium development goal. To achieve this, detailed knowledge about the spacial distribution and type of poverty is needed. A possible way to gain knowledge about the spacial distribution is to apply small area estimation (SAE) methods. These models typically need survey and census data with common covariates. To investigate the type of poverty in the target area it is necessary to use multidimensional poverty measures. One approach for estimating such multidimensional deprivations is to consider them as a multinomial outcome.
Most SAE methods work by first estimating a statistical model on survey data in which the target variable is known and obtaining predictions by making use of more numerous census data. This process is called borrowing strength and is useful whenever the target areas' sample sizes become too small to obtain valid estimates by direct estimation methods. In this procedure the predictive power of the underlying model is crucial for succeeding in obtaining valid estimates, while its interpretability only plays a subordinate role. Methods originating from machine learning (ML) often combine these two characteristics. They are considered black box methods with high predictive performances.
In this work we use survey unit level data from Mexico and focus on multidimensional poverty, with the two dimensions, economic welfare and social deprivation. Our target variable takes five unordered categories of poverty. The target variable is only present in the survey; hence we conduct a design-based simulation study by treating the survey as a pseudo population and draw multiple samples from it. A typical objective is to estimate prevalence of types of poverty in each area. To do this, we use tree-based methods to predict the probability of an individual belonging to a certain class of poverty. In a second step these individual predictions are aggregated to the target area level and the methods performances are compared. Finally, a mean squared error estimation is discussed.

30 Small area estimation in forest inventories: Overview of methods and challenges in practical applications

Anne-Sophie Stelzer

Forest Research Institute Baden-Württemberg (FVA), Germany

Forest Inventories (FI) like the German National Forest Inventory provide a foundation for gaining knowledge about the condition and structure of forests at national and regional levels. These surveys are based on the terrestrial data collection at permanent sample plots in a systematic grid. Using this data, estimates of forest attributes can be derived by means of simple-random-sampling (SRS) estimators, which are generally unbiased and do not require a high computational effort.
Particular challenges arise, when the focus moves to smaller sub-populations like municipalities or even single forest stands. Because of the design featuring permanent sample plots, these small sub-populations often entail only few (or even no) observations. As a consequence, estimation of quantities of interest becomes imprecise at best, or may even not be possible at all. In such settings, small area estimation (SAE) techniques may be used to improve the precision of an estimate. However, this comes at the cost of potentially introducing bias and being computationally more expensive for the variances. In the context of forest inventories, SAE methods build on auxiliary data that is correlated with the forest attribute of interest and that is available for the entire area under investigation. Remote sensing data, for instance delivered by aerial photographs from image flights, can provide such area-covering information, for example in form of the digital surface model.
Beyond the description of the SRS estimator, the ideas and characteristics of model-assisted and model-based estimators are introduced. All three types of methodology are illustrated by estimating standing stock volume with data from an enterprise FI in Baden-Wuerttemberg and associated challenges are outlined.

31 Generative Adversarial Imputation Nets for Small Area Estimation

Marcel Neunhoeffer, Richard Traunmüller
University of Mannheim, Germany

We introduce Generative Adversarial Imputation Nets (GAINs) as a new flexible method for Small Area Estimation (SAE). Producing reliable estimates for small geographic units or socio-demographic cells from sparse survey data is a common problem in many applied research problems. We conceptualize SAE as a missing data problem. A small survey sample is treated as observed data and combined with large scale, partially missing census data using multiple imputation. The imputed census data is then aggregated to the desired small regional unit or socio-demographic cells. We apply a flexible, state of the art imputation algorithm: GAINs which are an extension of Generative Adversarial Nets (GANs). In this innovative neural network architecture, a generator function produces simulated samples that mimic the real data. At the same time, a discriminator function learns to distinguish between real and simulated data. A Generative Adversarial Net is successful in producing simulated data if a discriminator is maximally uncertain about the origin of the data. This way, GANs can sample from arbitrary complex target distributions. We provide Monte Carlo evidence that our approach produces valid Small Area Estimates without prior knowledge or assumptions about the functional forms or structural relationships in the data. In addition, we demonstrate the utility of our approach in two empirical applications: the estimation of political preferences in small geographic units in the US and mental health outcomes of small socio-demographic groups in Switzerland.

Statistical Literacy and Statistical Education I

32 Data science education at school level – Conceptions, examples and experience from a pilot project

Rolf Biehler
Universität Paderborn, Germany

The ProDaBi project (www.prodabi.de), a collaborative effort of statistics and computer science educators, is developing a year-long experimental curriculum for data science in secondary education (grade 12), based on a design-based research paradigm. The project was initiated and is supported by Deutsche Telekom Stiftung.
The course includes detective work with data, machine learning with data driven algorithms (decision trees and neural networks), and project work in data science in cooperation with partners from industry and the local government. Cultural and societal reflection of these data and technology uses will be done throughout the course. The project is based on Biehler et al. (2018a, 2018b) where relevant trends and fundamental ideas of data science relevant to secondary education are discussed.
We use Codap (codap.concord.org) as an easy-entry tool for exploratory data analysis and for constructing simple decision trees but then gradually move to Jupyter notebooks with Python as the central tool, where students can analyze and design (parts of) data science algorithms. Jupyter notebooks are also used as a teaching tool.
The presentation will include examples from the ongoing course with 20 students in the schoolyear 2018/2019, which will show the data sets and introductory problems we used and how we elementarized topics from data science for secondary students. Examples from students' work (including their project presentations) will be discussed as well as a revised version of fundamental ideas for data science based on the practical experiences gathered so far.
Biehler, R., Budde, L., Frischemeier, D., Heinemann, B., Podworny, S., Schulte, C., & Wassong, T. (Eds.). (2018a). Paderborn Symposium on Data Science Education at School Level 2017: The Collected Extended Abstracts. Paderborn: Universitätsbibliothek Paderborn. http://dx.doi.org/10.17619/UNIPB/1-374
Biehler, R., & Schulte, C. (2018b). Perspectives for an interdisciplinary data science curriculum at German secondary schools. In R. Biehler et al., 2018a, pp. 2 -14

33 Causal Modelling in Introductory Statistics?

Karsten Lübke[1], Matthias Gehrke[1], Jörg Horst[2], Sebastian Sauer[1]
[1]FOM University of Applied Sciences, Germany
[2]Bielefeld University of Applied Sciences, Germany

The Guidelines for Assessment and Instruction in Statistics Education (GAISE, 2016) recommend to Teach statistical thinking"by Teach[ing] statistics as an investigative process of problem-solving and decision-makingänd to "Give students experience with multivariate thinkingänd also Ïntegrate real data with a context and purpose". Similar ideas include the claim to focus on modelling (e.g. Stigler and Son, 2018) and to support students to think with data"(e.g. Pruim et al., 2017).
However, multivariate modelling can be mislead through the presence of confounding variables such as the well-known Simpson's or Berkson's Paradox. Given the fact of ubiquitous data in recent times and the fact that the field of applied statistics has changed a lot since Fisher's days, we think that today's students should learn to think even more thoroughly about the data generating process in order to draw conclusions from data.
The basic ideas of Causal Inference like e.g. Directed Acyclic Graphs, the difference between observing and manipulating data, and counterfactual evaluation may foster a deeper understanding of what can and - maybe even more important - what cannot be deduced by (observational) data analysis; in a similar vein, we will discuss the assumptions of such modelling. Moreover, knowledge of ideas of causal modelling may help to refrain from over-simplified conclusion based on "Big-Dataänalysis. We argue that even in introductory statistic courses we can and should move beyond (and give meaning to) the mantra Correlation does not imply Causation".
Through the Focus on conceptual understanding"(GAISE, 2016) by teaching techniques such as Simulation Based Inference (Bootstrapping, Permutation Test) and by deemphasizing more traditional (asymptotic or approximate and sometimes flawed) inference techniques, free space is added to the curriculum. Therefore, our proposed curriculum puts data and context even more in its centre. The notion to integrate causal modelling in introductory statistics is supported by many, e.g. Ridgway, 2016, Angrist and Pischke, 2017, Kaplan, 2018, or the ASA Causality in Statistics Education Award".
In this talk, we would like to discuss how Causal Inference can serve as means towards this end in introductory statistics courses. In addition, we will provide an instructional rationale and some preliminary evaluation data will be presented.
References:
- Angrist, J.D., Pischke, J.S.: Undergraduate econometrics instruction: Through our classes, darkly. Journal of Economic Perspectives 31(2), 125–144 (2017).
- GAISE College Report ASA Revision Committee: Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016 (2016).
- Kaplan, D.: Teaching stats for data science. The American Statistician 72(1), 89–96 (2018).
- Pruim, R., Kaplan, D.T., Horton, N.J.: The mosaic package: Helping students to 'think with data' using R. The R Journal 9(1), 77–102 (2017).
- Ridgway, J.: Implications of the data revolution for statistics education. International Statistical Review 84(3), 528–549 (2016).
- Stigler, J.W., Son, J.Y.: Modeling first: A modeling approach to teaching introductory statistics. In: M.A. Sorto, A. White, L. Guyot (eds.) Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (2018).

34 Flipped Classroom Implementation in Large Statistics Lectures

Constantin Weiser[1], Manuel Förster[1], Florian Heiss[2], Sigbert Klinke[3], Andreas Maur[1], Thorsten Schank[1], Kirsten Winkel[1]

[1]Johannes Gutenberg University Mainz, Germany
[2]Heinrich Heine University Duesseldorf, Germany
[3]Humboldt University Berlin, Germany

Considering the challenges of manifold misconceptions, statistics anxiety and low cognitive activation in large statistics lectures in higher education, we would like to present the design and materials of a flipped classroom which we implemented in the context of a large field study with a treatment-control-design at two universities. The materials we developed and used in the flipped classroom setting consist of video tutorials, live voting, interactive demonstrations, electronic quizzes & final exams. We begin with a short introduction on the theoretical and practical significance of flipped classroom teaching in general and, in our specific case, in large statistics lectures. We will then introduce our diverse material and explainhow they relate to each other within our conceptual framework. And last, we present preliminary results for the ongoing evaluation.

35 shinyExample - ein R Paket zur Unterstützung der Entwicklung einfacher Shiny-Apps

Sigbert Klinke

Humboldt-Universität zu Berlin, Germany

Im Rahmen des BMBF Projektes "Förderung statistischer Lehr- und Lernprozesse in Großveranstaltungen mittels eines Flipped-Classroom-Designs" (FLIPPS) werden die Lehrmaterialien für die Statistikgrundvorlesungen überarbeitet. Für unser Wiki und die Lehrmaterialien wurden eine Vielzahl von R Programmen geschrieben sowohl um Grafiken und Tabellen als auch R Output zu zeigen.
Mit der interaktiven und web-basierten Software shiny lassen sich in R interaktive Programme entwickeln, jedoch erfordert es etwas Aufwand. Mit der Shiny-App "examples" aus dem Paket "shinyExample" lassen sich Beispielprogramme mehrfach ausführen, z.B. zur Demonstration welche Konsequenzen das Ziehen einer neuen Stichprobe hat, das zentral für die induktive Statistik ist.
Aus den vorliegenden R-Programmen kann mit "makeShinyApp" eine einfache interaktive Shiny App erzeugt werden. Sowohl die Ersetzung von fünf Grafiken im Wiki zur Darstellung des Bravais-Pearson-Korrelation-Koeffizient als auch die Visualisierung der Scagnostics-Koeffizienten für einen multivariaten Datensatz mit einer interaktiven Shiny-App konnte in sehr kurzer Zeit durchgeführt werden. Verschiedene Werkzeuge, z.B. zur Reskalierung von Grafiken, unterstützen die Programmautoren bei der Entwicklung.
Weblinks: Wiki (https://wikis.hu-berlin.de/mmstat/),
shinyExample (https://github.com/sigbertklinke/shinyExample) und
examples-App (https://shinyapps.wiwi.hu-berlin.de/examples/).

36 Die Prognose des Studienerfolgs auf Basis individueller Studienveräufe im Fach Wirtschaftswissenschaft

Ulrich Rendtel, Sören Pannier
Freie Universität Berlin, Germany

Über das individuelle Studierverhalten im Zeitablauf eines Studiengangs ist vergleichsweise wenig bekannt. Zwar gibt es Musterpläne für die Organisation der Studiengänge. Inwieweit die Studierenden diesen Empfehlungen folgen, ist jedoch weitgehend unbekannt. Auch ist unbekannt, ob bestimmte vom Studienplan empfohlenen Modulsequenzen gleichmäßig abgearbeitet werden oder ob an bestimmten Stellen Verzögerungen oder Abweichungen auftreten. Auch der Umgang mit dem Nichtbestehen einer Modulprüfung ist unbekannt. Werden die Möglichkeiten einer zeitigen Nachprüfung konsequent genutzt oder führen nicht bestandene Prüfungen zu Verzögerungen im Studienablauf oder teilweisen Umorientierung im Studienverlauf?
Im Rahmen eines Pilotprojekts am FB Wirtschaftswissenschaft der FU wurden auf Basis anonymisierter Matrikelnummern individuelle Studienverläufe analysiert. Da die Prüfungsverläufe allein wenig über die Studienmotivation und den sozialen Hintergrund verraten, wurde eine gesamte Kohorte zu Anfang des zweiten Semesters nach ihren bisherigen Studienerfahrungen befragt. Zusätzlich wurde um die schriftliche Erlaubnis gebeten, die Notenverläufe mit den Ergebnissen der Befragung zu verknüpfen. Von besonderem Interesse sind Indikatoren, die einen erfolgreichen Studienabschluß innerhalb der ersten 8 Semester der BA-Studiums vorhersagen. Der Vortrag berichtet über erste empirische Ergebnisse dieser Langzeitstudie.

Advanced Regression Modeling II (Nonparametric Regression Beyond the Mean)

37 Goodbye moments, hello expectiles

Paul Eilers
Erasmus MC, The Netherlands

The method of moments was invented by Karl Pearson, at the end of the 19th century, as a procedure for estimating parameters of statistical distributions. The basic idea is simple: equate empirical and theoretical moments, where the moments are expected values of powers of y, the variable under study. An example is the estimation of mean and variance of a (normal) distribution from averages of the observations themselves and their squares. A disadvantage of moments is that averages of third and higher powers of the observations are extremely sensitive to outlying values. That can make the method of moments unreliable.
Here I explore expectiles as alternatives for moments. Expectiles, based on asymmetric least squares, have pleasant properties and are easy to compute. One parameter, the asymmetry, can be varied continuously between 0 and 1, allowing a set of many expectiles to be computed.
In many applications, conditional distributions are of prime interest; growth curves, of height or weight with age, provide a familiar example. Smooth expectile curves have been used to characterize growth data. The curves themselves already give a good impression of changes. Expectile matching makes it easy to fit parametric distributions afterwards, at any chosen age. The quality of the fit can be judged by comparing observed expectiles with theoretical ones, for a range of asymmetries, similar to the familiar Q-Q plot for quantiles. In contrast to the latter, expectiles are uniquely defined for continuous *and* for discrete distributions.
Several examples illustrate expectile matching, one of them being the worrying trends in earthquake frequencies in the North of the Netherlands, caused by the massive exploration of natural gas reserves.

38 Generalized Expectile Regression with Flexible Response Function for patient reported outcomes

Elmar Spiegel[1,2], Thomas Kneib[1], Petra von Gablenz[3], Inga Holube[3], Fabian Otto-Sobotka[4]

[1]University of Goettingen, Germany
[2]Helmholtz Zentrum München, Germany
[3]Jade University of Applied Sciences Oldenburg, Germany
[4]Carl von Ossietzky University Oldenburg, Germany

Instruments that comprise patient reported data are usually analyzed using summation scores. A score is typically viewed as a metric response variable. The values for the Speech, Spatial and Qualities of Hearing Scale in its German short form SSQ17, for example, are computed as the mean of 17 hearing related questions answered on 11-point Likert scales. Although the values come with a limited range starting at zero points up to a maximum of 10 points, these limits are seldomly modeled in the final data analysis. Often an ordinary mean regression is carried out to analyze the impact of covariates on the score. Such a linear model depends on several assumptions including homoscedasticity. Expectile regression is one alternative that allows for heteroscedasticity and also does not need the specification of the underlying distribution. This model class can be seen as a mixture of quantile regression and least squares regression. Similarly to quantile regression the whole conditional distribution of a response variable can be modeled, but expectile regression is based on the more accessible L2-norm like the ordinary linear model rather than the L1-norm used to define quantile regression.
As expectile regression is essentially estimated via iteratively weighted least squares, all smooth covariate structures of the linear model can easily be included. However, even if no distribution is assumed in expectile regression, the model is still constructed in such a way that the relationship between the response and the predictor is linear. If the true underlying relationship is nonlinear, biased estimates can occur. Furthermore, many outcomes of interest (like the SSQ17) are limited in the range of their values. Classical expectile regression should in theory stick to these constraints like ordinary mean regression. However, in practice we observe violations. Therefore, we propose to include a response function for the predictor similar to generalized linear models. However, including a fixed response function would imply an assumption on the shape of the underlying distribution function. We want to avoid this assumption in expectile regression. Therefore we propose to estimate the response function and the covariate effects jointly. We design the response function as a monotonically increasing P-spline that may also contain constraints on the target set. These constraints on the response function are integrated as linear constraints on the coefficients of B-splines. During the estimation we iterate between estimating the covariate effects and the response function.
First, the validity of the point estimates is assessed in a simulation study. We then model the complete conditional distribution of the SSQ17 using a population-based sample of $n = 1708$ questionnaires and measurements from the northwest of Germany. The estimated response function is limited to the range of the score which increases the interpretability of some nonlinear covariate effects dramatically, as they do not cross the maximum of 10 points anymore and therefore return a valid distributional regression estimate.

39 Bayesian Conditional Transformation Models

Manuel Carlan[1], Thomas Kneib[1], Nadja Klein[2]

[1]University of Goettingen, Germany
[2]Humbolt University of Berlin

Recent developments in statistical regression methodology focus on regression models that are able to establish a relationship between higher moments of the response distribution and the explanatory variables. Conditional transformation models (CTMs), as introduced in Hothorn et al. (2014), go one step further and aim to infer the conditional distribution directly. They are based on the idea of applying a transformation function h to the response y in order to transform the conditional response distribution $F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y)$ to a simple yet arbitrary reference distribution F_Z that is specified a priori:

$$P(Y \leq y|\boldsymbol{X} = \boldsymbol{x}) = F_{Y|\boldsymbol{X}=\boldsymbol{x}}(y) = F_Z(h(y|\boldsymbol{x})).$$

Thus, CTMs allow not only variance, kurtosis and skewness but the whole conditional distribution function to depend flexibly on the covariates via an unknown monotonically increasing transformation function h.
We propose a Bayesian version of the versatile model class of CTMs that relies on an extended penalized (tensor) spline approach for the estimation of h. This spline approach allows for nonlinear interactions in the conditional transformation function while at the same time ensuring monotonicity by using modified basis functions and coefficients. In this way, we are able to make use of the blessings of the whole Bayesian toolkit such as the direct ability of uncertainty quantification via the posterior distribution and adaptable prior specifications.
As a benchmark, we compare our results to the maximum likelihood-based approach to the estimation of CTMs in Hothorn et al. (2018) and the Bayesian structured additive distibutional regression framework (Klein et al. 2013) in a simulation study.
References:
Hothorn, Torsten, Thomas Kneib, and Peter Bühlmann. Conditional transformation models."Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76.1 (2014): 3-27.
Hothorn, Torsten, Lisa Möst, and Peter Bühlmann. „Most likely transformations.“ Scandinavian Journal of Statistics 45.1 (2018): 110-134.
Klein, Nadja, Thomas Kneib, and Stefan Lang. Bayesian structured additive distributional regression. No. 2013-23. Working Papers in Economics and Statistics, 2013.

40 Flexible regression for probability densities in Bayes spaces

Almond Stöcker[1], Eva-Maria Maier[1], Bernd Fitzenberger[2], Sonja Greven[1]
[1]LMU Munich, Germany
[2]HU Berlin, Germany

We present a flexible regression framework for functional compositional data, i.e. functional additive regression models (FAMs) for the case that functional response variables are probability density functions (PDFs). The special nature of PDFs – in particular their non-negativity and property to integrate to one – prohibits direct application of usual functional regression models. Instead, we formulate FAMs for PDFs in a Bayes Hilbert Space. The isometry given by the so called centered log-ratio transform allows us to carry over the flexibility of previous FAMs to Bayes Space models. Thus, we are able to provide a wide range of different types of categorical, metrical and functional covariate effects. We apply the model framework to analyze the distribution of the relative income within couples earned by the wife since 1984 in the German socio-economic panel (SOEP). For estimation we use Gradient Boosting, which enables us to automatically select between different flexible covariate effects including temporal effects and the effects of geopolitical variables, the presence of children in the households and potentially relevant political reforms. Our approach can deal with discrete or continuous densities or a mixture of the two, as relevant for the SOEP data due to the positive probability mass on zero and one in the relative income.

Causal Inference I (Modelling Causal Structures)

41 How can graphical Markov models aid causal inference?

Nanny Wermuth
Chalmers University of Technology, Sweden

The search for causal relations motivates much of empirical research: experiments, and observational studies are designed and analysed to improve our understanding of conditions under which expected effects occur, are likely to be enhanced, stabilised, slowed down, reversed or completely prevented.
Given such a broad view of possible causal inquiries and corresponding inference, one would not speak of, say, 'causal models', 'causal graphs', 'causal parameters' or 'causal calculus' even though choosing and using appropriate models, graphs, measures of dependence, and algorithms may substantially advance reliable inference.
W.G. Cochran stated in 1965: 'A claim of proof of cause and effect must carry with it an explanation of the mechanism by which this effect is produced.' One interpretation is that an underlying data generating process has to be scientifically explainable given the available knowledge in the given research field. For this purpose, the currently most flexible class of generating processes is for recursive sequences of single and joint responses in which each response may depend only on variables that occurred in its past and in which the study context is captured by an undirected probabilistic graph.
In this lecture, I summarise some of the important results, obtained quite independently, for graphs, models and measures of dependence, show some of their applications and point to open questions for further theoretical research.

42 How to Deal With Reverse Causality Using Panel Data? Recommendations for Researchers Based on a Simulation Study

Tobias Wolbring[1], Lars Leszczensky[2]
[1]FAU Erlangen-Nürnberg, Germany
[2]University of Mannheim & MZES Mannheim

Does X affect Y? Answering this question is particularly difficult if Y may in turn affect X. Often facing the possibility of such reverse causality, many social scientists turn to panel data to address questions of causal ordering. Yet even in longitudinal analyses reverse causality threatens causal inference by biasing results obtained from conventional panel models. Having long recognized this problem, the econometric and statistical literature has suggested various alternative approaches to deal with reverse causality. However, these approaches have faced many criticisms, chief among them to be very sensitive to the correct specification of temporal lags. Applied researchers are thus left with little guidance. Seeking to provide such guidance, we compare how different panel models perform under a range of different data-generating conditions. Our Monte Carlo simulations reveal that while most conventional panel models fail to account for reverse causality, Arellano-Bond estimators and a cross-lagged panel model with fixed effects offer protection against bias arising from reverse causality under a wide range of conditions and help to circumvent the problem of misspecified temporal lags. Based on the simulation results, we provide researchers with recommendations on how to analyze panel data if causal inference is threatened by reverse causality.

43 Measurement, Causal Models and Treatment Effects

Marco Steenbergen, Lukas F. Stoetzer
University of Zurich, Switzerland

In recent years, causal identification has been front and centre in the social science. In comparison, interest in measurement is tepid and extremely rare are the efforts that link causal inference and measurement. Still, such a linkage is long overdue. Many causal designs, survey experiments being a prime example, rely on measurement instruments that are meant to shed light on latent attitudes. Yet, latent causal estimands have been mostly absent from the literature. What is more, it is possible that treatments exert an effect not only on the latent variable but also on the measurement instruments. This raises the possibility of differential item functioning (DIF) in treatment and control groups, including the possibility that the very meaning of concepts changes as a result of the treatment. In this paper, we develop a framework for estimating causal effects with latent variables. Working within the potential outcomes framework, we propose new causal estimands based on latent variables. We also explore the impact of DIF on estimating causal effects, as well as diagnostic tools and strategies that limit DIFs potential biasing effect. We illustrate our ideas and procedures using data from survey experiments and discuss the broader implications for the causal inference literature.

Design of Experiments and Clinical Trials II (Adaptive Designs I)

44 Optimal adaptive two-stage designs for normally distributed outcomes

Maximilian Pilz[1], Kevin Kunzmann[1], Carolin Herrmann[2], Geraldine Rauch[2], Meinhard Kieser[1]

[1]Institute for Medical Biometry and Informatics, University of Heidelberg, Germany
[2]Charité - Universitätsmedizin Berlin, Germany

In clinical trials, the choice of an adequate sample size is a crucial issue. While traditionally a design with a fixed sample size is applied, flexible strategies with one or several interim analyses are becoming increasingly popular.
When performing a trial in an adaptive two-stage design, one has to fix upfront the first-stage sample size n_1, the second-stage sample size n_2 as well as decision boundaries c_f (early stopping for futility), c_e (early stopping for efficacy) and c_2 (rejection boundary for the final analysis). In many classical adaptive designs, the decision boundaries are chosen according to a particular type of combination test in order to control the type one error rate, e.g., Fisher's combination test [1] or the inverse normal method [2]. The choice of the sample sizes is often based on (conditional) power considerations [3]. To protect patients and resources, the sample size should be as low as possible while at the same time guaranteeing a sufficiently high power for the trial.
Our general approach neither presumes a specific sample size recalculation rule nor a combination test but regards the described design variables as tuning parameters which can be chosen in an optimal way. We determine the design parameters such that the expected sample size under the alternative hypothesis is minimized. To guarantee type one error rate control and a sufficiently large power, these constraints are included as inequalities in the optimization framework.
We solve the resulting optimization problem with techniques of the calculus of variations. This approach yields to completely novel designs which differ from classical approaches and are superior to them with respect to expected sample size while controlling the specified type one and type two error rates. Additionally, knowledge of an optimal design allows to explore the efficiency of classical adaptive strategies such as combination tests. Statistical as well as practical properties of these new optimal designs are presented. Our methods are illustrated by means of a clinical trial example.
References:
[1] Bauer P., Köhne K.. Evaluation of experiments with adaptive interim analyses. Biometrics. 1994;50(4):1029–1041.
[2] Lehmacher W., Wassmer G.. Adaptive Sample Size Calculations in Group Sequential Trials. Biometrics. 1999;55(4):1286-1290.
[3] Proschan M. A., Hunsberger S. A.. Designed extension of studies based on conditional power. Biometrics. 1995;51(4):1315–1324.

45 Incorporating historical two-arm data in clinical trials with binary outcome

Manuel Feißt, Meinhard Kieser
Institute of Medical Biometry and Informatics

The feasibility of a new clinical trial may be increased by incorporating historical data of previous trials. In the particular case where only data from a single historical trial are available, there exists no clear recommendation in the literature regarding the most favorable approach.
A main problem of the incorporation of historical data is the possible inflation of the type I error rate[1]. A way to control this type of error is the so-called power prior approach[2]. This Bayesian method does not "borrow" the full historical information but uses a parameter 0 ? ? ? 1 to determine the amount of borrowed data.
Based on the methodology of the power prior, we propose a frequentist framework that allows incorporation of historical data from two-armed trials with binary outcome, while simultaneously controlling the type I error rate.
It is shown that for any specific trial scenario a value ? $>$ 0 can be determined such that the type I error rate falls below the pre-specified significance level. The magnitude of this value of ? depends on the characteristics of the data observed in the historical trial. Conditionally on these characteristics, an increase in power as compared to a trial without borrowing may result. Similarly, we propose methods how the required sample size can be reduced. The results are discussed, and application is illustrated by a clinical trial example.
1. Viele, K., et al., Use of historical control data for assessing treatment effects in clinical trials. Pharm Stat, 2014. 13(1): p. 41-54.
2. Ibrahim, J.G., et al., The power prior: theory and applications. Stat Med, 2015. 34(28): p. 3724-49.

46 A new rule for sample size recalculation based on resampling in an adaptive design setting

Geraldine Rauch[1], Carolin Herrmann[1], Maximilian Pilz[2], Meinhard Kieser[2]
[1]Charité Universitätsmedin Berlin, Germany
[2]University of Heidelberg, Germany

It is intuitive that the correct choice of the sample size is of major importance for an ethical justification of a trial and a responsible spending of resources. In an underpowered trial, the research hypothesis is unlikely to be proven, resources are wasted and patients are unnecessarily exposed to the study-specific risks. If the sample size is too large, the market approval is prolonged and later recruited patient in the control arm are exposed to a treatment already known to be less effective.
The parameter assumptions required for sample size calculation should be based on previously published results from the literature and on aspects of clinical relevance. In clinical practice, however, historical studies for the research topic of interest are often not directly comparable to the current situation under investigation or simply do not exist. Moreover, the results of previous studies often show a high variability or are even contradictory.
Calculating the 'correct' sample size is thus a difficult task. On the other side, the consequences of a 'wrong' sample size are severe.
A variety of sample size recalculation strategies have been proposed. Most frequently, these rules are based on conditional power arguments (e.g. [1], [2],[3]). This approach assumes implicitly that the true treatment effect is equal to the effect observed at the interim analysis. The conditional power approach is often criticized for this unrealistic assumption as the available information at the interim stage is usually limited and thus the treatment effect estimate shows a rather high variability resulting in a highly variable sample size. We present a new sample size recalculation strategy based on resampling which uses the interim effect as the expectation of a distribution rather than assuming that the interim effect is the true one.
[1] Lehmacher W, Wassmer G. Adaptive sample size calculations in group sequential trials. Biometrics 1999; 1286-1290.
[2] Mehta CR, Pocock SJ. Adaptive increase in sample size when interim results are promising: A practical guide with eamples. Stat. Med. 2011;30:3267-3284.
[3] Jennison C, Turnbull BW. Adaptive sample size modification in clinical trials: start small then ask for more? Stat. Med. 2015, 34: 3793–3810.

47 Reestimation of the prevalence in a confirmatory diagnostic accuracy study

Maria Stark, Antonia Zapf
Universitätsklinikum Hamburg-Eppendorf, Germany

In a confirmatory diagnostic accuracy study, sensitivity and specificity are considered as co-primary endpoints. The sample size calculation is done in three steps. First, the individual sample size of the sensitivity and of the specificity is calculated. Second, the total sample size of both endpoints is calculated by dividing the individual sample size of the sensitivity by the prevalence and by dividing the individual sample size of the specificity by one minus the prevalence. Third, the maximum of these both sample sizes represents the final sample size of the study. Often, the total sample size of the sensitivity and of the specificity differ which leads to an unbalanced design.
In a confirmatory diagnostic accuracy study, the sample size of each endpoint is often calculated with an individual power of 90% to reach an overall power of at least 80%. However, in the case of a low or high prevalence, the empirical overall power is noticeably larger than 80% due to the unbalanced design.
To prevent such an overpowered study, a method for an optimal sample size calculation for the comparison of a diagnostic index test with the gold standard is proposed. This method is based on Obuchowski's formulas for the sample size calculation in diagnostic studies (Obuchowski, 1998) and ensures an overall power of 80% which is independent of the prevalence.
As the sample size calculation is based on the true prevalence of the population, an incorrect assumption about the prevalence will lead to an over- or underestimated sample size. Therefore, an approach for the reestimation of the prevalence in a diagnostic accuracy study is presented. In this context, the assumption about the prevalence can be corrected and the sample size can be recalculated. The type I error rate, the overall power and the sample size of the interim analysis are compared with those of a fixed design.
Reference:
Obuchowski, N. A. (1998). Sample size calculations in studies of test accuracy. Statistical Methods in Medical Research, 7(4), 371-392.

Empirical Economics and Applied Econometrics II

48 Structural Interpretation of Vector Autoregressions with Incomplete Identification: Revisiting the Role of Oil Supply and Demand Shocks

Christiane Baumeister[1], James D. Hamilton[2]

[1]University of Notre Dame, United States of America
[2]University of California, San Diego, United States of America

Traditional approaches to structural vector autoregressions can be viewed as special cases of Bayesian inference arising from very strong prior beliefs. These methods can be generalized with a less restrictive formulation that incorporates uncertainty about the identifying assumptions themselves. We use this approach to revisit the importance of shocks to oil supply and demand. Supply disruptions turn out to be a bigger factor in historical oil price movements and inventory accumulation a smaller factor than implied by earlier estimates. Supply shocks lead to a reduction in global economic activity after a significant lag, whereas shocks to oil demand do not.

49 Identification of Structural Vector Autoregressions by Stochastic Volatility

Dominik Bertsche, **Robin Braun**

University of Konstanz, Germany

We propose to exploit stochastic volatility for statistical identification of Structural Vector Autoregressive models (SV-SVAR). We discuss full and partial identification of the model and develop efficient EM algorithms for Maximum Likelihood inference. Simulation evidence suggests that the SV-SVAR works well in identifying structural parameters also under misspecification of the variance process, particularly if compared to alternative heteroskedastic SVARs. We apply the model to study the interdependence between monetary policy and stock markets. Since shocks identified by heteroskedasticity may not be economically meaningful, we exploit the framework to test conventional exclusion restrictions as well as Proxy SVAR restrictions which are overidentifying in the heteroskedastic model. The tests indicate statistical evidence against employing conventional short run restrictions to identify a monetary policy shock. However, there is no evidence against using a certain long-run restriction and no evidence against identification by one of two external instruments included into the analysis.

50 Identification of independent structural shocks in the presence of multiple Gaussian components

Simone Maxand

University of Helsinki, Finland

Several recently developed identification techniques for structural VAR models are based on the assumption of non-Gaussianity. So-called independence based identification provides unique structural shocks (up to scaling and ordering) under the assumption of at most one Gaussian component. While non-Gaussianity of certain interesting shocks appears rather natural, not all macroeconomic shocks in the system might show this clear difference from Gaussianity. Identifiability can be generalized by noting that even in the presence of multiple Gaussian shocks the non-Gaussian ones are still unique. Consequently, independence based identification allows to uniquely determine the (non-Gaussian) shocks of interest irrespective of the distribution of the remaining system. Furthermore, studying settings close to normality or with multiple Gaussian components highlights the performance of normality diagnostics and their applicability to decide on the identifiability of the structural shock components. In an illustrative five dimensional model the identified monetary policy and stock price shock confirm the results of previous studies on the monetary policy asset price nexus.

Latent Variable Modelling I

51 An Approach to Addressing Multiple Imputation Model Uncertainty Using Bayesian Model Averaging

David Kaplan
University of Wisconsin, United States of America

This paper considers the problem of imputation model uncertainty in the context of missing data problems. We argue that so-called "Bayesianly proper" approaches to multiple imputation, although correctly accounting for uncertainty in imputation model parameters, ignore the uncertainty in the imputation model itself. We address imputation model uncertainty by implementing Bayesian model averaging as part of the imputation process.
Bayesian model averaging accounts for both model and parameter uncertainty, and thus we argue is fully Bayesianly proper, in the sense of Schafer (1997). We apply Bayesian model averaging to multiple imputation under the fully conditional specification approach. An extensive simulation study motivated by real data considerations is conducted comparing our Bayesian model averaging approach against choosing the imputation model with the highest
posterior model probability, and against normal theory-based Bayesian imputation not accounting for model uncertainty. The results reveal a small but consistent advantage to our Bayesian model averaging approach under MCAR and MAR in terms of Kullback-Liebler divergence. No procedure works well under NMAR. A small case study is also presented. Directions for future research are discussed.

52 Estimation of a Nonparametric Multidimensional Item Response Model Using Dirichlet Process Mixtures

Felix Naumann
LMU München, Germany

Parametric Item Response models do not always show acceptable fit to the data obtained from psychological tests. In these cases, one option is to resort to more flexible nonparametric models. Peress (2012) provides an identification proof for a very general Item Response model, which can be viewed as a multidimensional compensatory model with nonparametric Item Characteristic Curve (ICC).
The subject of this talk is the application of Bayesian nonparametrics to the estimation of the ICC and the item- and person parameters of this model. A reparameterisation is proposed, which allows for a Bayesian formulation of the model. The parameter space of the ICC is a function space and a Dirichlet Process Mixture of normal cumulative distribution functions can be chosen as a prior on this space. This allows for the derivation of all full conditionals of the joint posterior distribution of the parameters and thus for an implementation of a Gibbs Sampler.

53 Dirichlet Clustering in Onyx

Timo von Oertzen

Universität der Bundeswehr München, Germany

To successfully use latent variable models with normally distributed variables, homogenous data sets are required, which is often achieved by mixture models or other clustering algorithms. However, there are not many clustering algorithms that allow to include prior information of the clustering or result in a posterior distribution of clusterings instead of a single, fixed clustering. Dirichlet Clustering is a nonparametric Bayesian clustering method that provides these features. In this presentation, we will give a quick demonstration how to use Dirichlet Clustering in the graphical SEM software Onyx to easily identify subgroups of participants with different parameter solutions.

Mathematical Statistics I

54 Log-concave density estimation

Oliver Feng[1], Aditya Guntuboyina[2], Arlene Kim[3], **Richard Samworth**[1]
[1]University of Cambridge, United Kingdom
[2]Univeristy of California, Berkeley
[3]Sungshin Women's University, South Korea

In recent years, density estimation via log-concave maximum likelihood estimation has emerged as a fascinating alternative to traditional nonparametric smoothing techniques, such as kernel density estimation, which require the choice of one or more bandwidths. I will outline some of the attractive properties of this technique, with a focus on new results on adaptivity properties and the estimation of high-dimensional log-concave densities.

55 Spectral thresholding for the estimation of Markov chain transition operators

Matthias Loeffler[1], Antoine Picard[2]
[1]University of Cambridge, Cambridge, United Kingdom
[2]Ecole Normale Supérieure, Paris, France

We consider estimation of the transition operator P of a Markov chain and its transition density p where the eigenvalues of P are assumed to decay exponentially fast. This is for instance the case for periodised multi-dimensional diffusions observed in low frequency. We investigate the performance of a spectral hard thresholded Galerkin-type estimator for P and p, discarding most of the estimated eigenpairs. We show its statistical optimality by establishing matching minimax upper and lower bounds in L^2-loss. Particularly, the effect of the dimension d on the nonparametric rate improves from $2d$ to d compared to the case without eigenvalue decay.

56 Adaptive confidence sets for kink-location and kink-size in nonparametric regression

Viktor Bengs[1], Hajo Holzmann[2]

[1]Paderborn University, Germany
[2]Philipps-Universität Marburg, Germany

Kinks of a regression function are irregularities which indicate for instance an abrupt change of the function similar to classical change points. Therefore, it is of special interest to locate such irregularities as well as to estimate their magnitudes in order to analyze the cause and the impact of these points. Such kinks correspond to jumps in the first derivative of the regression function and various estimation methods are available for such discontinuities. Nevertheless, the literature is sparse concerning construction of confidence sets for kink-locations or their size.
In this talk, we first concentrate on the optimal estimation in the minimax sense of the location and of the size of the jump in the γ-th derivative of a regression curve, which is assumed to be Hölder smooth of order $s \geq \gamma + 1$ away from the kink. In the course of this, we establish optimal convergence rates as well as the joint asymptotic normal distribution of estimators based on the so-called zero-crossing-time (ZCT) technique known in the statistical literature. After that, we construct joint as well as marginal asymptotic confidence sets based on a Lepski-choice of the bandwidth-resolution for the ZCT technique and an appropriate control of the bias. Thanks to the Lepski-choice, the resulting confidence sets are on the one hand, honest and adaptive with respect to the smoothness parameter s over subsets of the Hölder classes, and on the other hand, their construction is mainly data-driven. The finite-sample performance of the method is illustrated by a real data application.

Preclinical and Pharmaceutical Statistics II (Estimands)

57 Estimand framework in Oncology drug development – impact and opportunities

Kaspar Rufibach[1], Evgeny Degtyarev[2], Jonathan Siegel[3], Viktoriya Stalbovskaya[4], Steven Sun[5]

[1]F. Hoffmann-La Roche, Switzerland
[2]Novartis, Switzerland
[3]Bayer Pharmaceuticals, US
[4]Merus, Netherlands
[5]Janssen, US

A draft addendum of the ICH E9 guideline on Statistical Principles for Clinical Trials was released in August 2017 and introduced an estimand framework. The new framework aims at aligning trial objectives with design and statistical analyses by requiring a precise definition of the inferential quantity of interest, the estimand. The addendum is anticipated to have a major impact on drug development, as the choice of estimands will drive the trial design, sample size, data collection, trial conduct, analysis, and interpretation. An industry working group for estimands in oncology was formed in February 2018, with members from 19 companies. The focus areas of the working group are treatment switching, censoring, hematology and solid tumor case studies, and causal estimands in the time-to-event setting. In this talk we will review common estimands of interest and intercurrent events proposed in oncology regulatory guidelines and applications. Several strategies to handle intercurrent events were described in the ICH E9 addendum. These strategies generally interpret intercurrent events as informative or counterfactual outcomes rather than noninformative "missing" data. We will embed those in the time-to-event framework discussing the differences and highlighting the consequences for study design, data collection, analysis and interpretation. Since estimation methods targeting estimands using the proposed strategies often require strong assumptions, the guideline emphasizes sensitivity analyses to justify these. We will discuss sensitivity analyses for key estimands. The concepts will be illustrated using case studies and we will provide recommendations of the industry working group for practical implementation of the estimand framework.

58 Estimation of Principal stratum effects, an overview and potential applications in oncology

Bjoern Bornkamp[1], Audrey Boruvka[2], Evgeny Degtyarev[1], Vera Kuehnl[9], Feng Liu[5], Yi Liu[6], Emily Martin[8], Devan Mehrotra[7], Satrajit Roychoudhury[3], Kaspar Rufibach[2], An Vandebosch[4]

[1]Novartis, Switzerland
[2]Roche
[3]Pfizer
[4]Janssen
[5]AstraZeneca
[6]Nektar
[7]Merck
[8]EMD Serono
[9]AbbVie

A draft addendum of the ICH E9 guideline on Statistical Principles for Clinical Trials was released in August 2017 and introduced an estimand framework. The new framework aims at aligning trial objectives with design and statistical analyses by requiring a precise definition of the inferential quantity of interest, the estimand. The addendum is anticipated to have a major impact on drug development, as the choice of estimands will drive the trial design, sample size, data collection, trial conduct, analysis, and interpretation. The addendum suggests five strategies to deal with so-called intercurrent events. Among those is the principal stratum strategy, which consists of targeting the effect in the population in which the intercurrent event would not occur (or would occur) on both test and control treatments (or only the test treatment depending on the clinical question of interest). Limited experience exists with such type of estimands in the regulated context of drug development.

In this presentation we would like to highlight situations, where the principal stratification strategy could be of interest, with a focus towards oncology drug development and time-to-event endpoints. We will also provide an overview of different approaches on how to estimate a principal stratum effect, drawing on developments in causal inference literature in the past 10-15 years. In addition we will outline strategies for sensitivity analyses with respect to the assumptions typically required for the principal stratum strategy.

This presentation is given on behalf of the causal subteam of the EFSPI oncology estimand working group.

59 Estimands in the presence of treatment switching

Viktoriya Stalbovskaya[1], Juliane Manitz[2], Marie-Laure Casadebaig[3], Emily Martin[2], Rui Sammi Tang[4], Godwin Yung[5], Vincent Haddad[6], Fei Jie[7], Christelle Lorenzato[8], Jiangxiu Zhou[9], Evgeny Degtyarev[10]

[1]Merus
[2]EMD Serono
[3]Celgene
[4]Servier, DahShu
[5]Takeda Pharmaceuticals
[6]AstraZeneca
[7]Astellas Pharma
[8]Sanofi
[9]GlaxoSmithKline
[10]Novartis, Switzerland

A draft addendum of the ICH E9 guideline on Statistical Principles for Clinical Trials was released in August 2017 and introduced an estimand framework. The new framework aims to align trial objectives and statistical analyses by requiring a precise definition of the population quantity of interest, the estimand. This definition should explicitly account for intercurrent events, i.e. events which occur after baseline but before observing the endpoint of interest such as start of new therapy.
The EFSPI Estimands in Oncology working group was initiated to foster a common understanding and consistent implementation of the relevant framework in oncology clinical trials. This presentation is given on behalf of the Treatment Switching Sub-team of this Working Group.
Treatment switching is considered one of the key intercurrent events in oncology clinical trials. Traditionally the main analysis of overall survival in the context of confirmatory study is performed ignoring treatment switching (treatment-policy estimand) and may therefore underestimate the survival benefit that would have been seen had patients not switched. Causal inference methodology such as rank-preserving structural failure time models or inverse probability weighting has been applied in oncology while analyzing overall survival and accounting for treatment switching (hypothetical estimand), providing further perspectives on the added value of novel therapies, e.g. for payers. We will present different choices of estimands using case studies, and illustrate the impact of the estimand choice on study design, data collection, trial conduct, analysis, and interpretation.

60 Implementation of the ICH E9 addendum: A case study in hematology

Hans-Jochen Weber[1], Marie-Laure Casadebaig[2], Emily Butler[3], Satrajit Roychoudhury[4], Kaspar Rufibach[5], Viktoriya Stalbovskaya[6], Steven Sun[7]

[1]Novartis
[2]Celgene
[3]GlaxoSmithKline
[4]Pfizer
[5]Roche
[6]Merus
[7]Johnson & Johnson

The draft addendum to the ICH E9 guideline was released in 2017 to introduce the estimand framework. This framework targets to align the objectives with the design and statistical analyses of a clinical study by providing detailed definitions of the inferential quantity of interest, the estimand.
The European Federation of Statisticians in the Pharmaceutical Industry (EFSPI) Estimands in Oncology working group was initiated to foster a common understanding and consistent implementation of the relevant framework in oncology clinical trials. This presentation is given on behalf of the Hematology Subteam of this Working Group.
Clinical trials in hematology often present unique challenges for study design due to possible administration of a stem cell transplant or a treatment sequence across different phases (induction, consolidation, maintenance). Different treatment effects and its estimation could be of interest in such situations and will be discussed using a case study. We present the considerations on how the estimand framework can support the design of clinical trials in hematology and how it can be implemented.

Small Area Analysis and Spatial Statistics II

61 Comparing designs for prediction based on stationary vs. non-stationary space-time covariance functions

Helmut Waldl
Johannes Kepler University Linz, Austria

Modeling space-time-data with spatio-temporal models requires the choice of a spatio-temporal covariance function. Using a separable covariance function facilitates parameter estimation, in many situations, however, it is a simplification and strong idealization of reality.
Furthermore choosing a stationary space and time covariance function imports an unnecessary oversimplification. In many practical applications the data show strong evidence of a spatially and temporarily non-stationary covariance structure.
Using stationary covariance functions for finding optimal designs for prediction results in more or less similar space filling designs. Good designs for prediction should pump as much information as possible from the available data. As variability of the data changes over space an time, also optimal designs should be adapted.
Some positions of a field hold more information than other positions just as it is plausible that the locations with high information change over time. It is intuitively clear that measurements should be taken at times and locations with high variability. On the other hand a trade-off has to be made between greedy information hunting and non-neglecting large regions with low variation. Using a kriging model generalized for a non-stationary covariance structure this trade-off is made automatically if we use the kriging variance as design criterion.
A computer simulation experiment based on spatio-temporal data compares the prediction performance of the design based on a stationary covariance function with the performance of a design optimized for a non-stationary covariance structure.

62 Identifying spatial dependence structures with copulas and generalized additive models

Marc Hüsch[1], Bruno U. Schyska[2,3], Lueder von Bremen[2,3]

[1]Technische Universität Dortmund, Germany
[2]DLR-Institut für Vernetzte Energiesysteme e.V., Germany
[3]Carl von Ossietzky Universität Oldenburg, Germany

We propose a flexible method to identify spatial dependence structures in spatio-temporal datasets with respect to different geographic properties.The approach bases on first selecting different pairs of spatial points which are representative for the entire spatial domain. For all selected pairs, the parameters of a suitable bivariate copula are estimated to characterize the dependence structure between the data points observed at the respective locations. Generalized additive models (GAMs) are used subsequently to model the estimated copula parameters subject to possible influencing factors such as spatial distance or different geographic conditions.
In a test case, the approach successfully identifies that spatial dependence of wind power forecast errors in Europe is more pronounced for longer forecast horizons. The approach further highlights spatial dependence structures that vary over different orographic conditions. In particular, spatial dependencies are mostly pronounced in low-terrain areas, whereas only small dependencies occur in high mountain areas like the Alps. The results further indicate that very large wind power forecast errors tend to occur jointly in a spatial context, which leads to high aggregated forecast errors in entire spatial regions. In an additional analysis, we notice that this finding is mostly prominent in areas with a high amount of installed wind power capacity, which makes it relevant for various stakeholders in the energy industry such as grid operators and energy traders.

63 Spatio-Temporal Smoothing of Drinking Water Contamination Data

Jonathan Rathjens[1], Eva Becker[1,2], Arthur Kolbe[2], Katharina Olthoff[3], Sabine Bergmann[3], Jürgen Hölzer[2], Katja Ickstadt[1]

[1]TU Dortmund University, Germany
[2]Ruhr-University Bochum, Germany
[3]NRW state environmental agency LANUV, Germany

The exposure to Perfluorooctanoic Acid (PFOA) is suspected to have various adverse effects on human health. For the general population of large parts of North Rhine-Westphalia (NRW), Germany, contaminated drinking water can be regarded as an important source of PFOA.
In order to estimate the state-wide exposure surface in the course of time, we formulate Bayesian spatio-temporal models based on local PFOA measurement data of diverging completeness, some vague information about (non-)contamination, and knowledge about spatial correlations affected by rivers.
PFOA concentrations in drinking water have been derived from an extensive monitoring programme of the NRW state environmental agency LANUV and by further acquisition from water supply companies. For many sites, there exist proper time series for several years; others have but a single measurement, usually below the limit of detection; for some regions, data are completely missing. Samples have been drawn from both the water supply stations and the network divided in water supply areas.
Therefore, data are given at two spatial levels, whose connections are generally neither injective nor surjective. The proportions (weights), how much of an area's water stems from the respective stations, are numerically estimated based on rough data of their supplied and demanded water amounts.
At both levels, we formulate Gaussian and gamma likelihoods and a linear map between the expected values of stations and areas using the weights, also including a random effect. The correlation of spatially and temporally adjacent measurements is handled in two ways for comparison: by a Markov random field and by smoothing approaches. We define spatial distances based on the stations' connectivity via rivers.
As spatial covariates, we use our so-defined spatial and temporal distances from polluted sites, particularly from a PFOA contamination incident near the city of Arnsberg prior to 2006.

64 Multilevel Conditional Autoregressive models for longitudinal data nested in geographical units with dynamic characteristics

Dany Djeudeu[1,2], Susanne Moebus[2], Katja Ickstadt[1]

[1]Faculty of Statistics, TU Dortmund University

[2]Centre for Urban Epidemiology (CUE), Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, University Duisburg-Essen

In analyzing the association between individual health outcomes and individual environmental exposures based on geographically hierarchical data sets, we can expect (or at least allow for) two types of dependence or spatial effects: spatial autocorrelation and the spatial heterogeneity.
To provide more efficient and accurate model estimation for regression coefficients by acknowledging the hierarchical structure of the data, traditional multilevel models well developed in the literature can be used.
These traditional multilevel models use geographical unit-level random effects to account for spatial heterogeneity and help to answer some important research questions but do not account for possible spatial dependence between spatial units.
Existing models for cross-sectional data combine the multilevel structure and the advantages of models like the **Besag-York-Mollié** and **Leroux**
for the random effect at the geographical unit to jointly model the spatial autocorrelation and the spatial heterogeneity, as these models also take advantages of the Conditional Autoregressive (CAR) prior specification of random effects. This combination leads to estimates of random covariate effects that are robust and have higher precision, which is of particular importance for areas with small sample size.
The overall aim of this work is to extend the idea of combining the Multilevel structure and the advantages of models like the **Besag-York-Mollié** and **Leroux** for longitudinal data on individuals nested in geographical units such as participants nested in districts, thereby accounting for the dynamic of group characteristics. In longitudinal studies involving environmental exposures, not only individuals are changing over time, so are geographical unit (characteristics).
We then simulate data sets with different spatial structures; heterogeneity only, spatial dependence only and both spatial heterogeneity and spatial dependence to analyze the association between a health outcome defined at individual level and an environmental exposure adjusting for confounders given at all levels in a defined study area. For each of the spatial structure, we consider the case where the strength of the spatial effect is changing over time.
Results suggest that the extended model yields more accurate coefficient estimates and reliable credible intervals compared to traditional spatial growth models formulated as a (three level) multilevel model with repeated measurements over time nested in participants and participants nested in spatial units.

Statistical Literacy and Statistical Education II

65 Statistical Literacy and Statistical Education

Laura Martignon
PH Ludwigsburg, Germany

We all agree that Statistical Education has become an indispensable component of young citizens' upbringing, not just as a means for understanding statistical data but also as a support of critical thinking and, above all, for decision making under uncertainty. Statistical Education consists in fostering Statistical Literacy, which, as will be described in the talk, can be seen as composed by specific competencies.
Aspects of statistical literacy between competency measures and indicators for conceptual knowledge, as well as students' views related to chance variability will be discussed based on empirical results. Risk Literacy as a support for Decision Making will also be a core topic.
Decision Models based on features extracted from Data will be presented.
A special emphasis will be given to fast and frugalßtatistical models - as emerging from Herbert Simon's paradigm of Bounded Rationality - and their possible role in Statistical Education. Their construction based on measures of "goodnessör "diagnosticityöf features will be illustrated. These fast and frugalmodels can be seen as heuristics which do not attempt to optimally fit parameters to a given environment; rather, they have simple structural features and "bet" that the environment will fit them. By not attempting to optimize, these heuristics can save time and computations, and demand only little knowledge concerning a situation. Such models have been designed for various tasks, including choice, numerical estimation, and classification. The talk will focus on some simple, naive, linear and lexicographic heuristics whose performance will be compared with that of standard statistical strategies that weigh and combine many reasons, such as multiple regression or complex Bayesian Networks. Contrary to common intuition, more reasons and greater complexity are not always better, especially when training sets are small.
In the more empirical part of the talk, the performance of the different approaches to making inferences under uncertainty (i.e., out of sample) will be evaluated in medical data sets in terms of Receiver Operating Characteristics (ROC) diagrams and predictive accuracy. Results will confirm that the heuristic approaches, Fast and Frugal trees in particular, generally outperform models that are normative when fitting known data. The success of fast-and-frugal heuristics is grounded in their ecological rationality: Their construction principles can exploit the structure of information in the data sets.
A possible educational syllabus for Statistical Literacy that includes Decision Making under Uncertainty based on simple but robust models will be an issue of the concluding remarks.

66 Civic Statistics: Big Ideas, Needs and Challenges. Why we need a new subdiscipline

Joachim Engel

PH Ludwigsburg, Germany

Effective citizen engagement with social issues requires active participation and a broad range of skills, including the understanding of data and statistics about society and our natural and social environment. We provide an overview of a subfield we call Civic Statistics that had been explored by a recent strategic partnership of six European universities under the Erasmus+ program of the EU. Civic Statistics is statistics about important societal trends and about topics that matter to the social and economic well-being of citizens. It sits at the crossroads of multiple disciplines including social science, education and statistics.

Hence, a multidisciplinary educational perspective is needed, stepping outside the comfort zone of traditional statistics instruction. Understanding Civic Statistics is needed for participation in democratic societies, but involves data that often are open, official, multivariate in nature, and/or dynamic, which are usually not at the core of regular statistics instruction. Many statistics classes and educational curricula are not designed to teach relevant skills and improve learners' statistical literacy, despite the importance of engaging learners and future citizens with data about social issues and their connections to social policy. This talk gives an overview to characteristics of Civic Statistics and provides guidance to specific teaching materials developed by the ProCivicStat project.

67 A New Approach for Developing Statistical Thinking

Ana Kolar

Tarastats, Statistical Consultancy, Finland

We extend the definition of statistical thinking as provided by Ben-Zvi and Garfield (2005), Box et al. (1978), Britz et al. (1996, 2000), Moore (1990), and Snee (1990) in order to make statistical thinking applicative also for those who are exposed to the digital (data) world, but professionally do not deal with statistical data analysis. We define statistical thinking as a conscious thought process that is based on statistical concepts of sampling theory, handling missing data, descriptive and inferential statistics, in order to analyse subjects, environments, or situations objectively also outside of statistical data analysis field. We define causal thinking as one of the crucial components of statistical thought process, because it assists us in recognising and acknowledging different causal (mental) models based on which a topic of investigation is studied. We define causal thinking processes in two parts: (i) as 'understandings' of what is casual conceptually, i.e., physical versus factual cause; and (ii) as 'understandings' of what is causal according to the statistical definition of causality. We claim that also universal objectivity is conditional on some causal (mental) model. The new approach for developing statistical thinking thus help us with developing observational skills, which allow for different subjective realities to be recognised and an objective reality approached. The paper presents reasoning of the derived definition and an application of the new approach to develop statistical thinking skills.

Keywords: statistical thinking, causal thinking, objectivity

Ben-Zvi, D., and Garfield, J. (eds.), (2005), The challenge of Developing Statistical Literacy, Reasoning and Thinking, Springer Science + Business Media, Inc.

Box, G. E. P., Hunter, W. G., and Hunter, J. A. (1978), Statistics for Experimenters, New York: John Wiley and Sons.

Britz, G.C., et al. (Spring 1996), Statistical Thinking, A special publication of the ASQ Statistics Division Newsletter.

Britz, G.C., Emerling, D.W., Hare, L.B., Hoerl, R.W., Janis, S. J., and Shade, J.E. (2000), Improving Performance Through Statistical Thinking, Quality Press, Milwaukee, WI.

Moore, D. S. (1990), "Uncertainty," in On the Shoulders of Giants, ed. L. A. Steen, National Academy Press, 95-173.

Snee, R. (1990). Statistical Thinking and its Contribution to Quality. The American Statistician, 44(2), 116-121.

Computational Statistics and Statistical Software II (Omics)

68 Comparison of different preprocessing methods for the analysis of metabolite data

Janine Wiebach, **Miriam Sieg**, Jochen Kruppa

Institute of Biometry and Clinical Epidemiology, Charité - Universitaetsmedizin Berlin, Germany

From all of the typical "omics" fields like genomics or proteomics, metabolomics is the nearest to the phenotype or outcome of interest. Hence, the examination of metabolites allows to draw a direct conclusion to the phenotype. In this time the most important techniques to analyze metabolites are liquid chromatography–mass spectrometry (LC-MS) and nuclear magnetic resonance (NMR) spectroscopy. The complexity of the data analysis, consisting of many depending steps of data processing, still remains a major challenge. For a high quality of metabolite annotation and correct calculation of their corresponding intensities, raw data preprocessing is a crucial step. The number of open access published methods in R according to this matter increased rapidly in the last few years. However, the decision which method to use depends strongly on the researcher and their experience with this type of data analysis. If the experience is low with this data analysis, in general a brute force approach would be used to determine the best workflow. Therefore different preprocessing methods are tested on the biological data to decide which of them works best and has the highest accuracy.
There are studies that compare a small number of methods [1, 2], but a comprehensive comparison of published methods with limitations and advantages on in silico data sets is still missing.
In my talk I want to present our strategy for creating a ranked list with published methods in R including their limitations and a recommendation which method to use for specific parameters.Therefore an in silico reference data set of a chosen metabolite pattern will be generated as a basis for the simulation of a huge amount of test data. This data will be very similar to biological data but varying in certain parameters like noise, peak width, peak shape and the structure of the baseline. The structure of our test data will allow the comparison of methods fast and automatically.
References:
Hao L, Wang J, Page D, et al. Comparative Evaluation of MS-based Metabolomics Software and Its Application to Preclinical Alzheimer's Disease. Scientific Reports. 2018;8(1).
Myers OD, Sumner SJ, Li S, Barnes S, Du X. Detailed Investigation and Comparison of the XCMS and MZmine 2 Chromatogram Construction and Chromatographic Peak Detection Methods for Preprocessing Mass Spectrometry Metabolomics Data. Analytical Chemistry. 2017;89(17):8689-8695.

69 Benchmarking survival prediction methods using 18 multi-omics datasets from the „The cancer genome atlas“ (TCGA)

Philipp Probst, Moritz Herrmann, Roman Hornung, Vindi Jurinovic, Anne-Laure Boulesteix

LMU Munich, Germany

A large number of “multi-omics datasets” (i.e., datasets that include different types of high-dimensional molecular variables – such as gene expression data, copy number data, etc. - in addition to classical clinical variables) are freely available from the webpage of The Cancer Genome Atlas (TCGA). Because each of the involved data types can provide valuable information for the prediction of clinical outcomes, using multi-omics data as covariate data in outcome prediction is promising. However, using the predictive information contained in the covariates effectively is challenging, because the structure of multi-omics data is complex. Against this background, we performed a benchmark analysis using TCGA datasets to compare various prediction methods with respect to their performance in outcome prediction using multi-omics covariate data. We selected 18 datasets including several types of omics data. The (disease-free) survival time was considered as the outcome of interest to be predicted. In our benchmark experiment we compared several classical and modern survival prediction methods including a simple Cox regression model based on the clinical variables, Lasso regression, random forest, different gradient boosting methods as well as specific methods such as IPF-Lasso and priority-Lasso that take the block structure of the multi-omics variables into account rather than throwing all covariates together in the same pot. As a baseline we used a simple Kaplan-Meier estimator that reflects the prediction accuracy that would be obtained if the covariates had no effect on the outcome. We performed our benchmark experiment using the R package ‘mlr’ and 10 times repeated 5-fold cross validation and compared the methods using Uno‘s C-Index and the integrated Brier score as performance metrics.

70 Block Forests: random forests for blocks of clinical and omics covariate data

Roman Hornung[1], Marvin N Wright[2,3]

[1]Institute for Medical Information Processing, Biometry and Epidemiology, University of Munich, Germany
[2]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Bremen, Germany
[3]Section of Biostatistics, Department of Public Health, University of Copenhagen, Denmark

In the last years more and more multi-omics data are becoming available, that is, data featuring measurements of several types of omics data for each patient. Examples of different omics data types include gene expression data, mutation data, and copy number variation measurements. While using multi-omics data as covariate data in outcome prediction is promising, it is also challenging due to the complex structure of such data. Random forest is a prediction method known for its ability to render complex dependency patterns between the outcome and the covariates. Against this background, we developed five candidate random forest variants tailored to multi-omics covariate data. These variants modify the split point selection of random forest to incorporate the block structure of multi-omics data and can be applied to any outcome type for which a random forest variant exists, such as categorical, continuous and survival outcomes. Using 20 publicly available real multi-omics data sets with survival outcome we compared the prediction performances of the block forest variants, using random survival forest as a reference method. We also considered the common special case of having clinical covariates and measurements of a single omics data type available.
We identify one variant termed "block forestthat performed significantly better than standard random survival forest (adjusted p-value: 0.023). The two best performing variants have in common that the block choice is randomized in the split point selection procedure. In the case of having clinical covariates and a single omics data type available, the improvements of the variants over random survival forest were larger than in the case of the multi-omics data. In the former case four of the five variants performed significantly better than random survival forest. The degrees of improvements over random survival forest varied strongly across data sets.
To conclude, the new method block forest for outcome prediction using multi-omics covariate data can significantly improve the prediction performance of random forest. Block forest is particularly effective for the special case of using clinical covariates in combination with measurements of a single omics data type.

71 Sparse-group lasso variants for whole-genome regression models in livestock

Jan Klosa[1], Noah R. Simon[2], Volkmar Liebscher[3], Dörte Wittenburg[1]

[1]Leibniz Institute for Farm Animal Biology (FBN), Germany
[2]University of Washington, USA
[3]University of Greifswald, Germany

For the genomic evaluation of domestic animals, whole-genome regression methods are applied which use extensive information about genomic markers (e.g. single nucleotide polymorphisms; SNPs). As the number of model parameters increases with a still growing number of SNPs, multicollinearity between covariates can affect the precision of marker-effect estimates. Furthermore, it is desired to select those markers that are relevant to trait expression. Hence, selection-and-shrinkage approaches are a promising option to generate sparse solutions with higher precision. The objective of this study is to develop a statistical method following the sparse-group lasso, which builds a model that automatically selects a sparse set of predictor variables. Because a particular source of dependence between SNPs is due to linkage and linkage disequilibrium (LD), an extension is developed that considers the spatial genome structure and allows grouping according to those measures. Furthermore, a novel variant is investigated where grouped fitted values are penalized instead of grouped covariates. The sparse-group lasso variants select a solution that is sparse among and within groups. In addition, when appropriate for a given phenotype, these methods can promote the inclusion of SNPs from contiguous genomic regions. Thus, genomic regions that affect a trait can be more effectively identified. As the span of linkage and LD strongly depends on the population structure, this study focusses on a population consisting of half-sib families which is typical, for instance, in dairy cattle. All methods were implemented in Fortran using Proximal Gradient Descent, step widths were determined via Backtracking Line Search, and a fast grid search for optimal penalization parameters was carried out by Warm Starts. Methods were evaluated using simulated data resembling a dairy cattle population. The new methods were more effective than a standard lasso regression, which ignores relevant structural information. More specifically, the median of mean squared errors of estimated SNP effects was reduced up to 57% in simulated scenarios. Future work will investigate options to further generalize the penalty term.

Design of Experiments and Clinical Trials III (Adaptive Designs II)

72 Blinded continuous information monitoring of recurrent events endpoints with time trends

Tobias Mütze[1], Susanna Salem[2], Norbert Benda[2,3], Heinz Schmidli[1], **Tim Friede**[2,4]

[1]Novartis Pharma AG, Switzerland
[2]University Medical Center Göttingen, Göttingen, Germany
[3]Federal Institute for Drugs and Medical Devices (BfArM), Bonn, Germany
[4]DZHK (German Centre for Cardiovascular Research), partner site Göttingen, Göttingen, Germany

In clinical trials with recurrent event endpoints, misspecified assumptions of event rates or the dispersion can lead to under- or overpowered trials. Specification of the overdispersion is often a particular problem, as it is usually not reported in clinical trial publications. To mitigate the risks of inadequate sample sizes, internal pilot study designs for clinical trials with recurrent events have been proposed, with a preference for blinded sample size re-estimation procedures as they generally do not affect the type I error rate and maintain trial integrity [1]. However, the re-estimated sample size can have considerable variance, in particular with early sample size reviews. Friede et al. (2018) [2] addressed the issue of variable re-estimated sample sizes by proposing a blinded continuous monitoring of information for clinical trials with recurrent events modelled by a homogeneous Poisson process with a Gamma frailty. However, the assumption of a time-independent event rate in a homogeneous Poisson process does not always hold. For example, Nicholas et al. (2011) [3] showed that the relapse rate in clinical trials in multiple sclerosis changes over time. In this presentation, we study the robustness of the continuous information monitoring procedure proposed by Friede et al. (2018) [2] towards recurrent events with time trends. Moreover, we propose a blinded continuous information monitoring procedure for recurrent events with time trends. We show that our proposed monitoring procedure does maintain the integrity a clinical trial by controlling the type I error rate and that the proposed procedure results in adequately powered clinical trials.

References

1. Friede, T., Schmidli, H. (2010). Blinded sample size reestimation with negative binomial counts in superiority and non-inferiority trials. Methods of Information in Medicine, 49:618-624.

2. Friede, T., Häring, D., Schmidli, H. (2018). Blinded continuous monitoring in clinical trials with recurrent event endpoints. Pharmaceutical Statistics.

3. Nicholas, R., Straube, S., Schmidli, H., Schneider, S., & Friede, T. (2011). Trends in annualized relapse rates in relapsing–remitting multiple sclerosis and consequences for clinical trial design. Multiple Sclerosis Journal, 17:1211-1217.

73 Adaptive designs for drug combination informed by longitudinal model for the response

Tobias Mielke[1], Vladimir Dragalin[2]

[1]Janssen, Germany

[2]Janssen Pharmaceuticals

Focus of the presentation are adaptive designs for drug combination studies in the Phase II of drug development. We consider that the safety-tolerated two-dimensional dose space of the two drugs has been established in previous phases, such that the objective is to estimate the efficacy response surface in this region and to select the most efficient dose-combination for the final Phase III clinical trial. In contrast with the dose escalation designs for Phase I trials, in Phase II studies subjects can be allocated upfront to all dose-combinations in the acceptable dose-combination region. The problem is then to find the optimal design that allocates subjects to these dose-combinations in order to maximize the efficacy information obtained in the trial. A binary endpoint will be considered in this presentation as a measure of efficacy. The practical situation when the timing of the endpoint assessment period on the subject level is considerably longer relative to the inter-arrival time of subjects will be considered. This poses some implementation challenges for the adaptive designs. All patients might be enrolled, at the time when an interim analysis can be conducted. We propose a application of time-to-event model as a particular type of longitudinal response models todraw information from partial follow-up on subjects at the interim analysis, while the final analysis will be based on a binary response model. The interim and final analysis model in this situation will differ, what could complicate the problem of optimally allocating patients to the available dose combinations. Different designs optimization approaches will be presented and the robustness of the proposed designs under model missspecification will be evaluated.

74 An Alternative Log-Rank Test for Adaptive Survival Trials

Laura Kerschke, Andreas Faldum, Rene Schmidt

Institute of Biostatistics and Clinical Research, University of Münster, Germany

When survival curves of two treatment groups are compared within a clinical trial the two-sample log-rank test proposed by Mantel [1] and Peto and Peto [2] is usually applied. It has favorable properties, as it is optimal under the proportional hazards condition and can easily be generalized to handle more than two groups. However, methodological difficulties arise in adapting the log-rank test to more complex study designs as e.g. platform trials where treatments might adaptively be dropped or added during the course of the trial. This is due to the fact that the common two-sample log-rank test statistic is obtained from pooled data of both treatment groups and cannot be written as the difference of two independent random variables derived from non-overlapping populations. To overcome this issue, we propose an alternative two-sample log-rank test such that the underlying test statistic is similar to that of an unpaired z-test with known variance. On this basis the well-known methodology for comparisons of means can immediately be transferred to the survival setting. The proposed method relies on asymptotic distributions. We study its performance by simulation. Our simulations support validity of the distributional approximations as well as adequate type I and type II error rate control.

[1] Mantel, N. (1966). Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemotherapy Reports. 50, 163-170.

[2] Peto, R. and Peto J. (1972). Asymptotically efficient rank invariant test procedures. Journal of the Royal Statistical Society, Series A. 135, 185-207.

75 Combining Parallel Adaptive Seamless Phase 2/3 Trials

Cornelia Ursula Kunz[1], Nigel Stallard[2]

[1]Boehringer Ingelheim Pharma GmbH & Co. KG, Germany
[2]Warwick Medical School, University of Warwick, UK

In order to obtain approval by, for example, the FDA, the manufacturer has to provide substantial evidence of the effectiveness of the new drug. This is often interpreted as conducting at least two Phase III trials. However, exemptions can be made where one trial will suffice [1].
Preliminary work has been undertaken by Shun et al. [2] on the difference between conducting one versus two trials in Phase III. However, they only focus on single stage trials. We therefore extended their considerations to more flexible trial approaches such as adaptive seamless Phase 2/3 designs where the aim is to combine one Phase 2 with one Phase 3 trial [3]. We investigate different ways to combine the data obtained in Phase 2 with the data from the two Phase 3 trials.
By deriving expressions for frequentist error rates and sample sizes for innovative design approaches, optimal design strategies can be considered, both for a single homogeneous population and when there are differences between the patient populations in which different trials are conducted.
[1] FDA (1998): Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products.
[2] Shun Z, Chi E, Durrleman S and Fisher L (2005): Statistical consideration of the strategy for demonstrating clinical evidence of effectiveness—one larger vs two smaller pivotal studies. Statistics in Medicine, 24:1619-1637.
[3] Kunz CU, Friede T, Parsons N, Todd S and Stallard N (2015): A comparison of methods for treatment selection in seamless phase II/III clinical trials incorporating information on short-term endpoints. Journal of Biopharmaceutical Statistics, 25:170-198.

Latent Variable Modelling II

76 A Recent Perspective on Differential Item Functioning and its Implications in the Rasch model

Carsten Szardenings[1], Anna Doebler[2], Philipp Doebler[1]
[1]TU Dortmund, Germany
[2]Universität Mannheim

Differential item functioning (DIF) means that an item's response function depends on the population. Since observations alone do not identify the difficulty (or the response function) of an item, the diagnosis of DIF requires knowledge about the (differential) levels of the latent variable between populations. Bechger and Maris (2015) and Doebler (2018) proposed methods for measuring DIF or rather an analogue concept that does not rely on this knowledge. Instead of looking at individual items, both look at differences of differences between items and populations. While Bechger and Maris (2015) provided a statistical test, Doebler (2018) derived descriptive measures to do quantify the amount of differential item heterogeneity within a test. We applied the test and these measures to data from the National Educational Panel Study (NEPS) and assessed the implications of DIF for parameter estimation and further analyses in a simulation study.

77 Adaptive Bayesian SLOPE – High-dimensional Model Selection with Missing Values

Malgorzata Bogdan[1], **Wei Jiang**[2,3], Julie Josse[2,3], Blazej Miasojedow[4], Veronika Rockova[5]
[1]University of Wroclaw, Poland
[2]Inria Saclay, France
[3]Ecole Polytechnique, France
[4]University of Warsaw, Poland
[5]University of Chicago Booth School of Business

Model selection with high-dimensional data becomes an important issue in the last two decades. With the presence of missing data, only a few methods are available to select a model, and their performances are limited. We propose a novel approach – Adaptive Bayesian SLOPE, as an extension of sorted l_1 regularization but in Bayesian framework, to perform parameter estimation and variable selection simultaneously in high-dimensional setting. This methodology in particular aims at controlling the False Discovery Rate (FDR). Meanwhile, we tackle the problem of missing data with a stochastic approximation EM algorithm. The proposed methodology is further illustrated by comprehensive simulation studies, in terms of power, FDR and bias of estimation.

78 Statistical methodologies for handling ordinal longitudinal responses with intermittent missingness

Omololu Stephen Aluko, Birhanu Ayele
Stellenbosch University, South Africa

The survival rate of human immunodeficiency virus infected persons continue to improve with the utilization of highly active antiretroviral therapy (HAART), but the prevalence of axial emphysema distribution is increasing unabated. Axial emphysema distribution was a latent ressponse variable, but associated with intermittent missingness. Each method applied jointly with multiple imputation to analyze and predict the rate of prevalence on HIV-infected persons. The methods include ordinal negative binomial, mixed-effects proportional odds, multiple imputation based generalized estimating equation, and direct likelihood. The usefulness of ordinal negative binomial is demonstrated through the simulation study and empirical dataset; and performs creditably well

Mathematical Statistics II

79 Super-Consistent Estimation of Points of Impact in Nonparametric Regression with Functional Predictors

Dominik Poß[1], **Dominik Liebl**[1], Alois Kneip[1], Hedwig Eisenbarth[2], Tor Wager[3], Lisa Feldman Barrett[4,5]

[1]University Bonn
[2]Victoria University of Wellington
[3]University of Colorado Boulder
[4]Northeastern University
[5]Harvard Medical School

Predicting scalar outcomes using function-valued predictor variables is a classical problem in functional data analysis. In many applications, however, only specific locations or time-points of the functional predictors have an impact on the outcome. The selection of such points of impactconstitutes a particular variable selection problem, since the high correlation in the functional predictors violates the basic assumptions of existing high-dimensional variable selection procedures. In this paper we introduce a nonparametric regression model with functional predictors evaluated at unknown points of impact which need to be estimated from the data. We propose a threshold-based and a fully data-driven estimator and derive the convergence rates of our point of impact estimators. The finite sample properties of our estimators are assessed by means of a simulation study. Our methodology is motivated by a psychological case study in which the participants were asked to continuously rate their emotional state while watching an affective online video on the persecution of African albinos.

80 Simultaneous confidence bands for the covariance kernel of Banach space valued functional data

Melanie Birke[1], Christoph Reihl[1], Hajo Holzmann[2]

[1]Universität Bayreuth, Germany
[2]Philipps-Universität Marburg, Germany

There exist already some approaches for constructing confidence regions for the mean, the covariance kernel or eigenfunctions of functional data. Those methods are developed for dense as well as sparse observational schemes. But must of the methods are based on the weak convergence in a Hilbert space.
In this talk we present a method for constructing uniform confidence bands for the covariance kernel of Banach space valued functional data, that is where functional data are situated in the space $(C([0,1]), ||\cdot||_\infty)$ and the covariance kernel is in the space $(C([0,1]^2), ||\cdot||_\infty)$. Starting in a model where functional data are observed on a dense grid with additional observation errors in each grid point we construct a local linear estimator for the covariance kernel. As main result we show the weak convergence of this estimator in $(C([0,1]^2), ||\cdot||_\infty)$ from which the construction of an asymptotic simultaneous confidence band directly follows. But this is not the only possibility for application of the weak convergence result. It can e.g. also be used for constructing statistical tests.
Besides the theoretical results we also show a simulation study to show the finite sample behavior. Because of the difficult structur of the asymptotic distribution and because the asymptotics might not be valid for moderate sample sizes we propose a bootstrap method in this part.

81 Fused Density Estimation on Infrastructure Networks

Robert Bassett[1], James Sharpnack[2]

[1]Naval Postgraduate School
[2]University of California Davis

We introduce a new method for nonparametric density estimation on geometric networks. By penalizing maximum likelihood estimation with a total variation penalty, we avoid overfitting and the dirac curse. We provide results which reduce the search space for the estimator from infinite dimensional function space to the finite-dimensional setting, and further demonstrate its computational tractability. We then focus on the asymptotic convergence rate of this density estimation method. Lastly, we review applications to infrastructure networks.

82 Empirical Regularized Optimal Transport: Statistical Theory and Applications

Marcel Klatt, Carla Tameling, Axel Munk
University of Göttingen, Germany

The theory of optimal transport (OT) has a long history in physics, mathematics, economics and related areas and dates back at least to the work of the French mathematician Gaspard Monge during the 18th century. OT is concerned with the problem to find the most efficient way to transport mass from one location to another. In fact, this basic principle has paved the way to reconsider several concepts in statistics, e.g. multivariate quantiles or distances on the space of probability distributions. To illustrate the latter, given two probability distributions, one which can be thought of as a pile of sand and the other as a hole, the distance between them is the minimal total amount of work required to transport all the sand into the hole. This distance is commonly known as the OT or Wasserstein distance between probability distributions and has several advantages compared to competing distances like total variation.
However, despite its conceptual appeal and practical success, the routine use of OT is still hampered by its computational complexity. Hence, regularized OT has encountered a growing interest and encouraged the development of various surrogates which are computationally better accessible. Among others, the most prominent proposal is entropy regularization of OT that serves to define a regularized OT distance also known as Sinkhorn divergence.
We derive limit distributions for certain empirical regularized OT distances between probability distributions supported on a finite metric space and show consistency of the (naive) bootstrap. In particular, we prove that the empirical regularized transport plan itself asymptotically follows a Gaussian law. The theory includes the entropy regularization and hence a limit law for the widely applied Sinkhorn divergence.
Our approach is based on an application of the implicit function theorem to necessary and sufficient optimality conditions for the regularized transport problem. The asymptotic results are investigated in Monte Carlo simulations. We further present computational and statistical applications, e.g. confidence bands for colocalization analysis of protein interaction networks based on regularized optimal transport.

Small Area Analysis and Spatial Statistics III

83 Joint spatial modelling of disease outcomes of Chilean survey data

Anna Schritz[1], Andrew Lawson[2], Gloria Aguayo[1]

[1]Luxembourg Institute of Health, Luxembourg
[2]Medical University of South Carolina, USA

We propose a general approach to the analysis of multivariate health outcome data where geo-coding at different spatial scales is available.
Individual participants data of the Chilean National Health Survey (CNHS) as well as outcomes aggregated to Chilean province level where used to estimate prevalences of multiple disease outcomes in a multi-scale Bayesian model. All outcomes, diabetes, obesity, hypertension and elevated LDL, are related to the metabolic syndrome and therefore, correlations between the outcomes were assumed. A joint model approach was used to add links between the diseases and also to allow for correlation between areas. Different random, common random and spatial effects were therefore added to the models. This means, that models for each outcome, once on the individual data level and once on the aggregated data level, were all run during the same Markov Chain Monte Carlo iterations. At the same time, survey weights were used for adjustment and missing values in outcome and predictor variables were imputed using predictive distributions and suitable prior distributions.
Estimated disease prevalences were visualized separately on a map of Chile and disease hotspots were observed using exceedance probability.

84 Local economic impact of universities

Britta Stöver

Institut für Statistik an der LUH, Germany

Universities are important economic actors in affecting the local economy in two ways: on the demand side they consume labour and materials for the provision of education and administrative tasks inducing direct and indirect multiplier effects. On the supply side they generate human capital or highly qualified manpower respectively by educating students. Additionally, they support innovation and development with their research.
The quantification of the significance and the economic value of universities for their respective region is of interest for local politicians as well as the universities themselves: it helps to legitimate public funds and the use of tax money, can be used as image campaign or supports reform or investment programmes. As a consequence, there already exist a variety of different publications evaluating the economic importance – and especially the labour market effects – of universities in Germany.
However, the published studies are almost all single case studies with limited comparability: They focus on selected single or groups of universities, apply different definitions of the spatial dimensions and operate with different statistical and empirical methods. Thus, the comparison and classification of the single results is a complex task and therefore seldom conducted. Another drawback is that the majority of the studies focus on the demand side neglecting the positive impact on qualification and innovation.
The aim of this study is to quantify, compare and classify the different economic demand and supply side contributions of each university location within Lower-Saxony.
The economic impact of the university locations given by the demand side is analysed by three different perspectives: importance, dynamics and interaction. The importance of the university locations for their respective local economy and in comparison with each other is assessed by an indicator. The dynamic and change of the importance of the different university locations is shown using a shift-share analysis. Input-output-based employment multipliers are estimated to display the interaction of the university locations with the local economy.
For the supply side effects different regressions based on a panel model are conducted that establish the link between the university locations and the respective labour market, economic performance as well as innovation and knowledge.
On the whole, the applied measures provide the opportunity to analyse and compare the significance of university locations and help to classify different types of university locations.

85 MikroSim – Sektorenübergreifendes kleinräumiges Mikrosimulationsmodell

Markus Zwick[1], **Ralf Münnich**[2], Johannes Kopp[2], Petra Stein[3], Rainer Schnell[3]
[1]Statistisches Bundesamt
[2]Universität Trier
[3]Universität Duisburg-Essen

In der heutigen Zeit ist der Bedarf an statistischen Daten und Auswertungsmethoden so groß wie nie zuvor. Phänomene wie der demographische Wandel, regionale, überregionale und internationale Wanderungsbewegungen wirken sich erheblich auf ländliche sowie städtische Gebiete aus und ergeben enorme Herausforderungen im Bereich Policy Oriented Research. Methoden der Mikrosimulation werden bereits seit vielen Jahren erfolgreich erforscht und zur umfassenden Untersuchung komplexer gesellschaftlicher Entwicklungen sowie zur Wirkungsanalyse politischer Maßnahmen eingesetzt. Nach nunmehr fast zwei Jahrzehnten findet Mikrosimulation in Deutschland wieder vermehrt Beachtung. In diesem Kontext wird eine neue Forschungsgruppe namens Sektorenübergreifendes kleinräumiges Mikrosimulationsmodell (MikroSim)“ von der Deutschen Forschungsgemeinschaft gefördert. Das Projekt wird gemeinsam von den Universitäten Trier und Duisburg-Essen unter Beteiligung des Statistischen Bundesamts durchgeführt. Es läuft seit 1. September 2018 zunächst für drei Jahre.
Ziel der DFG Forschungsgruppe MikroSim ist es, eine umfassende Mikrosimulationsinfrastruktur für Deutschland aufzubauen. Hierbei wird eine geeignete Datenbasis geschaffen, die es erlaubt, Auswertungen regional tief gegliedert durchzuführen. Darüber hinaus werden Übergangswahrscheinlichkeiten geschätzt, die es erlauben, die Datenbasis über die Simulation soziodemographischer Ereignisse, in die Zukunft zu projizieren. Dabei wird auf die Daten von bislang 36 verschiedenen Erhebungsprogrammen zurückgegriffen. Darüber hinaus werden konkrete Fragestellungen zu den Themenkomplexen der Integration von Migrantinnen und Migranten auf dem Arbeitsmarkt und zur Pflegebedürftigkeit unter Berücksichtigung familiärer Prozesse über die geschaffenen Strukturen erörtert.
Eines der Anwendungsfelder des Forschungsvorhabens ist die Modellierung der zukünftigen Entwicklung beruflicher Integration von Migranten. Zur Erfassung von Integrationsentwicklungen muss das Zusammenspiel individueller Einflussfaktoren mit der Entwicklung regionaler Disparitäten erforscht werden. Da sich kontextuellen Bedingungen wandeln können, ist eine regionaldynamische Modellierung und Fortschreibung sowie eine Verbindung dieser Entwicklungen mit dem demographischen Wandel unabdingbar. Demographisch bedingte Veränderungen können nämlich – aufgrund von Kompositionseffekten – den Integratiosstand zusätzlich zu kausalen und regionalen Faktoren beeinflussen. Diese Effekte sind sowohl auf gesamtdeutscher Ebene, als auch im regionalen Kontext zu modellieren, da sie je nach betrachteter räumlicher Einheit, unterschiedlich ausfallen können. Die aus der empirischen Analyse dieser Entwicklungsdynamiken gewonnenen Erkenntnisse sollen die Entwicklung realitätsnaher zukünftiger Integrationsszenarien ermöglichen, die dann mithilfe der Mikrosimulation in die Zukunft projiziert werden.
Ein weiteres Anwendungsfeld von MikroSim stellt der Bereich der Familie und Pflege, mit besonderem Fokus auf familialen Pflegepotentialen, dar. Um diese Problematik empirisch zu behandeln, müssen die verschiedensten sozialen Prozesse parallel in den Blick genommen werden.

Dass Familie und Pflege nicht zufällig gemeinsam untersucht werden, wird verständlich, wenn man in Rechnung stellt, dass immer noch ein Großteil der pflegebedürftigen Menschen in Deutschland ganz oder zumindest teilweise von Familienangehörigen versorgt und gepflegt wird. Will man also den zukünftigen Bedarf an Pflegekräften, entsprechenden Einrichtungen oder die Kosten für Pflege in Deutschland auch nur einigermaßen belastbar voraussagen, so kann das nur unter Berücksichtigung sich wandelnder familialer Strukturen und dem aus ihnen resultierenden informellen Pflegepotentials geschehen. Zudem soll mit Hilfe der sektorenübergreifenden kleinräumigen Mikrosimulationsmodelle auch simuliert werden können, welche Veränderungen verschiedene (politische) Maßnahmen, wie beispielsweise eine steigende finanzielle Subventionierung von Familien, bei den beschriebenen Phänomenen auslösen.

86 Data-driven Transformations for the Estimation of Small Area Means

Nora Würz[1], Timo Schmid[1], Nikos Tzavidis[2]

[1]Freie Universität Berlin, Germany
[2]University of Southampton, United Kingdom

For many surveys, the problem of small sample sizes within (certain) subpopulations arises. Small area estimation is a powerful tool to overcome this problem. As small area models rely on linear mixed models, for example, the Gaussian assumption of the error terms must hold. In real applications for many variables, like income, this assumption is often not satisfied. Therefore, this work focuses on tackling the potential lack of validity of the model assumptions by using transformations for the dependent variable in the context of restricted data access.
When the register covariates are available on unit-level ad-hoc chosen and data-driven transformations for adjusting the underlying data have been used in literature. However, in many applications, like in Germany, the register covariates are only available on the aggregated level. Therefore, we propose small area methods with (data-driven) transformations in situations when aggregated register information is only available. We have to consider a bias-correction a) due to the back-transformation of the dependent variable and b) due to the aggregated register information (covariates). If the covariates underlie a normal distribution, we can directly estimate for a given mean and standard derivation the bias-correction for the area-level covariates. However, in most real situations we cannot assume that the underlying distribution for the covariates is known. Therefore, we estimate the distribution of the covariates from the available sample. Next to point estimation, Mean Squared Error estimation of the proposed approach is also discussed. Extensive model-based simulations are used for comparing the presented methodology to alternative unit-level methodologies for estimating small area means. Finally, the need for such methods and their application is then illustrated in a poverty mapping using real survey and census data from Mexico.

Statistics in Agriculture and Ecology I

87 Point processes — abstraction and practical relevance in ecology

Janine Baerbel Illian
University of St. Andrews, United Kingdom

All statistical modelling of complex data structures involves an abstraction to the essential properties of interest into quantifiable units and associated random variables. In addition, it also often goes along with simplifying assumptions as part of the abstraction process, typically for practical reasons. As a result, methodology can tend to be far removed from reality and hence be of little practical relevance.
In the context of point process modelling, the usual abstraction reduces the available information to locations of individuals — or points — in space, whose spatial structure is analysed. Classical simplifications often concern assumptions of homogeneity, isotropy and known detection probabilities, often for computational reasons. Recent computational improvement however, allows us to relax some of these assumptions.
This talk provides a number of examples of how we have been able to relax these classical assumptions along with associated abstractions, leading to increased practical relevance for ecological data structures. In particular, I will discuss how this increased practical relevance has also caused an increasing demand for the development of new methodology that has previously played a rather minor role.

88 A continuous-time multi-state capture-recapture model for the annual movement of bottlenose dolphins on the east coast of Scotland

Sina Mews[1], Roland Langrock[1], Nicola Quick[2], Ruth King[3]

[1]Bielefeld University, Germany
[2]University of St Andrews, Scotland
[3]The University of Edinburgh, Scotland

Our modelling approach is motivated by individual sighting histories of bottlenose dolphins off the east coast of Scotland. Specifically, due to ongoing offshore development, conservation managers seek to better understand the temporal movement patterns of the dolphin population between different sites. Typically, the Arnason-Schwarz model is fitted to such multi-state capture-recapture data, assuming a first-order Markov chain in discrete time for the state process, which here corresponds to the location (site) of a given individual. In our case, however, the capture occasions are not regularly spaced in time, rendering the standard capture-recapture methods inapplicable as they address the more commonly found regular sampling protocols. Therefore, we consider a continuous-time model formulation instead. The capture-recapture setting can be regarded as a special case of a (partially) hidden Markov model (HMM), with the observed capture history of an individual as the state-dependent process and the state process corresponding to the true but only partially observed movement of the individual between the sites. In particular, we can exploit the convenient and efficient HMM-based forward algorithm for evaluating the likelihood and hence for parameter estimation. Further inferential tools that become applicable by embedding the capture-recapture setting in the HMM framework include the Viterbi algorithm and the forward-backward algorithm, which can be used to decode the underlying states (sites).
The main aim of the present analysis was to investigate how the dolphins' movement rates between two sites, expressed as state transition intensities in our model, depend on the time of year. The incorporation of time-varying covariates into the continuous-time Markov state process is however rather challenging as the corresponding likelihood function then becomes intractable in general. We suggest an approximation using piecewise constant state transition intensities, which renders the likelihood evaluation feasible. The approximation can be made arbitrarily accurate by using an increasingly fine resolution of the approximating step function. The suggested approach is applied to investigate the annual movement of bottlenose dolphins between their main sites on the east coast of Scotland, revealing seasonal patterns which can help to inform conservation management. Our modelling approach can easily be transferred to other scenarios and is hence a general method for irregularly sampled capture-recapture data subject to switches in underlying states.

89 A Coefficient of Determination (R2) for Generalized Linear Mixed Models

Hans-Peter Piepho

Universität Hohenheim, Germany

Extensions of linear models are very commonly used in the analysis of biological data. Whereas goodness of fit measures such as the coefficient of determination (R2) or the adjusted R2 are well established for linear models, it is not obvious how such measures should be defined for generalized linear and mixed models. There are by now several proposals but no consensus has yet emerged as to the best unified approach in these settings. In particular, it is an open question how to best account for heteroscedasticity and for covariance among observations induced by random effects. This paper proposes a new approach that addresses this issue and is universally applicable. It is exemplified using three biological examples.

Survival and Event History Analysis I (Non-standard Sampling)

90 Semiparametric Modeling of Doubly Truncated Lifetimes in Registry Data

Achim Dörre

University of Rostock, Germany

Doubly truncated lifetimes occur in registry data when individuals are sampled conditional on a terminating event, e.g. disease onset of a patient, insolvency of a company or machine failure. There are multiple potential applications in areas including medicine, economics and engineering. This particular sampling scheme is studied within a multidimensional Poisson process modeling framework, where the emergence of individuals over calendar time is thought of as a birth process with unknown shape. We overcome specific drawbacks of existing nonparametric methods for this data type by using a semiparametric approach. In particular, weighted Gamma processes are used to construct a viable Bayesian model. We illustrate the proposed methods via real datasets.

91 Left-censoring in survival analysis: An application to dementia incidence

Rafael Weissbach, Achim Doerre, Anne Fink, Gabriele Doblhammer

University of Rostock, Germany

We obsere a simple sample of 250.000 persons over a study period of 10 years. Our aim is dementia incidence and its development over cohorts, adjusted for age and sex. Persons not developing dementia within the study period are right-censored, persons already having dementia at the beginning of the period are left-censored. With the parameters of a multiplicative hazard model, parameters for censoring and of covariates are nuisance. The parameters of the covariates are eliminated by conditioning, the parameters of censoring are elininated by the factorization of the conditional likelihood. The asymptotic covariance matrix for M-estimation simplifies due to the conditional information matrix equality, so that confidence intervals are easily available. However, some effort goes into the proof of the required estimator consistency as neither results for the GLM, nor a convex likelihood do not apply.

92 Time-simultaneous inference in general nested case-control designs

Jan Feifel[1], Dennis Dobler[2]

[1]Institute of Statistics, Ulm University, Helmholtzstrasse 20, 89081 Ulm, Germany
[2]Department of Mathematics, Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1881 HV Amsterdam, The Netherlands

Hospital-acquired infections propose a major burden to clinicians and patients, but also statisticians face an interesting task as usually rare events within large cohorts are of interest.
In time-to-event analyses, information on the actual event times is only provided by uncensored patients. If the outcome is rare or if interest lies in evaluating expensive covariates, nested case-control designs are attractive. The appeal is that those designs restrict for all uncensored patients to a small number of controls. These controls are representative for all patients at risk just prior to the observed event time. As a result, the observed outcomes are oversampled and the nested case-control sample is no longer independent identical distributed. Therefore, common resampling procedures relying on Efron's bootstrap are inappropriate. In the special case of nested case-control designs with simple random sampling, confidence bands for the cumulative baseline hazard function have been proposed.
However, the martingale structure behind nested case-control designs allows for more powerful subset sampling designs. While respecting this martingale structure and exploiting its merits, we are able to propose time-simultaneous inference methods based on wild bootstrap resampling procedures within this general class of designs. We present several transformed confidence bands for an investigation of the risk factors for hospital-acquired pneumonia in Germany. First simulations show that confidence bands outperform the confidence intervals when being interpreted time-simultaneously. These results underpin the value of time-simultaneous procedures for functions like the cumulative baseline hazard in sound statistical inference.

93 On modeling complex longitudinal and survival data with a terminal trend

Kwun Chuen Gary Chan

University of Washington, United States of America

Recurrent event processes with marker measurements are mostly studied with forward time models starting from an initial event. Interestingly, the processes could exhibit important terminal behavior during a time period before occurrence of the failure event. A natural and direct way to study recurrent events prior to a failure event is to align the processes using the failure event as the time origin and to examine the terminal behavior by a backward time model. I will present some recently published models and their extensions, with emphasis on semiparametric models with unspecified baseline functions in multiple time scales.

Statistical Literacy and Statistical Education III

94 Increase in the speed of medical decisions due to natural frequencies

Leah Braun[1], Karin Binder[2]

[1]Ludwig Maximilian University of Munich, Germany
[2]University of Regensburg, Germany

When physicians are asked to determine the positive predictive value from the a priori probability of a disease and the sensitivity and false positive rate of a medical test, it often comes to misjudgments with serious consequences. Two strategies have proven helpful in such Bayesian tasks: 1) Provide statistical information in the format of natural frequencies instead of probabilities (Gigerenzer & Hoffrage, 1995) and 2) Visualizations, e.g., tree diagrams (Binder, Krauss & Bruckmaier, 2015). For visualizations that contain numerical information, the format of the numerical information is decisive: While frequency trees improve understanding, probability trees are of little help (Binder et al., 2015, 2018). This is astonishing because both medical textbooks and school textbooks predominantly contain explanations with probabilities.

In daily clinical practice, however, it is not only important that doctors and patients receive a tool with which they can correctly judge – the speed of these judgments is also a crucial factor, as there is little time left in everyday medical practice for decision-making. The decisive factor is therefore diagnostic efficiency, which is the quotient of the correctness of the diagnosis and the speed of the diagnosis.

In the lecture we will present a study with 111 medical students (Braun & Binder, in preparation), which shows that judgments with natural frequencies and natural frequency trees are not only more frequently correct, but also faster. The diagnostic efficiency of doctors is thus favored by natural frequencies and natural frequency trees by two factors: Through more correct judgements (which has already been investigated many times in the past) and faster judgments (which is an innovative question).

As in previous studies, higher solution rates are observed in the natural frequency versions and in the versions with a visualization. Furthermore, with regard to time on task, it can also be seen that both natural frequencies and frequency trees support a fast solution finding. If both parameters are combined and the diagnostic efficiency is calculated from the correctness of the solution and the speed of judgement, the two helpful factors "natural frequencies" and "natural frequency trees" are reinforced.

It is quite time-consuming to teach natural frequencies and natural frequency trees.

But when considering how much time is spent in schools and universities teaching probabilities and probability trees, and how little it helps with Bayesian thinking, it becomes clear that the journey is worth it. Frequency trees can help to better understand probabilities and enable not only more accurate but also faster judgments in Bayesian reasoning.

Furthermore, improving correctness and efficiency of medical students in Bayesian tasks might help to decrease over-diagnosis and over-treatment in daily clinical practice, which on the one hand cause cost and on the other hand might endanger patients' safety.

95 The development of epiLEARNER: an innovative e-learning project by and for medical students

Ursula Berger[1], Cornelia Oberhasuer[1,2], Michaela Coenen[1,2]

[1]Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie, LMU München, Germany

[2]Lehrstuhl für Public Health und Versorgungsforschung, LMU München, Germany

A successful medical education includes epidemiology and statistics. In order to understand the results of medical studies and to correctly interpret treatment effects of different therapies, but also to be able to assess risks and to make evidence based diagnostic and therapeutic decisions, physicians need a basic understanding of statistics and of the epidemiological and statistical methods most commonly used in medical research. Good statistical literacy is also crucial for the informative communication with patients and the patients briefing. In medical curricula, in addition to preclinical and clinical education, science orientation and the training of scientific skills are becoming increasingly important. However, it is usually difficult to get medical students interested in epidemiological and statistical topics.

To make epidemiology and statistics more accessible for medical students and to spark their interest in these topics, the e-learning app epiLEARNER was developed as part of a project at the Ludwig-Maximilians-Universität (LMU) in Munich. The idea of the project was to provide a comprehensive and practice-oriented, modular learning platform as a supplement to classroom courses, which enables medical students to develop good statistical literacy. The focus was thereby on concrete practical patient examples, which motivate the relevance of statistical knowledge in the medical profession.

The epiLEARNER consists of three core elements: (1) Interactive case studies with quizzes demonstrate and train the use of statistical skills in medicine. (2) A concise summary of the theory helps to understand and to deepen different epidemiological and statistical topics. (3) Multiple-choice questions allow to test the knowledge and to prepare for the exam.

The talk will present the concept of the epiLEARNER, will give an insight into its structure and discuss first evaluation results.

96 Challenges in teaching Medical Data Science

Michael Gabel, Marietta Kirchner, Lorenz Uhlmann, Maximilian Pilz, Dorothea Weber, Meinhard Kieser

Institut für medizinische Biometrie und Informatik, Germany

Our modern time has seen a tremendous increase of virtually all kinds of data. For many fields, the availability of Big Data holds the promise to answer questions that would have been out of reach just a couple of years ago. The analysis of Big Data, however, requires expert knowledge and skills, which is why data science is becoming an increasingly important topic. This is especially true for data evaluation in medicine, where statistical methods for data analysis need to be applied to an already complex and heterogeneous field. While there are many courses available that explain the basics of data science, teaching the relevant techniques with real-life application to medical data is currently still a niche subject.

To fill this gap, the Institute of Medical Biometry and Informatics at the University of Heidelberg will be offering a certified course that introduces and deepens the essentials of medical data science. The course is structured into four different modules, in which the participants will learn useful methods, such as statistical modelling and machine learning, to answer clinically relevant questions by analyzing Big Data. As both the statistical and medical topics are quite complex, providing an effective learning approach that is suitable for a heterogeneous audience is challenging. In our talk, we will therefore focus on the difficulties that arose during the development of the course. We will highlight the process of how we came up with satisfying teaching strategies that transmit theoretical knowledge in an applied context.

97 Regression model building in medical statistics

Paul Bach[1], Lorena Hafermann[1], Geraldine Rauch[1], Nadja Klein[2]

[1]1 Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Institut für Biometrie und Klinische Epidemiologie, Berlin, Germany
[2]Humboldt-Universität zu Berlin, Berlin, Germany

Motivation. In medical research such as clinical trials and observational cohorts, multivariable regression models are routinely applied 1) to identify markers associated with specific disease outcomes, 2) to adjust for confounders and 3) to predict specific disease outcomes. In the last decades, the statistical research regarding model building strategies has developed rapidly. In medical applications, however, this statistical knowledge is insufficiently integrated. Most often, only linear relations between the independent variables and the outcome of interest are considered, possible interactions between different variables are ignored, too many variables are included in a model with respect to the effective sample size and suboptimal variable selection techniques are applied [1,2]. Bearing in mind that erroneous conclusions may have severe consequences in the medical context, this is particularly inacceptable. Therefore, there is an urgent need to transfer the statistical knowledge into applied routine. This can be achieved by providing understandable guidance and recommendation documents published in the medical literature to guarantee easy access.
Aim & Method. To systematically assess which aspects of model building strategies are communicated to applied medical researchers, we conducted a systematic review with the aim to identify statistical series within medical journals which intend to train applied medical researchers with respect to their statistical knowledge. Within the identified statistical series, we searched for articles and single paragraphs providing recommendations towards model building strategies.
Results. We identified a clear gap between the statistical state of the art and the recommendations transferred to more applied researchers in the field of regression model building. Hardly any easy accessible guidance could be identified in the screened articles. The results of our systematic review show in detail which topics are ignored or unsatisfactorily addressed by the literature. We conclude with a very detailed demand on future guidance development.
Conclusion. There is an urgent need to close the gap between the statistical state of the art in regression model building and the common practice in applied medical research. Easy accessible and understandable guidance on regression model building needs to be provided.
References
1. W. Sauerbrei; M. Abrahamowicz; D. G. Altman; S. Le Cessie; J. Carpenter (2014): STRengthening Analytical Thinking for Observational Studies: the STRATOS initiative. Statistics in Medicine. 33(30): 5413–5432.
2. G. W. Sun, T. L. Shook, G. L. Kay (1996): Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. Journal of Clinical Epidemiology, 49: 907-916.

Time Series Analysis I (Time Series Econometrics)

98 On the Sensitivity of Granger Causality to Errors-in-Variables, Linear Transformations and Subsampling

Brian D O Anderson[1], **Manfred Deistler**[2], Jean-Marie Dufour[3]

[1]Australian National University, Canberra, Australien
[2]Technische Universität Wien, Osterreich
[3]McGill University, Montreal, Canada

This article studies the sensitivity of Granger causality to the addition of noise, the introduction of subsampling, and the application of causal invertible filters to weakly stationary processes. Using canonical spectral factors and Wold decompositions,we give general conditions under which additive noise or filtering distorts Granger-causal properties by inducing (spurious) Granger causality, as well as conditions under which it does not. For the errors-in-variables case, we give a continuity result, which implies that: a „small“noise-to-signal ratio entails „small“distortions in Granger causality. On filtering, we give generalnecessary and sufficient conditions under which „spurious“causal relations between (vector) time series are not inducedby linear transformations of the variables involved. This also yields transformations (or filters) which can eliminate Grangercausality from one vector to another one. In a number of cases, we clarify results in the existing literature, with a number ofcalculations streamlining some existing approaches.
Keywords: Granger causality; sensitivity; signal-to-noise ratio; errors-in-variables; measurement errors; filtering; subsampling. MOS subject classification: 62M10.

99 Fractional trends in unobserved components models

Tobias Hartl[1,2], Rolf Tschernig[1], Enzo Weber[1,2]

[1]University of Regensburg, Germany
[2]Institute for Employment Research

We develop a generalization of permanent transitory decompositions that avoids prior assumptions about the long-run dynamic characteristics by modeling the permanent component as a fractionally integrated process. The model neither requires stationarity nor orthogonal permanent and transitory shocks and is cast in state space form. In a multivariate setup, fractional trends may exhibit different orders of integration but depend on the same stochastic shocks, leading to a cointegrated system with different integration orders. For both, the univariate and the multivariate case, we derive the likelihood function. In an application to U.S. economic data, we find that a fractional process better describes the long-run dynamics of interest rates and inflation than a random walk, whereas income and consumption are I(1). In a multivariate setup, we find that interest rates and inflation are cointegrated, but exhibit different fractional integration orders.

100 The Asymptotic Validity of "Standard" Fully Modified OLS Estimation and Inference in Cointegrating Polynomial Regressions

Oliver Stypka[1], **Martin Wagner**[1,2], Peter Grabarczyk[1], Rafael Kawka[1]
[1]TU Dortmund, Germany
[2]Institute for Advanced Studies, Vienna

The paper considers estimation and inference in cointegrating polynomial regressions, i.e., regressions that include deterministic variables, integrated processes and their powers as explanatory variables. The stationary errors are allowed to be serially correlated and the regressors to be endogenous. We show that estimating such relationships using the Phillips and Hansen (1990) fully modified OLS approach developed for linear cointegrating relationships by incorrectly considering all integrated regressors and their powers as integrated regressors leads to the same limiting distribution as the Wagner and Hong (2016) fully modified type estimator developed for cointegrating polynomial regressions. The only restriction for this result to hold is that all integrated variables themselves are included as regressors. Key ingredients for our results are novel limit results for kernel weighted sums of properly scaled nonstationary processes involving powers of integrated processes and a functional central limit theorem involving polynomials of Brownian motions as both integrand and integrator. Even though simulation results indicate performance advantages of the Wagner and Hong (2016) estimator that are partly present even in large samples, the results of the paper drastically enlarge the useability of the Phillips and Hansen (1990) estimator implemented in many software packages.

101 Multivariate Testing for Fractional Integration

Marina Balboa[3], **Paulo M. M. Rodrigues**[1], Antonio Rubia[3], Robert Taylor[2]
[1]Universidade Nova de Lisboa, Portugal
[2]University of Essex
[3]University of Alicante

This paper introduces a new approach to test for (common) fractional integration in a multivariate time series context. The setting extends the results of Robinson (1994), Agiakloglou and Newbold (1994), Tanaka (1999) and Breitung and Hassler (2002), to a multivariate framework. The test procedures proposed can be easily implemented in empirical settings and are flexible enough to accommodate a broad spectrum of long- and short-memory specifications, including among others, VAR and/or multivariate GARCH-type dynamics, generalizing in this way also the approaches in Nielsen (2004, 2005). Furthermore, these tests have power against different types of alternative hypotheses and enable inference to be conducted under critical values drawn from a standard Chi-square distribution, irrespective of the long-memory parameters.

Young Statisticans

102 Selection Effects in Bayesian Hierarchical Models Bachelor Thesis in Cooperation with Boehringer Ingelheim Pharma GmbH & Co. KG

Martina Schlenker
University of Ulm, Germany

Bayesian hierarchical modeling is currently considered as a useful technique to incorporate borrowing into the analysis of several strata which may differ e.g. in indications or treatment doses. Such approaches could also be advantageous in a trial with two stages, where patients' responses are observed in several cohorts in the first stage, but only the cohort with the highest observed response rate is continued to the second stage. If the aim of the trial is to estimate the true response probability of the selected cohort using data from both stages, the selection procedure at interim may introduce bias.
The aim of this thesis is to evaluate and compare Bayesian hierarchical models as well as a Bayesian non-hierarchical method and a naive maximum likelihood approach with respect to bias and mean squared error of their resulting estimates in such a trial. For this purpose, a simulation study was conducted, considering various scenarios with different numbers of cohorts, response rates and sample sizes.
The bias is shown to come from the first stage data and increase for a growing number of strata, especially for those cohorts with a lower mean response rate. The findings also confirm the assumption that hierarchical modelling, particularly the model by Berry et al. (2013), which uses target correction, can improve estimation in terms of both bias and mean squared error.
However, the specification of appropriate prior parameters and targets rates is necessary to benefit from these approaches. Moreover, more sophisticated frequentist approaches may also reduce the selection bias. In addition, the applied selection rule is quite simple and could be extended in various ways.

103 Risks and benets of autologous stem cell transplantations in treating elderly patients with multiple myeloma: Competing risks analyses

Kaya Miah

Faculty of Statistics, Technical University of Dortmund

Multiple myeloma (MM) is a malign tumour disease of the blood building system. High-dose therapy (HDT) with melphalan followed by autologous stem cell transplantation (ASCT) is the standard of care for patients with newly diagnosed MM younger than 65 years. However, the median age at diagnosis is above 65 years and so far, there is only little evidence for the effectiveness and safety of HDT-ASCT in those patients. In particular, safety and risk assessment of ASCT for elderly patients is of great interest for the Medical Service of German Statutory Health Insurance providers. Thus, the main aim of the thesis is the evaluation of the impact of ASCT in treating elderly patients with newly diagnosed MM. To this end, prospective data of a randomised multicentre phase III trial comprising 604 patients with previously untreated MM are analysed with respect to a comparison of three age groups (>60 years, 61-65 years and 66-70 years). The primary endpoint of interest is progression-free survival (PFS), defined as time from randomisation to tumour progression or death from any cause, whichever occurs first. Additionally, retrospective data of an observational study of patients with newly diagnosed MM conducted at the University Hospital of Heidelberg should be used as supplementary external data and evaluated accordingly. With regard to the statistical analysis, we have to face at least four major challenges. First, ASCT was not performed by randomisation. Second, the treatment effect of ASCT is superimposed by the effect of the subsequent maintenance therapy. Third, the risks and benefits of ASCT must be assessed against each other. And fourth, the number of events is expected to be small due to reduced sample sizes in subgroups. With respect to non-randomised comparisons, we have to consider to include possible confounders in a multivariate Cox regression model in order to evaluate the effect of ASCT on PFS while accounting for differences between age groups. Regarding the risks of ASCT, the competing events of time to progression and non-relapse mortality as a possible consequence of ASCT have to be taken into account. Thus, modelling of competing risks as a generalisation of standard survival analysis of the combined endpoint PFS is of particular interest. In the models we consider maintenance therapy as a time-dependent intervention.As a result, conclusions are drawn from the combined analysis of prospective clinical trial data and retrospective observational data in order to evaluate the impact of ASCT in treating elderly patients with MM.

104 Challenging the commonly used log-link in statistical models for count data with an application to infection disease data

Aisouda Hoshiyar

Georg-August University Göttingen, Germany

A response function is an essential part of any generalized linear model, but its choice is rarely questioned. In particular, if the modeled expected value is restricted to be greater than zero, the choice often falls on the exponential function. Even for a response variable, for which the exponential function corresponds to the canonical link, there is no indication that this is the true response function in general. Therefore, we propose to take the softplus function as response function into consideration. The softplus function, which is technically used in the context of neural networks, enables the modeling of the conditional mean in an additive way and therefore ensures a linear interpretation of the regression coefficients while respecting the positivity boundary of the conditional mean at the same time. The central research question to be discussed in this study is: Does the softplus activating function represent an adequate substitute of the commonly used log-link with an application to infectious diseases? In the first step, a simulation study gives insight into the robustness of the estimated coefficients under various circumstances.Furthermore, the framework for the analysis of multivariate infection disease data yield by Held et al. (2005) is self-implemented via the open source software R. By doing so, the softplus function is introduced to the model class applied. The estimation results from Held et al. (2005) are reproduced and compared to those concerning the softplus link function with respect to the predictive quality. One-step-ahead-predictions build the basis for mean-squared prediction errors and coverage frequencies of the upper prediction limits. The results have been obtained using general optimisation routines via maximum likelihood estimation.

Acknowledgement

I thank Paul Wiemann, Georg-August-University Göttingen, Germany for his support during the master's thesis.

References

Held, L., Höhle, M., and Hofmann, M. (2005). A statistical framework for the analysis of multivariate infectious disease surveillance counts. Statistical modelling, 5(3):187-199.

105 Flexible instrumental variable distributional regression

Guillermo Briseño Sanchez[1], Maike Hohberg[2], Andreas Groll[1], Thomas Kneib[2]

[1]Technische Universität Dortmund, Germany
[2]Universität Göttingen, Germany

This work improves instrumental variable regression in nonlinear settings by adding distributional regression techniques. Standard instrumental variable regression as it is applied in both experimental and observational studies, suffers from two main limitations: Estimation is restricted to the conditional mean of the outcome, and the relationship between explanatory variables and outcome is assumed to be linear. More exible regression approaches that could solve those limitations have already been developed but have not yet been adopted in causality analysis. This work therefore develops an instrumental variable estimation procedure building on the framework of generalised additive models for location, scale and shape (GAMLSS). This allows for exibly modelling all distributional parameters of the potentially complex response distribution and nonlinear relationships between explanatory variables, instrument and outcome. A simulation study shows superior performance of the approach over the conventional two stage least squares estimator in nonlinear settings, and shows that the bias can be substantial when nonlinearities in the functional form of the covariates are ignored. We apply the method to challenge and complement the results of a seminal study on rural electrification that estimates the effect of electrification on female and male employment in the South African region of KwaZulu-Natal. Contrary to positive mean effects in the original over all communities, we find that the conditional distribution of an average community has a higher mean only for female employment and a decreased variance for both groups. The applied approach does not only close a gap from a statistical point of view but can provide an added value for policy recommendations.

Data Fusion and Meta-Analysis I

106 Clinical Prediction Models and the role of Evidence Synthesis

Thomas P.A. Debray
Julius Center, UMC Utrecht, The Netherlands

Clinical prediction models are an important tool in contemporary medical decision making and abundant in the medical literature. These models estimate the probability/risk that a certain condition is present or will occur in the future by combining information from multiple variables (predictors) from an individual, e.g. predictors from patient history, physical examination or medical testing. Unfortunately, many prediction models perform much worse than anticipated during their development. A major reason for unsatisfactory performance and limited use in clinical practice is that they are typically developed from relatively small datasets, and subsequently used in populations/settings too different from the original development population/setting, without proper validation and adaptation to the new situation. In this talk, I will discuss how we can investigate, quantify and improve the generalizability of prediction models by adopting formal strategies for evidence synthesis. I will highlight the potential advantages of undertaking a systematic review, and present statistical methods to build upon published evidence or multiple sources of individual participant data when developing or validating a prediction model.

107 Summray concordance index for meta-analysis of prognostic studies with survival outcome

Satoshi Hattori
Osaka University, Japan

Prognostic studies are widely conducted to investigate whether biomarkers are associated with patient's prognoses, and play important roles in mdeical decisions. To obtain more reliable findings from prgnostic studies, meta-analysis seems to be a promising approach. However, only limited number of methods are available to this end. Suppose we are interested in meta-analysis of prognostic studies with a time-to-event outcome for a continuous biomarker. In such prognostic studies, a study-specific cut-off value is defined and simple Kaplan-Meier and logrank analysis for two-sample comparison are often applied and reported in medical journals. In this research, we propose a method to estimate the concordance index for time-to-event based on such reported information. Application to some real data will be presented to illustrate the proposed method.

108 A nonparametric approach for meta-analysis of diagnostic accuracy studies with multiple cut-offs

Cornelia Frömke[1], Mathia Kirstein[1], Antonia Zapf[2]

[1]Hochschule Hannover, Germany
[2]Universitätsklinikum Hamburg-Eppendorf

The accuracy of a diagnostic test depends on two criteria: a) the sensitivity, being the proportion of positively diagnosed patients among all diseased patients under consideration and b) the specificity, being the proportion of negatively diagnosed patients among all non-diseased patients participating in the study.
According to the CEBM levels of evidence (https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/) it is not sufficient to test sensitivity and specificity in a single diagnostic accuracy study. Several studies are necessary and the results of these individual studies are summarized with a meta-analysis in a systematic review.
In contrast to therapeutic studies focussing most commonly on one endpoint, in a meta-analysis of diagnostic accuracy studies two criteria (both sensitivity and specificity) have to be taken into account. This is aggravated by the fact, that sensitivity and specificity are correlated and preferably the meta-analysis should model this correlation. Furthermore, a diagnostic test is often based on a biomarker. Hence, results for sensitivity and specificity are presented for multiple cut-offs within one study and – on top of that – the reported cut-offs vary among the individual studies.
As far as it is known to the authors, the present statistical approaches for the meta-analysis of diagnostic accuracy studies either a) do not take into account the multiple cut-offs both within and among the individual studies or the correlation between sensitivity and specificity, b) they are restricted to distributional assumptions or c) are iterative approaches and do not compute estimates for sensitivity and specificity when specific data structures occur.
Here, a nonparametric method for the meta-analysis of diagnostic accuracy studies considering the correlation structure of the two criteria as well as the multiple cut-offs within and among the individual studies will be presented.

Latent Variable Modelling III

109 A copula-based multivariate hidden Markov model for modelling momentum in football

Marius Ötting[1], Roland Langrock[1], Antonello Maruotti[2]
[1]Bielefeld University, Germany
[2]Libera Universita Maria Ss. Assunta, Italy

Sports commentators and fans frequently use vocabulary such as momentum", momentum shiftör related terms to refer change points in the dynamics of a match. Usage of such terms is typically associated with situations during a match where an event – such as a shot hitting the woodwork in a football match – possibly changes the dynamics of the match, e.g. in a sense that a team prior to the event had been pinned back in its own half suddenly seems to dominate the match. Using minute-by-minute in-game statistics of Bundesliga matches, we investigate whether such momentum shifts actually do exist in a football match and what kind of events lead to a shift.
For that purpose, multivariate time series $\{\mathbf{y}_{mt}\}_{t=1,2,\dots,T_m}$ are considered, where $\mathbf{y}_{mt} = (y_{mt1}, \dots, y_{mtK})$ is the vector of variables observed at time t (in minutes) during match m, $m = 1, \dots, 34$, with T_m denoting the total number of minutes played in match m. In our analysis, $K = 2$ variables are considered, namely the number of shots on goal and the number ball touches.
We consider hidden Markov models (HMMs) for modelling the minute-by-minute bivariate time series data, as they naturally accommodate the idea of a match progressing through different phases, with potentially changing momentum. HMMs involve an unobserved Markov chain with N possible states, denoted by $\{s_{mt}\}_{t=1,2,\dots,T_m}$, and an observed state-dependent process, whose observations are assumed to be generated by one of N distributions according to the Markov chain. Furthermore, within these (multivariate) HMMs, we allow for within-state correlation of our variables $\mathbf{y}_{mt}$ by formulating a bivariate distribution as state-dependent distribution using a copula, i.e.:

$$F(\mathbf{y}_{mt} \mid s_{mt}) = C\big(F_1(y_{mt1} \mid s_{mt}), F_2(y_{mt2} \mid s_{mt})\big),$$

where F_1 and F_2 are marginal distributions and C is a copula. Both the number of shots on goal and the number of ball touches are count variables, the marginal distributions of which we model using the Conway-Maxwell-Poisson distribution to account for possible over- and underdispersion. To account also for possible negative within-state correlation in $\mathbf{y}_{mt}$, the Frank copula is chosen here, but in principle other copulas are also possible. Using the forward algorithm to recursively evaluate the likelihood, we fit this copula-based bivariate HMM to data using numerical maximum likelihood estimation.
As a case study for assessing the potential of the multivariate-copula-HMM to investigate momentum in football, we analyse Bundesliga data from Bayer Leverkusen (season 2017/18) with $N = 3$ states. Preliminary results suggest a high persistence in the states, together with state-dependent distributions which refer to different playing styles (defensive only, predominantly defensive, and high pressure play). Current research focuses on including covariates in the state process, such that the probabilities of switching between the underlying states depend on (e.g.) the intermediate score of the match and the strength of the opponent. In addition, implementing different copulas and selecting the number of states is also currently investigated.

110 A Nonlinear Dynamic Latent Class Structural Equation Model

Augustin Kelava[1], Holger Brandt[2]

[1]Eberhard Karls Universität Tübingen, Germany
[2]University of Kansas, Lawrence, KS, USA

In this talk, we propose a nonlinear dynamic latent class structural equation model (NDLC-SEM). It can be used to examine intra-individual processes of observed or latent variables. These processes are decomposed into parts which include individual- and time-specific components. Unobserved heterogeneity of the intra-individual processes are modeled via a latent Markov process that can be predicted by individual-specific and time-specific variables as random effects. We discuss examples of sub-models which are special cases of the more general NDLC-SEM framework. Furthermore, we provide empirical examples and illustrate how to estimate this model in a Bayesian framework. Finally, we discuss essential properties of the proposed framework, give recommendations for applications, and highlight some general problems in the estimation of parameters in comprehensive frameworks for intensive longitudinal data.

111 A new varying threshold approach to model response styles in the IRT framework

Mirka Henninger
University of Mannheim, Germany

Responses to rating scale items do not only capture the respondent's trait, but the way he or she uses the rating scale. Examples of so-called response styles are extreme and midpoint response style as well as acquiescence. There exist a variety of IRT modeling approaches aiming at measuring and correcting for the heterogeneous use of the response scale. These models differ in their theoretical assumptions on response styles, and also in terms of their identification strategies. Promising IRT model candidates capture response tendencies by varying thresholds. However, in this group of models, random effects of threshold are restricted to be independent from each other. As covariances between varying thresholds are crucial in the modeling of response styles, this assumption is likely to be violated in empirical data and furthermore limits the interpretation of varying thresholds and their relation to the content trait. I therefore propose and test a new identification constraint for a varying threshold model that allows for the estimation of the full covariance matrix of random effects. In this model, the conditional probability of choosing category k when the response is either in category k or $k-1$ is given by:

$$\gamma_{nik} = \frac{\exp\left(\theta_n + \delta_{nk} - b_{ik}\right)}{1 + \exp\left(\theta_n + \delta_{nk} - b_{ik}\right)}$$

where θ_n is the respondent's trait parameter, δ_{nk} is a respondent-specific threshold shift reflecting response styles, and b_{ik} is an item-specific category parameter. In order to disentangle threshold variances δ_{nk} from trait parameters θ_n, a sum-to-zero constraint for δ_{nk} across thresholds within persons is imposed: $\sum_{k=1}^{K} \delta_{nk} = 0$. The new sum-to-zero constraint removes redundancy between θ_n and δ_{nk} and enforces interdependence in thresholds that is typically found in empirical data when, for example, extreme or midpoint response tendencies are present.

The results of two simulation studies demonstrated the applicability of the model with the new sum-to-zero constraint. First, when data in the population matched the identification strategy of the modeling approach, the simulation study showed that the new model with sum-to-zero constraint was a suitable alternative model comparable to an existing varying threshold approach that fixed covariances to zero. Second, when there were covariances in the generated data, the new model with sum-to-zero constraint outperformed the model with zero-covariance constraint in terms of parameter recovery. Furthermore, an analysis of empirical data supported the relevance of the new approach in real data settings: not only did the new model fit the data better than its competitor, but there were also substantial covariances between varying thresholds and between the traits and varying thresholds in the data that were captured by the new approach.

Taken together, the new varying threshold model with sum-to-zero constraint adds to the literature of IRT response style models. The model has an increased informative value as it offers the possibility to interpret the estimated covariances between content traits and varying thresholds. Most importantly, it is theoretically sound as it accounts for individual differences in response tendencies without a shift of the respondent's location on the latent continuum.

112 Identifying inattentive responses using dynamic latent class modeling

Holger Brandt[1], Zachary Roman[1], Mark Anderson[1], Augustin Kelava[2]
[1]University of Kansas
[2]University of Tuebingen

In this talk, a dynamic latent class modeling approach is presented that can be used to address inattention during testing procedures. Inattention is a serious problem that affects the validity of studies in general. It occurs if participants do not follow the instructions and do not respond carefully to items or questions. Instead, they might try to finish the questionnaire as fast as possible and use as few cognitive resources as possible. Recent methods that tried to account for inattention included presentation of bogus items, the investigation of the response pattern, and latent class analysis. The main limitation of these methods is that they assume that inattention is a constant trait throughout the testing procedure. This, however is implausible, and in most situations, persons might start attentive but become inattentive throughout the testing itself. The problem of inattention becomes more aggravated in long questionnaire batteries as they are often used in large scale data sets, in online platforms such as Amazon's Mturk, or in situations where participants are measured repeatedly (intensive longitudinal data). In this talk, a new method will be presented that is based on the dynamic latent class modeling framework. This method combines structural equation models with mixture modeling and time series analyses. It allows researchers to model inattention as a dynamic process that changes during testing procedures. Simulation studies and empirical data will be used to illustrate that this modeling approach can minimize bias due to inattention in a variety of scenarios including standard questionnaires and intensive longitudinal data. The results indicate that the method is sensitive to both which participants become inattentive and at which time point during testing.

Mathematical Statistics III

113 The sufficiency principle: How to teach it, and what does it entail?

Lutz Mattner
Universität Trier, Germany

We will recall and illustrate classical but apparently neglected ideas of Basu and others on the sufficiency principle: Every statistical procedure should be strictly nonrandomized, that is, depending deterministically on the data through a least sufficient statistic. While the emphasis will be on (counter-)examples, hopefully useful for teaching, we also prove the simple theorem: To every consistent estimator sequence there corresponds a strictly nonrandomized consistent estimator sequence. This becomes false when ëstimatorïs replaced by level α test".

114 A bounded quantile-based measure of kurtosis

Paul Jacobus van Staden
Department of Statistics, University of Pretoria, South Africa

A new quantile-based measure of kurtosis, named the kappa-functional, is defined. This shape functional is related to the ratio-of-spread functions, proposed by MacGillivray & Balanda (1988), but has the added advantage of being a bounded measure of kurtosis, hence simplifying interpretation. Together with the gamma-functional of MacGillivray (1986), a bounded measure of skewness, the kappa-functional can be used to describe and compare theshape characteristics of univariate distributions and data sets. This is illustrated via shape functional diagrams in which the kappa-functional is plotted against the gamma-functional for various distributions.

115 To choose or not to choose a prior. That's the question!

Fatemeh Ghaderinezhad, Christophe Ley
Ghent University, Belgium

The first challenging question in Bayesian statistics is how choosing the prior can affect the posterior distribution. How much can posteriors derived under different priors be similar as nowadays more and more data are collected? One of the newest instruments to answer this question is Stein's method. This crafty method gives the lower and upper bound to measure the Wasserstein distance between two posteriors derived from different priors (even improper priors) at fixed sample size.

116 Parameter Estimation for Lotka-Volterra Switching model

Houda Yaqine[1,2], Hamid El Maroufy[2], Christiane Fuchs[1,3]

[1]Bielefeld University, Germany.
[2]Faculty of Sciences and Technology of Beni Mellal, Morocco.
[3]Institute of Computational Biology, Helmholtz Zentrum München, Germany.

In this work the Lotka-Volterra with switching rates governed by a continuous time Markov chain (CTMC) is investigated. This model is a particular case of the Switching diffusion models, for which we have determined the likelihood function. In the study, the model is given in a parametric form, where the parameters to be estimated are the different rates as well as the transition rates of the CTMC. Along the estimation process, filtering equations have been used to approximate the conditional expectation of the CTMC given the observations in time of the Switching diffusion process.

Official Statistics and Survey Statistics I

117 Using Big Data in Official Statistics

Piet J.H. Daas
The CBS, Netherlands

Using Big Data for official statistics comes with many challenges. A considerable number of these challenges are of a methodological nature. This presentation will focus on these topics and discusses them with real-world examples. Topics touched upon are: ways to include Big Data in official statistics, working in a data-driven way, the (un)importance of a paradigm shift, the need to deal with new types of data, the role of populations, the importance of quality checks, the need for a continuous critical research attitude, and the link with Data Science. This provides an overview of the current state of the art in the area of Big Data methodology, resulting in a 10-step approach for the production of beta-products based on Big Data. In the presentation a considerable number of Big Data based application are shown. These results were obtained at the Center for Big Data Statistics of Statistics Netherlands and during the ESSnet on Big Data; a European project lead by Statistics Netherlands.

118 Smart Business Cycle Statistics

Markus Zwick, Clara Schartner
Federal Statistical Office Germany, Germany

Business cycles are important economic phenomena. These cycles have an enormous influence on society's welfare and well-being. Therefor the economic parameters that are responsible for the cycles are of core interest to politicians who need high quality and up-to-date information about the state of the business cycle. Reporting different indicators can help to detect the state of the economic activity. Eurostat and Destatis report these indicators via a dashboard, which contains the business cycle clock. However, the reporting process is complex and introduces a time lag of several weeks to publication. The goal of the project 'Smart Business Cycle Statistics' is to reduce this time lag by deriving indicators from economic activities, which are visible in satellite images. Satellite images are available with a short delay of only a few hours. The processing of the data and the detection of economic activities can also be done comparatively fast and thus allows a publication of economic indicators with a delay of only a few days. The economic activities need to be detectable from space, which means that they have to take place outdoors or leave traces outside, such as containers or ships at harbours. The satellites which are used, need to have a resolution high enough to detect these objects and observe them frequently enough to estimate reliable indicators based on a time series. The freely available Sentinel-2 images have a spatial resolution of 10 m and can be used to detect certain objects. Several suitable commercial satellites exist which can be used to get images with a higher spatial and temporal resolution, but this is very expensive at the moment. However, this field is developing fast and it can be expected that inexpensive data with the required spatial and temporal resolution will be available in the future and therefore, the smart business cycle statistics should become more accurate, affordable and reliable in the near future.

119 Enriching an Ongoing Panel Survey with Mobile Phone Measures: The IAB-SMART Study

Florian Keusch[1], Mark Trappmann[2,3], Georg-Christoph Haas[1,2], Sebastian Bähr[2], Frauke Kreuter[1,2,4]

[1]University of Mannheim, Germany
[2]Institute for Employment Research, Germany
[3]Bamberg University, Germany
[4]University of Maryland, USA

The annual panel study "Labour market and social security" (PASS) has been a major data source for labor market and poverty research in Germany since 2007. In January 2018 a subsample of PASS participants was invited to install a research app on their smartphones. The IAB-SMART app combines short questionnaires with passive data collection using the built-in smartphone sensors. Among others, questions are triggered when participants spend a predefined amount of time in a specific geographic location, allowing us to enrich annual retrospective information from PASS interviews with data collected immediately after an event has happened (e.g., a visit to the local job center). Passive data collection (e.g., geographic location, app use, activity, call and text message logs) allows us to collect innovative measures, e.g., on the integration into social networks or job search behavior, that will also complement traditional survey measures. Furthermore, the additional smartphone measures create the potential to address new research questions related to the labor market and technology use (e.g., digital stress, home office performance). Finally, the study provides new insights in the day structure and coping behavior of unemployed persons and thus replicate aspects of the classic Marienthal study from the 1930s with modern means. In this presentation, we will provide an overview of the study, its implementation, and we will share our experiences in conducting such a project. We will also focus on data protection issues and participation behavior in the study.

Preclinical and Pharmaceutical Statistics III (Preclinical studies)

120 A Novel Approach to Outlier Identification in Bioassays

Hannes Buchner[1], Robert Reidy[1], Michael Matiu[1], Johannes Solzin[2], Alexander Berger[2], Armin Boehrer[2], Erich Bluhmki[2,3]

[1]Staburo GmbH, Munich Germany
[2]Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach a.d. Riss, Germany
[3]Hochschule Biberach, University of Applied Sciences, 88400 Biberach, Germany

Introduction:
Human-based identification of unusual or out-of-place measurements from bioassays can suffer from bias as well as inconsistency. In this subjective outlier detection approach, there can be large inconsistencies between outlier identification both between analysts and even within analysts.
To minimise the intrinsic subjectivity in outlier detection, statistical methods should be applied as a classification system in order to achieve consistency and reduce subjectivity [1]. However, classical statistical outlier testing procedures (ROUT, Roesner, Grubbs, Dixon, k-SD, Hampel) suffer from the inability to distinguish between relevant and irrelevant differences as statistical hypotheses test do. This leads to a significant number of irrelevant cases where the human expert needs to overrule the outlier tool. Our proposed approach is a mixed approach in which outliers are identified using all major existing methods in combination with a derived 'action limit' which identifies a region where observations are not assessed for irregularity.
Method:
The variability between two external and experienced reviewers was assessed using a total of 62 plates from 5 assays with 2-4 replicates per plate of standard and samples. The necessity of the statistical approach in addition to the invoked action limit rule was investigated with the above-mentioned data as well as the necessity of the action limit to reliably estimate EC50.
The action limit approach defines an exception region, centred on a fitted four parameter logistic model of the bioassay measurements by concentration. The action limit itself is the margin around the model mean, calculated directly from an average variance within a relevant range of assay values. This variance is obtained using historical data from each assay of interest.
Results:
Independent reviewer 1 identified 27 outliers while reviewer 2 identified 53. The intersection of these sets of points was 17 while the union of the sets amounts to 63 identified outliers. Not applying the action limit leads to 5-6% additional points identified as outliers depending on the outlier method. 40 points fell outside of the action limit with between 32% and 66% categorised as statistical outlier depending on the detection method. When averaging over models which resulted in at least 1 outlier, the mixed approach resulting in a 14-38% average difference in the EC50 estimate depending on the detection method while the absence of the action limit resulted in a 4-23% average change. When averaging over all fitted models, the average change in the EC50 estimate was 1.3-2.7% and 1.7-2.8% for the mixed and non-mixed approach respectively.

Conclusions:
As expected, traditional reviewer-based approaches to outlier detection lead to greatly inconsistent interpretations. The proposed mixed method is a new strategy that significantly reduces the false outlier identification rate. This, in turn, minimises subjectivity as compared to the known standard methods. This approach better reflects the USP requirements [1] thereby also reducing costs.
References:
[1] The Unites States Pharmacopeial Convention – Article 1034

121 How to handle deviating control values in dose-response curves

Franziska Kappenberg[1], Jan Hengstler[2], Jörg Rahnenführer[1]

[1]TU Dortmund, Germany
[2]Leibniz Research Centre for Working Environment and Human Factors at the Technical University of Dortmund (IfaDo), Germany

In many toxicological assays a response variable is repeatedly measured under different conditions, for example for a negative control and increasing concentrations of a compound. A fitted dose-response curve (DRC) can be used to determine the concentration where a specific effect level is attained. Usually data is normalized before curve-fitting in order to have an inital value of 100% corresponding to the response value of the negative control.
In some cases problems arise from the fact that the response value of the control does not fit the left asymptote of the fitted DRC and therefore the asymptote does not correspond to an effect of 100%. This leads to the inability to properly interpret the concentration where the curve attains a given value.
In a simulation study we analyse different methods for dealing with the problem of deviating control values. A decision rule is derived which of the presented methods should be used, depending on different parameters, such as the number of concentrations, the number of replicates per concentration, the variance of the replicates and the difference between the left asymptote and the control value.
Results from the simulation study are applied to a toxicity assay in which the effect of a compound on the vitality of cells is measured for several replicates of increasing concentrations for three donors.

122 New Approaches for Bivariate Quantitative Dose-Response — A Screening Study from Hormone-Research and Development

Reinhard Meister

Beuth Hochschule für Technik Berlin, Germany

This paper addresses an important question in hormone research. Since 90 years, the question of safely administering hormones like e.g. estrogen orally has not been solved, as the hormone is immediately metabolized at every passage of blood through the liver. The metabolites of estrogen are compounds aggressively inducing adverse effects. Therefore, the concept of prodrugs, enabling a safe transfer of a substance without being metabolized in the liver is very attractive. We have contributed to a recent publication (Elger et.al. 2016), where results of a screening.series of dose-response experiments with estrogens carrying prodrugs have been reported. The statistical challenges and proposals for solutions will be presented.
In summary, we will give a very dense look at new dose response models and, in addition, a promising extension of the Bland-Altman ideas (see e.g. Bland JM, Altman DG 1986) to dose-response setups and a representation of such models in a function space.
I turns out, that our approach enables the comparison of benefit-risk ratios of a treatment and a competitor in a general setting without the assumption of parallel dose-response curves as commonly assumed in calculating relative efficiencies.
Our proposal goes back to the roots of bioassay (see e.g. Finney (1952)) along the pragmatic attitude of Fisher (1947) who advocated the log transform as a convenient way for transforming inference on ratios into one of inference on differences.
References
Fisher, R.A. (1947) The analysis of covariance method for the relation between a part and the whole. Biometrics, 3, 65–68.
Finney, D. J., Ed. (1952). Probit Analysis . Cambridge, England, Cambridge University Press.
Bland J.M„ Altman D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. Lancet, i, 307-310.

123 Adaptive designs in preclinical dose finding studies

Konrad Neumann[1], Samuel Knauss[2], Ulrike Grittner[1]

[1]Institute of Biometry and Clinical Epidemiology, Charité, Universitätsmedizin Berlin, Germany

[2]Department of Experimental Neurology, Charité Universitätsmedizin Berlin, Germany

Preclinical research is widely criticized for its wastage of test animals. The recently established Charité 3R center tries to address growing concerns in society. "3R" is an animal protection principle and stands for "Replace", "Reduce" and "Refine". Novel study designs could contribute to the efforts to reduce the number of animals in preclinical trials.

The need for a great deal of test animals is particularly high in dose finding studies since they are usually performed with many treatment arms. Moreover, differences in effect size between the experimental groups are small entailing the need for even a larger sample size.

In the talk, we will propose an adaptive two-stage design that has the potential to reduce the needed number of test animals.

The study design starts with four groups (control group and three groups with different dose levels) in the first stage. After an interim analysis, at least one, but possibly two dose levels are dropped or the study is terminated for futility. The second and final stage has then only two or three study groups. The basic idea of adaptive testing goes back to Fishers's p-value combination criterion [1]. Here we generalize and modify it for the needs of the proposed experimental design.

Preclinical trials often lack power. That could be one reason why translation of promising preclinical findings into clinical application often fails (c.f. [2] for a discussion of this issue in neurological stroke research). We demonstrate that the proposed adaptive design reduces the number of animals needed particularly in the case of underpowered studies that unfortunately occur so often in preclinical animal experiments.

Reference

1. P. Bauer; J. Röhmel: An Adaptive Method for Establishing a Dose-Response Relationship; Statistics in Medicine, Vol 14; 1595-1607 (1995).

2. O'Collins VE; Macleod MR; Donnan GA; Horky LL; van der Worp BH; Howells DW: 1,026 experimental treatments in acute stroke; Annals of neurology (2006) Mar, 59 (3), 467-477.

Robust and Nonparametric Statistics & Computational Statistics and Statistical Software

124 Comparison of Dependence Coefficients in Presence of Outliers, A Simulation Study

Ahmed R.M. Alsayed[1,2], Giancarlo Manzi[1], Sek Siok Kun[2]

[1]University of Milan, Italy

[2]University Science of Malaysia

The aim of this paper is to examine the performance of a recently proposed measure of dependence – namely the Monotonic Dependence Coefficient (MDC) – with respect to the Pearson's r and Spearman's rho correlation coefficients, using non-contaminated and contaminated datasets as well as a real dataset. The comparison is evaluated using Monte Carlo simulation and especially aimed at checking whether these coefficients are able to detect dependence relationships between two or more variables when outliers are present.

Several scenarios are created for the evaluation of the three measures, in particular contemplating multiple values of the coefficients, different contamination percentages, various simulation data patterns, or a combination of these settings. The basic dataset is generated from a bivariate or a multivariate standard normal distribution. Then, the data contamination is generated from different distributions: exponential, power-transformed and lognormal. The main findings tend to favor Spearman's rho when the contamination is taken into account and for most of the scenarios, whereas MDC performs well in non-contaminated data.

125 Comparison of some normality tests in the presence of outliers

Mustafa Çavus, Berna Yazici, Ahmet Sezer

Eskisehir Technical University, Turkey

There are many studies on normality tests which are compared in terms of power and type I error rate under the various alternatives such as light tailed alternatives as exponential, lognormal and gamma distribution, heavy tailed alternatives as Cauchy, Laplace, t3, t5 and logistic distributions and short tailed alternatives as beta and uniform distributions. In addition to these distributions, distortion of normality caused by outliers is the most common situation. This study investigates the sensivity of the normality tests to outliers. Performances of Shapiro-Wilk (1965), Lilliefors (1967), Shapiro-Francia (1972), Anderson-Darling (1986), Jarque-Bera (1987), Bonett-Seier (2002), Gel-Miao-Gastwirth (2007), Gel-Gastwirth (2008), Brys-Hubert-Struyf (2008) and Dornik-Hansen (2008) normality tests are compared for different sample sizes and outlier setups. The results are given in a comprehensive Monte-Carlo simulation study with the penalized power properties of the normality tests instead of classical setup is consist of power and type I error probability.

126 Removing Outliers: Effects on Statistical Inference and Suggestions for Choosing Exclusion Boundaries

Patrick Schenk

[1]IAB (Institute for Employment Research), Nuremberg, Germany
[2]Ludwig-Maximilians-Universität, Munich, Germany

Outliers influence the ability to estimate descriptive statistics or regression parameters of a particular statistical model / data generating process (DGP). The reason underlying the occurrence of outliers in this talk is that some data points do not come from the DGP of interest (called "good" below), but from (two) contaminating DGPs. As these have different parameters or distributional form, naïve analysis of all observations with standard methods produces biased results.
This situation, well-known to empirical researchers and computer scientists, is also located at the intersection of several statistical fields: (i) regression diagnostics, excelling in the identification of (potential) outliers but not necessarily offering optimal guidelines on how to treat them; (ii) robust statistics, offering methods that can cope with keeping the outliers in the analyzed data; (iii) finite mixture models which attempt to model all (three) DGPs simultaneously. Regression diagnostics suffer from the "catch-22" dilemma that for best identification of outliers (i.e. observations that are unusual relative to the DGP of interest) one would have to know the model parameters already, but ideal estimation would necessitate removing the outliers first. Applying the more sophisticated methods of robust statistics and finite mixture models can be difficult for very large data sets, for already complex models and procedures, or if they are not well-established in the tradition and training of a particular research area; their performance can also depend on the quality of initial values.
For all these reasons and more, researchers striving for unbiased parameter estimates often choose to remove outliers beyond certain boundaries – e.g., based on heuristics such as "beyond 1.96 standard deviations from the mean" ("1.96 SD-rule") – before applying standard methods. This talk addresses the consequences of this approach.
First, it is demonstrated that several common procedures do not perform well in removing observations originating from the contaminating DGPs: they may worsen data composition and the greater the problem (i.e., they greater the need for outlier treatment) the worse the performance problem can be.
Second, implications for statistical inference (standard errors and p-values) are considered. In removing observations – inadvertently not only contaminating data points but also good ones – for instance by the "1.96 SD-rule", sample size is reduced, motivated by the idea of improved sample composition yielding less biased parameter estimates. It can be shown that the sample size effect is small. However, among the retained (good) observations the model fit is artificially high (standard errors are shrunk) as the most extreme good observations are eliminated. The amount of shrinkage, depending on the level of outlier removal, is derived and a simple, easy to implement, and understandable correction is offered.
Third, it is shown that residuals in a linear model are no longer homoscedastic and independent of the regressors.
Fourth, a method for choosing the boundaries for outlier exclusion is presented and attempt is made to combine it with robust statistics and finite mixture models.

Statistics in Agriculture and Ecology II

127 Testing Multiplicative Terms in AMMI and GGE Models for Multienvironment Trials

Waqas Malik, Hans-Peter Piepho
Universität Hohenheim, Stuttgart, Germany

The additive main effects and multiplicative interaction model and genotype main effects and genotype-by-environment interaction model are commonly used for the analysis of multienvironment trial data. Agronomists and plant breeders are frequently using these models for cultivar trials repeated across different environments and/or years. In these models, it is crucial to decide how many significant multiplicative interaction terms to retain. Several tests have been proposed for this purpose, however, all of them assume that errors are normally distributed with a homogeneous variance. Here, we propose nonparametric resampling-based methods for multienvironment trial data, which are free from these distributional assumptions. The methods are compared with competing tests. In an extensive simulation study based on two multienvironment trials, it was found that the proposed methods performed well in terms of Type-I error rates regardless of the distribution of errors.

128 How to detect imprinted loci using estimated parent-of-origin effects and simple gene counts only

Inga Blunk[1], Manfred Mayer[1], Henning Haman[2], Norbert Reinsch[1]

[1]Institute of Genetics and Biometry, Leibniz Institute for Farm Animal Biology (FBN), Dummerstorf, Germany

[2]State-Office for Geo-Information and Rural Development, Geodata-Center, Kornwestheim, Germany

Depending on their parental origin, the alleles of genomically imprinted genes are fully or partially inactivated by epigenetic mechanisms such as DNA methylations and histone modifications. Established in mammals, plants and insects, the resulting effects are assigned to the broader class of parent-of-origin effects (POEs) which are statistically defined as the difference of an individual's transmitting ability under a male and a female expression pattern. For the mapping of imprinted loci in association or linkage analyses phenotypes as well as marker genotypes must be available for the same individuals. Moreover, the parental allele origin of a marker genotype must be known (ordered marker genotypes). For these reasons, large resources of phenotypic data of individuals without own genotypes and genotypic data of individuals without own phenotypes often remain unconnected and without benefit for such genomic imprinting analyses. We now developed a new methodology with which the above mentioned limitations can be overcome. This methodology first includes the estimation of POEs (ePOEs) from population data using all available phenotypes and a suitable mixed model. In another step these ePOEs are then employed as dependent variables to be regressed on un-ordered marker genotypes in a genome scan. That only imprinted loci give a signal was first demonstrated in a simulation study and is based on the theoretical consideration that the regression of a true POE on the number of Q-alleles at an imprinted biallelic locus (representing the un-ordered genotype) equals exactly the imprinting effect. To provide a practical application example of our approach the ePOEs for 1,857 Brown Swiss bulls were used to map imprinted quantitative trait loci expressed in their fattening progeny. One locus was associated with ePOEs estimated in net BW gain assuming a genome-wide false discovery rate of 5% on the Bos taurus autosome 11. Weaker signals were found for muscularity and fatness traits assuming chromosome-wide false discovery rates of 5% and 10%. Hence, in association studies phenotypic information from e.g. progeny without genotypes can be related to un-ordered marker genotypes of parents. In this way, ePOEs provide a new opportunity to cost-effectively detect imprinted loci and allow the exploitation of data, which otherwise would remain unused for imprinting analyses.

129 Application of Multivariate Statistical Methods in Water Pollution Footprinting

Hayal Boyacioglu[1], **Hulya Boyacioglu**[2]

[1]Ege University, Turkey
[2]Dokuz Eylul University, Turkey

In the study, multivariate statistical techniques were performed to fingerprint water pollution sources and investigate temporal/spatial variations in water quality. Data was obtained from three water reservoirs in Turkey and was subjected to statistical analysis comprasing factor analysis, Student's t-test and discriminant analysis.
Factor analysis helped to identify the factors/sources responsible for variations in reservoir water quality at three different sites. Student's t-test used to determine significant parameters having difference between summer and winter periods. Furthermore, discriminant analysis gave the best results to investigate spatial 0differences. For three reservoirs it yielded an important data reduction. Results showed that overall water quality was mainly governed by „natural factors"in the whole region. A parameter that was the most important in contributing to water quality variation for one reservoir was not important for another. Between summer and winter periods, difference in arsenic concentrations were statistically significant in the Tahtal?, Ürkmez and iron concentrations were in the Balçova reservoirs. Observation of high/low levels in two seasons was explained by different processes as for instance, dilution from runoff at times of high flow seeped through soil and entered the river along with the rainwater run-off and adsorption. Three variables „boron, arsenic and sulphate"discriminated quality among Balçova & Tahtal?, Balçova & Ürkmez and two variables „zinc and arsenic"among the Tahtal? & Ürkmez reservoirs. The results illustrated the usefulness of multivariate statistical techniques in ecological studies.

Time Series Analysis II

130 Detecting multiple location shifts under long-memory stationary errors

Mustafa Kilinc
WHU - Otto beisheim school of management, Germany

We consider detecting multiple structural breaks when the error of the regression model contains long memory. The general-to-specific model selection approach is extended by adding step indicators for every observation to detect breaks: see Castle, Doornik, Hendry and Pretis (2015). We refine their approach to handle not only short memory but long memory as well. Monte Carlo experiments show its good performance and the capability to detect multiple breaks. An empirical illustration of the procedure for the realized volatility of the S&P 500 is shown.

131 Long Memmory Conditional Heteroscedasticity in Count Data

Mawuli Segnon, **Manuel Stapper**
WWU Münster, Germany

This paper introduces a new class of integer-valued long memory processes that are adaptations of the well-known FIGARCH(p, d, q) process of Baillie (1996) and HYGARCH(p, d, q) process of Davidson (2004) to a count data setting. We derive the statistical properties of the models and show that reasonable parameter estimates are easily obtained via conditional maximum likelihood estimation. An empirical application with financial transaction data illustrates the practical importance of the models.

132 Estimation and Inference in Adaptive Learning Models with Slowly Decreasing Gains

Alexander Mayer
WHU – Otto Beisheim School of Management, Germany

This paper develops techniques of estimation and inference in a prototypical macroeconomic adaptive learning model with slowly decreasing gains. A sequential three-step procedure based on a 'super-consistent' estimator of the rational expectations equilibrium parameter is proposed. This procedure is asymptotically equivalent to first estimating the structural parameters jointly via ordinary least-squares and then using the so-obtained estimates to form a plug-in estimator of the rational expectations equilibrium parameter – an aproach reminiscent of the well known two-step approach to estimating co-integrating vectors and associated error correction models.
In spite of failing Grenander's conditions for well-behaved data, limiting normality of all estimators centered at their true parameter values is established; resorting to a Central Limit Theorem which allows for potentially degenerate variances. A trade-off between the rate at which the agent learns and the convergence rate of the estimators becomes apparent. Furthermore, the learning rate of the agent determines also the degree of time-series dependence which can be allowed for.
It is then shown that, notwithstanding potential threats to inference arising from non-standard convergence rates and a singular variance-covariance matrix, classical hypothesis tests involving single, as well as joint restrictions remain valid. Specifially, exploiting its transformation non-invariance, local asymptotic power of the classical Wald statistic is derived. As an interesting by-product of this local power exercise, the superiority, in terms of local power, over its counterpart equipped with a generalized inverse is established.
Monte-Carlo evidence confirms the accuracy of the asymptotic theory for the finite sample behaviour of the estimators and test statistics discussed here. Finally, the proposed procedure is applied to the estimation of the hybrid U.S. New-Keynesian Phillips curve.

133 Parameter estimation for time series models based on the simulated characteristic function

Thiago do Rego Sousa[1], Richard Davis[2], Claudia Klüppelberg[1]
[1]Technical University of Munich, DE
[2]Columbia University, New York, USA

Estimating the parameter of a time series by matching empirical and true characteristic functions for block time series vectors has been proposed in Knight and Yu (2002), but it requires the true characteristic function (CF) to be known in closed form.
For models where the true CF is not available in closed form, we borrow the simulation principle of Indirect Inference to approximate it and apply this method to estimate the parameter of a time series model. We show consistency and asymptotic normality of the parameter estimates under strong mixing, moment conditions and smoothness of the time series with respect to its parameter.
Since the estimates depend directly on the quality of the approximation of the true CF, we propose the use of control variates for reducing the variance of its Monte Carlo estimate. We show that this results in much better approximations of the true CF on a set of points which are most relevant for parameter estimation. We then propose a new estimator based on the control variates approximation of the true CF that is applicable to a large class of time series models. Our simulation study shows the superiority of the control variates based estimator for Poisson driven time series of counts.
This is joint work with Richard Davis and Claudia Klüppelberg.

Verleihung der IBS-DR Nachwuchspreise

134 Sample Size Calculation in Time-To-Event Trials with Non-Proportional Hazards Using GESTATE

Jasmin Rühl

Ulm University, Germany

Most methods used for time-to-event analysis rely on the assumption of proportional hazards (PH), which is frequently violated in practice. Although the log-rank test is not optimal in that case, it is often requested by regulatory authorities. Furthermore, there are only few approaches for trial planning if the PH assumption is wrong. GESTATE is an R platform that can be used to design, simulate and analyse time-to-event trials with non-proportional hazards. It is based on an average-hazard ratio approach that is weighted by the number of events. The idea behind the concept is to predict parameters by using PH methods (such as the Schoenfeld formula), without deploying a PH model for the estimation of the hazard ratio. In order to investigate the generalisability of GESTATE, we integrated log-logistic, Gompertz and generalised gamma distributions, and compared the predictions of hazard ratio and power to simulations. The results showed decent performance in a variety of common trial settings, although extreme conditions seem to diminish accuracy.

135 Statistical Approaches to Characterize and Compare Networks of Microbiome Data

Stefanie Krügel[1,2,3], Prof. Dr. Anne-Laure Boulesteix[2], Dr. Martin Depner[1]

[1]Institute for Asthma and Allergy Prevention; Helmholtz Zentrum München; Germany
[2]Institute for Medical Information Processing, Biometry, and Epidemiology; Ludwig-Maximilians-Universität München; Germany
[3]Department of Statistics; Ludwig-Maximilians-Universität München; Germany

The rapid development of high-throughput sequencing techniques offers new possibilities for investigating the human microbiome and provides the opportunity to discover relationships between the composition of microbial communities and certain diseases. Network analysis methods are widely used to investigate relations between microorganisms. The analysis of a single microbiome network, however, often does not lead to a conclusion about the relevance of microbial interactions for a specific disease. This master thesis arose from the idea to compare microbiome networks between two states of a feature variable (e.g. between diseased subjects and healthy controls or between different environments). The presented methods aim to uncover organisms or functional groups within a microbial community that might be associated with the occurrence of a disease.
The construction of microbiome networks comes along with some difficulties due to the special characteristics of read count data, such as sparsity or compositionality, which have to be taken into account when associations between taxa are estimated. Several approaches that seek to handle these problems have been published during the last decade. Furthermore, existing studies (mainly from other research fields) can be found where specific characteristics of networks are compared between groups. What is still missing, however, is an overall workflow – and in particular a software tool – for the group comparison of microbiome networks, which includes all steps from estimating associations based on count data, to the process of selecting edges of interest, up to the comparison of network properties in a graphical and a quantitative way (including statistical tests). This master thesis attempts to close this gap by giving an overview of existing methods, comparing them with one another and adapting them for the application on read count data if necessary. Furthermore, an own R package named "NetCoMi" is presented, which implements the described methods in the statistical software R and will be published soon.
The methods included in this R package are finally applied to read count data (i.e. bacteria) that have been collected in the course of the PASTURE project. Comparing networks of the intestinal microbiome of infants, which are constructed based on these data, aims to promote the process of understanding the relationship between the microbial composition of the gut and the development of asthma or allergies. The results reveal, for instance, organisms whose importance in the network (expressed via appropriate centrality measures) is different between the considered groups and, in addition, whether these differences are significant. It might also be of interest for researchers if the network hubs differ between the two groups or whether the detected clusters are different, which is also answered by the presented methods.

136 Group sequential designs with robust semiparametric recurrent event models

Tobias Mütze[1,3], Ekkehard Glimm[1,2], Heinz Schmidli[1], Tim Friede[3,4]

[1]Statistical Methodology, Novartis Pharma AG, Basel, Switzerland
[2]Institute for Biometrics and Medical Informatics, Medical Faculty, Otto-von-Guericke-University Magdeburg, Magdeburg, Germany
[3]Department of Medical Statistics, University Medical Center Göttingen, Göttingen, Germany
[4]DZHK (German Centre for Cardiovascular Research), partner site Göttingen, Göttingen, Germany

Robust semiparametric models for recurrent events have received increasing attention in the analysis of clinical trials in a variety of diseases including chronic heart failure. In comparison to parametric recurrent event models, robust semiparametric models are more flexible in that neither the baseline event rate nor the process inducing between-patient heterogeneity need to be specified in terms of a specific parametric statistical model. However, implementing group sequential designs is complicated by the fact that the sequence of Wald statistics in the robust semiparametric model does not follow asymptotically the canonical joint distribution. In this presentation, we propose two types of group sequential procedures for a robust semiparametric analysis of recurrent events. The first group sequential procedure is based on the asymptotic covariance of the sequence of Wald statistics and it guarantees asymptotic control of the type I error rate. The second procedure is based on the canonical joint distribution and does not guarantee asymptotic type I error rate control but is easy to implement and corresponds to the well-known standard approach for group sequential designs. Moreover, we describe how to determine the maximum information when planning a clinical trial with a group sequential design and a robust semiparametric analysis of recurrent events. We contrast the operating characteristics of the proposed group sequential procedures in a simulation study motivated by the ongoing phase 3 PARAGON-HF trial (ClinicalTrials.gov identifier: NCT01920711) in more than 4600 patients with chronic heart failure and a preserved ejection fraction. We found that both group sequential procedures have similar operating characteristics and that for some practically relevant scenarios the group sequential procedure based on the canonical joint distribution has advantages with respect to the control of the type I error rate.

137 A simple signaling rule for variable life-adjusted display derived from an equivalent risk-adjusted CUSUM chart

Philipp Wittenberg[1], Fah Fatt Gan[2], Sven Knoth[1]

[1]Department of Mathematics and Statistics, Helmut Schmidt University, Hamburg, Germany

[2]Department of Statistics and Applied Probability, National University of Singapore, Singapore

The variable life-adjusted display (VLAD) is the first risk-adjusted graphical procedure proposed in the literature for monitoring the performance of a surgeon. It displays the cumulative sum of expected minus observed deaths. It has since become highly popular because the statistic plotted is easy to understand. But it is also easy to misinterpret a surgeonâ€™s performance by utilizing the VLAD, potentially leading to grave consequences. The problem of misinterpretation is essentially caused by the variance of the VLAD's statistic that increases with sample size. In order for the VLAD to be truly useful, a simple signaling rule is desperately needed. Various forms of signaling rules have been developed but they are usually quite complicated. Without signaling rules, making inferences using the VLAD alone is difficult if not misleading. In this paper, we establish an equivalence between a VLAD with V-mask and a risk-adjusted cumulative sum (RA-CUSUM) chart based on the difference between the estimated probability of death and surgical outcome. Average run length analysis based on simulation shows that this particular RA-CUSUM chart has similar performance as compared to the established RA-CUSUM chart based on the log-likelihood ratio statistic obtained by testing the odds ratio of death. We provide a simple design procedure for determining the V-mask parameters based on a resampling approach. Resampling from a real data set ensures that these parameters can be estimated appropriately. Finally, we illustrate the monitoring of a real surgeonâ€™s performance using VLAD with V-mask.

Advanced Regression Modeling III (Distributional Regression/ GAMLSS)

138 Neural Network Distributional Regression

Nikolaus Umlauf[1], Nadja Klein[2], Thorsten Simon[1], Achim Zeileis[1]
[1]Universität Innsbruck, Austria
[2]Humboldt University of Berlin, Germany

During the last decades there has been an increasing interest in distributional regression models that allow to model all distributional parameters, such as location, scale and shape and thereby the entire data distribution conditional on covariates. In particular, the framework of structured additive distributional regression models enables to specify different types of effects such a linear, non-linear or interaction effects on all the distribution parameters hence providing a very flexible and generic framework suited for many complex real data problems. However, when it comes to the question of variable selection, establishing a reasonable and 'good' distributional model is hard work. In addition, the exact functional forms and possible interactions are often hard to fix in advance even with advanced expert knowledge. To overcome this drawback, we propose an extension of the structured additive regression predictors by a feedforward neural network that allows to learn the functional forms and potential complex interactions of dependent variables from the data with the algorithm. We propose an efficient implementation and illustrate the usefulness by an application of NWP-based lightning prediction.

139 Random Function Responses in Distributional Regression

Hannes Riebl[1], Nadja Klein[2], Thomas Kneib[1]

[1]University of Göttingen, Germany
[2]Humboldt University of Berlin, Germany

We propose a new class of regression models with conditionally independent random functions as responses motivated from an example on tree growth dynamics. Our model is embedded into the framework of structured additive distributional regression models, which have been popularized as Generalized Additive Models for Location, Scale, and Shape (GAMLSS). The distributional regression framework allows us to link all distributional parameters of the random functions to structured additive predictors. In the simplest case, these parameters are a mean, a variance, and a range parameter of a correlation function, which are constant over the individual random functions, thus rendering them stationary. Extensions to non-stationary random functions with more complicated mean functions and a higher number of parameters are considered as well. The proposed models are closely related to mixed models with within-group correlation structures, but the distributional regression framework allows for additional modeling flexibility and different inferential methods.
We perform inference in a Bayesian setting via an efficient MCMC sampling scheme with stochastic Newton proposals. Some performance pitfalls and numerical difficulties with our inference procedure and potential workarounds for these issues are discussed. A small simulation study provides evidence that the procedure works reliably. Finally, the application to tree growth data is presented, where the growth dynamics of the individual trees are modeled as random functions conditional on various explanatory variables like the forest mixture type, the tree species, and the tree age.

140 Bayesian Effect Selection in Structured Additive Distributional Regression Models

Nadja Klein[1], Manuel Carlan[2], Thomas Kneib[2], Stefan Lang[3], Helga Wagner[4]
[1]Humboldt University of Berlin, Germany
[2]University of Goettingen, Germany
[3]University of Innsbruck, Austria
[4]University of Linz, Austria

We propose a novel spike and slab prior specification with scaled beta prime marginals for the importance parameters of regression coefficients to allow for general effect selection within the class of structured additive distributional regression. This enables us to model effects on all distributional parameters for arbitrary parametric distributions, and to consider various effect types such as non-linear or spatial effects as well as hierarchical regression structures. Our spike and slab prior relies on a parameter expansion that separates blocks of regression coefficients into overall scalar importance parameters and vectors of standardised coefficients.Hence, we can work with a scalar quantity for effect selection instead of a possibly high-dimensional effect vector, which yields improved shrinkage and sampling performance compared to the classical normal-inverse-gamma prior. We investigate the propriety of the posterior, show that the prior yields desirable shrinkage properties, propose a way of eliciting prior parameters and provide efficient Markov Chain Monte Carlo sampling. Using both simulated and three large-scale data sets, we show that our approach is applicable for data with a potentially large number of covariates, multilevel predictors accounting for hierarchically nested data and non-standard response distributions, such as bivariate normal or zero-inflated Poisson.

141 Proper imputation for GAMLSS inference

Tobias Hepp[1], Angelina Hammon[2,3]

[1]University of Bonn, Germany
[2]University of Bamberg, Germany
[3]Leibniz Institute for Educational Trajectories

The problem of missing data occurs in many areas of research.
Depending on the proportion of missing values and the mechanism responsible for the missingness, using only complete cases for the analyses may lead to a substantial loss of power and even biased estimates.
In the last decades, various imputation methods have been proposed to address these problems.
Due to its ability to preserve a necessary amount of uncertainty regarding the imputations, Multiple Imputation (MI) has proven itself as the most useful approach when dealing with missing data.
Generalized models for location, scale and shape (GAMLSS) extend the original framework of generalized additive models by using additional additive predictors to model the dependency of up to four distribution parameters of an outcome on a set of (possibly distinct) covariates.
Although the concept of MI experienced a lot of further development and refinements in the last decades to accommodate a variety of different data structures, most techniques are based on modeling the dependence structure only with respect to the expectation of the distribution of the variables to be imputed.
As a consequence, estimates of the predictors of a GAMLSS derived after imputation focusing on parameters such as the variance or dispersion of the Gaussian and negative binomial distribution, respectively, will be severely biased.
In this talk, this problem is first demonstrated on different simulation scenarios and imputation methods.
Afterwards, a possible solution is presented which incorporates the required information into the modelling of the missingness pattern.

142 Multivariate Functional Additive Mixed Models

Alexander Volkmann, Almond Stöcker, Fabian Scheipl, Sonja Greven
Ludwig-Maximilians-Universität München, Germany

Functional data are often multivariate, i.e. they simultaneously measure different functional aspects of a process. So far, few regression methods have been developed to efficiently handle the full amount of information provided by multivariate functional data.
We develop a multivariate functional additive mixed model (MFAMM). The dependency structure between the different dimensions is incorporated using multivariate functional principal component analysis. The model accounts for correlation within the functions, between the multivariate functional dimensions as well as potentially further between-function correlation - which is often induced by the study design - via multivariate functional random intercepts.
Multivariate functional data generated in a speech production study with a crossed study design are analyzed. The analysis is more parsimonious compared to fitting independent univariate models to the data and generates insight into the dependency structure between acoustic and articulatory processes. Application results also suggest that estimated confidence regions might be more efficient for the MFAMM than for the univariate approach.

Computational Statistics and Statistical Software III(Open science and reproducibility)

143 Open Science and statistics

Anne-Laure Boulesteix, Felix Schönbrodt
LMU Munich, Germany

The recently created interdisciplinary LMU Open Science Center (OSC) has the mission to promote and foster open science practices at LMU Munich and beyond. In the first part of my talk, I will give a very brief overview of the concept of open science with a special focus on aspects which are directly relevant to applied statisticians. They include technical issues such as the publication of data and code for the purpose of reproducibility, the challenges related to personal data in this context, and quality assurance in software development as well as methodological issues such as the prevention of p-hacking and related questionable research practices (for example through the publication format "registered reports"), the appropriate representation of uncertainties related to the data analysis strategy, or the problem of publication bias and how to correct for it. In the second part of my talk, I will argue that the above mentioned methodological issues are not only relevant to statisticians working as statistical consultants in projects from various disciplines, but also, in a perhaps more subtle and not very well-understood way, to methodological statisticians in their own research projects devoted to the development of new methods. This second part will be illustrated through three of our empirical meta-research projects.

144 Research Software Engineers

Heidi Seibold
LMU, Germany

In recent years there has been a push for researchers to make their research results and the corresponding data analysis pipeline openly available. Researchers, however, will only make their work open, if they are confident that their results can be reproduced and for making their work reproducible they need the right software tools and skills.
Not every researcher has the time and resources to refine their software skills, but there is help in sight: Research software engineers (RSEs). RSEs focus on helping researchers with software. In the field of statistics RSEs may help with preparing data for long term storage, preparing analysis pipelines in a way that they can be run by everyone, or supporting methods developers in best practices for software development (e.g. R package development) so that they can make their methods available for users.
In this talk I will discuss the need for RSEs and highlight what needs to change for RSEs in order to tap their full potential in improving the quality of resarch.

145 Reproducible Methodological Research and Scientific Publishing

Bernd Bischl[1], Benjamin Hofner[2], **Fabian Scheipl**[1]

[1]LMU München

[2]Paul-Ehrlich-Institut, Langen

On the face of it, perfect reproducibility seems easy to achieve for statistical methodological research. Conclusions are typically drawn from in sillico simulation experiments which are trivial to replicate based on shared code and common execution environments. In practice, the lack of wide-spread coding expertise as well as sufficient computing infrastructure and common standards combined with complications due to intellectual property and privacy concerns often impede progress towards transparent and reproducible methodological research. In this talk, we summarize experiences gained from almost 10 years of – leniently – enforced reproducibility at the Biometrical Journal, describe the most important obstacles to open science encountered in practice, and discuss possible ways to overcome them.

146 Correcting for bias in the literature: A comparison of meta-analytic methods for bias-correction

Felix D. Schönbrodt

Ludwig-Maximilians-Universität München, Germany

Publication bias and questionable research practices in primary research can lead to badly overestimated effect sizes in meta-analysis. Methodologists have proposed a variety of statistical approaches to correcting for such overestimation. However, it is not clear which methods work best for data typically seen in practice. We conducted a comprehensive simulation study to examine how some of the most promising meta-analytic methods perform on data typical of psychological research. Along with plain random effects models, we examined the following bias-correcting methods: Three-parameter selection model, four-parameter selection model, trim-and-fill, p-curve, p-uniform, PET, PEESE, or PET-PEESE, and WAAP (weighted average of adequatly powered studies). Furthermore, we tried to mimic realistic scenarios by simulating several levels of questionable research practices (p-hacking"), publication bias, and heterogeneity, using study sample sizes empirically derived from the psychological literature. In a nutshell, results show that every method fails in some conditions, some methods are on average (but not always) better than others, and a naive random effects meta-analysis probably is the worst what can be done in the presence of bias and/or p-hacking. An online tool is provided and demonstrated that allows to perform a method performance check"which allows to assess whether a certain meta-analytic method can be trusted in plausible conditions for the task at hand. Such an a-priori performance check allows to preregister which meta-analytic method will be given the largest credit in case that multiple methods disagree about the size or presence of an effect. I will show examples how such a method performance check can be done in practice, and how to handle the case that different methods lead to diverging conclusions. Given the difficulties of correcting, we strongly recommend that researchers continue their efforts on improving the primary literature and conducting large-scale, pre-registered replications instead of trying to correct a biased existing literature.

147 The multiplicity of possible analysis strategies and how it is handled across scientific disciplines

Sabine Hoffmann, Felix Schönbrodt, Ralf Elsas, Simon Klau, Anne-Laure Boulesteix
Ludwig-Maximilians-Universität München, Germany

In a large number of disciplines, an important part of research projects consists in the generation of numerical results through computational analyses (a phrase to be understood in a broad sense), for instance through the statistical analysis of empirical data or through the application of a physical model. In the following, we will denote the specification of the whole computational analysis pipeline as „analysis strategy". For a given research question, there is usually a large variety of possible analysis strategies that are acceptable according to the scientific standards of the field. These choices are also referred to as researcher degrees of freedom. When analyzing data from an observational study in psychology or epidemiology, for instance, there are numerous judgements and choices to be made concerning data-preprocessing, including the definition of predictor and outcome variables, but also data inclusion and exclusion criteria and the treatment of outliers and missing values. After these data-preprocessing steps, a probability model is typically chosen to describe the association between the outcome of interest and an uncertain number of predictor variables. After these decisions, there are still many judgments that have to be made concerning the method and the method setting to use in order to conduct statistical inference for the specified probability model. In light of concerns that researcher degrees of freedom may play an important part in the non-replicability of research findings by systematically leading to an increase in false positive results and inflated effect sizes, a number of approaches for the handling of the multiplicity of possible analysis strategies have been proposed in psychology and in epidemiology. In other disciplines, including climatology, ecology and risk analysis, there is a long-standing tradition of accounting for the multiplicity of possible analysis strategies. As the non-replicability of research findings is a problem touching many fields in a similar way, the aim of this work is to take an interdisciplinary view and to compare approaches and ideas that have been proposed in the different fields to handle the multiplicity of possible analysis strategies.

Design of Experiments and Clinical Trials IV (Adaptive Designs III)

148 Novel designs for trials with multiple treatments and subgroups

James Wason

Newcastle University, United Kingdom

Clinical trials provide a gold-standard evaluation of the safety and efficacy of a new intervention. However they are increasingly expensive and often result in not finding a beneficial effect. Developing and applying new statistical methods for improving the efficiency of clinical trials is a priority for medical research.
Three ways of addressing the challenges above are: 1) to incorporate multiple experimental intervention arms into a single 'multi-arm' trial; and 2) to better account for patient subgroups that may be associated with the efficacy of a treatment; 3) to use adaptive methods that allows use of accruing information in a trial to improve efficiency.
Multi-arm trials are increasingly being recommended for use. This is because they allow a shared control group, which considerably reduces the sample size required compared to separate randomised trials. Further gains in efficiency can be obtained by an adaptive design (multi-arm multi-stage, MAMS trials). At a series of interim analyses, a variety of modifications are possible, including changing the allocation to different treatments, dropping of ineffective treatments or stopping the trial early if sufficient evidence of a treatment being superior to control is found. These modifications allow focusing of resources on the most promising treatments, and thereby increase both the efficiency and ethical properties of the trial.
In this talk I will describe some different types of MAMS designs and how they may be useful in different situations. I will also discuss the design of trials that test efficacy of multiple treatments in different patient subgroups. I propose a design that incorporates biological hypotheses about links between treatments and biomarker subgroups effects of treatments, but allows alternative links to be formed during the trial. The statistical properties of this design compares well to alternative approaches available. I will illustrate the methods through real trials that I have been involved with designing.

149 Optimal designs for multi-arm phase II/III drug development programs

Stella Preussler[1], Marietta Kirchner[1], Heiko Götte[2], Meinhard Kieser[1]

[1]Institute of Medical Biometry and Informatics, University of Heidelberg, Germany
[2]Merck KGaA, Darmstadt, Germany

In drug development programs, phase II trials are often conducted as multi-arm studies with the aim of identifying the treatment(s) with the best benefit-to-risk profile(s). Subsequent confirmatory phase III trials then attempt to demonstrate efficacy and safety of the selected treatments(s). In the context of optimal planning of phase II/III drug development programs with regard to sample size allocation and go/no-go decision rules (applied to decide whether to stop or to proceed to phase III), this leads to more complex go/no-go decision rules as compared to two-arm studies [1]. For example, one has to decide whether to conduct the phase III trial with a single (the most promising; Strategy 1) treatment only, or with multiple treatments (if sufficiently promising; Strategy 2). The former strategy is less expensive as fewer patients are included in the trial. However, lower success probabilities are to be expected. Thus, it is not clear which strategy results in a higher expected benefit.

Optimal sample size and go/no-go decision rules are presented for time-to-event outcomes and different scenarios, where at least one treatment needs to show efficacy in the final stage of the drug development program. We consider a three-arm phase II trial which in case of a "go" decision is followed by a two-arm phase III trial (Strategy 1) or by a two- or three-arm phase III trial (Strategy 2). Due to the time-to-event endpoint setting, the testing procedures and power calculations are based on number of events (Log-rank test and Dunnett test for a two- and three-arm phase III trial, respectively). However, to calculate the patient costs the sample size for the trial needs to be computed. By assuming control arm event rates (which can be approximated by the Simpson's rule and/or using information of previous trials), choosing appropriate treatment effect sizes and fixing the allocation ratio, the trial event rates (i.e. proportion of patients in the trial with event) for each phase can be calculated [2]. For calculating the trial event rates for phase II and III, the assumed true treatment effects and the treatment effect estimates of phase II may be regarded as "appropriate" treatment effect sizes, respectively. Hence, the sample size for phase II and III can be calculated by dividing the number of events by the trial event rates.

The proposed method takes into account costs of the program, expected benefit, and development risk (success probability) to deliver insights on how optimal design parameters (e.g. sample size allocation, go/no-go decision rule, number of arms in phase III) change with varying benefit and assumptions about the true treatment effect and control arm event rate. The usage of the approach is facilitated by a user friendly R shiny App which will also be presented.

[1] Kirchner, M., Kieser, M., Götte, H., & Schüler, A. (2016). Utility?based optimization of phase II/III programs. Statistics in Medicine, 35(2), 305-316.

[2] Schoenfeld, D. A. (1983). Sample-size formula for the proportional-hazards regression model. Biometrics, 499-503.

150 Comparison of the efficacy of Bayesian and frequentist designs for oncological phase II basket trials.

Maja Krajewska, Geraldine Rauch

Charité - Universitätsmedizin Berlin, Institute of Biometry and Clinical Epidemiology

In oncology, the choice of treatment has, traditionally, been determined based mainly on the anatomical location of the tumor, while in recent years a shift to additionally taking into account its genetic mutations took place. In order to take this shift into account in phase II proof of concept trials, so-called basket trials can be performed. In oncological settings, these trials are based on the accrual of patients exhibiting tumors in different anatomic locations but the same genetic mutation, who are then allocated into subgroups (so-called "baskets") based on the anatomical location of their tumor. Propositions for the evaluation of this type of trials have been mainly based on the concept of hierarchical Bayesian modeling [1, 2, 3], which allows for the exchange of information among patient subgroups exhibiting homogeneous treatment effects. However, frequentist approaches have also been proposed [4, 5], which are based on the Simon two-stage design [6] and can incorporate a homogeneity analysis at the point of an interim analysis. In this work, we aim to compare the efficiency of such frequentist designs for oncological phase II basket trials to Bayesian ones. In order to do so, we perform Monte Carlo simulations of a basket trial examining the response to a treatment in $K = 5$ patient subgroups in R [7] for both Bayesian and frequentist designs. We consider all possible scenarios of effect homogeneity and heterogeneity among patient subgroups while assuming the true underlying response rate to be either at an ineffective null response rate θ_0 or at an effective alternative response rate θ_a. We examine the efficacy of the designs and compare them based on Specificity, Sensitivity, Type I error rates, Type II error rates and expected trial sample size. Additionally, we discuss the advantages and disadvantages of the application of each design as well as the consequences of various prior choices for the discussed Bayesian designs.

References:

1. Berry, S.M., Broglio K.R., Groshen S., Berry D.A. (2013) Bayesian hierarchical modeling of patient subpopulations: efficient designs of phase II oncology clinical trials. Clinical Trials, 10:720-734.

2. Neuenschwander, B., Wandel, S., Roychoudhury, S., Bailey, S. (2015) Robust exchangeability designs for early phase clinical trials with multiple strata. Pharmaceutical Statistics; 15(2):123-134.

3. Simon, R., Geyer, S., Subramanian, J., Roychowdhury, S. (2016) The Bayesian basket design for genomic variant driven phase II trials. Seminars in Oncology; 43(1):1–6.

4. Cunanan, K., Iasonos, A., Shen, R., Begg, C.B., Gonen, M. (2017) An efficient basket trial design. Statistics in Medicine; 1036(10):1568-1579.

5. Krajewska, M., Rauch, G. (2018) Homogenitätsanalysen in frequentistischen Ansätzen der Auswertung von Basket Studien. Oral Presentation at the 63. Annual Conference of the German Association for Medical Informatics, Biometry and Epidemiology (GMDS), Osnabrück, Germany; doi: 10.3205/18gmds031.

6. Simon, R. (1989) Optimal two-stage designs for phase II clinical trials. Controlled Clinical Trials; 10: 1-10.

7. R Core Team. (2017) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria; URL https://www.R-project.org/.

151 Geometric representation of master protocols

Deepak Parashar

University of Warwick, United Kingdom

In recent years, clinical trials have gone beyond the objective of just finding a drug or therapy superior to the existing standard of care for all patients, to the objective of offering targeting therapies to specific sub-population of patients most likely to benefit. Such a classification is generally based on whether or not the patient has a particular molecular aberration or mutation or biomarker leading to the disease, coupled with a plan for the interventional drug to target that mutation. In cancer, however, due to the large number of tumour types and mutations, the data collected at the end of the trial will no longer be small but potentially large and multidimensional and hence standard statistical techniques could be of limited use. On the other hand, aspects of topological and geometric approaches have been useful to understand structural features of large datasets. In particular, hypersurfaces embedded in ambient Euclidean space provide a natural framework to capture the multiple dimensions. This presentation explores relevant features of hypersurfaces that could provide the structural representation of multidimensional stratified clinical trial data vis-à-vis 'master protocols' comprising multiple tumour types, multiple mutations and multiple drugs. Basket and Umbrella trial designs then emerge as subspaces of such hypersurfaces.

Data Fusion and Meta-Analysis II

152 Meta-analysis of full ROC curves: A parametric model based on flexible distributions of diagnostic test values

Annika Hoyer[1,2], Oliver Kuss[2]

[1]Department of Statistics, Ludwig-Maximilians-University Munich, Germany
[2]German Diabetes Center, Institute for Biometrics and Epidemiology

Diagnostic accuracy studies are often used to evaluate diagnostic tests at several thresholds, aiming to determine optimal thresholds for use in practice. While methods for the meta-analysis of the resulting full receiver operating characteristic (ROC) curves have already been proposed, they still come along with deficiencies as, for example, the numbers and values of thresholds have to be identical across studies or the concrete threshold values are ignored. Recently, we proposed a parametric approach that uses bivariate time-to-event models for interval-censored data to this task which compensates disadvantages of other approaches [1]. As an extension of this model, which increases the flexibility and also addresses the open point of model selection, we here suggest to use the generalized F family of distributions for the underlying diagnostic test values [2]. This flexible approach includes currently used distributions for the bivariate time-to-event model as special cases. We illustrate the model by the example of population-based screening for type 2 diabetes mellitus.

References

[1] Hoyer, A., Hirt, S., Kuss, O. (2018) Meta-analysis of full ROC curves using bivariate time-to-event models for interval-censored data. Research Synthesis Methods, 9(1):62-72
[2] Cox C. (2008) The generalized F-distribution: An umbrella for parametric survival analysis. Statistics in Medicine 27;4301-4312

153 The dark side of the force: multiple testing issues in network meta-analysis and how to address them

Orestis Efthimiou[1], Ian White[2]

[1]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland
[2]MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK

Standard models for network meta-analysis simultaneously estimate multiple relative treatment effects. This naturally raises concerns regarding a possible inflation of the Type I error. In this paper, we present theoretical arguments as well as results from simulations to illustrate the problem. We discuss how the multiple testing problem can be addressed via multi-level Bayesian modelling, where treatment effects are modelled exchangeably, and hence estimates are shrunk away from large values. We present a set of alternative models for network meta-analysis, and we show in simulations that in several scenarios these models can give much better type I error rates and power than the usual network meta-analysis model.

154 The importance of a study for treatment estimates in network meta-analysis

Gerta Rücker[1], Adriani Nikolakopoulou[2], Theodoros Papakonstantinou[2], Guido Schwarzer[1]
[1]Faculty of Medicine and Medical Center - University of Freiburg, Germany
[2]Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland

In pairwise meta-analysis, the contribution of each study to the pooled estimate is given by its inverse variance weight. For network meta-analysis (NMA), the contribution of direct (and indirect) evidence is easily obtained from the diagonal elements of the hat matrix. It is, however, not fully clear how to generalize this to the percentage contribution of each study to a NMA estimate. Several proposals exist in the literature, based on different approaches. Most of them come with some limitations; their results also do not in general agree (Krahn 2013, 2014; Salanti 2014; Riley 2018; Papakonstantinou 2018). In this talk, after briefly discussing these approaches, we want to question whether it is possible to obtain unique percentage contributions and discuss another approach. We define the importance of each study for a NMA estimate by the reduction of the estimate's variance when adding the given study to the others. An equivalent interpretation is the relative reduction of the precision when the study is removed. These numbers in general do not add to one and thus cannot be interpreted as percentage contributions". However, they generalize the concept of weights in pairwise meta-analysis in a natural way. Moreover, they are uniquely defined, easily calculated, and have an intuitive interpretation. We give some real examples for illustration.

155 On ranking multiple health interventions

Georgia Salanti[1], Adiani Nikolakopoulou[1], Dimitris Mavridis[3]

[1]University of Bern, Switzerland
[2]University of Ioannina, Greece

Ranking multiple competing treatments for the same health condition has been presented as one of the advantages of network meta-analysis (NMA). Several of the applied NMAs include in their aims the estimation of a treatment hierarchy according to several efficacy and safety outcomes. Estimated probabilities that each treatment assumes a certain rank, median and mean ranks of treatments and the surface under the cumulative ranking curve (SUCRA) have been proposed in the literature as statistical metrics that can be used to obtain a hierarchy of the competing treatments [1]. Non-probabilistic ranking can be obtained by sorting the relative treatments effects against a common comparator treatment.

Several articles call for caution when interpreting the ranking of treatments obtained from network meta-analysis. Kibret at al. performed a simulation study and concluded that «decisions should not be made based on rank probabilities especially when treatments are not directly compared as they may be ill-informed»[2]. An empirical study of the change in treatment hierarchy when data in the network changes concluded about ranking metrics that «... their interpretability is limited by the fact that they are driven predominantly by the estimated effect sizes, and that standard errors play an unduly small role in determining their position»[3]. Mbuagbaw et al used a published ??? where the evidence about the high-ranked treatments is of poor quality to conclude that this is «the most compelling reason to potentially mistrust rankings in general and SUCRA in particular» [4]. Veroniki et al state a few examples of published networks where the wide uncertainty around SUCRA values is interpreted as «the ranking statistic values might be unstable».

In this presentation we aim to explain why ranking metrics are not misleading per se. We propose separating the objectives of NMA (to answer a clearly defined treatment hierarchy question) from the estimated quantities used to answer this question (the ranking metric used to obtain the treatment hierarchy). We suggest that it is neither useful nor appropriate to compare the treatment hierarchies obtained by the various methods without first defining what the treatment hierarchy question is. Then we explain that a source of the apparent confusion in the literature is because it is not widely acknowledged that each ranking metric aims to answer a different treatment hierarchy question and because researchers fail to define what "the best treatment" means in each setting. Through theoretical examples we show that each ranking metric encompasses differently the uncertainty in the estimation of the relative treatment effects, and this can lead to remarkably different treatment hierarchies. We discuss that none of the different hierarchies obtained is right or wrong, as each one addresses a different question.

1. Salanti G et al. J ClinEpidemiol. 2011;64:163–71.
2. Kibret T et al. Clin Epidemiol. 2014;6:451–60.
3. Mills EJ et al. BMJ. 2013;347:f5195.
4. Mbuagbaw L et al. Syst Rev. 2017;6:79.

156 Using flow decomposition to estimate the contribution of studies in network meta-analysis

Theodoros Papakonstantinou[1], Adriani Nikolakopoulou[1], Gerta Rücker[2], Anna Chaimani[3], Guido Schwarzer[2], Matthias Egger[1], Georgia Salanti[1]

[1]University of Bern Institute of Social and Preventive Medicine (ISPM), Switzerland
[2]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center, University of Freiburg, Freiberg, Germany
[3]Paris Descartes University, INSERM, UMR1153 Epidemiology and Statistics, Sorbonne Paris Cité Research Center (CRESS), METHODS Team

It has been suggested that in order to infer upon study limitations in a network meta-analysis (NMA) effect, study limitations judgements on observed direct effects shall be considered jointly with their contribution in the NMA effect (1). The key instrument in this approach is the percentage contribution matrix: a matrix that shows how much each direct treatment effect contributes to the estimation of the NMA treatment effect.
To derive the percentage that each direct treatment effect contributes to each NMA treatment effect we use ideas from graph theory. We start with the 'projection' matrix in a two-step NMA model, called H matrix, which is analogous to the hat matrix in a linear regression model. A previous attempt to translate H entries to percentage contributions by normalising them was incorrect (1). We developed a novel method to translate H entries to percentage contributions based on the observation that the rows of H can be interpreted as flow networks (2). We present an algorithm that identifies the flow of evidence in each path and decomposes it to direct comparisons.
To illustrate the methodology, we use a network of interventions of topical antibiotics without steroids for underlying eardrum perforations (comparing no treatment (x), quilone antibiotic (y), non-quilone antibiotic (u) and antiseptic (v)). Focusing on the NMA effect x vs y, there are three different paths from x to y, one based on direct evidence, xy, and two based on indirect evidence, xv,vy and xv,vu,uy. The algorithm identifies their flows and derives the percentage contributions of each direct comparisons by splitting equally paths' flows to the involved comparisons. We will show how this idea is expanded in complex network structures.
The derivation of the percentage contributions of direct to NMA treatment effects is of great importance when examining the impact of particular study characteristics to the findings of a NMA.
1. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JPT. Evaluating the quality of evidence from a network meta-analysis. PloS One. 2014;9(7):e99682.
2. König J, Krahn U, Binder H. Visualizing the flow of evidence in network meta-analysis and characterizing mixed treatment comparisons. Stat Med. 2013 Dec 30;32(30):5414–29.

Measurement and Measurement Error I (Measurement Error)

157 Measurement error and misclassification of variables in observational epidemiology: basic knowledge and practical guidance

Laurence S. Freedman[1], Paul Gustafson[2], Pamela Shaw[3], Raymond J. Carroll[4], **Veronika Deffner**[5], Kevin Dodd[6], Victor Kipnis[6], Ruth Keogh[7], Helmut Küchenhoff[5], Janet Tooze[8]

[1]Gertner Institute for Epidemiology and Health Policy Research, Tel Hashomer, Israel
[2]The University of British Columbia, Vancouver, Canada
[3]University of Pennsylvania Perelman School of Medicine, Philadelphia, USA
[4]Texas A&M University, College Station, USA
[5]Ludwig-Maximilians-Universität München, Munich, Germany
[6]National Cancer Institute, Bethesda, USA
[7]London School of Hygiene and Tropical Medicine, London, Great Britain
[8]Wake Forest School of Medicine, Winston-Salem, USA

Measurement error and misclassification of variables frequently arise in observational epidemiology and are inherent in data collection. Although appropriate methods are available, these data uncertainties are often not or not sufficiently considered in the statistical analysis (Shaw et al., 2018), possibly leading to biased results. Topic Group 4 of the STRATOS initiative (STRengthening Analytical Thinking for Observational Studies) have developed guidance to facilitate the suitable handling of measurement error and misclassification in epidemiological studies. We present some essential aspects of our guidance paper.

We differentiate between errors of continuous (measurement error) and categorical (misclassification) variables, between errors of outcome variables and covariates, between different types of dependence between the error-prone and the precise variable (classical measurement error, linear measurement error, Berkson error) and between error-prone variables which provide (differential error) or do not provide (non-differential error) additional information beyond the precise variable.

Biased estimates of regression coefficients result from non-differential classical measurement error of a covariate, non-differential linear measurement error, non-differential Berkson error of an outcome variable and differential error. In some situations, inference regarding the null hypothesis is not valid. Lower and upper percentiles of the distribution of a continuous variable are biased when estimated from a variable that has measurement error, as is the mean of a continuous variable when there is linear measurement error.

We provide theoretical and practical guidance for suitably handling measurement error and misclassification regarding three aspects: ancillary studies to provide information on the error, statistical analysis methods and software. Since information about the measurement error is essential for its consideration in the statistical analysis, ancillary studies are necessary in addition to the main study to assess the nature and magnitude of measurement error. Further, we introduce methods for adjusting the statistical analysis for measurement errors like regression calibration, moment reconstruction and moment-adjusted multiple imputation, multiple imputation, likelihood methods, Bayesian methods and simulation-extrapolation (SIMEX). An overview of available software in various software packages for dealing with measurement error and misclassification is presented.

158 Estimation methods to address correlated covariate and time-to-event error

Pamela Ann Shaw[1], Eric Oh[1], Bryan Shepherd[2], Thomas Lumley[3]

[1]University of Pennsylvania, United States of America
[2]Vanderbilt University, United States of America
[3]University of Auckland, New Zealand

Electronic health records (EHR) data are increasingly used in medical research, but the data, which typically are not collected to support research, are often subject to measurement error. These errors, if not addressed, can bias results in association analyses. Methodology to address covariate measurement error has been well developed; however, methods to address errors in time-to-event outcomes are relatively underdeveloped. We will consider methods to address errors in both the covariate and time-to-event outcome that are potentially correlated. We develop an extension to the popular regression calibration method for this setting. Regression calibration has been shown to perform well for settings with covariate measurement error (Prentice, 1982; Shaw and Prentice, 2012), but it is known that this method is generally biased for nonlinear regression models, such as the Cox model for time-to-event outcomes. Thus, we additionally propose raking estimators, which will be unbiased when an unbiased estimating equation is available on a validation subset. Raking is a standard method in survey sampling that makes use of auxiliary information on the population to improve upon the simple Horvitz-Thompson estimator applied to a subset of data (e.g. the validation subset). We demonstrate through numerical studies that raking can improve upon the regression calibration estimators in certain settings with failure-time data. We will discuss the choice of the auxiliary variable and aspects of the underlying estimation problem that affect the degree of improvement that the raking estimator will have over the simpler, biased regression calibration approach. Detailed simulation studies are presented to examine the relative performance of the proposed estimators under varying levels of signal, covariance, and censoring. We further illustrate the methods with an analysis of observational EHR data on HIV outcomes from the Vanderbilt Comprehensive Care Clinic.

159 SIMEX for Box-Cox transformed measurements

Timm Intemann[1], Kirsten Mehlig[2], Stefaan De Henauw[3], Alfonso Siani[4], Tassos Constantinou[5], Luis A. Moreno[6], Dénes Molnár[7], Thomas Veidebaum[8], Iris Pigeot[1]

[1]Leibniz Institute for Prevention Research and Epidemiology - BIPS, Germany
[2]Sahlgrenska Academy at University of Gothenburg, Institute of Medicine, Department of Public Health and Community Medicine, Sweden
[3]Ghent University, Department of Public Health, Belgium
[4]Institute of Food Sciences, National Research Council, Italy
[5]Research and Education Institute of Child Health, Cyprus
[6]University of Zaragoza, GENUD (Growth, Exercise, Nutrition and Development) Research Group, Instituto Agroalimentario de Aragón (IA2), Instituto de Investigación Sanitaria Aragón (IIS Aragón), Centro de Investigación Biomédica en Red Fisiopatología de la Obesidad y Nutrición (CIBERObn), Spain
[7]University of Pécs, Medical School, Department of Pediatrics, Hungary
[8]National Institute for Health Development, Department of Chronic Diseases, Estonia

In nutritional epidemiology we have to cope with varying requirements regarding the development of statistical methods due to different types of measurement error (ME). Each new dietary assessment tool requires a specific ME model and an appropriate correction method since dietary data is usually error-prone. Otherwise, the naïve approach would lead to biased effect estimators when investigating the association between dietary exposures and health outcomes. By now the 24-hour dietary recall (24HDR) has become the preferred tool to assess diet in large epidemiological studies. 24HDR data are characterised by skewed distributions with high intra-individual variance. For such data a nonlinear mixed effects ME model using the Box-Cox transformation was proposed following the assumption that single recalls are unbiased for the individual usual intake (1). The corresponding correction method, the NCI-method, is based on the regression calibration approach. However, a recent review revealed that studies with inadequate treatment of ME remain commonplace in nutritional epidemiology (2).
We propose a new correction method for 24HDR data based on the simulation extrapolation (SIMEX) approach. The idea of SIMEX is (i) to generate remeasurements of the original data with additional ME in order to obtain even more biased estimates and (ii) to estimate the functional association between the level of ME and effect estimations which is then extrapolated to the hypothetical case of zero ME.
Assuming the error model of the NCI-method we define an algorithm that ensures that the remeasurements of the recalls are unbiased on the original scale while the additional error is normally distributed with variance $\zeta\sigma_\varepsilon^2$ on the Box-Cox transformed scale. We give the justification for the proposed algorithm by showing that the key property of SIMEX holds, i.e. that the mean squared error (MSE) of the individual remeasurements converges to zero if the level of additional ME converges to the hypothetical value of $-\sigma_\varepsilon^2$.
To demonstrate its practical feasibility, we apply this method to real data from the I.Family study (3) to correct the estimated association between salt intake and blood pressure. Furthermore, in a simulation study we compare the proposed SIMEX algorithm with the NCI-method. The SIMEX algorithm leads to estimates with either lower MSE or to nearly unbiased estimates.

The proposed SIMEX algorithm is an alternative method for modelling associations between health outcomes and dietary exposures assessed with a 24HDR. Its advantages are the easy implementation, a lower MSE compared to the NCI-method and the provision of a comprehensible illustration showing consequences of ME which could be beneficial especially for non-statisticians.

References

1. Kipnis, V. et al. (2009). Modeling data with excess zeros and measurement error: Application to evaluating relationships between episodically consumed foods and health outcomes. Biometrics 65, 1003-1010.

2. Shaw, P. et al. (2018). Epidemiologic analyses with error-prone exposures: Review of current practice and recommendations. Ann Epidemiol 28, 821-828.

3. Ahrens, W. et al. (2017). Cohort profile: The transition from childhood to adolescence in European children - how I.Family extends the IDEFICS cohort. Int J Epidemiol 46, 1394-1395j.

160 Accounting for misclassification in automated disease diagnosis based on medical image data

Felix Günther[1,2], Caroline Brandl[2], Iris M. Heid[2], Helmut Küchenhoff[1]
[1]Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Germany
[2]Department of Genetic Epidemiology, University Regensburg, Germany

Recently, there is a rising interest in the automated classification of medical image data for disease diagnosis using machine learning techniques. There is a substantial number of published algorithms for the diagnosis of various diseases based on different types of images; many are based on convolutional neural networks. Some examples are the classification and detection of skin cancer based on images of skin lesions or the diagnosis of eye diseases based on image data from retina photography or optical coherence tomography. Such approaches are highly promising to accelerate diagnosis in a clinical setting and to facilitate outcome assessment in large-scale epidemiological studies: while a manual classification of disease based on the images can be very time consuming and require especially trained personnel, an automated analysis of images can speed up outcome assessment substantially facilitating large data association analyses, e.g. to better understand risk factors of diseases. The diagnoses based on such algorithms are, however, assumed to be imperfect and can yield error-prone results. This is particularly relevant, if the images to be diagnosed differ structurally from the learning data. A subsequent naive analysis of the resulting data that ignores the existing misclassification leads in general to biased results. Methods for bias adjustments rely on information about the misclassification (e.g. sensitivity and specificity). This information is usually available for the automated procedures based on methods like cross-validation or the evaluation of the performance in independent test data.
We illustrate this issue based on data of the UK Biobank with a focus on the association of genetic variants with the retinal disease age-related macular degeneration (AMD). In this study, retinal images as well as genetic variants data are available for >60000 persons. The images have not been graded manually towards AMD; there is only information on self-reported AMD, which might be unreliable for early AMD stages. We therefore apply a published convolutional neural network ensemble to classify the available fundus images towards AMD status and evaluate the performance of the algorithm based on a subset of images that were manually graded by an experienced ophthalmologist. We quantify the misclassification in the predictions. The subsequent association analysis of genetic variants and the predicted AMD status suffers from response misclassification (misclassification in the AMD status). We investigate consequences of ignoring this misclassification by comparing the results of a naive association analysis ignoring the existing misclassification with a maximum likelihood approach that considers misclassification in the responses and discuss assumptions associated with different analysis strategies.

Official Statistics and Survey Statistics II

161 Sampling in Times of High Immigration: The IAB-BAMF-SOEP Survey of Refugees

Simon Kühne[1], Jannes Jacobsen[2], Martin Kroh[1,2]

[1]University of Bielefeld

[2]The German Institute for Economic Research DIW Berlin

Over the course of 2013 to 2016, over one million refugees arrived in Germany, around 890,000 of them in 2015 alone. The growing refugee population posed a major challenge for Germany's policy makers, civic administrators, and society at large, in finding new approaches to registration procedures, housing, and social and economic integration. To design policies and programs that meet these needs, government administrators, politicians, and the public require robust analyses of the accompanying social and demographic changes based on timely, valid, and reliable empirical data. Yet despite the urgent need for quantitative data on this target group, survey organizations and data collection agencies had little experience gaining access to the target population and approaching and surveying them effectively.

In late 2015, when the influx reached its peak, the Institute for Employment Research (IAB), the Research Department of the Federal Office for Migration and Refugees (BAMF-FZ), and the Socio-Economic Panel (SOEP) joined together in a cooperative longitudinal project to survey a nationwide random sample of refugee households in Germany: the IAB-BAMF-SOEP Survey of Refugees. In the first wave in 2016, a total of 3,336 households were interviewed, resulting in 4,527 face-to-face interviews with individual adult respondents. An enlargement sample in 2017 added an additional 1,519 households and 2,252 individuals.

In this presentation, we summarize the sampling and fieldwork design as well as the challenges faced in the IAB-BAMF-SOEP Survey of Refugees. We discuss the sequential strategy applied for sampling the group of refugees and asylum seekers who arrived in Germany in 2015 and 2016, a period in which large numbers of refugees were arriving daily, and in many cases their initial accommodations were only temporary. Moreover, the presentation discusses alternative survey instruments introduced for the difficult-to-interview population of the IAB-BAMF-SOEP Survey of Refugees, including translated questionnaires and audio files.

162 European Union Minorities and Discrimination Survey (EU-MIDIS II) - Surveying immigrants and ethnic minorities in the 28 EU Member States

Rossalina Latcheva, David Reichel, **Ursula Till-Tentschert**
EU Agency for Fundamental Rights, Vienna

FRA's (EU Agency for Fundamental Rights) Second European Union Minorities and Discrimination Survey (EU-MIDIS II) surveyed 25,500 people from different ethnic minority and immigrant backgrounds across all 28 EU Member States in 2015 and 2016. It presents one of the largest international surveys covering groups of immigrants collecting information on experiences of discrimination in different areas of life (labour market, education, housing, health and other services), criminal victimisation (including hate crime), social inclusion and societal participation. The groups selected are immigrants and descendants of immigrants – identified via country of births and parents' country of birth – from Turkey, Sub-Saharan Africa, North Africa, South Asia and Asia, as well as recent immigrants from all countries outside the EU, Roma and the Russian minority. In each of the 28 EU countries one to three groups were selected and surveyed. Further developing the methods used in EU-MIDIS I from 2008, EU-MIDIS II achieved probability sampling methods in all countries (except Luxembourg), using the best possible sampling method per group and country. In the absence of registers, multi-stage sampling, with innovative methods to increase the efficiency of screening the population for the target groups, were employed in most countries. In addition, location sampling methodology was applied, if more traditional sampling methods were not deemed feasible. In some countries a combination of methods was used. The presentation provides an overview of the methods employed and lessons learned concerning sampling and surveying immigrants in the EU, which supports any future efforts on sampling immigrants and ethnic minorities in EU.

163 Building a Sampling Frame for Migrant Populations via an Onomastic Approach – One or more lessons learned from the Austrian Immigrant Survey 2016

Dimitri Prandner[1], Martin Weichbold[2]

[1]Johannes Kepler University JKU Linz
[2]Universität Salzburg

Immigrants are traditionally viewed as hard to survey. Their number is often too small to be analysed via data gained in general population surveys and registers to identify them are often missing or incomplete. Therefore, researchers are forced to use alternative strategies to draw a sample.
For the Austrian Immigrant Survey 2016, an onomastic (name based) approach was used, establishing a sampling frame in a 2-step procedure, with the goal to recruit 300 individuals from Turkish as well 300 individuals from ex-Yugoslavian origin for telephone interviews.
The first step was based on selecting names with a high probability to come from the chosen target populations. In this case given names with a high (>80%) probability to match the target population were included. In a second step the Austrian phone register was screened for those names and a random sample could be drawn.
In theory this would be advantageous as the onomastic component allows for a studies with a hard to reach population to feature a random sample; as long as a comprehensive list including names can be accessed (Fernandez et al., 2006).
However, despite its wide use and adoption for migrant surveys, there is still no consensus if it is an adequate strategy, when it comes to building representative survey samples for specific populations (Font & Méndes, 2013).
The planed presentation describes the concept and the implementation of the sampling and evaluates the sample that could be realized. And results show a mixed picture. Firstly, while it was indeed possible to build a heterogeneous sample for both populations, the criteria of representation could not be matched – especially when it comes to the markers of citizenship and education. Secondly, there is an – expected – lack of information when it comes to non-contacts in the field phase.
Additionally the results hint at the fact that the built frame may have had some structural issues resulting in a contact bias.

164 Integration of migrant populations into health monitoring in Germany - Results from a feasibility study

Marie-Luise Zeisler, **Johannes Lemcke**, Leman Bilgic, Claudia Santos-Hövener, Patrick Schmich

Abteilung für Epidemiologie und Gesundheitsmonitoring, Robert Koch-Institut, Berlin

Background:
Over the last years Germany has become one of the most popular migration destinations in the world. It is therefore important to collect reliable data on migrants' health. To improve inclusion of migrant populations in German Health Monitoring, the Robert Koch Institute (RKI) has launched the project Improving Health Monitoring in Migrant Populations (IMIRA). One of the project's aims was to test different recruitment strategies to increase participation of people with migration background (PMB) in national health interview surveys.
Methods
A multilingual feasibility study was conducted between January and May 2018 in two German federal states. A sample from population registers was used. The target populations were persons with Turkish, Polish, Romanian, Syrian and Croatian citizenship living in Germany (n=9,068). Age and gender strata were applied. Multilingual online questionnaires, a multilingual study hotline and home visits with bilingual interviewers were offered in a sequential mixed-mode design. To evaluate usability and effectiveness an experimental design was applied for Turkish and Syrian migrants.
Results
Overall response rate was 15.9% (n=1,190). Response rate differences between the groups were statistically significant ($?^2$= 218.58; p=0.00), ranging from 8.6% in the Turkish to 24.3% in the Syrian group. Little differences between the groups were found concerning the use of the different modes of participation, but the home visits led to a remarkable increase (+5.4% in the Turkish group; +7.3% in the Syrian group). The majority of respondents (57.1%) used the translated questionnaire. This finding is related with lower age, lower educational level and poorer subjective health status.
Conclusion
The response rates of PMB can be increased by personal contact. A multilingual questionnaire showed to be effective to reach PMB with lower educational level and poorer subjective health conditions. The results will be implemented in further health monitoring at RKI.

Statistics in Behavorial and Educational Sciences I (Educational Sciences)

165 Using timing information to model missing values in test data

Steffi Pohl

Freie Universität Berlin, Germany

In test data of low-stakes assessments missing values occur due to different reasons. Test takers omit items, do not reach the end of the test due to time limits or due to quitting. For adequate modeling of missing responses, a thorough understanding of the nonresponse mechanisms is vital. Due to the growing use of computer-based assessment, a rich body of additional data becomes easily accessible. These additional data may contain valuable information on the examinees' test taking behavior and, thus, on the missing data process. In my work I bring together research on missing values with approaches for modeling timing data and simultaneous model response and nonresponse behavior. I propose different models for different missing mechanisms. The proposed models allow a) to model different kinds of missing responses and b) to gain a deeper understanding of the examinees' nonresponse behavior and the processes underlying item-level nonresponse. As test takers use different test taking strategies, which are reflected in different timing data and occurrence of missing values, instead of reporting just one competence score, I suggest reporting a profile of different aspects that describe the performance of the test takers. I will discuss the implications of this choice of missing data treatment as compared to tradition approaches for competence assessment in large-scale assessments.

166 Paradoxical properties of parameter estimates in multidimensional models

Pascal Jordan

University of Hamburg, Germany

The application of multidimensional models in educational/psychological testing for the classification of test takers can lead to scoring methods which not only are counterintuitive, but which moreover can pose a threat in terms of social acceptability. For instance, it was shown by Hooker, Finkelman & Schwartzman (2009) in the context of multidimensional item response models, that changing correct into incorrect responses can lead to an increasing ability estimate in some domain. Surprisingly, this paradoxical effect is not an artefact but appears in every multidimensional model of compensatory type. This effect (and its high prevalence) is less known among researchers and I aim at a) introducing the phenomenon to researchers not familiar with it; b) providing a mathematical explanation of the underlying effect and c) discussing recent results on the paradoxical scoring effect.

167 A Hierarchical Latent Response Model for Inferences about Examinee Engagement in Terms of Guessing and Item-Level Nonresponse

Esther Ulitzsch[1], Matthias von Davier[2], Steffi Pohl[1]

[1]Freie Universität Berlin, Germany
[2]National Board of Medical Examiners

In low-stakes large scale assessments (LSAs), test performance comes with little or no consequences for examinees themselves, so that examinees may not to be fully engaged when answering the items: Instead of engaging in solution behavior, disengaged examinees might randomly guess or generate no response at all. When ignored, examinee disengagement poses a severe threat to the validity of results obtained from LSAs. Up to now, in educational measurement statistical modeling approaches have been proposed that account for nonresponse or for guessing, but did not consider both types of disengaged behavior simultaneously. We bring together research on modeling examinee engagement and research on missing values and present a hierarchical latent response model for identifying and modeling the processes associated with examinee disengagement jointly with the processes associated with engaged responses. To that end, we employ mixture models that identify disengagement on the item-by-examinee level by assuming different data-generating processes underlying item responses, omissions as well as response times associated with engaged and disengaged behavior. By modeling examinee engagement within a latent response framework, the model allows for a) assessing how examinee engagement relates to ability and speed, b) identifying items that are likely to evoke disengaged test-taking behavior as well as c) retrieving less biased and more reliable ability estimates. Parameter recovery of the proposed model is studied within a simulation study. An illustration of the model by means of an application to real data is presented.

168 Revisiting Dispersion in Count Data Item Response Theory Models: The Conway-Maxwell-Poisson Counts Model

Boris Forthmann[1], **Daniela Gühne**[2], Philipp Doebler[2]
[1]WWU Münster, Germany
[2]TU Dortmund, Germany

Count data naturally arise in several areas of cognitive ability testing, e.g., processing speed, memory, verbal fluency, and divergent thinking. Contemporary count data item response theory models, however, are not flexible enough, especially to account for over and underdispersion at the same time. For example, the Rasch Poisson counts model assumes equidispersion (conditional mean and variance coincide) which is often violated in empirical data. This work introduces the Conway-Maxwell-Poisson counts model that can handle underdispersion (variance lower than the mean), equidispersion, and overdispersion (variance larger than the mean) in general and specifically at the item level. A simulation study revealed satisfactory parameter recovery at moderate sample sizes and mostly unbiased to conservative standard errors for the proposed estimation approach. In addition, plausible empirical reliability estimates resulted, while those based on the Rasch Poisson counts model were biased downwards (underdispersion) and biased upwards (overdispersion) when the simulation model deviated from equidispersion. Finally, verbal fluency data were analyzed and the Conway-Maxwell-Poisson counts model with item-specific dispersion parameters fit the data best. Dispersion parameter estimates indicated underdispersion for three out of four items. Overall, these findings indicate the feasibility and importance of the suggested flexible count data modeling approach.

Time Series Analysis III (Change Points)

169 Fixed-Bandwidth CUSUM Tests Under Long Memory

Kai Rouven Wenger, Christian Leschinski
Institute of Statistics, Leibniz University Hannover, Germany

To test for structural change under long memory, we propose a family of self-normalized CUSUM tests. These apply non-parametric kernel-based fixed-b and fixed-m long-run variance estimators. The test statistics have well-defined limiting distributions that only depend on the long-memory parameter. A Monte Carlo simulation shows that these tests provide finite sample size control while outperforming competing procedures in terms of power.

170 Backward CUSUM for Testing and Monitoring Structural Change

Sven Otto, Jörg Breitung
University of Cologne, Germany

A drawback of the conventional CUSUM test for structural breaks of Brown et al. (1975) and its monitoring version of Chu et al. (1996) is the low power and the huge detection delay. The CUSUM statistic cumulates recursive residuals sequentially from the beginning to the end of the sample. The residuals that are cumulated in the beginning of the CUSUM sequence are the ones from the time period before the potential break point. However, the residuals that contain relevant information about the break point are the ones of the post-break period.
We propose two detector statistics, the Backward CUSUM and the Stacked Backward CUSUM. Both detectors sequentially cumulate the recursive residuals in reversed order. While the Backward CUSUM starts cumulating at the end of the sample, the Stacked Backward CUSUM backwardly cumulates the residuals while the endpoint is moving together with the time the detector is evaluated. Therefore the Stacked Backward CUSUM is also suitable for monitoring, while the Backward CUSUM can only be applied to retrospective testing. We derive the limiting distributions of their maximum statistics for different boundary functions under the null hypothesis of no break and under local alternatives of a single break in the constant. The distributions are derived for retrospective testing, fixed endpoint monitoring and infinite horizon monitoring. Critical values are obtained via simulation of the corresponding functional of Brownian motions. In the retrospective setting, the local power of both tests is substantially higher than for the conventional CUSUM test if the break occurs after one third of the sample size. In the case of monitoring, the detection delay of the Stacked Backward CUSUM under local alternatives is shown to be much lower than for the monitoring CUSUM detector by Chu et al. (1996).
Another problem is that the CUSUM test has no power if the mean of the regressor is orthogonal to the structural change. In order to obtain tests that have power against these breaks we extend the existing invariance principle for recursive residuals to a multivariate version. For the Backward CUSUM and the Stacked Backward CUSUM we replace the recursive residuals with normed weighted score vectors and derive limiting distributions as well as critical values. The theoretical size and power results are confirmed in finite sample simulations.

171 Consistent change point detection in a nonparametric time series regression model

Maria Mohr, Natalie Neumeyer
University of Hamburg, Germany

A weakly dependent time series regression model with multivariate covariates and univariate observations is considered, for which we develop a procedure to detect whether the nonparametric conditional mean function is stable in time against change point alternatives. Our proposal is based on a modified CUSUM type test procedure, which uses a sequential marked empirical process of residuals. We show weak convergence of the considered process to a centered Gaussian process under the null hypothesis of no change in the mean function and a stationarity assumption. This requires some sophisticated arguments for sequential empirical processes of weakly dependent variables. As a consequence we obtain convergence of Kolmogorov-Smirnov and Cramér-von Mises type test statistics. The proposed procedure acquires a very simple limiting distribution and nice consistency properties, features from which related tests are lacking. A simulation study is conducted to investigate the finite sample performance of our tests.

172 Change-point tests based on self-normalization and subsampling for LRD data

Annika Betken, Martin Wendler
Ruhr-Universität Bochum, Germany

We consider robust change-point tests based on (self-normalized) rank statistics to identify changes in the mean of long-range dependent data. Unlike common testing procedures an application of self-normalized tests does not require the estimation of an unknown standardization. Approximating the distribution of the self-normalized test statistics by subsampling procedures additionally bypasses estimation of parameters in the limit distribution. It can be shown that the so-called ßampling window methodïs valid for any statistic applied to long-range dependent subordinated Gaussian processes which satisfy mild regularity conditions.

173 Robust change point tests using bounded transformations

Alexander Dürre, Roland Fried
Technische Universität Dortmund, Germany

Classical moment based change point tests like the cusum test are very powerful under Gaussian time series with no more than one change point but behave poorly under heavy tailed distributions and corrupted data. A new class of robust change point tests based on cusum statistics of robustly transformed observations is proposed. This framework is very flexible, depending on the used transformation one can detect amongst others changes in the mean, scale or dependence of a possibly multivariate time series. The calculation of p-values can be simplified by using asymptotics which yields a computational complexity of $Tlog(T)$ where T is the number of observations. We apply our general approach to detect changes in the covariance structure of a multivariate time series. Simulations indicate high power under Gaussianity as well as heavy tails.

Data Fusion and Meta-Analysis III

174 Count Outcome Meta-Analysis with Mixed Arm Information

Dankmar Boehning[1], Patarawan Sangnawakij[2], Holling Heinz[3]

[1]University of Southampton, United Kingdom
[2]Thammasat University, Thailand
[3]University of Münster, Germany

Typically, meta-analysis proceeds in the following 2-stage process. From published (or otherwise collected) studies the effect measure of interest is retrieved and a meta-analytic data set of all collected evidence including the effect measure of interest, the risk ratio, say, is compiled. This is then the basis for meta-analytic inference (in a 2-stage meta-analysis).
When studies report count outcome data frequently only one of the two (or more) groups to be compared has the report count of interest. To proceed in a 2-stage analysis only studies are considered which allow computaiton of the effect measure of interest, the relative risk, say.
We suggest here a direct approach using all available information in a mixed Poisson or binomial model. The approach will also allow diagnosing any biasing effect of the inclusing of single group infomation. In general, the approach provided increased efficiency in dependence of the amount of mixed information. We consider this as a hypbrid 1-stage meta-analysis, settling between individual person and 2-stage meta-analysis.

175 Robust covariance estimation in mixed-effects meta-regression models - A simulation study

Thilo Welz, Markus Pauly
Universität Ulm, Germany

One fruitful way to explain heterogeneity in a meta-analysis is to introduce study specific covariates. Researchers may then consider statistical tests regarding the significance of such moderators. In doing so, special consideration should be made regarding a possibly heteroscedastic covariance structure of the data. In this talk we consider tests based on different heteroscedasticity consistent (HC) robust covariance estimators of Jackknife- and White-type in the setting of a mixed-effects meta-regression model. In an extensive simulation study we compare these with the powerful Knapp-Hartung method in a multitude of settings, also studying the effect of deviations from the usual normality assumption for the model random effects. Special attention is given to the scenario of few studies. We close with a recommendation for which covariance estimator should be used in which setting.
References:
(1.) Cribari-Neto, F., Souza, T. C., and Vasconcellos, K. L. (2007). Inference under heteroscedasticity and leveraged data. Communications in Statistics Theory and Methods, 36(10):1877–1888.
(2.) Jackson, D. and White, I. (2018). When should meta-analysis avoid making hidden normality assumptions? Biometrical Journal, Volume 60, Issue 6,
doi: 10.1002/bimj.201800071.
(3.) Knapp, G. and Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. Statistics in Medicine, 22(17):2693–2710.
(4.) MacKinnon, J. G. and White, H. (1985). Some heteroscedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of Econometrics, 29(3):305–325.
(5.) Pauly, M. and Welz, T. (2018). Contribution to the discussion of "When should meta?analysis avoid making hidden normality assumptions?" Biometrical Journal, Volume 60, Issue 6, doi: 10.1002/bimj.201800184.
(6.) Viechtbauer, W., Lopez-Lopez, J. A., Sanchez-Meca, J., and Marin-Martinez, F. (2015). A comparison of procedures to test for moderators in mixed-effects meta-regression models. Psychological Methods, 20(3):360.

176 Classification of tail-adjusted heterogeneity priors in the Bayesian meta-analysis estimated by bayesmeta

Malgorzata Roos, Sona Hunanyan, Leonhard Held
University of Zurich, Switzerland

The Cochrane strongly recommends meta-analysis to aggregate scientific knowledge. Frequently, however, only a very small number of studies is available for meta-analysis [Friede et al., 2017], and thus the results can be misleading due to unobserved heterogeneity. The Bayesian methodology mitigates this problem by incorporating priors on heterogeneity standard deviation.
Recently, the Bayesian meta-analysis became easily accessible through the R package bayesmeta [Roever, 2018], which focuses on Bayesian inference for a normal-normal hierarchical model (NNHM). Among others, bayesmeta offers Half Normal (HN), Half Cauchy (HC), Exponential (EXP) and Lomax (LMX) heterogeneity priors.
Although it is well known that heterogeneity priors can greatly affect the posterior inference, a systematic account of their impact on NNHM is lacking. It is unclear to which extent the posteriors are informed by the data. Moreover, a formal sensitivity assessment [Roos et al., 2015] for NNHM is missing.
In the context of a medical meta-analysis we apply HN, HC, EXP and LMX priors, which are made comparable through a suitable tail-adjustment. We show that heterogeneity priors affect the resulting posteriors differently, depending on the mathematical form of the density function. For example, the impact of HN and EXP is comparable but clearly differs from HC and LMX. Heterogeneity priors' properties can be classified with respect to the ability to let the data speak for themselvesänd the resulting posterior sensitivity. An excellent agreement of the bayesmeta-driven results with those generated by other Bayesian general-purpose software systems ensures the validity of our findings in a much broader context. Our approach has the potential to be implemented within the bayesmeta package, providing additional user-friendly features.
References
T. Friede, C. Roever, S. Wandel, and B. Neuenschwander. Meta-analysis of two studies in the presence of heterogeneity with applications in rare diseases. Biometrical Journal, 59(4):658-671, 2017.
M. Roos, T.G. Martins, L. Held, and H. Rue. Sensitivity analysis for Bayesian hierarchical models. Bayesian Analysis, 10(2):321-349, 2015.
C. Roever. Bayesian random-effects meta-analysis using the bayesmeta R package. (arXiv:1711.08683v1 [stat.CO] 23 Nov 2017), 2018.

177 Recovery of IPD inferences from key IPD summaries only: application to distributed computing under privacy constraints

Federico Bonofiglio, Harald Binder, Martin Schumacher

Institut für medizinische Biometrie und Statistik, Uniklinik Freiburg/Universität Freiburg, Germany

We consider fusion of original Individual Participant Data (IPD) between different locations/servers under disclosure constraints. That is, locations cannot share original IPD but only anonymous IPD proxies or summaries, that are later pooled. This procedure is known as distributed computing (also implemented in the DataSHIELD infrastructure) and obtaining original IPD inferences while protecting privacy can be complicated here. For example, in DataSHIELD original IPD GLM regressions can be anonymously obtained but adjusting for a between locations random-effect can be already challenging. Here our goal is to extend the inferential scope with the following approach. Instead of focusing on privacy-complying inferences we first generate a private version of the original IPD and then compute IPD inferences on it. Private IPD reconstructions are generated via a copula inversion technique that uses original empirical IPD marginal moments and correlation matrix, sent by each location, as input data only. We show we can well recover fixed and latent effect estimates of an original IPD multi-variate Logistic regression from such private IPD reconstructions, but the scope of inferential recovery can be much larger here. Application of the method to distributed computing is readily possible.

Measurement and Measurement Error II (Measurement Error and missing data)

178 Comparing cohorts from distinct sources: The issue of differently operationalized predictor variables

Dominikus Stelzer[1], Julia Ortner[2], Louis Velthuis[2], Reyn van Ewijk[3], Anita Arslanow[4,5], Michael Nagel[4,6], Marc Nguyen-Tat[4,6,7], Peter R. Galle[4,6], Frank Lammert[5], Erik Farin-Glattacker[1], Harald Binder[1], Erika Graf[1]

[1]Institute of Medical Biometry and Statistics (IMBI), Faculty of Medicine and Medical Center – University of Freiburg, Freiburg
[2]Department of Controlling, Johannes Gutenberg University Mainz, Mainz
[3]Department of Economics, Johannes Gutenberg University Mainz, Mainz
[4]Department of Internal Medicine I, University Medical Center Mainz, Mainz
[5]Department of Internal Medicine II, Saarland University Medical Center, Homburg
[6]Cirrhose Centrum Mainz (CCM), University Medical Center Mainz, Mainz
[7]Department of Internal Medicine II, Medical Center Osnabrück, Osnabrück

In medical research, the randomized controlled trial (RCT) is regarded as the "gold standard", because it generally provides the best evidence. However, in some research areas like health care, randomization of patients is often not possible, for example due to ethical considerations or time and money constraints. On the other hand, large samples of secondary data such as routine healthcare data may be obtained with relative ease.
For this reason many projects financed by the Innovation Fund of the Federal Joint Committee (G-BA) use a quasi-experimental design: For the intervention cohort data is acquired prospectively, while information regarding the control cohort is obtained from routine healthcare data. Unfortunately, combining data sets from different sources necessitates merging variables, both outcome and predictors of outcome, which may be operationalized differently or measured with different errors. In the literature this is sometimes called differential measurement error (Carroll et al., 2006) and can result in substantial bias, if not taken into account.
We discuss this issue in one of our latest applications, the SEAL ("Structured Early Assessment of Asymptomatic Liver Fibrosis and Cirrhosis", grant no. 01NVF16026) study. The primary aim of this study is to investigate to what extent the early diagnosis of chronic liver diseases can be improved upon by introducing an early detection program, implemented as part of a nationwide primary care program (the "Check-up 35"). Throughout the SEAL program phase, data of around 16.000 program participants, who constitute the intervention cohort, will be collected by physicians using electronic case report forms (eCRF). To evaluate the rate of early diagnoses under the SEAL program, we use a historical control cohort consisting of Check-up participants for whom only generic healthcare data is available.
Due to the non-randomized study design, there are of course several confounders that should be taken into account, which could be problematic considering the distinct nature of the data. To investigate potential effects of differently operationalized outcome predictors in the SEAL study, we present a simulation study and discuss possible remedies that may be found in the measurement-error or sample survey literature. In particular, the simulation study investigates the bias introduced by differential alcohol consumption data in both cohorts with respect to the marginal intervention odds ratio for early detection. Considerations of this type can be relevant for many current German health care projects.

Carroll, R. J.; Ruppert, D.; Stefanski, L. A.; Crainiceanu, C. M. (2006).
Measurement Error in Nonlinear Models: A Modern Perspective (Second Edition).
London: Chapman & Hall.

179 Measurement for better public administration research (and better theory, too)

Xavier Fernández-i-Marín
LMU Munich, Germany

The interest for measurement in public administration and management sciences is still in its infancy. Regular use of measurement models in public administration and management research still has to come. Political science in general is a latecomer to measurement, and still far from natural sciences in general. And amongst political science, public administration is still even further away.
Several concepts that are underdeveloped in public administration and management theory can benefit from a wider use of systematic models. This includes either very broad concepts such as governance, accountability, independence / autonomy, administrative tradition; and also more specific features of regimes and organizations such as complexity of an institution, strength of a board, accountability, etc.., to name but a few.
The traditional approach to creating indices is based on simple averages of the intermediate variables that are used as *indicators* of this latent trait or concept. The paper argues that the current state of the art of statistics in general and measurement in particular in public administration and management research is underdeveloped. Developing measurement models is needed to produce better measures of the concepts of the discipline, instead of relying on indices based on simple aggregations of items.
Using Bayesian techniques to carry out the inference process using Markov Chain Monte Carlo methods (MCMC) increases the added value of all the previous advantages. Bayesian inference does not rely in infinite and repeated sampling, which is usually a really strong assumption done in public administration and management research: there are no infinite countries or institutions in the world that we can sample.
For concepts that involve indices of countries, the paper reanalyzes a dataset on business regulations and shows that simple aggregations of variables and measurement models can lead to different results and produce other valuable information (Ease of Doing Business).
For concepts that involve measuring traits of institutions and organizations, the paper shows how measurement models can improve the understanding of independence of regulatory agencies (Gilardi). In addition, an example on multidimensional characteristics of public institutions is also introduced.

180 Systematic Review on handling missing participant data in longitudinal studies

Dominik de Sordi, Fabian Otto-Sobotka, Antje Timmer
Universität Oldenburg, Germany

Background Ignoring missing participant data (MPD) leads to various errors including estimation errors. For every analysis, all available data points should be used for best possible reliability as well as for ethical reasons. This review aims to assess and compare methods on dealing with MPD as used in published studies with repeated measurement (RM) of continuous variables.
Objectives To show the range and frequency of methods used for handling longitudinal continuous variables with MPD in a medical context.
Search methods A systematic search was carried out in MEDLINE (January 2000 to June 2018). The search was restricted to publications in prespecified methodological journals relating to medicine. The search strategy included text words as well as relevant MeSH terms, such as longitudinal studies", Patient Dropouts", Lost to Follow-Up", Ïnterrupted Time Series Analysis", Treatment Refusaländ Patient Compliance". In addition, there were restrictions by publication type (cohort studies and clinical trials) and type of participants (humans).
Additional selection criteria Only cohort studies and clinical trials which consider more than one method, with continuous variables as main endpoint which were measured at least 5 times were included. Studies with Time-to-event or Survival endpoints were excluded.
Data collection and analysis Selection of articles and data extraction were performed independently by two biometricians. A standardized extraction form was used following pilot testing. Information was collected on study design, unit of analysis, types of outcome, times of measurement and used analysis method. For this review, the methods used to handle MPD were grouped by the extractors to at least one main statistical family (multiple imputation (MI), mixed models (MM), likelihood-based procedures and pattern mixture models (PMM), etc. (not mutually exclusive)).
Main results The initial search returned 1090 articles, 225 of them could be identified as possibly relevant due to examination of title and abstract for selection criteria. 110 (48.9%) articles were excluded after full paper access ("less than 5 measurement times" = 51 (22.7%), "not metric" = 28 (12.4%), "only one method presented" = 24 (10.7%), "time-to-event or survival endpoints" = 12 (5.3%)). 115 (51.1%) articles were included in the extended analysis.
The 115 reviewed articles reported more than 100 different methods for dealing with MPD. Between five and 10080 measurement times were reported (median: 7). The observation period was between 6 hours and 28 years (median: 51.4 weeks). The most frequently mentioned methods related to MI and MM with 40 articles (34.8%) each, followed by ML (21, 18.3%) and PMM (19, 16.5%). Single imputation procedures such as last observation carried forward (LOCF) were often used as reference (30, 26.1%).
Authors' Conclusion The review hit upon a great variety of methods for dealing with MPD. A more detailed presentation will follow. In a next step we will compare selected methods in a simulation study, varying scenarios, numbers of measurement times and proportions of MPD.

Official Statistics and Survey Statistics III

181 An alternative measure of income inequality over successive surveys

Murray Aitkin
University of Melbourne, Australia

In Australia there has been a recent major argument over the claim by the Australian Bureau of Statistics that inequality had not worsened in Australia over the period 2014-2016. The then Commonwealth Government Treasurer (now Prime Minister), Scott Morrison, gave a speech at the time to the Australian Industry Group, in which he said: Änalysis of the more recent census data for the 2016 census shows the Gini coefficient based on gross household income has declined from 0.382 to 0.366 since 2011.Ïn a separate comment, he said The last census showed that on the global measure of inequality, which is the Gini coefficient – that is the accepted global measure of income inequality around the world – and that figure shows that it hasn't got worse, inequality, that it's actually got better.Mr Morrison's figures were derived using gross income data taken from the census, and are based on internal, unpublished calculations. There are at least three problems with the Gini coefficient for income inequality comparisons.
The first is that countries or years with widely different income distributions may have the same Gini index. The second is that its calculation requires access to individual-level income data, to develop both the percentiles of the individual income distribution and the proportion of national income received by each income percentile group. These data are generally confidential to the national statistical office and are not publicly available, except by personal application through a recognized University. What is publicly available, at least in Australia, is the numbers of households receiving income in ABS-defined income intervals. This information is insufficient to compute the Lorenz curve, from which the Gini index is computed. The third and principal problem with the Gini index, or any other single number, is that it cannot represent variability in the income distribution. A Gaussian distribution can be summarized by two numbers, but only a single-parameter distribution, like the Poisson or exponential, can be summarized by one number. Recognition of this allows us to develop a statistical modelling approach to changes in income inequality over repeated surveys or censuses, using publicly available income data. I give a detailed description of the Lorenz curve and Gini index for non-publicly available individual household income data, and an example of publicly available total household income reported in percentile ranges from the Australian censuses of 2006, 2011 and 2016. The alternative analysis uses a four-moment distribution to model the income distribution and graduate its reported percentiles.
It is clear from the analysis that the reported income distribution changed very little from 2006 to 2011, but changed substantially - mean and variability both increasing - from 2011 to 2016. This would appear to represent a decrease in inequality from 2011-2016, but the voluntary response makes almost any conclusion doubtful.

182 KOALA: A new paradigm for election coverage - An opinion poll based "now-cast" of probabilities of events in multi-party electoral systems

Alexander Bauer[1], Andreas Bender[2], André Klima[1], Helmut Küchenhoff[1]

[1]Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Germany
[2]Nuffield Department of Clinical Medicine, University of Oxford, United Kingdom

Common election poll reporting is often misleading as sample uncertainty is addressed insufficiently or not covered at all. Furthermore, main interest usually lies beyond the simple party shares. For a more comprehensive opinion poll and election coverage, we propose shifting the focus towards the reporting of survey-based probabilities for specific events of interest. We present such an approach for multi-party electoral systems, focusing on probabilities of coalition majorities. A Monte Carlo approach based on a Bayesian Multinomial-Dirichlet model is used for estimation. Probabilities are estimated, assuming the election was held today ("now-cast"), not accounting for potential shifts in the electorate until election day ("fore-cast"). Since our method is based on the posterior distribution of party shares, the approach can be used to answer a variety of questions related to the outcome of an election. We also introduce visualization techniques that facilitate a more adequate depiction of relevant quantities as well as respective uncertainties. The benefits of our approach are discussed by application to the German federal elections in 2013 and 2017. An open source implementation of our methods is freely available in the R package coalitions.

183 Nonparametric Multiple Imputation for Bridging Between Different Industry Coding Systems

Jörg Drechsler[1], Birgit Pech[2]

[1]Institute for Employment Research, Germany
[2]Amt für Statistik Berlin-Brandenburg

Industry classifications such as the Statistical Classification of Economic Activities in the European Community (NACE) are regularly updated to ensure that all economic sectors are fully covered. While regular updates are desirable to ensure for example that emerging industries can be classified properly, the changes in the coding system can be a major problem in longitudinal analyses. In most cases a one-to-one mapping between the different versions is not possible which makes a consistent classification for all establishments difficult. In this talk we treat these changes as a missing data problem. The new code is missing for those establishments that only existed while the old code was used and vice versa. We use classification and regression trees (CART) to model the transition probabilities between the classification systems based on years for which both classification systems are available and use these models to impute the missing industry codes. We illustrate that this approach is superior to commonly used strategies such as setting the industry code to the most frequently observed successor/predecessor industry code.

184 Estimation of voter transitions in the immediate post-election period

André Klima

Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, Germany

The presentation of voter transitions is a relevant part of post-election reporting in Germany. A timely publication of the results is necessary in order to be noticed in the public discussion. The largest provider in Germany is infratest dimap, whose estimated voter transition is mainly based on survey data. Generally, the estimation problem in multiparty systems is highly dimensional and a large number of issues are known for the established estimation methods. In the case of individual level data, there are above all difficulties with accessibility and data quality. Alternative methods that only - or especially - use aggregate level data strongly depend on their assumptions. To make matters worse, the quality of the estimate cannot be objectively verified due to the lack of the necessary information.

For the 2018 parliamentary elections in Bavaria, various strategies for gathering individual level data and estimating voter transitions were used as part of the USBW18. This includes different methods for collecting individual level data: computer-assisted telephone interviews before and after the election, post-election interviews (exit-poll) and post-election online interviews. In addition, aggregate level data-based models were also used to estimate the transition, e.g. by the statistical office of Munich. This allows a direct comparison of the methods and a discussion of the issues arising in data collection. For the Bavarian federal-state election, the data shows the problem that in the pre-election phase many participants either refused to answer or were still undecided, as well as the usual bias comparing the individual level data with the election result, including a strong underestimation of non-voters. The post-election surveys (exit-poll) also show distortions compared to the election result, which implies a structural difference in the response. When comparing the estimates, differences between the aggregate level data-based models and the estimates based on individual level data are visible. However, there is also a great deal of uncertainty in the later estimates, easily recognizable by the differences between the Bavarian estimate of the USBW18 and the estimated voter transition published by infratest dimap.

Robust and Nonparametric Statistics I

185 An extension for smoothed empirical likelihood confidence intervals for extreme quantiles and small sample sizes

Oliver Thunich[1], Sebastian Schoneberg[2], Bertram Schäfer[2]
[1]TU Dortmund, Germany
[2]Statcon GmbH,Germany

In industry production one frequently wishes to draw conclusion about the performance of a relatively large, or conceptually infinite, number of further units. Of particular interest is calculating a process capability. In many applications one cannot assume that the produced units follow a parametric distribution. Typically, this kind of distribution-free tolerance interval requires relatively large sample sizes, especially for extreme quantiles which are close to zero or one respectively. For example to calculate a 99 % tolerance interval, even a sample size of 100 seems to be insufficient. To get acceptable coverage rates much larger sample sizes (ca. 500) are required. In industrial applications those sample sizes are often not achievable.
We constructed the tolerance intervals by calculating confidence intervals for quantiles. The approaches which we will present are based on smoothed empirical likelihood algorithms. Here the shape of the confidence regions automatically reflects emphasis in the observed data, by that smoothed empirical likelihood regions do not require scale or skewness estimation. Moreover, empirical likelihood regions are Bartlett correctable.
Empirical likelihood ratio functions are constructed by calculating the ratio of the nonparametric likelihood of a given distribution function and the likelihood of the empirical distribution function of the observed data which maximizes the nonparametric likelihood. Since both parts of the ratio are not continuous, the resulting empirical likelihood ratio is also not continuous. As for parametric likelihood ratio functions the Wilks theorem can be applied here. By that, asymptotic confidence regions can be constructed. Different approaches to smooth this ratio are published. Smoothing the empirical likelihood leads to smaller coverage errors of resulting confidence intervals. In the earliest version a kernel to smooth the function was introduced. Afterwards a simple linear interpolation and a weighted mean were proposed as alternative smoothing methods. We compare these methods with focus on extreme quantiles and small sample sizes.
The proposed methods just focus on confidence regions within the observed data. If one considers the situation of extreme quantiles and small sample sizes, the confidence regions of the quantiles are frequently outside of the observed data. Using the proposed methods either leads to small coverage rates or very large, usually infinite, confidence regions. Based on the smoothing approaches we introduce an extension of the likelihood function for values outside of the observed data, to improve the performance in those cases.
Based on simulations, we found that in certain situations, the extension grants good coverage rates without returning extremely large confidence regions for small sample sizes. Therefore, the extension seems to be superior to other distribution free intervals.
The presentation is part of a master thesis, which is supervised by the faculty of statistic, of the university Dortmund, the faculty of mechanical engineering of the university Ingolstadt and Statcon (company with focus on statistical consulting and teaching). By that there is a strong cooperation between basic and applied research.

186 Stochastic models for non-destructive testing in civil engineering

Markus Sebastian Doktor[1], Wolfgang Kurz[1], Peter Ruckdeschel[2], Jean-Pierre Stockis[1]
[1]University of Kaiserslautern, Germany
[2]University of Oldenburg, Germany

Markus Sebastian Doktor, Wolfgang Kurz, Peter Ruckdeschel, Jean-Pierre Stockis
Non-destructive testing methods became quite popular within the last decade, especially in mechanical and electrical engineering, whereas in civil engineering destructive testing is mainly used.
Currently, there are only destructive ways for safely testing the yield limit as well as determining the current stress level. However, steel beams incorporated in buildings are one main source of stability. Rise of ultrasonic and micro-magnetic tools for (non-destructive) measurements, which might still be error prone, allows the characterization of old steel bridges. This offers the possibility of ecological and economical maintenance for ageing infrastructure.
Mathematical tasks are the classification of e.g. inbuilt material, further quantification of the global stress and successive estimation of occurring internal forces. The presented approach uses techniques from nonparametric statistics such as statistical classification based on support vector machines and on the other hand (robustified) sieve and partition estimators.
Finally, we show exemplary applications for the approach with simulated as well as real datasets and demonstrate the use of robust diagnostics in the domain of civil engineering to gain additional information (e.g. statical systems, fatigue) in recorded data.

187 Applications of a minimum distance estimator for self-exciting counting processes

Mirko Alexander Jakubzik
Technische Universität Dortmund, Germany

In this contribution based on a paper by Kopperschmidt and Stute published in 2013, we study minimum distance estimation for self-exciting counting processes. In a semi-parametric modelling approach, we invoke the compensator of a counting process given by the Doob-Meyer decomposition to predict its qualitative behaviour. The introduced minimum distance estimator yields consistent and asymptotically gaussian distributed estimates for the parametric part of the predictor. While the main results concerning these asymptotic properties are due to Kopperschmidt and Stute, we augment the range of applications by discussing models convenient for practical usage.
This approach encompasses load-sharing systems commonly employed in civil engineering that also allow for damage accumulation.
The proposed models are debated in view of simulation studies as well as real data obtained through recent fatigue testing that took place at TU Dortmund University. We furthermore utilize the asymptotic properties of the minimum distance estimator as shown by Kopperschmidt and Stute to establish confidence sets for the corresponding parameters.
Finally, we discuss the emerging problems, e.g. the numerical issues occuring throughout simulation and application, and give an outlook on how to overcome them.

188 Regression based on medians with application to survey data

Beat Hulliger
FHNW School of Business, Switzerland

Outlier robust linear regression has been researched extensively and good methods are available (e.g. in the R-package robustbase). However, most of these methods rely on an iterative solution from a starting value, often the least squares estimates. Simple but robust starting values would be desirable and non-iterative solutions would enhance the acceptance of the method.
A median-based simple linear regression method has been proposed by John Tukey, his so-called resistant line. However, several versions of the resistant line have been implemented (e.g. in the function line() of base R) and other ways to use medians for simple linear regression are possible. This paper derives median-based estimators which also take into account weights. Starting from the weighted least squares solution when survey weights and heteroscedasticity is taken into account three ways of replacing weighted means by weighted medians are developed. These estimators are compared with resistant line versions.
In multiple least squares regression the coefficients can be estimated by sequential simple linear regression on the partial residuals of preceding regressions. Translating this sequential procedure into a median based regression was proposed David Andrews, using a particular version of median-based simple linear regression. Experience with this type of regression is scant and the application to survey data is not documented.
Median-based simple and multiple linear regression has been implemented in R functions. Bias, variance and robustness of these methods are studied with standard data sets and with data sets from complex surveys.

Survival and Event History Analysis II (Complex Modeling)

189 Detecting Deceleration in Old-Age Mortality Rates Using Focused Model Selection

Marie Böhnstedt[1,2], Jutta Gampe[1], Hein Putter[2]

[1]Max Planck Institute for Demographic Research, Rostock, Germany
[2]Department of Biomedical Data Sciences, Leiden University Medical Center, Leiden, The Netherlands

The exponential increase of human mortality with age, observed at the mid-adult and early old ages, is well described by the Gompertz model. However, at higher ages a downward deviation from the exponential hazard has been reported in many studies. Such mortality deceleration can be explained theoretically as an effect of selection in heterogeneous populations. The frail individuals with high mortality risks tend to die at younger ages, while the more robust individuals with lower death risks tend to survive to higher ages. This heterogeneity hypothesis can be formalized adequately in frailty proportional hazards models (Vaupel et al., 1979), such as the gamma-Gompertz model considered here. Still, in practice, the statistical methods commonly used to assess mortality deceleration may fail to detect the phenomenon, because it mainly affects the tail of the survival distribution.
In this work, we propose a focused information criterion (FIC; Claeskens and Hjort, 2003) for assessing mortality deceleration. Using an FIC, model choice is driven by the performance of the model with respect to a specific parameter of interest, the so-called focus parameter. This is achieved by constructing the FIC as an estimator of the limiting risk of the estimator of the focus parameter. Defining a focus parameter allows us to directly target the quantities which reveal mortality deceleration, e.g., the hazard at some advanced age. Yet, the original version of the FIC was derived under standard regularity assumptions, which are violated in the setting of choosing between the gamma-Gompertz model and the Gompertz model. The gamma-Gompertz model has only one additional parameter, the frailty variance, but in the case that mortality does not decelerate the variance takes value zero and, hence, lies on the boundary of the parameter space.
We present a modified FIC that is adapted to the nonstandard condition of a boundary constraint on the parameter. Two versions based on the mean squared error and the mean absolute error of the estimator of the focus are studied. The properties of the new method are investigated in a simulation study and illustrated in an application.
References
Claeskens, G., and Hjort, N.L. (2003). The focused information criterion. Journal of the American Statistical Association, 98, 900-916.
Vaupel, J.W., Manton, K.G., and Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography, 16, 439-454.

190 Tree-Structured Modeling of Time-Varying Coefficients for Discrete Time-to-Event Data

Moritz Berger[1], Marie-Therese Puth[1], Gerhard Tutz[2], Nils Heim[3], Matthias Schmid[1]
[1]University of Bonn
[2]Ludwig-Maximilians-Universität München
[3]University Hospital Bonn

Time-to-event models are a popular tool to analyse data where the outcome variable describes the time to the occurrence of a specific event of interest. We focus on the analysis of time-to-event outcomes that are measured on a discrete time scale, which is a likely scenario when events occur between pairs of consecutive points in time (e.g., between two follow-up visits of an epidemiological study).
In the literature there exists a variety of regression models for discrete time-to-event data, see for example [1], [2]. The main principle is to model the discrete hazard function $\lambda(t|\boldsymbol{x}) = P(T = t|T \geq t, \boldsymbol{x})$, $t = 1, 2, ..., k$, where T denotes the event time and $\boldsymbol{x}$ is a set of explanatory variables $\boldsymbol{x}^\top = (x_1, ..., x_p)$. Commonly it is assumed that the effects of the explanatory variables are constant over time, and the dependency on time is solely modeled by separate intercepts for each t, also referred to baseline coefficients. In many applications, however, this assumption may be too restrictive, as for example the effect of an explanatory variable on the hazard might be stronger at the beginning of the study than at later times.
A more general approach is to allow the effects to vary over time. Frequently it is natural to assume that the coefficients are represented by smooth functions of t. A popular strategy is to use P-splines propagated by [2], where the smooth functions are expanded by B-spline basis functions and an additional penalty term is used to obtain stable estimates.
Alternatively, it is often assumed that the effects of the explanatory variables do not vary over the whole range of t, but are constant over several adjacent points in time. That is, one assumes that the time-varying coefficients are represented by piecewise constant functions. These functions can be obtained by using recursive partitioning techniques. To address this issue, we propose to adapt the tree-based approach that was recently proposed by [3]. By iterative splitting in one of the explanatory variables (with regard to the effect modifier t) the method yields a tree for each variable that shows time-varying coefficients. Thereby, the algorithm itself identifies the coefficients (corresponding to an explanatory variable) that deviate from a constant and the corresponding partitioning of the observation times.
The proposed approach is illustrated using data collected by the center for dental, oral and maxillofacial surgery at the University Hospital Bonn. The aim of the study was to investigate factors (e.g., age, gender, presence of diabetes mellitus type 2) that tend to prolong the length of stay in hospital in the treatment of severe odontogenic infections.

References:
1 Tutz, G. and M. Schmid (2016). Modeling Discrete Time-to-Event Data. New York: Springer.
2 Berger, M. and M. Schmid (2018). Semiparametric regression for discrete time-to-event data. Statistical Modelling 18, 1-24.
3 Eilers, P. and B. Marx (1996). Flexible smoothing with B-splines and penalties. Statistical Science 11, 89-102.
4 Berger, M., G. Tutz and M. Schmid (2018). Tree-structured modelling of varying coefficients. Statistics and Computing. doi: 10.1007/s11222-018-9804-8.

191 The genesis and use of time-varying frailty models for representing heterogeneities in the transmission of infectious diseases

Steffen Unkel[1], Steven Abrams[2], Andreas Wienke[3], Niel Hens[2,4]

[1]University Medical Center Göttingen, Germany
[2]Interuniversity Institute for Biostatistics and statistical Bioinformatics, Hasselt University, Diepenbeek, Belgium
[3]Institute of Medical Epidemiology, Biostatistics and Computer Science, Martin Luther University Halle-Wittenberg, Halle, Germany
[4]Centre for Health Economics Research and Modelling Infectious Diseases, Centre for the Evaluation of Vaccination, Vaccine & Infectious Disease Institute and Unit of Epidemiology and Social Medicine, University of Antwerp, Wilrijk, Belgium

Frailty models are commonly used for representing and making inference on individual heterogeneities relevant to the transmission of infectious diseases, including heterogeneities that evolve over time. This talk takes as its focus time-varying frailty models for the infection-specific hazard rate, concentrating on the genesis of such models and how they can be derived from mechanisms underlying the endemic equilibrium equation for the infection of interest. We show that time-varying frailty models are a natural choice for capturing individual heterogeneities as they follow naturally from the underlying biology of the infections.
Multivariate frailty models with shared or correlated frailties are particularly useful for inducing association of infection times within individuals as well as variability among individuals. For time-varying shared and correlated frailty models, we discuss issues of identifiability and methods of estimation. Illustrations with real data from serological surveys are provided.

192 Weighting Expectile Regression for Survival Analysis with Right-Censoring

Alexander Seipp, Verena Jürgens, Antje Timmer, Fabian Otto-Sobotka
Division of Epidemiology and Biometry, Carl von Ossietzky University Oldenburg, Germany

Expectile regression generalizes conventional mean regression. We estimate least asymmetrically weighted squares with weights defined according to the expectile level. Quantile regression extends median regression along the same principle. Both quantile and expectile regression aim to describe the distribution of a response variable as a whole, instead of just the conditional mean. Expectile regression has advantages in comparison to quantile regression. It is computationally simple and can be more efficient. However, single expectiles are more difficult to interpret, and they are more sensitive to outliers than quantiles. Because of its benefits in the classical regression setting, extending expectile regression to other fields like time-to-event analysis can add substantial value. For one thing, we can complement mean or median survival times with prediction intervals, especially in the case of heteroscedasticity.
In survival analysis, we typically analyze right-censored data. Regression for right-censored data is often done with parametric Accelerated Failure Time Models. Distribution-free alternatives are also available. Koul et al. (1981) used a weighted data approach: Observed event times receive weights amounting to the inverse probability of censoring, whereas censored event times are excluded. The chosen estimator for the probability of censoring is similar to the product-limit estimator, which is why other authors in subsequent papers use the Kaplan-Meier estimator instead. Zhou (1992) uses this approach to apply the weights to the corresponding loss instead of the response itself. We extend this method to expectile regression using asymmetric squared loss. The described weights are easy to calculate, which we see as the main advantage, since the computational simplicity of expectile regression is thereby retained.
After the introduction of the method and inference, we present simulation results. We show evaluations of our method in comparison to quantile regression for censored data, as well as expectile regression in the uncensored case. Our main focuses are efficiency and robustness. Since we exclude censored observations, there might be significant information loss, which could result in poor efficiency. We also investigate simulated data with outliers. Expectile regression and the proposed weights can both be prone to outliers, hence the robustness of our combination should be examined. Since large uncensored observations receive substantial weights and the variance of the product-limit estimator increases in time, Koul et al. (1981) constrain their estimator and exclude uncensored observations which are too large.
We illustrate our method by applying it to overall survival of cancer patients. The data is provided to us by our clinical partners from a cooperating university hospital. In oncology, common endpoints are overall or progression-free survival. By choosing conventional analysis techniques, the implicit focus lies on the effect on the average patient. With the help of expectile regression, we can analyze the impact of a therapy on more progressed or mutated tumors with shorter survival times than the average.
References
Koul, H., Susarla, V., & Van Ryzin, J. (1981). Regression Analysis with
Randomly Right-Censored Data. The Annals of Statistics, 9(6), 1276-1288.
Zhou, M. (1992). M-Estimation in Censored Linear Models. Biometrika, 79(4), 837-841.
doi:10.2307/2337240

Statistics in Finance I

193 How do market participants contribute to market quality? A statistical approach

Mathieu Rosenbaum
Ecole Polytechnique, France

In this work we establish a methodology enabling us to assess the stabilizing or destabilizing role of market participants on a given market. To do so, we first build a relevant order book model and provide associated ergodic properties and limit theorems. We then use these results together with a thorough statistical analysis of market participants order flows thanks to full order book data. This allows us to measure the influence of each market participant on various important quantities such as the liquidity or the volatility. In particular we are able to say whether or not their trading strategies are supporting market quality. This is joint work with Othmane Mounjid and Pamela Saliba.

194 Measuring risks in a network of light-tailed financial objects

Claudia Klüppelberg[1], **Miriam Isabel Seifert**[2]
[1]Technical University of Munich, Germany
[2]Ruhr University Bochum, Germany

We investigate a financial network of agents holding portfolios of independent light-tailed risky objects with losses assumed to be asymptotically exponentially distributed with distinct tail parameters. The derived asymptotic distributions of portfolio losses refer to the class of functional exponential mixtures. We also provide statements for Value-at-Risk and Expected Shortfall measures as well as for their conditional counterparts CoVaR and CES. We establish important qualitative differences in the asymptotic behavior of portfolio risks under light tail assumption compared to heavy tail settings which should be accounted for in practical risk management.

195 Factor State-Space Models for High-Dimensional Realized Covariance Matrices of Asset Returns

Bastian Gribisch, **Jan Patrick Hartkopf**, Roman Liesenfeld
University of Cologne, Germany

We propose a dynamic factor state-space model for high-dimensional covariance matrices of asset returns. It makes use of observed risk factors and assumes that the latent integrated joint covariance matrix of the assets and the factors is observed through their realized covariance matrix with a Wishart measurement density. For the latent integrated covariance matrix of the assets we impose a strict factor structure allowing for dynamic variation in the covariance matrices of the factors and the residual components as well as in the factor loadings. This factor structure translates into a factorization of the Wishart measurement density which facilitates statistical inference based on simple Bayesian MCMC procedures making the approach scalable w.r.t. the number of assets. An empirical application to realized covariance matrices for 60 NYSE traded stocks using the Fama-French factors and sector-specific factors represented by Exchange Traded Funds (ETFs) shows that the model performs very well in- and out of sample.

Statistics of High Dimensional Data I

196 A Model-free Approach to Linear Least Squares Regression with Exact Probabilities and Applications to Covariate Selection

Patrick Laurie Davies[1], Lutz Dümbgen[2]

[1]University Duisburg-Essen, Germany
[2]University of Bern, Switzerland

Given data y of size n and a covariate x the degree of relevance of x can for measured by postulating a linear model $y = \beta_0 + \beta_1 x + \varepsilon$ and then calculating the P-value for the null hypothesis $H_0 : \beta_1 = 0$. If the errors ε are i.i.d. Gaussian the P-value is based on Fisher's F-distribution. We introduce a different method of quantifying the relevance of x by comparing x with a covariate Z whose components are i.i.d. standard Gaussian random variables. Denote the suõf squared residuals using x by ss_x and the sum of squared residuals using Z by ss_Z. Then Z is better than x if $ss_Z < ss_x$ with probability $P(ss_Z < ss_x)$ which we call the P-value based on Gaussian covariates. It can be shown that this P-value is the same as the P-value based on the F-distribution which is derived from the standard linear model. In contrast the Gaussian covariate P-value is model free as the distribution of ss_Z/ss_x is independent of y and x. The idea can be extended to the case of p covariates $x_i, i = 1, \ldots, p$, with a stepwise version for large p which allows for $p > n$. the stepwise version outperforms other covariate selection procedures such as lasso and knockoff in all respects: it is simpler, faster, it does not overfit or require any form of post selection analysis. Real data examples are given as well as the results of simulations.

197 Random coefficient model - model selection and estimation of first and second moments

Philipp Hermann, Hajo Holzmann

Philipps-Universität Marburg, Germany

We consider the linear regression model with random coefficients with a sparse mean vector μ and a sparse covariance matrix Σ. For a fixed number of coefficients we deduce conditions under which the first and second moments are identifiable. Furthermore we show sign-consistency of the Adaptive Lasso μ_n^{AL} for the mean vector and in addition we give an asymptotic normality result. Thereby we use the primal-dual witness characterization of the Adaptive Lasso introduced by Wainwright (2009). In a second step we conduct model selection with the Adaptive Lasso for the vectorization of the covariance matrix whereby we use μ_n^{AL} as approximation of the unknown mean vector. We show again sign-consistency and give an analogous normality.
We also provide some high-dimensional results for the Adaptive Lasso for the above model under a sub-Gaussian distribution assumption of the coefficient vector. These considerations are based on Zhou, van de Geer, Bühlmann (2009).

198 Shrinkage in Estimating High Dimensional Copulas

Vladimir Pyrlik
CERGE-EI, Czech Republic

As a combination of separate marginal distributions and a dependence structure, copulas have proved a convenient framework to synthesize joint distributions, including in high dimensions. Currently, the high dimensional settings analyzed with copulas contain at most several hundred variables, as higher dimensionality appears too demanding for either model selection or estimation. However, the practical application in many fields demands for higher dimensionality. In our paper, we suggest adopting recently developed techniques of large covariance matrices estimation for the task of copulas estimation, with the dimensionality of the data going well beyond that studied in the literature. We apply the large covariance matrices shrinkage estimators of Ledoit and Wolf (2004, 2017) to estimate some types of copulas in high dimensions. We consider Gaussian and Student's t copulas as well as their skewed versions and take the dimensionality of the data up to thousands of variables. A simulation study shows that the shrinkage estimation of the large matrix parameters of the copulas significantly outperforms the traditional estimators, including Kendall's rank corelation based estimator and sample corelation matrix of pseudo-observations.

199 Improving Estimation in Functional Linear Regression with Points of Impact: Insights into Google AdWords

Dominik Liebl[1], Stefan Rameseder[2], **Christoph Rust**[2]
[1]University of Bonn
[2]University of Regensburg, Germany

The functional linear regression model with points of impact is a recent augmentation of the classical functional linear model with many practically important applications. In this work, however, we demonstrate that the existing procedure for estimating the parameters of this regression model can be very inaccurate. The tendency to omit relevant points of impact is a particularly problematic aspect resulting in omitted-variable biases. We explain the theoretical reason for this problem and propose a new sequential estimation algorithm that leads to significantly improved estimation results. Our estimation algorithm is compared with the existing estimation procedure using an in-depth simulation study. The applicability is demonstrated using data from Google AdWords, today's most important platform for online advertisements.

Statistics in Science, Technology and Industry I

200 Some examples of handling uncertainty in industrial applications

Axel Gandy
Imperial College London, United Kingdom

In many technical systems, handling and controlling uncertainty is of paramount importance. We will consider several methods of how this can be done, including control charts, power system planning under climate uncertainty, transport planning and cyber security. Techniques involved will include bootstrapping and extreme value theory.

201 Fully automatic nonparametric intensity estimates for studying the microstructure of composite materials from 2d and 3d images

Jürgen Franke, Pak Hang Lo
Technische Universität Kaiserslautern, Germany

We consider inhomogeneous Poisson processes in dimension 2 and 3 as models for the locations of fibres or grains in composite materials. Its local intensity function, which is crucial for the quality of the material, may be estimated nonparametrically by local smoothing, e.g. by kernel estimates. They crucially depend on the choice of bandwidths as tuning parameters controlling the smoothness of the function estimate. We propose a fast algorithm for learning suitable global and local bandwidths from the data which is suitable for very large samples where common approaches like crossvalidation become computational expensive. In material science, in particular, it is very common to have several thousand up to several million points with data arising from 3d imaging, e.g. computer tomography and postprocessing.
We extend the ideas of Engel et al (1994) who developped an iterative fast plug-in algorithm for choosing the bandwidths of kernel density estimates. Our method is based on a detailed asymptotics of estimators of the intensity function and of its second derivatives and integrals of second derivatives which, then, replace the unknown quantities in formulas for asymptotically global resp. local optimal bandwidths. Using this asymptotics, we are also able to determine the exact number of iteration steps. For both global and local case, fewer than 10 iterations suffice leading even to a feasible local bandwidth selection in large samples. We illustrate the method with some apllications to test bodies from of fiber-reinforced high-performance concrete, clearly showing some inhomogeneity in the fiber intensity.
Engel, J., Hermann, E. and Gasser, Th. (1994). An iterative bandwidth selector for kernel estimation of densities and their derivatives. J. Nonpar. Statist. 4, 21-34
Lo, P.H. (2018). An Iterative Plug-in Algorithm for Optimal Bandwidth Selection in Kernel Intensity Estimation for Spatial Data. PhD Thesis, Technische Universität Kaiserslautern.
kluedo.ub.uni-kl.de/frontdoor/deliver/index/docId/5161/file/_Thesis_LoPakHang.pdf

202 Statistical Modelling and Design for Quality Control and Reliability Analysis in Power Semiconductor Manufacturing Processes

Jürgen Pilz[1], Natalie Vollert[2], Konstantin Posch[1]

[1]Alpen-Adria-Universität Klagenfurt, Austria
[2]Carinthian Tech Research AG, Austria

We give an overview of some of our recent contributions to statistical reliability modeling and analysis for power semiconductor manufacturing processes. In particular, we review advanced burn-in-strategies for early lifetime failure analysis and Bayesian experimental design strategies for reliability studies.
Furthermore, we report on the application of Bayesian design methods for computer experiments using additive Gaussian Process models for improved sensor position detection systems. Finally, we present some initial findings on characterizing uncertainty associated with Bayes deep learning models for applications in image analysis.

Time Series Analysis IV (Discrete and Functional Time Series)

203 Distance-based Analysis of Ordinal Time Series

Christian Weiß

Helmut-Schmidt-Universität Hamburg, Germany

We consider ordinal processes, i.e., categorical processes where the range consists of a finite number of ordered categories. The dissimilarity of the ordinal categories can be expressed with a distance measure. To express marginal properties of the ordinal process (location, dispersion, symmetry) and to measure the extent of serial dependence, a unified approach relying on expected distances is proposed. The expected-distance approach also has the advantage of leading to well-interpretable and easily estimated measures. For special types of distance, these analytic tools lead to known approaches for ordinal or real-valued processes; otherwise, we get tailor-made measures for the considered application scenario. We also analyze the sample counterparts of the proposed measures and derive asymptotic results for practically important cases in ordinal time series analysis. The talk concludes with a real application about the credit ratings of European countries.

204 Autoregressive-type time series models with bounded support

Lena Reichmann[1], Carsten Jentsch[2]

[1]Mannheim University, Germany
[2]TU Dortmund, Germany

In practice, many time series are bounded from below and above such that they can take only values in a certain interval $[a, b]$ instead of the whole real line. Such cases include rates, weights and proportions. Most modelling approaches for such time series rely on certain transformations that map the interval $[a, b]$ to the real line such that classical time series models, as e.g. autoregressive moving-average models, can be applied. However, such a transformation is clearly not uniquely determined such that the model will generally depend on the specific transformation that has been used. By considering a direct modeling approach that avoid transformations to capture the serial dependence, we propose a new autoregressive moving average type model class with nicely interpretable structure. To take the boundedness of such time series into account, the proposed models make use of beta distributions. In Rocha and Cribary – Neto (2008), the authors followed a similar path and proposed the beta ARMA model class, which also relies on beta distributions conditional on the past, but have to utilize also a suitable link function to transform the response to the unit interval.
For our new model class, we provide sufficient stationarity conditions and derive the stationary solution of the model equations. For the purely autoregressive case, we prove the Yule-Walker equations to hold which facilitate the task of parameter estimation in these new models as the whole toolbox for classical autoregressive models becomes applicable. Further, we discuss mixing properties and provide some simulation results.

205 Asymptotic Normality of Integrated Periodogram Operators

Daniel Constantin Rademacher[1], Jens-Peter Kreiß[1], Efstathios Paparoditis[2]
[1]Technische Universität Braunschweig, Germany
[2]University of Cyprus, Cyprus

Consider a strictly stationary (functional) process $(X_t)_{t\in Z}$ taking values in $L^2([0,1])$. A key element of a frequency domain framework for drawing statistical inference on the second-order structure of the process is the spectral density operator, which generalizes the notion of a spectral density matrix to the functional time series setting. As an integral operator, the spectral density operator is completely determined by its corresponding kernel, which can be estimated via a smoothed version of the periodogram kernel (i.e., the functional analogue to the periodogram matrix). Many interesting quantities of the data generating process, such as autocovariance operators or the spectral distribution operator can be represented as a weighted integral of the spectral density kernel. Estimators for such quantities are obtained by replacing the spectral density kernel by the periodogram kernel. Thus the class of *integrated periodogram operators* covers many familiar statistics, including empirical autocovariance and smoothed periodogram operators. We show that any finite collection of such estimators converges in distribution to (complex) random variables which are jointly (complex) normally distributed. As a side-result we obtain joint asymptotic normality for empirical autocovariance operators. Another relevant byproduct is the asymptotic (complex) normality of the spectral distribution operator. The results do not depend on specific model assumptions on the underlying functional time series, but only on functional versions of cumulant mixing conditions. In order to formulate these conditions precisely, we also introduce a general form of cumulants for random elements taking values in an arbitrary separable Hilbert space.

206 Semi-parametric hidden Markov models for time series of counts

Timo Adam[1], Roland Langrock[1], Christian H. Weiß[2]
[1]Bielefeld University, Germany
[2]Helmut-Schmidt-University Hamburg, Gemany

Hidden Markov models are popular tools for modeling time series where, at each point in time, a hidden state process selects among a finite set of possible distributions for the observations. The state process is typically modeled by a discrete-time, N-state Markov chain. In economic applications, for example, the states are often good proxies for market regimes such as periods of economic growth or recessions, while in animal ecology, they can regularly be linked to behavioral modes such as resting, foraging, or traveling. A typical example are corporate default counts observed on a monthly, quarterly, or yearly basis: In periods of economic growth, they may be generated by some distribution with relatively small mean, whereas during recessions, another distribution with relatively higher mean may be active. Although the market regime is not directly observable, it apparently drives the observed counts. Depending on the application at hand, potential aims which can be addressed using hidden Markov models are manifold, including the prediction of future values of a time series, decoding of the hidden states underlying the observations, and inference on the drivers for example on the state-switching dynamics. Specifically for time series of counts, a Poisson distribution often provides a natural choice fort the state-dependent (emission) distributions, where one rate parameter is estimated for each of the N states of the underlying Markov chain. While more flexible distributions such as the negative binomial can also be used, choosing an adequate class of (parametric) distributions generally remains a complex task, and an inadequate choice may have a severe negative impact on the resulting model fit and hence also on the predictive performance, and ultimately lead to invalid inferences made from the data at hand. To circumvent this problem, we propose semi-parametric hidden Markov models for time series of counts, where the distributions are estimated in a completely data-driven way without relying on any parametric assumptions. To avoid overfitting, a roughness penalty based on the second-order differences between adjacent count probabilities is added to the likelihood, which is demonstrated to produce smooth functional shapes of the estimated distributions. In addition, we demonstrate how the penalty term can be adjusted in presence of (for example zero-) inflated observations, where small differences between corresponding count probabilities and their respective neighbors are not neccessarily desired.
The feasibility of the suggested approach is assessed in simulation experiments, where we compare the performance of the proposed semi-parametric approach to its parametric counterpart. In addition, we present a real-data example, where we model acceleration counts of an oceanic whitetip shark (*Carcharhinus longimanus*) over time to illustrate how the underlying, behavioral modes can be inferred from the observations.

Clustering I (Copula and Genetics)

207 Model-based Clustering with R-vine copulas

Marta Nai Ruscone
LIUC Università Carlo Cattaneo, Italy

Finite mixtures are applied to perform model-based clustering of multivariate data. Existing models are not flexible enough for modeling the dependence of multivariate data since they rely on potentially undesirable correlation restrictions to be computationally tractable. We discuss a model-based clustering method via R-vine copula to understand the complex and hidden dependence patterns in correlated multivariate data. One of the advantages of this approach is that it accounts for the tail asymmetry of the data by using blocks of asymmetric bivariate copulas. We use real datasets to illustrate the proposed procedure.
References:
Banfield J. and Raftery A. (1993). Model-based Gaussian an non-Gaussian clustering. Biometrics, 49, 803-821
Kosmidis I. and Karlis D. (2016). Model-based clustering using copulas with applications. Statistics and Computing, 26, 1079-1099.
McLachlan G. and Peel D. (2000). Finite Mixture Models. Wiley, New York.

208 Dissimilarity functions for copula-based hierarchical clustering of continuous variables

F. Marta L. Di Lascio[1], Fabrizio Durante[2], **Sebastian Fuchs**[3]
[1]Libera Università di Bolzano, Italy
[2]Università del Salento, Lecce, Italy
[3]TU Dortmund, Germany

We introduce and formalize a copula–based notion of dissimilarity between continuous random variables. Such a concept aims at detecting rank–invariant dependence properties among random variables and, as such, it will be defined as a functional on the collection of all copulas.
We show how the provided definition includes previous dissimilarity measures considered in the literature like those derived from measures of association and tail dependence but also those of agglomerative hierarchical type. In the latter case, it turns out that the related clustering procedure does not consider the higher-dimensional dependencies among the involved random variables.
Finally, we compare novel proposed clustering algorithms (taking into account higher–dimensional dependencies) with classical agglomerative clustering methods.

209 Detection of Genetic Similarities using Unsupervised Random Forest

Cesaire Joris Kuete Fouodo, Inke R. König

Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Germany

Genome-wide association studies (GWAS) have in the past been successful in the identification of associations with well-defined phenotypes as well as in the establishment of supervised classification models based on univariable analyses. To perform multivariable genome-wide analyses, random forest (RF) have been shown to be fast and to produce good predictive performances in high dimensional classification problems. In the case of unsupervised genome-wide studies, principal components analysis (PCA), used to decompose the total genetic variation among individuals into components, is the commonly used approach. Besides PCA, unsupervised random forest (URF) has been proposed to compute genetic similarities. Our aim is to cluster individuals into genetically homogeneous groups, thus to detect possible novel subgroups of patients. To achieve better similarities results, we combine PCA with URF, i.e. we run URF on loadings obtained from the PCA, and incorporate a novel boosting idea into our approach. The evaluation of our approach is demonstrated in an application to samples from the Pan-Asia SNP Consortium database with the objective to validate the derived clusters against the known ethnical background. Additional coalescent data are simulated to support our findings, and results are compared to those from PCA and URF each considered separately, without combination. Our results show that combining PCA, URF and boosting improves clusters homogeneities.

210 Alternative splicing based clustering of genes

Claus-Dieter Mayer

Biomathematics & Statistics Scotland (BioSS), United Kingdom

Clustering techniques have been very popular in the analysis of high-dimensional omics data for the last two decades e.g. when trying to identify groups of genes that show similar gene expression patterns across different samples. With the arrival of RNA-seq technology it has become possible to not only calculate the total expression of a gene, but to split this up into the contributions of the various alternatively spliced transcript variants that correspond to this gene. When studying these splice variants the interest is not so much in their absolute abundance but rather their relative importance expressed as a percentage of total gene expression. This leads to a gene being represented by a vector of percentages, i.e. a compositional data structure that differs in length between genes.

Johnson and Purdom (Biostatistics. 2017 Apr 1;18(2):295-307.) consider the situation of clustering samples based on alternative splicing data, whereas we deal with the novel problem of clustering genes based on the percentage contributions of their sets of splice variants. Here each gene is represented by a $k \times n$ matrix auf percentages where k is the (gene-specific) number of variants and n is the sample size. We suggest to use the RV coefficient, a matrix correlation as a measure of similarity for each pair of genes, which can then be used as input for various clustering techniques. We will discuss and illustrate this problem in the context of an alternative splicing study of samples from 16 different tissue types of barley from the International Barley Genome Sequencing Consortium (IBGSC) project.

Official Statistics and Survey Statistics IV

211 Push-to-web recruitment of a probability-based online panel: Experimental evidence

Annelies Blom, Carina Cornesse, **Barbara Felderer**, Marina Fikel, Ulrich Krieger
University of Mannheim, Germany

Past research has shown that pushing respondents to the web is a successful way to increase response rates, reduce data collection costs, and produce representative outcomes. However, studies in that literature are usually limited to cross-sectional surveys on small and homogeneous target populations. Our study rises beyond this limited scope to a broad and, so far, unique application: We investigate the relative success of pushing respondents to the web compared to alternative survey design strategies across the recruitment stages of a probability-based online panel. In order to do this, we implemented a large-scale experiment into the 2018 recruitment of the German Internet Panel (GIP). In this experiment, we sampled 12,000 individuals from population registers and randomly assigned each individual to an experimental group: online-only, online-first, offline-first, or concurrent. Individuals in the online-only group received a postal mail invitation to participate in the web version of the GIP recruitment survey. Nonrespondents in the online-only group were followed up by invitations to the web version of the GIP recruitment survey again. Individuals assigned to the online-first group received the same invitation letter as the online-only group asking them to participate in the web version of the GIP recruitment survey. However, nonrespondents were followed up with a reminder letter containing a paper-and-pencil version of the GIP recruitment survey. Individuals in the offline-first group received the paper-and-pencil questionnaire with the initial invitation letter and were followed up with invitations to the web version of the GIP recruitment survey. Individuals in the concurrent group were given the choice between participating in the web version of the GIP recruitment survey or the paper-and-pencil version. In our presentation, we will show the results of this experiment and discuss our findings.

212 Are Paradata Worth the Effort? Using Adjusted Response Times and Other Paradata To Predict Data Quality in a Survey.

Patrick Schenk

[1]IAB (Institute for Employment Research), Nuremberg, Germany
[2]Ludwig-Maximilians-Universität, Munich, Germany

Paradata are a class of data captured during and about the data collection process itself: e.g., time and date, contact attempt histories, subjective interviewer observations, editing of responses, mouse movements and keystrokes. Suggested applications for monitoring, evaluating, improving, and modeling data quality include nearly every component of the Total Survey Error framework, most notably nonresponse and measurement error. They also hold promise for responsive design during a field phase – e.g., for identifying problematic items, interviewers in need of additional training, or respondent sub-group idiosyncrasies – and for adaptive design on the interview-level, e.g., offering varying incentives or adjusting duration.
However, capturing and preparing raw paradata can require substantial time, resources, and programming skills due to their complex structure.
We explore the predictive potential of paradata – one key prerequisite for their usefulness – in a German CATI survey from 2011 on employment and consumer behavior.
First, we discuss a particular type of paradata: item-level response times – the time it takes to complete a specific item in a specific interview. As widely accepted, we adjust them for characteristics of the item (e.g., number of words), the respondent (e.g., age), the interviewer (e.g., experience), the respondent-interviewer interaction, and other factors. We include novel several variables – that in other settings are either constant, unavailable, or might be too difficult to construct – and estimate flexible additive mixed models. As respondents were randomized to different design variants of the survey (placement and wording of items, response scales, type and placement of follow-up questions), we can discern effects that would be non-identifiable with a single fixed design, in particular on the item-level (such as content, location, and response scale of an item) and for characteristics that vary within a single interview (e.g., how often a respondent had been exposed to the same response scale before the current item). We exploit detailed, self-provided information about the interviewers and construct their prior experiences for every point in time (number of completed interviews and refusals, both overall and on the particular day).
Second, we use paradata (e.g., residual response times, editing of responses, interviewer and respondent experiences) and survey design variables along with respondent and interviewer characteristics to predict data quality: after linking survey data to high-quality administrative records (95% consent rate), we try to predict when responses deviate from "true values" for several common questions on employment histories and receipt of unemployment benefits. The performance of logistic regression is compared to that of standard machine learning algorithms.

213 Multi factor modelling of survey external validity by using statistic and administrative data

Andrei Veikher

National Research University Higher School of Economics, Russian Federation

External validity is the most powerful tool for proving sample survey representativeness. However, until recently, this method of evaluation was used rarely, although recognition of the possibility of obtaining reliable sample survey data came as a result of successful predictions of election outcomes, i.e. immediate and effective assessment of external validity

Most of the applications of external validity in the XX century consisted of searching in independent sources of data on indicators similar to those measured in the survey.

The development of the theory of validity in psychology created an examination of external validity in the format of a model that includes the theoretical combination of several parameters of the research object into a system that explains the formation of the target indicator.

The author used this Multifactor Model of Target Validity Indicator/MMTVI/ in the studies of 2007-2013 for representative surveys of the urban population (St. Petersburg) for study of socio-economic phenomena which are only partially registered officially: shadow wages, self-treatment of short diseases, payment for repair services in cash (reports at conference ISA-RC33, Sidney, 2012; congress ISA, Yokohama, 2014).

The possibility and necessity for its application is determined in several steps: 1) the identification of objective non-survey information about the parameters that have cause-effect relationships with the phenomenon that is meant for the survey (because otherwise it can not be fixed); 2) the creation of a relationship model for several parameters at different levels; 3) the creation of a questionnaire, including questions about several parameters identified in step 1); 4) implementation of a survey; 5) special processing of answers to the questions associated with the parameters chosen in step 1) - calculation of external validity indicators (distributions' mode and averages); 6) calculations from the obtained parameters of the model 2). The choice of the mathematical apparatus of calculations on the model depends on its complexity - the number of its parameters and levels.

The result of the implementation of such an algorithm is as follows: the estimation of the difference in the deviations of the distributions of various parameters measured by the survey, from their real distributions in the population, and their possible impact on the target indicators is cumulative or opposite, when the deviation of some is suppressed by the deviations of others. The algorithm proposed is based on the assumption: 1) high unamendable nonresponse rate violates representativeness in different social parameters and is not identical - there is no single error estimate, traditional ßampling error"; 2) along with a significant non-representativeness of distributions in some parameters, the representativeness of others can satisfy the requirements of this study.

Using of Big Data resources will provide information on many other factors to be included in the chain and multi-level calculations of MVTVI. This will significantly expand the range of statistical and administrative data to assess the external validity of local indicators – indicators "item validity"..

214 Mobilfunkdaten in der amtlichen Statistik

Sandra Hadam

Federal Statistical Office, Germany

Im Rahmen der allgemein fortschreitenden Digitalisierung ist die amtliche Statistik gefordert, neue Datenquellen zu erforschen und einzusetzen und ihre Prozesse und Verfahrensweisen entsprechend auszurichten. Durch die Nutzung solcher Neuen Digitalen Daten, wie z. B. Mobilfunkdaten, werden Optionen zur Ergänzung amtlicher Erhebungen, Erschließung neuer Themenfelder und eine Entlastung der Auskunftsgebenden gesehen. Um das Potenzial von Mobilfunkdaten und ihre möglichen Anwendungsfelder in der amtlichen Statistik zu verdeutlichen, werden die aktuellen Projekte und wichtigsten Ergebnisse der Machbarkeitsstudien vorgestellt.
Erste Machbarkeitsstudien erfolgen zur Fragestellung ob Mobilfunkdaten geeignet sind, die Bevölkerung abzubilden. Hierzu wurden die in INSPIRE-konforme Raterzellen regional tief gegliederten Mobilfunkdaten mit den analogen Zellen des Zensus 2011 abgeglichen. Die Ergebnisse zeigen vom Grundsatz, dass die Bevölkerung mit den vorliegenden Mobilfunkdaten teilweise gut abgebildet werden könnte. Beobachtbare Unterschiede in der Bevölkerungsdarstellung mittels Mobilfunkdaten und den Zensuswerten können teilweise durch die zeitliche Differenz zwischen den Mobilfunkdaten einer statistischen Woche aus verschiedenen Monaten des Jahres 2017 und den Zensusdaten aus dem Jahr 2011 erklärt werden. Aber auch das angewandte Hochrechnungsverfahren könnte hierfür ursächlich sein und muss wird Verbesserung geprüft. Zur Frage der Verzerrungen sowie zu den Selektivitäten erfolgen weitere Untersuchungen.
Zudem stößt die Pendlerrechnung in ihrer aktuellen Form an Grenzen. Die Nutzung von Informationen aus unterschiedlichen Statistiken, wie der Personalstandstatistik, der Beschäftigungsstatistik und dem Mikrozensus decken Pendlerströme unter anderem der Sozialversicherungspflichtigen, Beamten und geringfügig entlohnten Beschäftigten mehr oder minder adäquat ab. Schwierigkeiten entstehen dagegen beim Darstellen der Pendlerströme von Bildungspendlern, selbständig erwerbstätigen Personen sowie Grenzpendlern, also Einpendlern aus dem Ausland. In Kooperation mit IT.NRW erfolgen Analysen hinsichtlich der Pendlerrechnung im Projekt ‚Pendler Mobil'. Mit Hilfe von sogenannten ‚Origin Destination Matrizen', auch ‚Quelle-Ziel Matrizen' genannt, können Mobilfunkdaten dazu genutzt werden, Pendlerströme im Tagesverlauf abzubilden. Vor allem durch die Einbeziehung der Nationalität der SIM-Karte können grenzüberschreitende Einpendler möglicherweise erfasst und in die Pendlerrechnung miteinbezogen werden.
Weiterhin wurden im Rahmen des ESSnet Projektes ‚City Data from LFS and Big Data' Indikatoren der Arbeitskräfteerhebung mittels Small Area Verfahren unter Verwendung von Mobilfunkdaten auf kleinräumige Ebenen geschätzt. Durch dieses Verfahren konnte Gebieten ohne Beobachtungen verlässliche Schätzer ausgegeben werden und gleichzeitig die Unsicherheit in den bei der Schätzung der Indikatoren verringert werden.

Robust and Nonparametric Statistics II

215 Robustness and Stability of Kernel-Based Machine Learning

Andreas Christmann
University of Bayreuth, Germany

Statistical machine learning and methods for high dimensional and large data sets are important topics in current research. The fourth dimension of big data analysis is often called veracity, i.e. handling data in doubt. This concept is similar to the notion of statistical robustness. Certain questions of universal consistency, statistical robustness, and stability will be addressed for the case of kernel-based methods.

216 Fitting additive models with regularized kernel methods: methodology, robustness properties, and business applications

Robert Hable
Technische Hochschule Deggendorf, Germany

In business applications, fully nonparametric methods have the disadvantage that the output of these methods is often considered as being a black-box solution which can hardly be interpreted and comprehended. Therefore, these methods are not accepted in many business applications. In particular, this applies to prominent machine learning methods like neural networks and support vector machines. Additive models provide a good compromise between parametric and fully nonparametric models as they are more general than parametric methods but provide easily interpretable output.
In the talk, a class of methods for fitting additive models is presented which is a generalization of support vector machines. It is shown that these methods have good robustness properties, which is particularly important in business applications. E.g., in case of demand forcasts or in the analysis of data on inventories or machine downtimes, skewed distributions and the presence of outliers is rather the rule than the exception. In the talk, an application of the methodology is demonstrated on real data from an industrial project.

217 Empirical examination of the potential of robust regularized regression to examine genetic associations with circulating metabolite levels

Heike Deutelmoser[1,2], Dominique Scherer[1], Justo Lorenzo Bermejo[1]

[1]Institute of Medical Biometry and Informatics, University Hospital Heidelberg, Heidelberg, Germany

[2]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany

Multiple linear regression is often applied to investigate the association between circulating metabolite levels as response variable and large numbers of individual genotypes as explanatory variables. Regularized regression has been proposed as an alternative to identify the strongest inherited predictors of circulating metabolite levels. Regularized regression penalizes the number of explanatory genotypes by a tuning parameter λ that avoids potential model overfitting.

Results from both standard multiple linear regression and regularized regression models are considerably influenced by few observations departing from the majority of the sample, so called outliers. The identification of outliers is challenging in high-dimensional multi-omics data and, even if outliers can be identified, the exclusion criteria are often arbitrary. A much more promising approach relies on constraining the influence of outlying observations by robust statistical methods. Robust statistics aims to infer the best prediction model for the majority of the study population instead of the best model for any observation. Existing regularized regression methods have been adapted taking advantage of robust statistics.

We used genotype and metabolite data from 172 participants of the largest European prospective cohorts (Lifelines, HUNT, FINRISK, Estonian Biobank, ESTHER, EPIC and TwinGene) to compare the performance of standard and robust approaches in assessing potential associations between circulating metabolite levels and individual genotypes. In particular, we investigated the influence of outlying observations on (1) the complexity of the selected model, (2) the regularization parameter λ and (3) the estimated regression parameters. Individuals with outlying genotypes or metabolite measurements had a stronger influence on the results from the standard LASSO than on those from the robust (Huber) LASSO. Details on the methodology, our findings and practical recommendations will be presented at the meeting.

218 Bias Correction for Local Linear Regression Estimation Using Asymmetric Kernels via the Skewing Method

Benedikt Funke[1], **Masayuki Hirukawa**[2]

[1]Technical University of Dortmund, Germany

[2]Ryukoku University, Japan

This paper extends the skewing method by Choi and Hall (1998) that has been originally proposed as a bias correction device for local linear regression estimation using standard symmetric kernels to the cases of asymmetric kernels (see, e.g., Chen, 1999, 2000). The method is defined as a convex combination of three local linear estimators. We derive the analytical expressions of the leading bias and variance terms for the skewed regression estimator using asymmetric kernels. It is demonstrated that the estimator with properly chosen weights can accelerate the bias convergence from $O(b)$ to $O(b^2)$ under sufficient smoothness of the unknown regression curve while not inflating the variance in order of magnitude, where b is the smoothing parameter. As a consequence, the estimator has optimal pointwise convergence of $n^{4/9}$ when best implemented, where n is the sample size. It is noteworthy that these properties are the same as those for a local cubic regression estimator. A remarkable difference can be found in the choice of weights that can lead to a faster bias convergence of the skewed estimator when symmetric and asymmetric kernels are employed. While the weights are constant regardless of the position of the design point for symmetric kernels, they vary with the design point for asymmetric kernels. Finite-sample performance of the skewed regression estimator is examined via Monte Carlo simulations in comparison with a local cubic estimator, and an empirical application is considered.

Statistics in Behavorial and Educational Sciences II (Behavioural Sciences)

219 Dynamic Microsimulation Modelling of Care Needs in Germany

Christoph Frohn[1], Monika Obersneider[2]

[1]Katholische Hochschule Nordrhein-Westfalen, Germany
[2]Universität Duisburg-Essen, Germany

Computer simulations have proven to be a useful statistical tool in both technical and natural sciences as well as in economic sciences. In the social sciences, however, simulations do not belong to the standard repertoire of statistical methods. Hence, the exploration of complex societal mechanisms and their interplay for the future projection of social and behavioral developments have received insufficient attention. If simulations are used in the context of social science, they mainly concentrate on predicting the impact of political or fiscal interventions e.g. the ex-ante evaluation of tax policy changes on different population groups. Potentials to explain the emergence of social phenomena and the long-term dynamics in individual behaviour patterns, on the other hand, are not exhausted.

Against this background, this contribution attempts to unfold the advantages of using microsimulation modelling (Orcutt 1957) in the realm of social sciences. Due to the modular and stochastic structure of microsimulations, projections of the development of social phenomena become realistic and comprehensible (e.g. nursing, employment or poverty rates). At the same time, microsimulations are able to show in what way these societal developments can be attributed to individual actions, which in turn can influence future behaviour.

(Action-)Theories of social science on health, family, migration or demography in general can easily be transferred into the structure of a microsimulation. We present the results of a dynamic microsimulation on the development of health inequalities and nursing care needs in Germany. In addition to individual data on socio-economic aspects, the simulation focuses on family- and migration-specific factors. Although with reference to demographic developments, growth in the need for long-term care in Germany appears obvious, the simulation results highlight the importance of dynamics in family and household structures for the future development of the long-term care situation in Germany. Microsimulation not only helps to reveal the complex interactions between dimensions of social inequality and health in the population but also helps to project potential future developments.

Orcutt, G. (1957). A new type of socio-economic system. The Review of Economics and Statistics, 39, 116-123. (Reprinted in the International Journal of Microsimulation, 2007, 1(1), 3-9).

220 Elicited preferences of potential spontaneous unaffiliated on-site volunteers in the context of natural disasters

Christoph Herrmann, Susanne Kirschstein-Barczewski
Martin Luther University Halle-Wittenberg, Germany

Cases of natural disasters have shown that the support of spontaneous unaffiliated on-site volunteers (SUVs) is indispensable in catastrophes. SUVs are important in disaster management, but they are not integrated in the coordination planning system. The first step to improve SUV coordination is to investigate the main drivers for helping in catastrophes in order to anticipate the SUVs' behavior. For that reason, we use the random utility framework to evaluate the attributes of potential SUVs by means of discrete choice experiments (DCEs). For estimating SUV-preferences, we assess a set of attributes collected in a DCE survey.
To validate the appropriateness of our modeling approach, we collect two samples. Both samples consist of students in an introductory statistics course. The first sample is used to create and estimate a model of the decision making process to elicit SUV-preferences. The model is evaluated by forecasting preferences for the second dataset. Comparing the forecasted and empirical results of the second dataset is used to validate the estimates. The out-of-sample prediction shows a reliable fit of the data.
On the one hand, the results indicate that the personal involvement, an attribute determined by the individual, plays an important role for SUVs getting involved. On the other hand, the results show that attributes determined by the context, like the size and scale of the catastrophe or disasters' coverage in social media are main drivers for the decision to help. Therefore, the implementation of social media in the coordination planning system is helpful to deploy SUVs efficiently.

221 Tactical Voting and Ticket-Splitting in Mixed Electoral Systems: A Finite-Mixture Approach Applied to the Case of Germany

Martin Elff
Zeppelin Universität Friedrichshafen, Germany

Why voters split their votes between elections at different (state vs federal) levels or between ballots in mixed dual vote electoral systems has been subject of a debate for several years now. A natural explanation is that ticket-splitting in mixed electoral systems is the result of various voting strategies: wasted-vote avoidance in the plurality vote on the ballot, threshold insurance in the proportional vote on the ballot, or a combination of both. Uncovering such voting strategies however has posed a considerable challenge, as they cannot be observed directly.
Despite the importance of tactical voting, no consensus has yet been reached on how to measure it. The proposed paper applies a finite-mixture discrete choice model to the case of ticket-splitting in the context of the mixed electoral system of Germany. Based on the German Election studies of 2009 through 2017 it derives the proportion of ticket-splitting that can be attributed to indifference between parties, wasted-vote avoidance, the intention to bring about a preferred coalition, or the expression of a rank-order between the first- and second-most preferred party.

222 Bayes Factor: Inconsistency in Sequential Updating.

Patrick Michael Schwaferts

LMU, Germany

The enormous attention Bayes factors have recently attracted for instance in psychological research should not mask the fact that Bayes factors suffer from severe drawbacks. In particular, as the talk will show, they are dynamically inconsistent.
The Bayes Factor is a quantity within a Bayesian approach to hypotheses testing. Typically, there are two hypotheses, which specify a distribution of the parameter of interest. In addition, there are prior probabilities of the hypotheses as well, which do get updated by the observed data. In this context the Bayes Factor quantifies, how the probabilities of the hypotheses change.
If two data sets are observed, it is possible to calculate two Bayes Factor values and update the prior probabilities of the hypotheses two times sequentially. Alternatively, it is possible to merge both data sets and calculate one Bayes Factor value to update the prior probabilities of the hypotheses at once. Both approaches do not yield the same posterior probabilities of the hypotheses. On closer inspection, it appears that sequential updating discards information and therefore might be pictured as inconsistent updating.

Survival and Event History Analysis III (High-Dimensional Analysis)

223 Correlation-Adjusted Regression Survival Scores for High-Dimensional Variable Selection

Thomas Welchowski[1], Verena Zuber[2], Matthias Schmid[1]

[1]University Hospital Bonn, Germany
[2]Imperial College of Science, Technology and Medicine, London, UK

Background
The development of classification methods for personalized medicine is highly dependent on the identification of predictive genetic markers. In survival analysis it is often necessary to discriminate between influential and non-influential markers. It is common to perform univariate screening using Cox scores, which quantify the associations between survival and each of the markers to provide a ranking. Since Cox scores do not account for dependencies between the markers, their use is suboptimal in the presence highly correlated markers.
Methods
As an alternative to the Cox score, we propose the correlation-adjusted regression survival (CARS) score for right-censored survival outcomes. By removing the correlations between the markers, the CARS score quantifies the associations between the outcome and the set of "de-correlatedmarker values. Estimation of the scores is based on inverse probability weighting, which is applied to log-transformed event times. For high-dimensional data, estimation is based on shrinkage techniques.
Results
The consistency of the CARS score is proven under mild regularity conditions. In simulations, survival models based on CARS score rankings achieved higher areas under the precision-recall curve than competing methods. Two example applications on prostate and breast cancer confirmed these results. CARS scores are implemented in the R package carSurv.
Conclusions
In research applications involving high-dimensional genetic data, the use of CARS scores for marker selection is a favorable alternative to Cox scores even when correlations between covariates are low. Having a straightforward interpretation and low computational requirements, CARS scores are an easy-to-use screening tool in personalized medicine research.

224 Bayesian variable selection for Cox models with network-structured covariates

Katrin Madjar[1], Manuela Zucknick[2], Katja Ickstadt[1], Jörg Rahnenführer[1]
[1]TU Dortmund, Germany
[2]University of Oslo, Norway

We propose a new Bayesian approach to perform variable selection in the Cox proportional hazards model. The method is developed for situations where data is heterogeneous due to known subgroups of patients that may differ in their relationship between predictors and survival outcome. Specifically, we address the problem of building a separate risk prediction model for each subgroup while at the same time allowing sharing information between subgroups to increase power when appropriate. This is achieved by assuming a network that links predictors within and across different subgroups. The network structure is not known a priori and inferred simultaneously with the important variables for each subgroup, using a Gaussian graphical model. Network information is incorporated into variable selection via a Markov random field (MRF) prior that encourages selection of predictors with neighbors already included into the model. The proposed model is well suited for genomic applications as it helps identifying pathways of functionally related genes and genes that are simultaneously prognostic in different subgroups. In simulations, we compare the performance of the new method against standard approaches with independent Bernoulli priors for variable selection instead of the MRF prior. The results demonstrate that our proposed model can achieve improved prediction and variable selection accuracy, in particular when the sample size is small.

225 Adaptive LASSO Cox frailty models based on the full likelihood

Maike Hohberg[1], Andreas Groll[2]
[1]University of Goettingen, Germany
[2]TU Dortmund

We propose a regularized Cox frailty model that accommodates time-varying covariates as well as time-varying covariate effects and is based on the full likelihood. In previous simulation studies, it has been shown that using the partial likelihood compared to the full likelihood yields a loss in precision especially in small or moderate samples. Given that in many medical applications for example, the sample size is often rather small, it seems surprising that none of the established R routines are based on the full likelihood considering that for small datasets using the full likelihood does not drastically increase computing time. We provide a function coxlasso() that fits regularized Cox frailty models, show the function's performance compared to existing routines, and assess situations in which using the full likelihood might be most effective.

226 Smooth backfitting of additively structured hazard rates for in-sample forecasting

Stephan M. Bischofberger[1], Munir Hiabu[2], Enno Mammen[3], Jens P. Nielsen[1]

[1]Cass Business School, United Kingdom
[2]University of Sydney, Australia
[3]Heidelberg University, Germany

Smooth backfitting has been established in nonparametric regression and in density estimation as a very promising alternative to the classic backfitting method by Buja, Hastie and Tibshirani. We apply the concept to a survival model with additively structured nonparametric hazard. The model allows for very general censoring and truncation patterns occurring in many forecasting applications such as medical studies or actuarial reserving. A crucial point is that – in in contrast to classical backfitting – we do not assume independence between the covariates. Our estimators are shown to be a projection of the data into the space of multivariate hazard functions with additive components. Hence, our hazard estimator is the closest additive fit even if the actual hazard rate is not additive. Another big advantage of our additive model is that our estimators are straight forward to derive in theory including excellent properties as well as their simple implementation in practice even for high dimensional covariates. We provide full asymptotic theory for our estimators as well as a simulation study.

Statistics in Finance II

227 Testing for Daily Jumps in Risky Asset Returns: a novel approach based on Gini concentration measure

V Golosnoy, J Kellermann
Ruhr Universität Bochum, Germany

Detection of a daily jump component in the data generating process of financial returns is of crucial importance for measuring daily volatility of risky assets.
In the current paper we propose a new test for possible jumps during the day which is based on the Gini concentration index computed for absolute intraday returns. As part of our testing procedure, we suggest a new estimator for the standard error of the Gini index. We investigate the performance of our test in the Monte Carlo simulation where our approach shows to have power advantages compared to the existing testing procedures. Moreover, we provide an empirical illustration where we apply our testing approach to intraday stock market data.

228 Dynamic regular vine copulas with an application to exchange rates dependence

Alexander Kreuzer, Claudia Czado
Technische Universität München, Germany

Modeling dependence among financial assets is an important research topic as the dependence structure has high influence on the risk associated with a corresponding portfolio. Regular vine copulas have proven as a useful tool in this context. They allow for characteristics like asymmetric tail dependence, which cannot be modeled with a multivariate Gaussian or Student t copula. Usually it is assumed, that the dependence parameters of the regular vine copula remain constant as time evolves. We get rid of this assumption and propose dynamic regular vine copulas. In this dynamic model dependence parameters are described through latent AR(1) processes. Since maximum likelihood estimation is infeasible for these latent AR(1) processes we employ Markov Chain Monte Carlo within a sequential estimation procedure. The approach is illustrated with 25-dimensional exchange rates data, where we find clear evidence for dynamic dependence.

229 Portfolio Pretesting with Machine Learning

Ekaterina Kazak, Winfried Pohlmeier
Universität Konstanz, Germany

This paper exploits the idea of pretesting to choose between competing portfolio strategies.
We propose a strategy that optimally trades off between the risk of going for a false positive strategy choice versus the risk of making a false negative choice.
Various different data-driven approaches are proposed based on an optimal choice of the pretested certainty equivalent and Sharpe Ratio. Our approach belongs to the class of shrinkage portfolio estimators. However, contrary to previous approaches the shrinkage intensity is continuously updated to incorporate the most recent information in the rolling window forecasting set-up.
We show that the bagged pretest estimator performs exceptionally well, especially when combined with adaptive smoothing. The resulting strategy allows for a flexible and smooth switch between the underlying strategies and is shown to outperform the corresponding stand-alone strategies.

230 The risk function of the goodness-of-fit tests for tail models.

Ingo Hoffmann, Christoph J. Börner
Heinrich Heine University Düsseldorf, Germany

This paper contributes to answering a question that is of crucial importance in risk management and extreme value theory: How to select the threshold above which one assumes that the tail of a distribution follows a generalized Pareto distribution. This question has gained increasing attention, particularly in finance institutions, as the recent regulative norms require the assessment of risk at high quantiles.
Recent methods answer this question by multiple uses of the standard goodness-of-fit tests. These tests are based on a particular choice of symmetric weighting of the mean square error between the empirical and the fitted tail distributions. Assuming an asymmetric weighting, which rates high quantiles more than small ones, we propose new goodness-of-fit tests and automated threshold selection procedures. We consider a parameterized family of asymmetric weight functions and calculate the corresponding mean square error as a loss function. We then explicitly determine the risk function as the finite sample expected value of the loss function. Finally, the risk function can be used to discuss the question of which symmetric or asymmetric weight function and, thus, which goodness-of-fit test should be used in a new method for determining the threshold value.

Statistics of High Dimensional Data II

231 Detecting binding sites in PAR-CLIP data using a Bayesian hierarchical model

Eva-Maria Huessler[1], Martin Schäfer[1,2], Pablo Landgraf[3], Holger Schwender[1]
[1]Heinrich Heine University Düsseldorf
[2]German Rheumatism Research Centre Berlin
[3]University Hospital Cologne

MicroRNAs are small non-coding RNAs that play an important role in gene regulation, as they bind to target mRNA to initiate translational repression and mRNA destabilization. Such targeted mRNA sites can be detected using PAR-CLIP (Photoactivatable-Ribonucleoside-Enhanced Crosslinking and Immunoprecipitation). This method introduces T-to-C substitutions at sequenced cDNA that help to detect binding sites on mRNA. Nevertheless, detected T-to-C nucleotide exchanges can also be due to other reasons such as sequencing errors or single nucleotide polymorphisms (SNPs). Only few statistical methods have been developed to detect potential binding sites in PAR-CLIP data, most of which do not account for such errors and SNPs. None of these methods allow the incorporation of additional information helpful for distinguishing PAR-CLIP induced and other T-to-C substitutions. It is, e.g., well-known that binding sites occur most often in the 3'UTR (untranslated region) of the mRNA.
We have developed BayMAP [1], a fully Bayesian hierarchical model based on a three-component mixture model, that allows taking other sources of nucleotide exchange than the PAR-CLIP induced ones into account, e.g., sequencing errors and SNPs. Thanks to its hierarchical structure, BayMAP is the first method for the analysis of PAR-CLIP data that additionally allows the incorporation of further prior information relevant for the biology of microRNA binding sites such as the mRNA region.
In this presentation, we will demonstrate its good performance in comparison to other procedures for analyzing PAR-CLIP data in an extensive simulation study and its application to several real PAR-CLIP data sets. We will also show how BayMAP's ability of incorporating additional information can be used to investigate the influence of other potential biological factors than the mRNA region on the probability of having a binding site position.
BayMAP as well as other procedures, first, detect single high confident binding positions. In a second step, binding sites are derived by, e.g., sequence alignments of reads encompassing these high confident binding positions. The positions are, therefore, supposed to be mutually independent. However, it is likely that a binding site has more than one T-to-C substitution position and that the substitutions of two positions close to each other are not independent. We propose to model these dependencies employing a conditional autoregressive model (CAR). In our talk, we will show how to implement the CAR model in BayMAP and we will present results of its application to real PAR-CLIP data sets.
[1] Huessler EM, Schäfer M, Schwender H, Landgraf P (2018). BayMAP: A Bayesian hierarchical model for the analysis of PAR-CLIP data. Bioinformatics, doi: 10.1093/bioinformatics/bty904.

232 High-throughput DNA methylation analysis with reference-free cell type adjustment: method comparison in a real data application

Miriam Kesselmeier[1], Anke Hinney[2], André Scherag[1,3]

[1]Research Group Clinical Epidemiology, Center for Sepsis Control and Care (CSCC), Jena University Hospital, Jena, Germany
[2]Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, University Hospital Essen, University of Duisburg-Essen, Essen, Germany
[3]Institute of Medical Statistics, Computer and Data Sciences, Jena University Hospital, Jena, Germany

Differential DNA methylation is regarded as a key element for our understanding of complex traits and diseases. Given that DNA methylation is cell type dependent and that cell type distributions are sometimes associated with the trait of interest, cell type distribution is a possible confounder in DNA methylation analyses. Several methods developed to account for this possible confounding have been compared [e.g., 1-3].
Such a confounding must be considered when analysing DNA methylation in females suffering from anorexia nervosa (AN). In females with AN, starvation is a key clinical feature affecting both DNA methylation and cell type distribution. We investigated DNA methylation in females with AN compared to lean females without AN [4]. To account for a possible cell type effect, we applied both RefFreeEWAS [5] and FaST-LMM-EWASher [6]. As the results strongly depended on the analysis method, we searched for consensus CpG sites, i.e. sites identified by both methods (both p-values below 0.01). Shortly, we identified 26,769 differently methylated sites when applying RefFreeEWAS and 1,059 sites with FaST-LMM-EWASher. There were 51 consensus sites. These differences might be caused by the filter applied by authors' recommendation in FaST-LMM-EWASher that excludes probes with very low/high methylation values. However, even without this filter, this method might be better suited for situations in which avoiding false positive results is a major concern [3]. Hence, we will re-run the FaST-LMM-EWASher analysis without the filter and compare the results to the results of the original investigation with a special focus on differences and agreement.
References:
[1] McGregor K, Bernatsky S, Colmegna I, et al. An evaluation of methods correcting for cell-type heterogeneity in DNA methylation studies. Genome Biology (2016);17: 84.
[2] Kaushal A, Zhang H, Karmaus WJJ, et al. Comparison of different cell type correction methods for genome-scale epigenetics studies. BMC Bioinformatics (2017); 18(1): 216.
[3] Brägelmann J, Lorenzo Bermejo J. A comparative analysis of cell-type adjustment methods for epigenome-wide association studies based on simulated and real data sets. Briefings in Bioinformatics (2018); [Epub ahead of print].
[4] Kesselmeier M, Pütter C, Volckmar A-L, et al. High-throughput DNA methylation analysis in anorexia nervosa confirms TNXB hypermethylation. The World Journal of Biological Psychiatry (2018); 19 (3): 187-199.
[5] Houseman EA, Molitor J, Marsit CJ. Reference-free cell mixture adjustments in analysis of DNA methylation data. Bioinformatics (2014); 30: 1431-1439.
[6] Zou J, Lippert C, Heckerman D, et al. Epigenome-wide association studies without the need for cell-type composition. Nature Methods (2014); 11:309-311.

233 Forecasting with Supervised Factor Models

Simon Lineu Umbach

University of Cologne, Germany

A popular approach to cope with large predictor spaces in forecasting applications is to estimate a small set of latent factors by Principal Component Analysis and then relate them in a separate step to the forecasting target in a regression framework. This study analyzes under which circumstances gains in forecast accuracy can be achieved by incorporating some form of supervision in the factor estimation process. Specifically, Principle Covariate Regression (PCovR) is considered. For the problem of choosing a value for the supervision parameter in PCovR an information criterion is proposed. The information criterion is shown to be an appropriate means to find a good balance between predictor space compression and target orientation of the estimated factors. A simulation study and an empirical application on a macroeconomic dataset show that supervised factors can improve the forecasting accuracy of factor models by incorporating relevant information of the regressor space that is neglected or only captured in some minor principal components by the unsupervised factor model.

234 Impact of Population Stratification on Polygenic Risk Score Approaches

Anke Huels, Michael P. Epstein

Department of Human Genetics, Emory University, Atlanta, Georgia, USA

Since most complex diseases are influenced by several genes each having a small effect on its own, polygenic risk scores (PRS) that model the genetic basis en masse often provide more insight into the variation of complex traits than single variant approaches. A study constructs a PRS in a test sample of interest using weights derived from effect estimates from a large independent training set of samples, e.g. genome-wide association studies (GWAS) when analyzing genotype data or epigenome-wide association studies (EWAS) when analyzing DNA methylation data. PRS analyses implicitly assume no systematic differences between the training and the test sample. However, this assumption is likely violated both for genotype data (test and training datasets may differ by confounders like ancestry) and methylation data (test and training datasets may differ by confounders like ancestry, age, tissue and cell-type composition). Ignoring these systematic differences between test and training datasets might lead to misleading inference.
The objective of this study is to investigate the impact of population stratification on power, type I error and prediction capacity of PRS approaches and to show how its performance improves by modeling the subgroups.
In our simulation studies, we analyze scenarios in which associated (epi)genetic markers differ between subgroups of the training as well as test datasets or in which the effects are smaller or not existent in a subgroup of the data. Associations between (epi)genetic markers and disease are estimated in the training data using elastic net regression as well as the more common linear regression models with p-value thresholds for marker selection. The resulting effect estimates are subsequently used as weights to construct the PRS and to estimate the association between PRS and disease in the test data. We compare the performance of PRS both ignoring and modeling the subgroups, which are defined by homogenous effects within subgroups and heterogeneous effects between subgroups. To model subgroups, we matched training and test samples by subgroup and then estimated weights and constructed PRS within each subgroup. We then either tested the PRS in the entire dataset (PRSsub) or separately in each subgroup followed by a fixed-effects meta analysis combining results across subgroups (PRSmeta). In addition, our approaches will be applied to genotype as well as DNA methylation data from the Multi-Ethnic Study of Atherosclerosis (MESA) to investigate the genetic and epigenetic bases of cognitive function.
Ignoring population stratification in PRS led to a very low power as well as a poor prediction capacity. The PRSsub approach with elastic net regression achieved the highest power and prediction capacity with a well-controlled type I error: For instance, if the effect sizes in half of the study population are only 10% of the effect sizes in the rest of the population, the power of PRSsub is 0.97, whereas it drops to 0.31 when ignoring the population stratification and is 0.45 when using the PRSmeta approach.
In conclusion, our study clearly shows the need for PRS approaches that can model and handle systematic differences in subgroups between test and training datasets.

Statistics in Science, Technology and Industry II

235 Inference and Change Detection for LSHD Time Series and Applications to Ozone Monitoring

Ansgar Steland
RWTH Aachen University, Institute of Statistics

In data science applications such as ozone sensor monitoring computations and analysis of the covariance matrix often has to be done in a low-sample-size-high-dimensional (LSHD) regime. Especially, if the dimension is larger than the sample size, classic methods fail and need to be replaced by procedures which are designed for high-dimensional data.
We present recent results on LSHD asymptotics of bilinear forms of the sample covariance matrix. This approach allows us to detect and infer changes in the dependence structure. The theoretical results hold without the need to constraint the dimension relative to the sample size. For the statistical estimation of unknowns one often needs a (large) learning sample. To circumvent this, we propose in-sample estimators not requiring a learning sample. Simulations show that the proposed methods work reliable for realistic mathematical models.
As a real world application the method is applied to analyze monitoring data from ozone sensors. The sensor data is compressed by projecting it onto sparse principal directions obtained by a sparse principal component analysis (SPCA). It turns out that the SPCA automatically learns the spatial locations of the sensors and leads to a spatial segmentation. Analyzing the projections for a change-point provides a mean to detect changes in the spatial dependence structure of the sensor network measuring ozone.
Financial support from DFG, grant #1034/11-1, is gratefully acknowledged.

236 Estimation of the Spatial Weighting Matrix for Spatiotemporal Data with Structural Breaks

Philipp Otto[1], Rick Steinert[2]

[1]Leibniz University Hannover, Germany
[2]European University Viadrina, Frankfurt (Oder), Germany

In this talk, we propose a two-step lasso estimation approach for the estimation of a full spatial weights matrix of spatiotemporal autoregressive models. In addition, we allow for an unknown number of structural breaks in the local means of each spatial location. The proposed approach jointly estimates the spatial dependence, all structural breaks, and the local mean levels.
Simultaneously, we address an important problem in spatial econometrics. The classical approach is to replace the unknown spatial dependence structure by a linear combination of a scalar and a predefined, non-stochastic weighting matrix describing the dependence. One might get an insight on the spatial dependence looking at the spatial covariogram or semivariogram (see, e.g., Cressie and Wikle 2011). In practice, however, this matrix can not easily be assessed, and is, therefore, estimated by maximizing a certain goodness-of-fit measure, like the log-likelihood, in-sample fits, information criteria, or cross validations over a certain classical weighting schemes. In contrast to this classical approach, we estimate all entries of this weighting matrix by a penalized regression approach. In addition, we suppose that there might be an unknown number of structural breaks in the data. These breaks can occur at different time points for each location and they can be of different magnitude. This yields nT possible structural breaks in the mean level of the data. Due to the spatial dependence and the high number of possible changes, it is hard to distinguish between effects resulting from mean-level changes or from the spatial dependence. In order to overcome this issue of interaction, we propose a two-step estimation approach.
In addition, the computation of the suggested estimators is simplified, due to the convex objective function resulting from a slight simplification. Through simulation studies, we will show the finite-sample performance of the estimators and provide practical guidance as to when the approach could be applied. Finally, the invented method will be illustrated by an empirical example of regional monthly real-estate prices in Berlin between 1995 and 2014. The spatial units will be defined by the respective ZIP codes. The new approach allows us to estimate local mean levels and quantify the deviation of the observed prices from these levels due to spatial spillover effects.

237 Statistical analysis of joint pausing in parallel spike trains

Matthias Gärtner, Solveig Plomer, Sevil Duvarci, Jochen Roeper, Michael Messer, **Gaby Schneider**

Goethe University Frankfurt, Germany

Periods with surprisingly few events, so-called 'pauses', have gained increasing attention in the analysis of parallel spike trains. In particular, simultaneous pausing may have an important impact on information processing. Available analysis methods that first identify pauses in individual spike trains are often threshold dependent and can fail to identify joint pauses in parallel processes that are visible to the eye.
We therefore introduce techniques that rely on the joint analysis of parallel spike trains. A statistic called pausiness is introduced that works with the superposition of two processes. This statistic measures the degree of synchronous pausing in spike train pairs and avoids threshold-dependent identification of individual pauses. A graphic termed the 'cross-pauseogram' compares the joint pausiness of two spike trains with its time shifted analogue and can thus indicate simultaneous pausing. When assessing statistical significance, we use a stochastic point process model that also comprises synchronous spiking events. Parameter estimates are obtained from auto- and cross-correlation functions. We also explore the use of jointly moving sum statistics. The techniques are applied to a data set of dopaminergic neurons recorded from freely moving mice, where joint pausiness was observed in about 20% of the investigated neuron pairs.
This work was supported by the Priority Program 1665 of the DFG.

238 Simultaneous optimization of several correlated response variables

Eva-Christina Becker-Emden, Sonja Kuhnt

Dortmund University of Applied Sciences and Arts, Department of Computer Science

In industrial applications, it is often of interest to reach certain target values for several quality features of a product at the same time. Robust parameter design looks for process settings, for which the mean of the quality features is on target with minimal variation. We use a loss based optimization strategy for multiple responses. Taking possible correlation into account statistical models are wanted, where the mean vector as well as the variance-covariance matrix depends on the process factors. We review recent approaches for multivariate normal responses and extend the basic ideas to wider classes of multivariate distributions. We use experimental data of a high velocity oxygen fuel (HVOF) spraying process for illustration. HVOF spraying is a method to apply protective coatings on surfaces. Properties, like hardness and deposition efficiency reflect the quality of the coating.
This work is supported by the Deutsche Forschungsgemeinschaft (SFB 823, project B1).

Advanced Regression Modeling IV (Longitudinal Data and Mixed Models)

239 Flexible Bayesian modelling of treatment effects on panel outcomes

Helga Wagner

Johannes Kepler University, Austria

Identification and estimation of treatment effects is an important issue in many application fields,e.g. to evaluate of the effectiveness of social programs, government policies or medical interventions. In contrast to randomized studies, where treatment effects can be estimated based on a comparison of the outcome of interest in treatment and control group, unobserved confounding due to endogeneity of treatment selection has to be taken into account for data from observational studies. In the Bayesian approach this is accomplished by specifying a joint model of treatment selection and the potential outcomes. For the estimation of dynamic effects of a binary treatment on a continuous outcome observed over subsequent time periods two models, the switching regression model and the shared factor model, have been suggested so far. We show that both impose restrictions on the joint correlation structure of treatment selection and the two outcomes sequences that can result in biased treatment effects estimates.
To achieve more flexibility we propose a new model that allows to separate longitudinal association of the outcomes from association due to endogeneity of treatment selection. We employ this model to analyse the effects of a long maternity leave on earnings of Austrian mothers. where we exploit a change in the parental leave policy in Austria that extended maternal benefits from 18 months since birth of the child to 30 months. Our analysis is based on data from the Austrian Social Security Register which contains individual employment histories since 1972 and also reports number of births and maternity and parental leave spells for all Austrian employees.

240 A D-Vine Copula-Based Model for Repeated Measurements Extending Linear Mixed Models with Homogeneous Correlation Structure

Claudia Czado, Matthias Killiches

Technical University Munich, Germany

We propose a model for unbalanced longitudinal data, where the univariate margins can be selected arbitrarily and the dependence structure is described with the help of a D-vine copula. We show that our approach is an extremely
flexible extension of the widely used linear mixed model if the correlation is homogeneous over the considered individuals. As an alternative to joint maximum–likelihood a sequential estimation approach for the D-vine copula is provided and validated in a simulation study. The model can handle missing values without being forced to discard data. Since conditional distributions
are known analytically, we easily make predictions for future events. For model selection, we adjust the Bayesian information criterion to our situation. In an application to heart surgery data our model performs clearly better than competing linear mixed models.

241 Solving separation in the mixed effects logistic regression model

Georg Heinze[1], Rok Blagus[2]

[1]Medical University of Vienna, Austria
[2]University of Ljubljana, Slovenia

Researchers analyzing studies with binary outcomes using logistic regression are often faced with the problem that finite maximum likelihood estimates of regression coefficients do not exist. This phenomenon is known as 'separation' and solutions to it were proposed and implemented in standard software, e.g., the Firth correction (Firth D, Biometrika 1993). However, so far no solution has been proposed for the case where such a study contains clustered (correlated) observations. Usually, these studies are analyzed with generalized estimating equations or mixed effects models. For the latter we propose a strategy to tackle the problem of separation in the fixed effects, which is based on the Firth correction. The approach is presented as a general algorithm where any likelihood approximation needed to estimate a mixed effects logistic regression model, e.g. the adaptive Gaussian quadrature, can be used. Accuracy of fixed effect estimates by our approach is demonstrated by a simulation study, and analysis of a real data set (relationship of the prevalence of parasites in birds with their nutritional behavior and migration habits) exemplifies some aspects of application.

242 Second order asymptotic biases of consistent estimators under many instruments

Stanislav Anatolyev

CERGE-EI, Czech Republic

We consider a linear homoskedastic instrumental variables model with many instruments. In an asymptotic framework where their number is proportional to the sample size, we derive the second order asymptotic biases of the jackknife IV, bias corrected 2SLS and LIML estimators, as well as the Fuller-type correction of LIML. The structure of second order asymptotic biases is similar to that of first order asymptotic variances. We compare their expressions whenever possible and elaborate on certain special cases as well as a specific numerical example.

Computational Statistics and Statistical Software IV(Software)

243 tidyfun: a new framework for representing and working with function-valued data

Fabian Scheipl[1], Jeff Goldsmith[2]

[1]LMU Munich, Germany
[2]Columbia University, New York

We present a new R package tidyfun (github.com/fabian-s/tidyfun) for representing and working with function-valued data that presents a unified interface for dealing with regularly or irregularly observed function-valued data.
The package provides new data types for representing functional data, arithmetic operators, descriptive statistics and graphics functions for such data and tidyverse-verbs for handling functional data inside data frames.
The package is aimed at lowering the barrier of entry for analysts in order to quickly and painlessly analyse and interact with functional data and, specifically, datasets that contain both scalar and functional data or multiple types of functional data, potentially measured over different domains. We discuss the available feature set as well as forthcoming extensions and show some application examples.

244 Controlling the false discovery rate for discrete data: New results and software

Sebastian Doehler[1], Guillermo Durand[2], Etienne Roquain[2]

[1]Hochschule Darmstadt, Germany
[2]Sorbonne Universite, France

The Benjamini-Hochberg procedure and related methods are classical methods for controlling the false discovery rate (FDR) for multiple testing problems. It is however well known that for discrete data these procedures may be unnecessarily conservative. Recently, new procedures were developed that remove this conservatism while still guaranteeing FDR control (see [1]). In this talk, we review these new procedures and present an implementation in the R-package 'DiscreteFDR' (see [2]).
Joint work with Etienne Roquain and Guillermo Durand.
[1] Döhler, S., Durand, G. and Roquain, E. (2018). New FDR bounds for discrete and heterogeneous tests, Electronic Journal of Statistics, Volume 12, Number 1 , 1867-1900
[2] Durand, G. and Junge, F. (2018). DiscreteFDR: Multiple Testing Procedures with Adaptation for Discrete Tests. R package version 1.2.

245 recent advances in deep reinforcement learning and the R implementation rlR package

Xudong Sun

LMU Munich, Germany

Deep reinforcement learning has gained great attention in the past years with its rapid advances in solving complex scenarios like Atari Games, the game of Go and continous robotic control, etc.
The idea of using deep neural network as a general function approximator is relatively new to both the Computer Science and Statistics field. This talk aims at briding the statisticians with the recent advances in deep reinforcement learning and shed light to some connections with statistics. For example, the relationship between Probablistic Graphical Models and reinforcement learning, the general Bayesian Inference framework using deep neural network which showed great efficiency in approximating arbitrary posterior, etc.
The talk will cover several deep RL algorithms from year 2015 till 2018 and some of its implementations in the R package rlR. Algorithms like Frozen Target Deep Q learning, Deep Deterministic Policy Gradient, Trust Region Policy Optimization, soft Q learning and Stein Varitional Inference will be covered. API of the rlR package will be introduced and examples will be showed.

246 A workflow for metabolomics using CRAN packages to demonstrate association between a covariate and multiple analytes (some with detection limit)

Ludwig A. Hothorn[1], Torsten Hothorn[2], **Paola G. Ferrario**[3]

[1]Retired from Leibniz University Hannover, Germany
[2]Universität Zürich, Switzerland
[3]Max Rubner-Institut, Karlsruhe, Germany

In metabolomics, to demonstrate an association between one clinical covariate of interest and multiple analytes from diverse metabolomics platforms is a recent problem. (e.g. Bassi et al.).
For this problem, we present a workflow using the CRAN packages multcomp (for mmm), tukeytrend and mlt, share its code and demonstrate it using the KarMeN cross-sectional data (Rist et al.). Specifically, the association between the clinical covariate age for the KarMeN sample and two selected analytes (without and with detection limit) is presented.
Generally, the relationship between a continuous covariate (also grouped in a randomized design) and multiple, differently distributed analytes (including those with observations under a detection limit) can be demonstrated by a Tukey trend test (Tukey et al.) based on previously transformed analytes by means of the most likely transformation concept (Hothorn et al.). The Tukey trend test is sensitive to different forms of association, allows both a global claim (by the smallest adjusted p-value), and also a local claim, i.e. at each individual analyte. It takes into account the correlation between the analytes and the models through the multiple marginal models concept (Pipper et al.). Using the most likely transformation approach, differently distributed analytes, including those with ties and/or censorship, are analyzed on a uniform, comparable scale.
A particularly interesting modification is the interpretation of an optimal odds ratio using continuous outcome logistic regression (Lohse et al.). By this modification, it is possible to consider the association between a continuous covariate and an arbitrarily distributed analyte, independently of certain cut-off for categorization of the covariate. These odds ratios can be evaluated for all potential values or cut-off of the covariate function, which allows the associations for different categorization types.
References:
Bassi R et al. (2017) PLOS ONE 12(1)
Rist M. et al. (2017) PLOS ONE 12(8)
Tukey JW et al. (1985) Biometrics, 41(1):295-301
Hothorn T et al. (2017) Scand J Stat 45(1):110-134
Pipper CB et al. (2012) J Royal Stat Soc 61:315-326
Lohse T et al. (2017) F1000Research 6:19-33

Design of Experiments and Clinical Trials V (Optimal Design II)

247 Optimal designs for frequentist model averaging

Kira Alhorn[1], Kirsten Schorning[2], Holger Dette[2]

[1]Fakultät Statistik, TU Dortmund, Germany
[2]Fakultät für Mathematik, Ruhr-Universität Bochum, Germany

We consider the problem of designing experiments for the estimation of a target parameter in regression analysis if there is uncertainty about the parametric form of the regression function. A new optimality criterion is proposed, which minimizes the asymptotic mean squared error of the frequentist model averaging estimate by the choice of an experimental design. Necessary conditions for the optimal solution of a locally and Bayesian optimal design problem are established. The results are illustrated in several examples and it is demonstrated that Bayesian optimal designs can yield a reduction of the mean squared error of the model averaging estimator up to 45%.

248 Optimal Bayesian design for model discrimination via classification

Markus Hainy[1], David Price[2], Olivier Restif[3], Christopher Drovandi[4]

[1]Johannes Kepler University, Austria
[2]University of Melbourne, Australia
[3]University of Cambridge, UK
[4]Queensland University of Technology, Australia

Performing optimal Bayesian design for discriminating between competing models is computationally intensive as it involves estimating posterior model probabilities for thousands of simulated datasets. This issue is compounded further when the likelihood functions for the rival models are computationally expensive. A new approach using supervised classification methods is developed to perform Bayesian optimal model discrimination design. This approach requires considerably fewer simulations from the candidate models than previous approaches using approximate Bayesian computation. Further, it is easy to assess the performance of the optimal design through the misclassification error rate. 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 apply our method to find optimal designs for discriminating between models in epidemiology and cell biology.

249 The adaptive Wynn-algorithm in generalized linear models with univariate response

Norbert Gaffke

University of Magdeburg, Germany

For a nonlinear regression model the information matrices of designs depend on the parameter of the model. The adaptive Wynn-algorithm for D-optimal design estimates the parameter at each step on the basis of the observed responses and employed design points so far, and selects the next design point as in the classical Wynn-algorithm for D-optimal design. The name 'Wynn-algorithm' is in honor of Henry P. Wynn who established the 'classical' algorithm for linear regression in his 1970 paper (Ann. Math. Statist. 5, 1655-1664). The asymptotics of the sequences of designs generated by the algorithm and of the adaptive maximum likelihood estimators is studied for an important class of nonlinear regression models: generalized linear models whose (univariate) response variables follow a one parameter exponential distribution. Crucial assumptions are the compactness of the experimental region and of the parameter space together with some natural continuity assumptions. The talk is based on joint work with Fritjof Freise (TU Dortmund) and Rainer Schwabe (University of Magdeburg).

250 Locally optimal designs for gamma models

Osama Idais

Otto-von-Guericke- Universität Magdeburg, Germany

The gamma model is a generalized linear model for gamma-distributed observations. We consider a first order gamma model with one or more quantitative factors. Locally complete classes of designs are established. On that basis locally D- and A-optimal designs are derived. The properties of particular optimal designs are determined via their effciencies

Survival and Event History Analysis IV (Competing Risks and Multistate Models I)

251 Consistent estimation in non-Markov multi-state models

Lucas Radloff, Rafael Weißbach
Universität Rostock, Germany

When using multi-state models in event history analysis the Markov assumption is attractive – there is a large amount of theory available and many computations simplify. However, statistical tests show that the Markov assumption is not sustainable in every practical case (see e.g. Lando and Skødeberg (2002)). Datta and Satten (2001) showed that the Aalen-Johansen-Estimator remains consistent in the non-Markov-case and Titman (2015) and Uña-Álvarez and Meira-Machado (2015) describe approaches for nonparametric estimation of transition-probabilities for non-Markovian histories. All of these estimate entities that are conditioned only on the current state of the process. However, for non-Markovian histories case the whole past of the process might affect its future. As a consequence Semi-Markov models have been used (see e.g. Koopman et al. (2008)). We discuss a nonparametric model for transition-intensities that are conditional on the complete past of the process, including the duration of occupancy in the current state.

DATTA, Somnath ; SATTEN, Glen A.: Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. In: Statistics & Probability Letters 55 (2001), No. 4, pp. 403-411
KOOPMAN, Siem Jan ; LUCAS, André ; Monteiro, André: The multi-state latent factor intensity model for credit rating transitions. In: Journal of Econometrics 142 (2008), No. 1, pp. 399-424
LANDO, David ; SKØDEBERG, Torben M.: Analyzing rating transitions and rating drift with continuous observations. In: Journal of Banking & Finance 26 (2002), No. 2-3, pp. 423-444
TITMAN, Andrew C.: Transition probability estimates for non-Markov multi-state models. In: Biometrics 71 (2015), No. 4, pp. 1034-1041
UÑA-ÁLVAREZ, Jacobo de ; MEIRA-MACHADO, Luís: Nonparametric estimation of transition probabilities in the non-Markov illness-death model: A comprehensive study. In: Biometrics 71 (2015), No. 2, pp. 364-375

252 Goodness-of-fit tests for the cure rate in a mixture cure model

Ursula U. Müller[1], Ingrid Van Keilegom[2]

[1]Texas A&M University, United States of America

[2]KU Leuven, Belgium

We consider models for time-to-event data that allow that an event, e.g., a relapse of a disease, never occurs for a certain percentage p of the population, called the cure rate. We suppose that these data are subject to random right censoring and we model the data using a mixture cure model, in which the survival function of the uncured subjects is left unspecified. The aim is to test whether the cure rate p, as a function of the covariates, satisfies a certain parametric model. To do so, we propose a test statistic that is inspired by a goodness-of-fit test for a regression function by Härdle and Mammen. We show that the statistic is asymptotically normally distributed under the null hypothesis that the model is correctly specified and under local alternatives. A bootstrap procedure is proposed to implement the test. The good performance of the approach is confirmed with simulations. For illustration we apply the test to data on the times between first and second birth.

253 Semiparametric Accelerated Failure Times Quantile and Expectile Regression using Auxiliary Likelihoods

Fabian Otto-Sobotka, Alexander Seipp, Verena Jürgens, Antje Timmer

Division of Epidemiology and Biometry, Carl von Ossietzky University Oldenburg, Germany

Survival times are often modelled in parametric Accelerated Failure Times (AFT) regression models. They usually rely on the assumption of a Weibull, lognormal or log-logistic distribution with a constant scaling parameter σ. We often find in classical mean regression setting that these assumptions seldomly hold up in real life data sets. For this case, distributional regression methods were developed that, at first, added a regression predictor to the scaling parameter as well to capture the heteroscedasticity in the data. Alternatively, quantile and expectile regression models allow for a nonparametric estimate of the complete conditional distribution of the response. They do not require any assumptions regarding the distribution that is estimated as they are only defined by their asymmetrically weighted absolute or squared loss functions. Those, however, will only incorporate complete observations and estimation will run into problems once censoring is possible.

We propose to construct auxiliary likelihoods from the least asymmetrically weighted squares loss function for expectiles and the so called check loss function for quantiles. The expectile loss function can be transformed to an asymmetric normal likelihood whereas the quantile loss function will result in the same estimate as the maximisation of an asymmetric laplace likelihood. Both distributions have been used as auxiliary likelihoods to successfully estimate Bayesian quantile and expectile regression models. The definition of a likelihood for our estimates allows us to also define a survivor function that can handle censored observations. Hence, we can analyse censored time-to-event data with nonparametric distributional regression. This also includes the extension to a penalised likelihood which enables very flexible semiparametric predictors of nonlinear or spatial effects via P-splines or Gaussian Markov random fields.

The quality of these estimates is evaluated in a simulation study with a variety of scenarios. We especially want to determine whether the use of auxiliary likelihoods leads to inefficiencies due to a possibly false specification of the response and residual distribution, respectively. We therefore compare our distribution estimates from a rich sequence of quantile or expectile regressions with correctly and falsely specified parametric AFT models. We also add Generalised Additive Models for Location, Scale and Shape (GAMLSS) for censored data to our study as a competitor. Both parametric alternatives can also comprise semiparametric predictors. We then illustrate our approach with an analysis of overall survival times from patients with mutated non-small cell lung cancer.

Statistics in Finance III

254 Goodness-of-fit tests for centralized Wishart processes

Gustav Alfelt, Taras Bodnar, Joanna Tyrcha
Stockholm University, Sweden

In this paper we present several goodness-of-fit tests for the centralized Wishart process, a popular matrix-variate time series model used to capture the stochastic properties of realized covariance matrices. The new test procedures are based on the extended Bartlett decomposition derived from the properties of the Wishart distribution and allows to obtain sets of independently and standard normally distributed random variables under the null hypothesis. Several tests for normality and independence are then applied to these variables in order to support or to reject the underlying assumption of a centralized Wishart process. In order to investigate the influence of estimated parameters on the suggested testing procedures in the finite-sample case, a simulation study is conducted. Finally, the new test methods are applied to real data consisting of realized covariance matrices computed for the returns on six assets traded on the New York Stock Exchange.

255 Matrixvariate Factor Model for Realized Covariances

Eugen Ivanov, Yarema Okhrin
Universität Augsburg, Germany

Modeling and forecasting covariance matrices of asset returns plays an important role in economic theory and finance. The increasing availability and accessibility of high frequency data shifted the focus of research community to realized measures of covariance, which are intraday counterparts of covariance matrices and were shown to have superior properties. Whereas the univariate case saw a multitude of proposed models, modeling realized measures of covariance lacks truly multivariate methods. One standard approach is to model entries of some matrix decomposition, which leads to loss of interpretability and introduces bias. Another approach considers vectorization of matrix, whereas additional measures have to be taken, in order to ensure positive definiteness. These methods are, however, difficult to estimate and work with in higher dimensions. We propose a matrix variate factor model, which preserves interpretability during estimation and properties of covariance matrix for forecasting. Estimation framework based on Expectation Maximisation algorithm (EM), where factors are estimated with Kalman smoother at each iteration, is presented. We further show the asymptotical consistency of factors estimates under true model parameters when the dimension of observable information is large. An application study presents empirical properties of the proposed model.

256 Dynamic Modeling of the Global Minimum Variance Portfolio weights

Fabian Krüger[1], Roman Liesenfeld[2], **Laura Reh**[2]
[1]Heidelberg University, Germany
[2]University of Cologne, Germany

We propose a novel dynamic approach to forecast the weights of the global minimum variance portfolio (GMVP). We exploit the fact that the GMVP weights can be obtained as the population coefficients of a linear regression of one benchmark return on a vector of return differences. This fact enables us to derive a consistent loss function from which we can infer the optimal GMVP weights without imposing any distributional assumptions on the returns. In order to capture time variation in the assets' conditional covariance structure, we model the portfolio weights through Generalized Autoregressive Score type dynamics. Sparse parameterizations ensure scalability with respect to the number of assets. An empirical application to daily returns from 40 NYSE traded stocks shows that the model performs very well in- and out-of-sample in comparison to existing approaches.

257 Detecting a hidden component in high-frequency yield curves using rank tests for the covolatility process

Lars Winkelmann
Freie Universität Berlin, Germany

This paper quantifies the importance of a hidden component in U.S. government bonds. The component is unspanned by (imperfectly correlated with) information about the shape of the yield curve. Using inflation-indexed bonds as an additional source of information, we show that under the conditions of the Mundell-Tobin effect the hidden component is detectable by rank tests on intraday covolatility matrices of the nominal and real yield curves. To determine the rank empirically, we apply tests on the spectral spot-covolatility matrix, directly estimated from nonsynchronous tick-by-tick data. Empirical evidence shows that the hidden component is detectable only on specific intraday time intervals. It explains locally up to 18% of the total variation of the yield curve. The detected intervals coincide with news releases about inflation expectations, hence identify times where markets significantly revise inflation expectations.

Statistics in Science, Technology and Industry III

258 Inference on the Second Moment Structure of High-Dimensional Sensor-Type Data in a K - Sample Setting

Nils Mause, Ansgar Steland
RWTH Aachen University, Germany

In this talk we consider K high-dimensional vector time series $\mathbf{Y}_{T,1}, \ldots, \mathbf{Y}_{T,K}$ generated at independent locations $j = 1, \ldots, K$ of sensors within a time interval $[0, T]$ for some $T > 0$. These sensors may collect and transmit data at different sampling frequencies $\omega_j \in [0, 1]$, such that the resulting sample sizes given by $N_j = \lfloor \omega_j T \rfloor$ may also be different. Projections $\mathbf{w}_T' \mathbf{Y}_{T,j}$ based on such high-dimensional time series with some appropriate weighting vector $\mathbf{w}_T$ appear naturally in many statistical procedures and applications, such as the Principal Component Analysis and Portfolio Optimization, and are a common method of dealing with high-dimensional data sets. Taking the data of all samples into account, bilinear forms based on the pooled sample variance-covariance matrix S_T need to be analyzed in order to draw inference on the variance of these projections.
Within the high-dimensional framework where not only the time horizon T shall go to infinity but also the dimension d_T of the data is allowed to grow with the time horizon we establish a new strong approximation result for bilinear forms and even an increasing amount, L_T, of bilinear forms based on the pooled sample variance-covariance matrix in terms of Brownian motions. These approximation results do not depend on any constraints on the ratio of dimension and sample sizes or the amount of bilinear forms L_T and thus also holds for situations where the dimension d_T grows much faster than the sample sizes N_j and the time horizon T.
As an application of the strong approximation results we will deal with a change in variance problem, where, under the null hypothesis of no change, the data $\mathbf{Y}_{T,j,1}, \ldots, \mathbf{Y}_{T,j,N_j}$ is supposed to form a stationary d_T - dimensional vector time series with mean zero and variance - covariance matrix $\mathbf{\Sigma}_{T,0}^{(j)}$, which can either be known or unknown. Lastly, we conducted a simulation study in order to illustrate the finite sample performance of our change-point test statistic.
Acknowledgement:
This is joint work with Ansgar Steland from the Institute of Statistics, RWTH Aachen University, Germany. It was supported by the German Research Foundation / Deutsche Forschungsgemeinschaft (DFG) [grant number STE 1034/11-1].
References:
Steland, A., von Sachs, R. (2017a). Large-sample approximations for variance - covariance matrices of high-dimensional time series. Bernoulli, 23 (4A), 2299 - 2329.
Steland, A., von Sachs, R. (2017b). Asymptotics for High-Dimensional Covariance Matrices and Quadratic Forms with Applications to the Trace Functional and Shrinkage. Stochastic Processes and Their Applications, to appear.

259 Statistical Process Monitoring to Improve Quality Assurance of Inpatient Care

Lena Hubig[1], Nicholas Lack[2], Ulrich Mansmann[1]

[1]Institute for Medical Information Processing, Biometry, and Epidemiology, Ludwig-Maximilians University, Munich, Germany
[2]Bavarian Institute for Quality Assurance, Munich, Germany

Statistical Process Control (SPC) in hospital benchmarking using control charts is a common instrument for monitoring clinical performance and early detection of quality deficits. The external quality assurance program (EQA) of German hospitals does not yet employ SPC. Previous work has failed to come up with suggestions for efficient application of SPC. There is also a lack of focus on the importance of preventing false positive signals.
In this contribution we study control limits for defined false signal probability and their dependence on specific features such as hospital volume, risk score and patient mix. We also determine the detection quality of specific control switches. We conduct simulation studies in order to investigate optimal designs for crude and risk-adjusted performance indicators of the log-likelihood CUSUM chart of Steiner et al. (Biostatistics 1.4 (2000), pp. 441-52). Examples are taken from the EQA in Bavaria, Germany.
Focusing on signal probability instead of average run length allows control of the false signal probability and performance evaluation of control charts. Thus it was possible to construct CUSUM charts for different hospital volumes and failure probabilities. We gained better understanding of the influence of control switches in constructing CUSUM charts. We also compare our results to run-length based control strategies.
The presented results are useful for regulatory decision making and help to implement CUSUM charts within EQA. We expect application of CUSUM control charts to significantly improve early detection of quality deficits with appropriate adjustment for different case mix and hospital volume.

260 On Steady-state Performance Characteristics of Control Charts – Meaning and Numerics

Sven Knoth

Institute of Mathematics and Statistics, Department of Economics and Social Sciences, Helmut Schmidt University Hamburg, Germany

At first sight, the definition of a reasonable steady-state measure of control chart detection performance seems to be both straightforward and well established. Classical results are provided by Darroch and Seneta (1965), Roberts (1966), Taylor (1968) and Crosier (1986). Moreover, there are some relationships to change point detection delay measures introduced in Pollak and Siegmund (1975) and Pollak (1985), and more recently in Pollak and Tartakovsky (2009). Here, we want to illustrate how careless several steady-state measures are used within SPC literature. Moreover, some useful integral equations (most of them are known) are provided and numerically solved. In particular, univariate and multivariate EWMA control charts are analyzed.

References

CROSIER, R. B. (1986). "A new two-sided cumulative quality control scheme". Technometrics, 28(3), pp. 187–194.

DARROCH, J. N. and SENETA, E. (1965). "On quasi-stationary distributions in absorbing discrete-time finite Markov chains". Journal of Applied Probability, 2(1), pp. 88–100.

POLLAK, M. (1985). "Optimal detection of a change in distribution". Ann. Stat., 13(1), pp. 206–227.

POLLAK, M. and SIEGMUND, D. (1975). "Approximations to the expected sample size of certain sequential tests". Ann. Stat., 3(6), pp. 1267–1282.

POLLAK, M. and TARTAKOVSKY, A. G. (2009). "Optimality properties of the Shiryaev-Roberts procedure". Statistica Sinica, 19(4), pp. 1729–1739.

ROBERTS, S. W. (1966). "A comparison of some control chart procedures". Technometrics, 8(3), pp. 411–430.

TAYLOR, H. M. (1968). "The Economic Design of Cumulative Sum Control Charts". Technometrics, 10(3), pp. 479–488.

Clustering II (Mixture Models)

261 Model-based clustering for cytometry

Jean-Patrick Baudry[1], Gilles Celeux[2]

[1]Sorbonne Université, France

[2]Inria Saclay-Île-de-France, France

High-dimensional flow and mass cytometry allow to measure the expression of several proteins on tens of thousands of immune cells of a patient. A common task is to predict patients disease status. This can be done based on characteristics of the cells clusters of each patient. Hence the need for clustering methods.

Some constraints make this problem challenging. The clusters of cells need be interpretable as biologically meaningful profiles. Also, interesting groups of cells are typically rare populations.

We propose a procedure relying on model-based clustering and merging of clusters. We discuss the choices made to make the clustering procedure fit the biological purpose: model, merging criterion and procedure... and compare our choices and results to those of a challenge-winning procedure in this field.

262 Model-based clustering in very high dimensions via adaptive projections

Bernd Taschler[1], Frank Dondelinger[2], Sach Mukherjee[1]

[1]German Center for Neurodegenerative Diseases

[2]University of Lancaster

We consider model-based clustering in the setting where the dimension p is large relative to sample size n and where either or both of means and covariances/graphical models may differ between the latent groups. In this setting, mixture models face statistical and computational difficulties due to high dimensionality.

Many standard clustering methods, such as K-means, are aimed at detecting mean signals. In contrast, suitably formulated model-based clustering methods can in principle detect both mean and covariance signals (at least in the large sample setting) since they fully model the underlying distributions. However, the higher dimensional setting remains challenging, particularly if covariance signals are of interest and particularly when $p \gg n$.

We propose an approach called Model-based Clustering via Adaptive Projections"(MCAP) that, instead of estimating mixtures in the original space, works in a low-dimensional space obtained by linear projection. The projection dimension plays a crucial role and governs a type of bias-variance tradeoff. If the target dimension q is too small, the relevant signals can be lost. On the other hand, if q is too large, the statistical (and computational) cost increases, rendering analysis ineffective and/or intractable. Therefore, in the proposed MCAP approach, the dimensionality of the projected space is set automatically in a data-adaptive manner using cluster stability as a selection criterion. The mixture modelling itself is done using a full covariance formulation in the lower-dimensional space and this, combined with the adaptive projection, allows detection of both mean and covariance signals in very high dimensional problems.

Real-world high-dimensional data often have rich covariance structure that is not necessarily identical across groups and that cannot be assumed to be diagonal. We show real-data examples in which covariance signals are reliably detected in problems with $p \sim 10^4$ or more. For example, in an analysis of gene expression data, MCAP is able to maintain performance even when the mean signal is entirely removed, leaving differential covariance structure in the high-dimensional space as the only signal. The experimental settings span a range of scenarios where we vary the number of samples (100 to over 2000), the dimensionality (100 to 10^6), the number of true clusters (2 to 10) and the strength of the mean signal.

Across this body of experiments, we find that variants of MCAP (including PCA-based as well as random projections) are able to automatically adapt to the data and match or outperform the best of a panel of existing methods, including, among others, a classical Gaussian mixture model, spectral clustering and a recently-proposed ℓ_1-penalized mixture (the latter specifically formulated for the high-dimensional setting). MCAP is computationally very fast, and scales well, both computationally and statistically, with dimension p.

263 Highly Multimodal Likelihood Functions of Mixture Distributions

Malte Jastrow, Claus Weihs

TU Dortmund University, Germany

When applying mixture distribution models, parameters are usually estimated via maximum likelihood. It is known that likelihood functions of mixture distributions can be highly multimodal and thereby presenting obstacles for optimization algorithms trying to find the global optimum. As pointed out by Hathaway (1985), the common case of mixing two equally weighted univariate normals with known variances leads to a multimodal likelihood function if a considerable difference in the variances of the two components exists.

In this talk, mixture distribution models are analyzed from an optimization point of view to find out which distributions lead to multimodal likelihood functions and how multimodality can be controlled by the mixture models parameters. Example functions are visualized and multimodality as well as the size of the attraction area of the global optimum are measured heuristically. Since convergence to the global optimum of these multimodal functions is not guaranteed for any optimization algorithm, the optimization performance is compared for different techniques including the highly favored expectation maximization (McLachlan and Peel, 2004) as well as a genetic algorithm (CMA-ES, Hansen, 2006) and model based optimization via Kriging (Jones et al., 1998). In this simulation scenario, functions with varying degree of multimodality are considered as well as different numbers of given observations in function generation.

A further aspect, beyond this talk, is to evaluate the functions' usefulness as test functions for black box optimization. Complementing the popular BBOB set consisting of artificial functions (Hansen et al., 2009) by these mixture likelihood functions directly derived from statistical application could be considered.

References:

N. Hansen (2006). The CMA evolution strategy: a comparing review. In Towards a new evolutionary computation (pp. 75-102). Springer.

N. Hansen, et al. (2009). Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. INRIA.

R. J. Hathaway (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. The Annals of Statistics, 13(2), 795-800.

D. R. Jones, M. Schonlau, W. J. Welch (1998). Efficient global optimization of expensive black-box functions. Journal of Global optimization, 13(4), 455-492.

G. McLachlan, D. Peel (2004). Finite mixture models. John Wiley & Sons.

264 Maximum Number of Modes of Gaussian Mixtures

Carlos Améndola[1], Christian Haase[2], Alexander Engström[3]

[1]TU München, Germany
[2]FU Berlin, Germany
[3]Aalto University, Finland

Gaussian mixture models are widely used for clustering. In this context, one way of identifying a potential cluster in the data is by looking at the modes of a resulting mixture density with Gaussian kernel, where the locations of the local maxima are found for example by the mean-shift algorithm. However, the relation between the number of components and the possible number of modes is far from well understood. In particular, it is not known how many modes a mixture of k Gaussians in d dimensions can have. In this talk we give a brief account of this problem's history and its relationship to the still elusive proof of convergence for the mean-shift algorithm. Then, we give improved lower bounds and the first upper bound on the maximum number of modes, provided it is finite.

Classification and Pattern Recognition I

265 Analyzing and Learning from Ranking Data: New Problems and Challenges

Eyke Hüllermeier
Universität Paderborn, Germany

The analysis of ranking data has a long tradition in statistics, and corresponding methods have been used in various fields of application, such as psychology and the social sciences. More recently, applications in information retrieval and machine learning have caused a renewed interest in the analysis of rankings and topics such as learning to rankänd preference learning. This talk provides a snapshot of ranking in the field of machine learning, with a specific focus on new problems and challenges from a statistical point of view. In addition to problems of unsupervised learning on ranking data and different types of ranking tasks in the realm of supervised learning, this also includes recent work on preference learning and ranking in an online setting.

266 Classification with stylized betweenness-relations allowing for regularization with uniform Vapnik-Chervonenkis-guarantees

Georg Schollmeyer

Lmu Munich, Germany

In this contribution, we present a new approach to classification that is guided by a relational perspective on data analysis:
In the spirit of the methodology of subgroup discovery, we search for subgroups of a given "population of data points" that is as pure as possible with respect to the observed class labels of the target variable. Then, a natural classification rule would be to classify every object according to the label of the "most pure" (and "significant" enough) subgroup it belongs to. (Of course, there is a subtle trade of between pureness and "significance"/size of the considered subgroup, which is also related to statistical aspects of the estimation of pureness, but which can be handled with our method of regularization, see below)
In contrast to classical subgroup discovery, which becomes computationally intractable in very high dimensions, in this contribution, we do not focus on subgroups that are directly described by combinations of attribute characteristics. Instead, we define subgroups implicitly in terms of a stylized notion of betweenness, which turns out to be computationally tractable with insights from lattice theory/formal concept analysis and methods of linear optimization.
Given a stylized notion of betweenness, in a first step we look at subgroups that contain with every two points p and q also every data point r, which lies (in a stylized sense) between p and q. Compared to classical methods like nearest neighbors, where one looks at subgroups that contain with every point p also every point r that is closer to a given center c, our approach could be seen as a sort of generalization of the method of nearest neighbors, which does not emphasize aspects of distance, but instead looks at aspects of incidence in the sense of (stylized) betweeness.
Of course, at the latest in high dimensions, the notion of betweenness becomes very sparse if not meaningless. This is the reason for dealing with a concept of stylized betweenness, where one looks at data points that lie "approximately" between other given data points. Concretely, (in a modification of the first step in order to reduce computational burdens) we deal here with stylized star-shaped sets, i.e. subgroups that contain a center point c such that every point that lies approximately between c and another point of the subgroup is also contained in the subgroup.
It turns out that locally (i.e., given a subgroup with center point c) one can "regularize" the stylized betweenness notion in such a way, that one can exactly control the (local) V.C.-dimension. By locally controlling the V.C.-dimension, one can uniformly control the capacity of the whole classifier in terms of the growth function.
All-together, we are thus able to build a classifier with theoretically accessible properties. Beyond this, we can show that this classifier does behave substantially different compared to e.g. nearest neighbor methods for example in the (very rare, but theoretically interesting) situation where nearest neighbor methods (and also e.g. kernel methods) do reveal the phenomena of antilearning.

267 Sensible functional linear discriminant analysis

Lu-Hung Chen[1], Ci-Ren Jiang[2]

[1]National Chung-Hsing University, Taiwan
[2]Academia Sinica, Taiwan

Fisher's linear discriminant analysis (LDA) is extended to both densely recorded functional data and sparsely observed longitudinal data for general c-category classification problems. An efficient approach is proposed to identify the optimal LDA projections in addition to managing the noninvertibility issue of the covariance operator emerging from this extension. To tackle the challenge of projecting sparse data to the LDA directions, a conditional expectation technique is employed. The asymptotic properties of the proposed estimators are investigated and asymptotically perfect classification is shown to be achievable in certain circumstances. The performance of this new approach is further demonstrated with both simulated data and real examples.

268 Talk and Action in the United Nations General Assembly - Vote-buying and the power to induce states to vote against their own preferences

Dennis Hammerschmidt

University of Mannheim, Germany

Countries more often than not say on thing in public speeches but do the other when they vote. The Philippines did so with regard to climate change, Laos in terms of diplomatic dispute settlement, and even the most ardent protectors of the world sea, the Federated States of Micronesia, voted against a resolution to prevent an oil slick. All of these examples occurred during the 71st UN General Assembly and all of them are rather the norm than the exception. Why is this the case? I argue that the diverging preferences expressed in states' votes and speeches is the result of vote-buying. While the exchange of aid for votes is commonly supported, it has not yet been directly tested whether aid actually induces countries to vote against their own preferences. I rely on UN General Debate (UNGD) speeches to overcome the problem of identifying how countries would have voted in the absence of vote-buying. I argue that votes and speeches of primarily aid-dependent countries differ given that they address two different audiences: While votes are used to signal compliance to their donor, UNGD speeches are addressed at a domestic audience, showing that the country advocates for their interests in the international arena. I use several measures from text-as-data to extract ideal points of countries' UNGD speeches and predict their voting behavior in the subsequent UN General Assembly session. If my theory holds, I see a pattern where I am systematically worse in predicting aid-dependent countries that increasingly align their voting behavior vis-a-vis the donor than in predicting the votes of donor countries. To further relate this to vote-buying, I use the results from the confusion matrix of how accurate my predictions were and analyze to what extent foreign aid explains the different patterns between donor and recipient. By providing the long-missing effect of aid on voting behavior, my study thus shows that vote-buying explains the difference in countries' talks and actions.

Design of Experiments and Clinical Trials VI (Optimal Design III)

269 Privacy sets for distance constraints

Eva Benkova[1], Radoslav Harman[2], **Werner G. Müller**[1]
[1]JKU Linz, Austria
[2]Comenius University Bratislava, Slovakia

In computer simulation experiments, which have now become a popular substitute for real experiments, one usually aims to spread out the measurements uniformly across the design space, yielding so-called space-filling designs. Most of the literature on space-filling designs attempts to achieve its aim by optimizing a prescribed objective measuring a degree of space-fillingness (see eg. Pronzato and Müller, 2012). These criteria are sometimes combined with an estimation or prediction oriented criterion. Let us label those as "soft" space-filling methods. In contrast "hard" space-filling methods ensure desirable properties by enforcing constraints on the designs, as for instance provided by privacy sets (see Benková et al. 2016), such that a secondary criterion can be used for optimization. External constraints such as on the design region or else can be incorporated in a similar manner. This talk provides a fresh look on the role of privacy sets for the construction of space-filling designs with new algorithms and new examples. In contrast to the privacy sets considered in our previous work, the new constraints guarantee some minimal distance between any two design points, which spreads out the measurements across the design space in a very natural way.
References:
Eva Benková, Radoslav Harman, and Werner G. Müller. Privacy sets for constrained space-filling. Journal of Statistical Planning and Inference, 171:1-9, 2016.
Luc Pronzato and Werner G. Müller. Design of computer experiments: space filling and beyond. Statistics and Computing, 22(3):681-701, 2012.

270 c- and ϕ_k-optimal designs for a class of nonlinear multiple regression models

Dennis Schmidt
Otto-von-Guericke Universität Magdeburg, Germany

A broad class of nonlinear multiple regression models with an arbitrary number of covariates is considered. This class includes proportional hazards models with both type I and random censoring, the Poisson and the negative binomial model. We analytically determine c-optimal designs for certain vectors $\boldsymbol{c}$, thereby showing that c-optimal designs with singular information matrices may be optimal. In certain cases the c-optimal design is not unique and additional c-optimal designs with regular information matrices exist. For the general ϕ_k-optimality criteria we compute optimal designs for an arbitrary number of covariates and obtain the D- and A-optimal designs as special cases. In a numerical example it is investigated how censoring affects the optimal designs for the different optimality criteria. It is shown that the optimal designs have a much better performance than the optimal designs in the case of ignoring the censoring.

271 Optimal inspection times for lifetime estimation based on interval-censored samples

Nadja Malevich, Christine Müller
TU Dortmund, Germany

In some experiments, continuous lifetime random variables can not be observed steadily but only at some number of predefined points in time (inspection times). The choice of inspection times affects the accuracy of the lifetime estimation. We determine optimal equidistant and optimal non-equidistant inspection times for exponentially distributed lifetime variables. For both cases, we show that there is no loss of information if the number of inspections is converging to infinity. Since optimal equidistant inspection times are easier to calculate and easier to handle in practice, we study the efficiency of optimal equidistant inspection times with respect to optimal non-equidistant inspection times. Moreover, since the optimal inspection times are only locally optimal, we also provide some results concerning maximin efficient designs.

272 Locally D-optimal Designs for Non-linear Models on the k-dimensional Ball

Martin Radloff, Rainer Schwabe
Otto-von-Guericke-Universität Magdeburg, Germany

We want to construct (locally) D-optimal designs for a wide class of linear and non-linear multiple regression models where the design region is a k-dimensional unit ball. This class includes, for example, Poisson regression, negative binomial regression and regression models with censored data and has been already investigated on a one-dimensional bounded interval and with k covariates on a k-dimensional cuboid. The latter case can be split into k marginal (one-dimensional) sub-problems with only one covariate.
In linear multiple regression on the k-dimensional unit ball it is known that the D-optimal design consists of the vertices of an arbitrarily rotated k-dimensional regular simplex, whose vertices lie on the unit sphere.
In our case, by using invariance and equivariance the multidimensional problem can be reduced to a (similar) one-dimensional marginal problem. So we obtain feasible results for generating exact D-optimal designs even for small to moderate sample sizes.

273 Optimal dose-finding for efficacy-safety-models

Eirini Renata Tsirpitzi, Frank Miller

Stockholm University, Sweden

Dose finding is an important part in clinical development of a new drug. The purpose of dose-finding studies is to determine a good dose for future development based on both efficacy and safety. Optimal experimental designs have already been used to determine the design for such kind of studies, however often it is focused on efficacy only. We will consider an efficacy-safety model which is a simplified version of the bivariate Emax model. We will use here the Clinical Utility Index (CUI) concept giving us a desirable balance between efficacy and safety. The dose that maximizes the utility of the patient should be estimated as good as possible in the dose-finding study. This desire will lead us to locally c-optimal designs. By replacing the dose with the logarithm of dose in the bivariate model, an algebraic solution for c-optimal designs has been determined for arbitrary c vectors.

Data Science

274 Robust algorithmics: a foundation for science?!

Joachim M. Buhmann
ETH Zürich, Switzerland

The ALGORITHM is the idiom of modern science, as Bernard Chazelle phrazed it. I like to go a step further in this talk by claiming that algorithmics lays the foundation of modern science. The scientific method of ßystematic observation, measurements, and experiments, as well as the formulation, testing, and modification of hypothesesrequires algorithms for knowledge discovery in complex experimental situations. Algorithms in data science map data spaces to hypothesis classes. Beside running time and memory consumption, such algorithms should be characterized by their sensitivity to the signal in the input and their robustness to input fluctuations. The achievable precision of an algorithm, i.e., the attainable resolution in output space, is determined by its capability to extract predictive information. I will advocate an information theoretic framework for algorithm analysis where an algorithm is characterized as a computational evolution of a posterior distribution on the output space. The method allows us to investigate complex data analysis pipelines as they occur in computational neuroscience and neurology as well as in molecular biology. I will demonstrate this design concept for algorithm validation with a statistical analysis of diffusion tensor imaging data. A theoretical result for sparse minimum bisection yields statistical hints why random combinatorial optimization problems are hard to solve.

275 Dynamic topological analysis of residential mobility

Ulrich Pötter[1], Ingrid Schockaert[2]

[1]Deutsches Jugendinstitut, Germany
[2]Statistics Flanders, Belgium

Our contribution discusses the analysis of a huge data set from the Belgium National Register on individual residential mobility. Place of residence is recorded on a yearly basis for the years from 2009 to 2012 up to the statistical sector. Statistical sectors are regional subdivisions of Belgium below the level of municipalities. There are 20,000 sectors so that sectors have a mean population of about 500. We analyse a random subsample of 100,000 persons living in Belgium.
The analysis is based on a combination of methods from the emerging field of topological data analysis that tackles two obstacles to comprehensive data summary: a) From a geographical perspective, the analysis needs to be able to cover simultaneously the very different scales of residential mobility both within and between cities. b) In order for the analysis to be effective, one needs a good statistical summary of the underlying dynamics of the process.
The capability to identify structure on different scales is one of the main advantages of the by now well established techniques of topological data analysis. On the other hand, dynamic aspects are not well represented within classical topological analysis. To amend the situation, we take as a starting point of our analysis not the usual point clouds and associated complexes, but directed graphs of movements evolving in time. This leads to an analysis based on so called Dowker complexes. Dynamic aspects of mobility can then be summarized by time-dependent topological characteristics of these complexes. We pay special attention to representational aspects emerging from our data set.

276 Dealing with complex patterns in mobile app and wearable device data

Daniela Zöller[1,2], Marlon Claaßen[1], Stefan Lenz[1,2], Martin Treppner[1,2], Harald Binder[1,2]

[1]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

[2]Freiburg Center of Data Analysis and Modelling, Mathematical Institute - Faculty of Mathematics and Physics, University of Freiburg, Freiburg, Germany

Mobile apps and wearable devices offer great opportunities to collect direct and longitudinal measurements in the context of clinical research even after hospital discharge. The technology can, e.g., be used to obtain patient-reported-outcomes in combination with activity data after an orthopedic operation. During the study planning phase, one must keep in mind that different research questions might require different software specifications to avoid biases. For example, a participant might need a reminder to collect patient-reported-outcomes in time, but a high frequency of reminders might have a negative effect on the compliance. To give guidance on the construction of mobile apps for clinical studies, we will present the results of a systematic literature review on the current state-of-the-art. We will identify the most prevalent types of potential research questions and the corresponding software requirements and will focus on the researchers' awareness of potential statistical biases, e.g., due to complex patterns of missing values. Based on this results, we will present simulation designs exhibiting typical features such as a large number of measurement points combined with informative measurement gaps, e.g., due to selective use, which can subsequently be used to support planning of clinical studies using mobile apps. As an example of a potentially relevant analysis method in this context, we will consider deep learning techniques to identify also unexpected data patterns in short time windows. Specifically, deep Boltzmann machines (DBMs) will be used, as these more easily allow for statistical inference, e.g., statistical testing with strict type 1 error control. We will demonstrate the approach using real data and simulated data with a corresponding structure. This exemplary result both shows the utility of the generated data to inform study planning, as well as the utility of the mobile app data for clinical uses.

277 Risk factors with a spike at zero in epigenome-wide association studies

Jochen Kruppa

Institute of Medical Biometrics and Clinical Epidemiology, Charité - Universitätsmedizin Berlin, Berlin

In the bioinformatical analyzes of methylation data in epigenome-wide association studies high amounts of zero's in the independent riskc variable can occur, like smoking package years. Normally, the continuous and independent variable is dichotomized into „non smoking"and „smoking". Therefore, all zero's are combined in one category, „non smoker"and all other contiguous values greater than zero, are assigned into the category „smoking". The dichotomization will cause lost of information. On the other hand, if the risk factor is modeled continuous, many patients are non smokers and therefore a spike at zero occurs. In case of the logistic regression different approaches to modeling a spike at zero in the independent have been proposed [1].
In bioinformatics data is normally analyzed in a pipeline like fashion. Hence, the data is processed in a string of methods, which is run one after one, often available in one software package. Often the focus is set to the preprocessing and data quality. Well known pipelines distinguish between categorical independent variables i.e. group comparisons or continuous independent variables i.e. a linear regression [2, 3]. Sometimes the differentially expression analysis is done internally by the R package limma and its inherent assumption on the independent variable and endpoint. Therefore the case of a spike at zero in a continuous variable is neglected and not correctly modeled.
In my talk I will compare different analysis pipelines like Champ and minfi with a manually done analysis in a simulation study [2, 3]. I will concentrate on the analysis of the differential expression. The aim of the simulation studies is to find the best approach to analyze a dataset with spikes at zero in epigenome-wide association studies. Further I reanalyzed a dataset on smoking and differential methylation.
References
1. Jenkner, C., Lorenz, E., Becher, H. and Sauerbrei, W. (2016). Modeling continuous covariates with a „spike"at zero: Bivariate approaches. Biometrical Journal, 58(4), 783-796.
2. Aryee, M. J., Jaffe, A. E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A. P., Hansen, K. D. and Irizarry, R. A. (2014). Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics, 30(10), 1363-1369.

3. Morris, T. J., Butcher, L. M., Feber, A., Teschendorff, A. E., Chakravarthy, A. R., Wojdacz, T. K. and Beck, S. (2013). ChAMP: 450k chip analysis methylation pipeline. Bioinformatics, 30(3), 428-430.

Epidemiology I (Causal inference methods)

278 Estimating per-protocol effects. Randomized trials analyzed like observational studies

Miguel Hernan

Harvard T.H. Chan School of Public Health, United States of America

Pragmatic trials are designed to guide real-world decisions by patients, clinicians, and other stakeholders. However, the analysis of pragmatic trials often relies exclusively on the intention-to-treat principle and therefore estimates the intention-to-treat effect. This talk will discuss problems that may arise from uncritical reliance on the intention-to-treat effect, which is the effect of assignment to treatment rather the effect of treatment itself, and will outline approaches to estimate the per-protocol effect, a complement to the intention-to-treat effect. Several examples from the literature will be presented.

279 A comparison of sequential and simultaneous Propensity Score matching in a study with three treatment groups

Dirk Enders

InGef - Institute for Applied Health Research Berlin, Germany

Background: Propensity Score (PS) matching is a well-known method to balance covariates between treatment groups in observational studies. While PS matching has been applied extensively in studies with two treatments, applications which simultaneously match three or more treatment groups are sparse [1,2]. Reasons might be a less clear definition of the matching distance and computational complexity. A naïve sequential matching of two treatment groups one after the other might be an easily implementable alternative. We aimed to compare a sequential with a simultaneous matching approach in an exemplary cohort study based on routine health care data.

Methods: The cohort included patients with chronic kidney disease and iron deficiency/anemia, who started one of three iron treatments (oral, low dose i.v. or high dose i.v.) in 2014. The probability for each of the three treatment groups (PS_1, PS_2, PS_3) was estimated by multinomial logistic regression with covariates obtained in the year before the start of treatment. In the sequential approach, we applied conventional PS matching with two treatment groups first, followed by a second PS matching with the remaining treatment group and one of the two matched groups. In the simultaneous approach, the distance according to Wang et al. [1] was calculated for each element in the cartesian product of the three treatment groups. In both approaches, nearest neighbor matching with a caliper of 0.2 times the pooled standard deviation of logit(PS_1), logit(PS_2) and logit(PS_3) was performed. The balance was assessed via the mean of the standardized mean difference (SMD) over all covariates and treatment group comparisons (oral vs. low dose i.v., low vs. high dose i.v., high dose i.v. vs. oral). A simulation based on the example study was performed to compare both approaches for different ratios of patients in the three treatment groups.

Results: The cohort comprised 1,889 patients with oral, 156 patients with low dose i.v. and 49 patients with high dose i.v. iron treatment. The sequential approach resulted in 38-40 matched patients in each treatment group (mean SMD:0.042 - 0.056) depending on the order of the treatment groups in the matching process. In the simultaneous approach 44 patients could be matched in each treatment group (mean SMD: 0.046). The simulation study confirmed the results.

Conclusion: Simultaneous PS matching achieved a higher percentage of matching partners than sequential PS matching while maintaining balance in the covariates. An explanation could be, that the sequential approach selects ideal matching partners for two treatment groups first without considering the distance to the patients in the third treatment group. The sequential matching does not suffer this limitation because all treatment groups are treated equally. Since the storage of the cartesian product requires large working memory, the simultaneous approach of Wang et al. [1] is appealing in studies of moderate size.

References:

Rassen et al. (2013). Matching by propensity score in cohort studies with three treatment groups. Epidemiology 24:401-409.

Wang et al. (2013). Optimal caliper width for propensity score matching of three treatment groups: A Monte Carlo study. PLOS One 8 (12): e81045.

280 Association of Obesity with Health Care Costs: Strengthening the Instrument in Mendelian Randomization Studies

Christoph Kurz, Michael Laxy

Helmholtz Zentrum München, Germany

Overweight and obesity are global public health concerns in terms of economic costs and effect on health. While life style factors and nutrition are important drivers of obesity, there is also a significant genetic component: at least 34% of body mass index (BMI) variation can be explained by genetic loci.
Many previous studies that estimated the association of obesity with health care costs arise from observational studies that are unable to fully account for unmeasured confounders.
Mendelian Randomization (MR) is a concept from genetic epidemiology that uses genetic variants as instrumental variables to overcome some of these limitations in cases where genetic polymorphisms have a well-known effect on modifiable risk factors like BMI. Because genotypes are assigned randomly when passed from parents to offspring, the population genotype distribution is assumed to be unrelated to confounders and can therefore serve as valid instruments. Recent literature suggested that MR can be a powerful tool for economic evaluation, but few studies have yet used it.
One challenge in using genetic instrumental variables (IVs) is that many genetic variants are only weakly associated with the risk factor of interest, which limits the power and precision of a study.
This implies that, even if a set of multiple instruments is valid, the two-stage least squares (2SLS) estimator can still be biased towards the conventional regression estimate.
In this study, we use matching to extract from a single large study with weak genetic instruments a more powerful, smaller study with stronger instruments.
Our findings indicate that obesity increases total medical spending by 1569 Euro. The same model, that doesn't use matching and therefore has weak instruments, only finds additional spending of 373 Euro for obese individuals compared with individuals of normal weight. This demonstrates the importance of having strong instruments in MR studies when using 2SLS. Weak instruments can cause bias, leading to severe underestimation of effects in our case. On the other hand, a MR study with strong instruments may detect more hidden bias than a non-IV analysis, leading to higher estimated effects.
This paper is one of the first that uses an MR approach to estimate the marginal health costs of a prevalent clinical condition. MR offers new opportunities for reliable causal inference in health economic research within the framework of observational research designs. We approach the problem of weak instruments that classically appears in the context of complex chronic diseases. We demonstrate the effect of matching to strengthen the instrument in MR studies and show that strengthening an instrument can have great impact on the marginal effect of interest in economic evaluation studies.

281 Propensity Scores aus hochdimensionalen Routinedaten und das DMP Koronare Herzkrankheit

Roland Weigand
AOK Bayern, Germany

Routinedaten der gesetzlichen Krankenversicherung werden regelmäßig zur Evaluation von Versorgungsprogrammen sowie für Beobachtungsstudien in der Epidemiologie herangezogen. Dabei spielt die Risikobereinigung eine essenzielle Rolle. Häufig finden hier grobe Krankheitsgruppen oder Morbiditätsscores aus dem klinischen Kontext Verwendung, etwa die Morbiditätsgruppen von Charlson oder Elixhauser. Auf die potentiell hochdimensionale Informationsmenge in Form von granularen Diagnosekodierungen und Arzneimittelverschreibungen wird in der Regel nicht zurückgegriffen, auch weil die traditionellen Regressions- oder Propensity-Score Methoden bei mehreren Tausend Confoundern an ihre Grenzen gelangen.
Als Ausweg hat sich das High-dimensional Propensity Score (hdPS) Verfahren von Schneeweiß et al. (2009, Epidemiology 20: 512–522) etabliert, wobei die Confounder in mehreren Stufen selektiert und transformiert und dann für die Propensity-Score-Schätzung verwendet werden. Der vorliegende Beitrag behandelt ein alternatives Framework, bei dem die Propensity Scores ohne Vorselektion und Aggregation mithilfe von Methoden des maschinellen Lernens (ML) wie dem Gradient Boosting geschätzt werden. Wir verwenden endstellige Diagnosecodes nach ICD10 aus der stationären und ambulanten Behandlung sowie endstellige Wirkstoffcodes der Arzneimittelverschreibungen nach der ATC-Klassifikation, um für die beobachtete Morbidität von Treatment- und Kontrollgruppe zu kontrollieren. Über 30.000 mögliche Confounder gehen somit in die Schätzung der Propensity Scores ein.
Grundsätzlich bieten ML-Verfahren mehrere Vorteile im Vergleich zum hdPS Algorithmus: Im Gegensatz zur hdPS-Methode werden auch seltene Diagnose- und Arzneimittelcodes berücksichtigt, falls sie einen deutlichen Effekt auf das Treatment haben. Die Anzahl der Variablen wird automatisch gewählt, so dass weniger Parameter diskretionär gesetzt werden müssen. Während bei der hdPS-Methode endstellige Codes der Performance schaden können, verbessern sie hier die Güte durch eine geeignete Kombination aus Modellselektion und Shrinkage. Zudem können metrische Confounder (wie die Höhe der Dosierung bei den Medikamenten) effizient genutzt werden, während der hdPS-Algorithmus diese ohne Bezug zum Treatment in Gruppen aufteilt.
Die Methoden werden angewandt zur Untersuchung der Wirksamkeit des deutschen Disease Management Programms (DMP) für Koronare Herzkrankheit. Die Datenbasis bilden neuerkrankte Versicherte der AOK Bayern von 2009 bis 2012, der Auswertungszeitraum reicht bis 2016. Zielgröße ist die 4-Jahres Mortalität. Die geschätzte Mortalitätsreduktion im DMP unterscheidet sich mit 4 Prozentpunkten (Elixhauser), 3 Prozentpunkten (hdPS) und 1 Prozentpunkt (ML) deutlich zwischen den Verfahren. Im Vergleich mit den Benchmarks zeigt sich eine viel bessere Prognosekraft der reinen ML-Verfahren, was in scheinbarem Widerspruch zur bestehenden Literatur steht. Auch in Bezug auf die Schätzung der Treatment-Effekte, evaluiert in sogenannten Plasmode-Simulationen, ist das reine ML-Verfahren überlegen. Dies steht ebenfalls im Widerspruch zur bisherigen Literatur und kann mit den dort vergleichsweise restriktiven Simulationssetups begründet werden.
Insgesamt zeigen die Ergebnisse eine verbesserte Risikobereinigung mit hochdimensionalen Routinedaten, insbesondere unter Nutzung reiner ML-Methoden. Unabhängig von der verwendeten Schätzmethode zeigt das DMP für Koronare Herzkrankheit einen günstigen Effekt auf die 4-Jahres-Mortalität.

Marketing and E-Commerce

282 Latent Class Analysis in Marketing: Drawing Inferences for Social Brand Personalities

Friederike Paetz
TU Clausthal, Germany

Nowadays, the classification of consumers into preference-based segments via conjoint analytic methods is well established in marketing theory and practice. Latent Class approaches, which simultaneously determine consumer preferences and assign consumers to certain segments have shown their superiority with regard to the identification of highly distinctive and manageable segments. Especially Latent Class-Multinomial Logit models are frequently applied in practical applications, because their estimation via Maximum Likelihood estimation is quite easy and quick.
The resulting preference-based segments are commonly used as a basis for further marketing-relevant issues. For example, segments could be profiled with demographic or psychographic variables of the associated segment members in order to strengthen the segments' accessibility with promotional efforts.
Currently, we observe the trend that companies enhance their (already established) products with social product attributes - like a Fair Trade label - to attract consumers. Social product attributes are important drivers for purchase decisions in social consumer segments. Hence, an all-encompassing knowledge of social segments is the key for companies to address such segments. Context-specific psychographic variables like a consumer's consciousness-for-fair-consumption have already proven to be good predictors for consumers' willingness-to-pay for social product attributes. However, little is known about the predictive power of more general psychographic variables like a consumer's personality.
Using an empirical conjoint choice study, we investigate the personality of social consumer segments by using the popular Five-Factor approach and draw inferences on price premia for the Fair Trade label attribute within the product category of denim jeans. Furthermore, we link the personality structure of social segments to brand personalities. This enables the derivation of suggestions for suitable brand personalities for companies that plan to enhance their products with social product attributes. Both the determination of price premia for social product enhancements and the promotion of suitable personality traits for social brands strengthen a company's market presence and profitability.

283 Statistical considerations on assessment of responsiveness of sales to salesforce effort: A Japanese pharmaceutical company's example

Toshifumi Sugitani

Astellas Pharma Inc., Japan

In pharmaceutical industry, salesforce activity or detailing, which is the interaction between the salesperson and the physician to inform about new and existing drugs, is the primary source of promotion.
Many companies spend significant amounts of money on the salesforce activity. Therefore, assessment of responsiveness of sales to salesforce effort is very important matter in the pharmaceutical companies to optimize resource allocation.
However, in Japan, there are numerous restrictions regarding the market data that can be used in pharmaceutical firms. For example, it is not possible to obtain physician-level prescription (or sales) data, but only retail-level (i.e., hospital or pharmacy level) sales data is available.
With these limitations in mind, we are trying hard to best extract useful information for marketing decisions from a given data, using statistical techniques such as state space modeling or hierarchical Bayes modeling.
In this talk, we will share some of key elements of statistical considerations from our practice.

284 Modeling price-sensitive demand: An application to continuous pricing

Felix Meyer

Swiss International Airlines, Switzerland

For service industries such as air transport, hotels, and car rentals, pricing drives demand. To optimise the price, firms have to predict real-time customer demand at the micro level and optimise the price. This paper contributes to the field of revenue management by introducing a nonparametric statistical approach to predict price-sensitive demand and its application to continuous pricing.
Continuous pricing lets service companies maximise revenue by exploiting customers' willingness to pay. However, it requires accurate demand estimations, in particular of customers' price-sensitivity. To estimate price-sensitivity, this paper introduces an augmented generalised additive model, which identifies substantial variations in price-sensitivity, exceeds the predictive performance of state-of-the-art alternatives, and controls for price endogeneity. This leads to a demand model with variable price-derivatives enabling continuous pricing.
The proposed approach offers a simple and efficient way to implement continuous pricing by a closed form solution. Our research also highlights the relevance of considering the problem of price endogeneity when estimating price-sensitive demand based on observations that resulted from prior pricing decisions.
By applying the approach to empirical airline ticket data, we demonstrate how continuous pricing is implemented. We document a field study, which shows a revenue increase of 6% on average.

285 Whether, when and which: modelling advanced seat reservations by airline passengers

Shuai Shao

Ludwig-Maximilians-Universität

Motivated by the growing importance of ancillary revenues in the tourism and transport industry, we propose a statistical model for the behaviour of booking ancillaries using the example of airline passengers making Advanced Seat Reservations (ASR). We focus on the questions of whether, when and which seats are selected. To address these questions, we employ a discrete time duration model, combined with a discrete choice model. Both employ unknown smooth covariate effects that are estimated using contemporary P-spline methodology. The model is applied to a large database of bookings on five intercontinental routes. By incorporating random effect terms to account for seat-specific heterogeneity, we find strong evidence of "middle seat avoiding" and "front seat preferring" effects. We also show that the willingness to pay for ASR depends on its price in relation to the ticket price, as well as on the distribution channel. These and other insights allow for product differentiation and variable pricing in ASR for each and every seat. In addition, the statistical model can also be used for other ancillary products, allowing dynamic pricing of ancillary products in general.

Machine Learning I

286 Stability Assessment for Trees and other Supervised Statistical Learning Results

Michel Philipp[1], Thomas Rusch[2], Kurt Hornik[2], Achim Zeileis[3], Carolin Strobl[1]
[1]Universität Zürich, Switzerland
[2]Wirtschaftsuniversität Wien, Austria
[3]Universität Innsbruck, Austria

Classification and regression trees are an example of a statistical learning algorithm that is known to be ünstable", because small changes in the learning data can lead to substantially different trees. Ensemble methods, like random forests, are more stable but lack the interpretability of a single tree. Therefore, from a user's perspective, the question is: When is it OK to interpret a single tree and when should it be considered with caution?
To address this question, we will review they key properties of tree-based methods and present a toolbox of summary statistics and plots for assessing their stability. Furthermore, we will show how these ideas can be generalized to a framework for measuring the stability of supervised statistical learning results in general. All presented methods are freely available in the R-package stablelearner.

287 Surrogate minimal depth as an importance measure for variables in random forests

Stephan Seifert, Sven Gundlach, Silke Szymczak
Institute of Medical Informatics and Statistics, Kiel University, Kiel, Germany

Is has been shown that the machine learning approach random forests [1] can be successfully applied to omics data, such as gene expression data for classification or regression, and to select variables that are important for prediction. However, the complex relationships between predictor variables, in particular with regard to the outcome, make the interpretation of currently applied variable selection techniques difficult.
Here we propose a new variable selection approach called surrogate minimal depth (SMD) that incorporates surrogate variables into the concept of minimal depth (MD) variable importance [2]. Applying SMD to simulated data we show that simulated correlation patterns can be reconstructed and that the increased consideration of variable relationships improves variable selection. Compared with existing state-of-the-art methods and MD, SMD has higher empirical power to identify causal variables and the resulting variable lists are equally stable. In conclusion, SMD is a promising approach to get more insight into the complex interplay of predictor variables and outcome in a high dimensional data setting.
The R package is available under:
https://github.com/StephanSeifert/SurrogateMinimalDepth.
References
1. Breiman, L. (2001) Random Forests. Mach. Learn., 45: 5-32.
2. Ishwaran H., Kogalur U., Gorodeski E., Minn A., Lauer M. (2010) High-Dimensional Variable Selection for Survival Data. J. Am. Stat. Assoc., 105: 205-217.

288 A classification tree for functional data

Annette Möller, Jan Gertheiss

Clausthal University of Technology, Germany

Functional data occur frequently in various fields of application and many standard tools for data analysis already have their functional counterparts tailored to the specific properties of such data. Specifically, there is a growing interest in classification methods that are designed for functional data and allow to utilize information about appearance and shape of the functional observations. Most of the procedures available, however, are rather inspired by "classical" statistical methods than machine learning approaches. Here we propose a novel classification tree designed to deal with functional predictors. Partitioning for a chosen predictor in a specific node of the tree is based on comparing each observational curve in that node to the class-specific mean curves. A metric suitable for functional data is used to measure the "amount of closeness/similarity" of an observed curve to each of the class-specific mean curves. A curve under consideration is assigned to the class to whose mean curve it is closest in terms of the chosen metric. To choose a predictor for the next split in the current node, common node impurity measures such as the misclassification rate or the Gini index are employed. The proposed functional classification tree may be used as a self-contained approach, but also as a base learner for an ensemble learning method such as a random forest or boosting. The performance of the functional classification tree is investigated on real as well as on simulated data and compared to alternative procedures such as a functional logistic regression model.

289 Measuring and optimizing machine learning interpretability

Christoph Molnar, Giuseppe Casalicchio, Bernd Bischl
LMU Munich

Supervised non-linear models such as random forests, gradient boosted trees and neural networks are widely used. However, one of the biggest disadvantages of these models is their lack of interpretability. Two solutions have been proposed: (1) intrinsically interpretable models and (2) post-hoc methods. Intrinsically interpretable models such as decision trees or linear regression models restrict users to that model class and often suffer from a suboptimal predictive performance. Post-hoc methods such as variable importance measures, partial dependence plots or LIME can be applied to any model but their reliability and usefulness depends on the complexity of the original model: The higher the number of variables used, the more verbose the interpretation of the model becomes; Interactions between variables make post-hoc methods like partial dependence plots or linear surrogate models unreliable; The more complex the marginal relationship between a variable and the prediction, the more difficult it becomes to describe it with a simple functional form (e.g. a linear curve).
We propose interpretability measures for arbitrary machine learning models. The higher the interpretability of a machine learning model according to these measures, the better the model can be interpreted using post-hoc interpretability methods. The measures quantify overall model complexity, based on the functional decomposition of the model's prediction function, taking into account the distribution of the data. These interpretability measures can be applied in automatic model-selection procedures (across hyperparameters or model classes) to optimize not only predictive performance but also interpretability.
We show an application of this principle in combination with multi-criteria optimization which leads to a set of models that capture different trade-offs between interpretability and predictive performance. Our proposed approach allows the user to visualize this trade-off and select one of the models.

290 A multiple testing framework for the efficient statistical evaluation of (machine-learned) prediction models

Max Westphal, Werner Brannath
University of Bremen, Germany

Major advances have been made in the past years regarding the utilization of machine learning methods for many areas of application. For instance, end-to-end deep learning techniques have been successfully applied to disease diagnosis and prognosis tasks involving complex and high-dimensional data.
However, despite all justified excitement, there are also several challenges. In particular, overfitting issues and overoptimistic claims concerning the predictive performance are common. This is especially true when acquisition of labeled data is expensive and datasets are thus only of modest size which is often the case in medical research.
The generally recommended safeguard against over-optimism in machine learning is the complete separation of model selection and evaluation by means of data splitting. In previous and our recent work, we investigated novel evaluation strategies where multiple models are evaluated simultaneously on the final test dataset [1, 2]. The necessary correction for multiplicity is based on work of Hothorn et al. (2008) and is applicable for a wide variety of performance measures [3].
In this talk we will showcase our general methodology and results of extensive simulation studies under least favorable and more realistic parameter configurations. The latter arise from training common machine learning algorithms (EN, RPART, SVM, XGB) on artificial prediction tasks. We compare different evaluation strategies regarding operating characteristics like type 1 error rate, statistical power and final model performance.
As a special, more complex case, we consider demonstrating a sufficiently high sensitivity and specificity of a binary classifier at once. This corresponds to the widely applied co-primary endpoint setting in (comparative) phase III diagnostic accuracy studies in medical diagnosis applications. In this regard, our approach extends the usual study design to the simultaneous investigation of several diagnostic procedures to increase statistical power while still controlling the type 1 error rate.
References:
[1] Westphal, Max and Werner Brannath. “Evaluation of multiple prediction models: A novel view on model selection and performance assessment.” Manuscript submitted for publication (2018).
[2] Westphal, Max and Werner Brannath. “A multiple testing framework for diagnostic accuracy studies with co-primary endpoints.” Manuscript in preparation (2018).
[3] Hothorn, Torsten, Frank Bretz, and Peter Westfall. SSimultaneous inference in general parametric models."Biometrical journal 50.3 (2008): 346-363.

Small Sample Statistics

291 Convex optimization methods for identifying predictors when n<p

Malgorzata Bogdan
University of Wroclaw, Poland

Convex optimization methods, like ridge regression, LASSO or SLOPE allow to identify important predictors in the data base when the sample size n is smaller than the number of observations. In this talk we will critically compare these methods and discuss their limitations in terms of the sparsity of the true model generating the data. We will also discuss some known and new solutions relying on thresholding estimates of regression coefficients or adaptive debiasing.
The talk will be illustrated with theoretical results, simulation study in the context of identifying genetic biomarkers for clinical trials and the real data Genome-Wide Association Study.

292 Designing pediatric phase I clinical trials in oncology by borrowing information from trials with adult patients

Dario Zocholl[1,2], Manuel Wiesenfarth[2], Geraldine Rauch[1], Annette Kopp-Schneider[2]

[1]Institute of Biometry and Clinical Epidemiology, Charité-Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, Germany, Berlin, Deutschland

[2]Department of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany

Pediatric phase I trials in oncology are rarely conducted, hence pediatric drug dosing often relies on arbitrary rules of thumb based on the dose for adults. Among others, one important reason for this are difficulties in patient accrual arising from low incidence rates and strong ethical concerns. Although dose-finding methods like the Continual Reassessment Method (CRM, O'Quigley et al., 1990) and other Bayesian approaches with one or two-parameters (Neuenschwander et al., 2008) are designed for dose-finding in small samples of around 30 patients, the requirements for pediatric trials are often even stricter. In many potential applications in oncology, the desired limit is distinctly below 20 pediatric patients which is impossible to achieve with standard dose escalation designs. Novel designs for dose-finding studies that require smaller sample sizes could therefore improve the situation.

We propose a design that uses the results from a preceding adult trial to derive a compressed pediatric dosing scheme which may allow drawing conclusions from trials with 8 to 15 patients. The method further enables borrowing information from the adult trial in form of an informative pediatric prior distribution which expressed prior beliefs about the position of the maximum tolerable dose according to the data from adults. The prior distribution's impact can be adjusted by applying a weighting parameter to adjust for varying degree of similarity between adult and pediatric patients. The degree of similarity can be specified beforehand based on prior knowledge or prior assumptions, or it can be estimated within a Bayesian hierarchical model (Cunanan and Koopmeiners, 2018). The performance of the method is tested in an extensive simulation by simulating adult and subsequent pediatric trials under various realistic settings and comparing common dose finding methods with one and two parameters. Real-world applicability of the design will be discussed.

References:

K M Cunanan and J S Koopmeiners. Hierarchical models for sharing information across populations in phase I dose-escalation studies. Statistical Methods in Medical Research, pages 1–13, May 2017.

B Neuenschwander, M Branson, and T Gsponer. Critical aspects of the Bayesian approach to phase I cancer trials. Statistics in Medicine, 27(13):2420–2439, June 2008.

J O'Quigley, M Pepe, and L Fisher. Continual reassessment method: a practical design for phase 1 clinical trials in cancer:. Biometrics, 46(1):33–48, March 1990.

293 Use of external information in clinical trials: What can be gained in terms of frequentist power?

Annette Kopp-Schneider, Silvia Calderazzo, Manuel Wiesenfarth
DKFZ, Germany

The research is motivated by a basket trial in precision medicine in which adults with a specific molecular tumor profile are treated with targeted therapy and response to therapy is assessed. The population of children with this specific molecular profile is too small to warrant a separate pediatric trial. This motivates the implementation of a pediatric stratum in the adult trial and the setting suggests that information from the adult trial should be used for the pediatric stratum as "historical information".
An overview of different methods for borrowing from historical data has been given by Viele et al. (2014). Several adaptive methods have been proposed that dynamically discount the amount of information borrowed from historical data based on the conformity between the historical and current data. Adaptive power priors represent one of the approaches suited for this situation where the discounting factor can be selected by, e.g., an empirical Bayes approach as suggested by Gravestock et al. (2017). Another approach is the use of robust mixture priors as proposed by, e.g., Schmidli et al. (2014).
However, even in case of dynamic borrowing no power can be gained when strict frequentist type I error control is required. We exemplify this finding in the case of the pediatric arm of an adult trial and a dichotomous outcome for various methods of dynamic borrowing. We discuss that this counter-intuitive limitation is true in any situation in which a uniformly most powerful test exists and show situations for which this applies. Nevertheless, incorporation of prior information can give a rationale for type I error inflation with benefit of a power gain.
References:
Gravestock, I., Held, L.; COMBACTE-Net consortium (2017). Adaptive power priors with empirical Bayes for clinical trials. Pharm Stat. 16(5): 349-360.
Schmidli H., Gsteiger S., Roychoudhury S., O'Hagan A., Spiegelhalter D., Neuenschwander B. (2014) Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics 70(4):1023-1032.
Viele, K., Berry, S., Neuenschwander, B., Amzal, B., Chen, F., Enas, N., Hobbs, B., Ibrahim, J.G., Kinnersley, N., Lindborg, S., Micallef, S., Roychoudhury, S., Thompson, L. (2014) Use of historical control data for assessing treatment effects in clinical trials. Pharmaceutical Statistics 13(1):41–54.

Clustering III (General clustering and classification)

294 Aspects of adaptive density-based cluster analysis

Ingo Steinwart
Universität Stuttgart, Germany

A central, initial task in data science is cluster analysis, where the goal is to find clusters in unlabeled data. One widely accepted definition of clusters has its roots in a paper by Carmichael et al., where clusters are described to be densely populated areas in the input space that are separated by less populated areas.
The mathematical translation of this idea usually assumes that the data is generated by some unknown probability measure that has a density with respect to the Lebesgue measure. Given a threshold level, the clusters are then defined to be the connected components of the density level set. However, choosing this threshold (and possible width parameters of a density estimator), which is left to the user, is a notoriously difficult problem, typically only addressed by heuristics.
In the first part of this talk, I show how a simple algorithm based on a density estimator can find the smallest level for which there are more than one connected component in the level set. For some classical density estimators I further establish rates of convergence and present a simple adaptive approach for selecting the width parameter.
Finally, the relation to DBSCAN and practical considerations for efficient implementations are discussed.

295 K-quantiles clustering

Christian Hennig, Cinzia Viroli, Laura Anderlucci
University of Bologna, Italy

A new cluster analysis method, K-quantiles clustering, is introduced. K-quantiles clustering can be computed by a simple greedy algorithm in the style of the classical Lloyd's algorithm for K-means. It can be applied to large and high-dimensional datasets. It allows for within-cluster skewness and internal variable scaling based on within-cluster variation. Different versions allow for different levels of parsimony and computational efficiency. Although K-quantiles clustering is conceived as nonparametric, it can be connected to a fixed partition model of generalized asymmetric Laplace-distributions. K-quantiles clustering is consistent for its canonical functional, and it is shown that K-quantiles clusters correspond to well separated mixture components in a nonparametric mixture. A simulation study has been carried out that shows in what situations K-quantiles clustering is superior to existing clustering methods, and it is applied to some real data if time allows.

296 Hybrid Image Classification using Captions and Image Features

Adalbert F.X. Wilhelm
Jacobs University, Germany

Methods for finding groups of similar objects in large data sets with the purpose of facilitating data interpretation play an important role in exploratory data analysis. However, classical cluster analysis methods do not scale well with an increased number of objects and/or dimensions. Recent work in the field has focused on designing algorithms that can overcome these difficulties while providing meaningful solutions. We propose a projection-based hierarchical partitioning method inspired by the OptiGrid algorithm. Given a data sample, the present algorithm searches for low-density points (local-minima) in selected low-dimensional projections, and partitions the data by a hyperplane passing through the best split point found. Measures such as iterative implementation, objects and dimensions sampling, and simplified search for projections and local minima, ensure the computational efficiency of the algorithm. A comparative evaluation of the algorithm is presented based on synthetic and reference data. Performance of the algorithm is explicated for some image analysis tasks.

Design of Experiments and Clinical Trials VII (Optimal Design IV)

297 Efficient design for longitudinal, cluster-randomized clinical trials with repeated measures

Florin Vaida, Kristen Hansen, Ming Tai-Seale

University of California San Diego, United States of America

We propose and analyze a novel study design, a cluster-randomized, longitudinal clinical trial with repeated measures. Units grouped within clusters are cluster-randomized to two or more groups. Units are observed longitudinally, at baseline and follow-up visits. Repeated measures for the response of interest (a continuous or binary variable) are obtained at each visit. We show that, counterintuitively, the best allocation schedule has fewer repeated measures at baseline than at follow-up, rather than an even allocation. The optimal degree of imbalance depends on the within-unit correlation, with highest power gains to be made for low within-unit correlation (up to 60% improvement). Robustness of the design is also considered. Statistical analyses are based on analytical derivations and simulation.

Five different analysis strategies are compared, depending on the treatment of the baseline measures: i) ignoring baseline; ii) analyzing change from baseline; iii) controlling for baseline as a covariate; iv) including baseline in a longitudinal analysis; and v) including baseline, and assuming no baseline differences. Approaches iii) and v) are effectively equivalent, and dominate the others.

This design was motivated by and applied to the Open & Ask Study. Open & Ask (NCT ID NCT03385512) is an on-going US large scale multi-center cluster-randomized controlled trial (RCT) assessing the comparative effectiveness of three interventions to improve engagement of healthcare providers (HP) with their patients. Twenty-one clinics (the clusters) are randomized to one of three educational interventions, with an average of 5 HP's within each clinic receiving same intervention. The study is longitudinal, with baseline, follow-up, and long-term follow-up visits. Continuous and binary endpoints of interest are based on patients' evaluation of their HP's interaction during the visit. Each HP sees a number of patients (the repeated measures) at baseline and follow-up. Different patients are seen at different visits. A pilot study (Tai-Seale et al., 2016) showed very low within-HP correlation for the measures of interest. This lead to the modification of the initial design with 25 patients seen at each visit, to a schedule with 10 patients at baseline and 40 patients at follow-up, currently being implemented. Under conditions similar to the pilot study the current design is more powerful.

298 Optimal Designs in Multiple Group Random Coeffcient Regression Models

Maryna Prus

Otto von Guericke University Magdeburg, Germany

Random coeffcients regression (RCR) models have been introduced in biosciences for selection purposes and are nowadays popular in many fields of statistical applications, for example in medical research and pharmacology. Optimal designs for the estimation of population (fixed) parameters are well discussed in the literature (see e. g. [2]). RCR models with known population (mean) parameters were investigated by [1]. [3] proposed solutions for optimal designs for the prediction of individual random parameters in models with unknowns population mean under assumption of the same treatment for all individuals. This talk presents analytical results for optimal designs for the prediction in multiple group RCR models, where different treatments are allowed for different groups.
References
[1] Gladitz J., Pilz J. Construction of optimal designs in random coeffcient regression models // Statistics, 1982, v. 13, p. 371-385.
[2] Entholzner M., Benda N., Schmelter T., Schwabe R. A note on designs for estimating population parameters // Biometrical Letters - Listy Biometryczne, 2005, v. 42, p. 25-41.
[3] Prus M., Schwabe R. Optimal designs for the prediction of individual parameters in hierarchical models // Journal of the Royal Statistical Society: Series B., 2016, v. 78, p. 175-191.

299 Efficient Designs for the estimation of mixed and self carryover effects

Joachim Kunert, Johanna Mielke

TU Dortmund, Germany

Biosimilars are copies of biological medicines that are developed by a competitor after the patent for the originator drug has expired. Extensive clinical trials are required to show therapeutic equivalence between the biosimilar and its reference product before a biosimilar can be sold on the market.
However, is it clear that, after admission, a patient currently being treated with the reference product can change to the biosimilar and vice versa without negative impact? Obviously, there is a risk of carryover effects. To ensure switchability, there should be no differences between the carryover effects of biosimilar and reference product.
Hence, we need designs that can efficiently estimate these carryover effects. The paper derives a bound for the A-criterion for the joint estimation of mixed and self carryover effects. We also determine a class of designs with good efficiencies compared to this bound.

300 Standardized Maximin D- and c-optimal Designs for Poisson Count Data with Gamma Block Effects

Marius Schmidt, Rainer Schwabe

Otto-von-Guericke-Universität Magdeburg, Germany

The Poisson model is frequently used to model 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 as a generalization of the Poisson model.
The information matrix for the Poisson-Gamma model is derived analytically and relations between the information matrices for the Poisson and Poisson-Gamma model are investigated. We consider the case of a single covariate and show that the D-optimality criterion in the Poisson-Gamma model is equivalent to a combined weighted optimality criterion of D-optimality and c-optimality for the slope parameter in the Poisson model. Since the Poisson-Gamma model is nonlinear, the D- and c-optimal designs for the Poisson-Gamma model depend on the unknown parameter vector. To obtain more robust designs regarding parameter misspecification standardized maximin D-optimal designs for a binary and continuous design region are derived, which maximize the worst efficiency with respect to a prespecified parameter set. Moreover, we show that the c-optimality criteria in the Poisson and Poisson-Gamma model coincide. Thus, using the decomposition of the D-optimality criterion, we obtain the standardized maximin c-optimal designs in the Poisson and Poisson-Gamma model as a special case.

Epidemiology II (Enviromental risks)

301 Exposure-lag response associations between lung cancer mortality and radon exposure in German uranium miners

Matthias Aßenmacher[1], Jan Christian Kaiser[2], Ignacio Zaballa[2], Antonio Gasparrini[3], Helmut Küchenhoff[1]

[1]Ludwig-Maximilians-Universität München, Germany
[2]Helmholtz Zentrum München, Germany
[3]London School of Hygiene and Tropical Medicine, United Kingdom

Exposure-lag response associations shed light on the duration of pathogenesis for radiation-induced diseases. To investigate such relations for lung cancer mortality in the German uranium miners of the Wismut company, we apply distributed lag non-linear models (DLNMs) which offer a flexible description of the lagged risk response to protracted radon exposure. Exposure-lag functions are implemented with B-Splines in Cox models of proportional hazards. The DLNM approach yielded good agreement of exposure-lag response surfaces for the German cohort and for the previously studied cohort of American Colorado miners.
A minimum lag of about 2 yr for the onset of risk after first exposure was found for both cohorts. Risk estimates from DLNMs were directly compared with estimates from both complex radio-epidemiological models and biologically-based mechanistic models. For age > 60 yr model predictions converge to similar estimates of the Excess Relative Risk (ERR). But at younger age marked differences appear as DLNMs predict ERR peaks which are not detected by other models. After comparing exposure responses for biological processes in mechanistic risk models with exposure responses for hazard ratios in DLNMs we propose a typical period of 15 yr for radon-related lung carcinogenesis. The period covers the onset of radiation-induced inflammation until cancer death. The DLNM framework provides a view on age-risk patterns supplemental to the standard radio-epidemiological approach and to biologically-based modeling.

302 Independent estimation of risk from smoking and radiation for different histologic lung cancer types using generalized additive models and biologically-based models of carcinogenesis

Noemi Castelletti[1], Kyoji Furukawa[2], Cristoforo Simonetto[1], Helmut Küchenhoff[3], Georgios T. Stathopoulos[4,5], Jan Christian Kaiser[1]

[1]Helmholtz Zentrum München, Deutsches Forschungszentrum für Gesundheit und Umwelt (GmbH), Ludwig-Maximilians Universität München, Germany
[2]Department of Statistics, Radiation Effects Research Foundation (RERF)
[3]Department of Statistics, Ludwig-Maximilians Universität (LMU) Munich
[4]Laboratory for Molecular Respiratory Carcinogenesis, Department of Physiology, Faculty of Medicine
[5]Comprehensive Pneumology Center (CPC) and Institute for Lung Biology and Disease (iLBD), University Hospital, Ludwig-Maximilian University (LMU) and Helmholtz Zentrum München, Member of the German Center for Lung Research (DZL)

Lung cancer is the deadliest cancer world-wide with adenocarcinoma (LADC) and squamous cell carcinoma (SQUAM) as most frequent histological subtypes. Both types are affected strongly by smoking and different fields of ionizing radiation. Mutational spectra of LADC and SQUAM tissue are markedly different and genomic alterations related to smoking have been identified for both types. Less is known about the biological effects of radiation in human carcinogenesis. Lung cancer risks have been estimated with epidemiological studies involving smoking and radiation as carcinogens. Based on the biologically substantiated assumption that the number of mutations is linearly related to the dose, in radiation epidemiology it is standard to model linear risk responses. Possible effects of interaction between smoking and radiation on lung cancer risk are largely unknown. To investigate these effects we apply three different classes of risk models to lung cancer incidence data 1958 – 1999 from the Life Span Study (LSS) of Japanese atomic bomb survivors. Full dosimetric information from a mixed field of photons and neutrons is available for the LSS and missing information for smoking behaviour has been completed by imputation. To obtain a comprehensive characterisation of the age-risk patterns we apply both biologically-based models of carcinogenesis and statistical generalised additive models (GAMs) which are rarely applied in radio-epidemiology. To put the results from these models into perspective, risk estimates have also been calculated using state-of-the-art radio-epidemiological models.
The mechanistic model for LADC is based on two independent molecular pathways to cancer which are influenced by either smoking or radiation. The model design is motivated by molecular analysis which revealed two genomic groups of transducer genes related to smoking and receptor genes related to radiation. Both GAMs and biologically based models support the hypothesis that smoking and radiation do not interact in the pathogenesis of LADC.

The vast majority of SQUAMs are caused by smoking so that a single pathway in the mechanistic model was sufficient to describe the incidence data. Interestingly, both GAMs and mechanistic models do not find a statistically significant risk response to (mostly) photon radiation.
Statistical risk modeling produced no evidence for interaction between radiation and smoking in LADC pathogenesis. These findings are supported independent statistical investigations using GAMs and biologically-based risk models, and by molecular analysis of LADC cancer tissue. SQUAM showed a strong risk response for smoking but the radiation risk was too small to achieve statistical significance.
Our analysis suggests that radiation risks from conventional analysis for all histologic types combined are difficult to interpret. To improve recommendations in radiation protection we strongly recommend to report type-specific radiation risks of lung cancer.

303 Application of weighted risk scores to estimate the relative contribution of environmental and genetic factors to skin aging

Claudia Wigmann[1], Anke Hüls[1,2], Jean Krutmann[1], Tamara Schikowski[1]

[1]IUF - Leibniz Research Institute for Environmental Medicine, Düsseldorf, Germany
[2]Department of Human Genetics, Emory University, Atlanta, Georgia, USA

There is substantial evidence that skin aging results from several environmental factors like solar radiation, air pollution, smoking, but also nutritional traits. In addition, this process is influenced by genetic factors. The extent, to which each factor contributes to skin aging, is currently unknown. We use data derived from the SALIA study cohort, in which air pollution as well as nutritional factors were previously shown to be associated with skin aging.
Risk factors are often assessed by several, in part highly correlated predictors. We use weighted risk scores, which can assess combined effects of several such predictors, and the Lindeman-Merenda-Gold measure of relative importance to quantify the contributions of various risk factors to the outcome. We determine the risk score weights via elastic net regression in a training dataset split off the sample data. The relative contributions to skin aging are estimated in the test dataset by their proportion of explained variance in a linear regression.
All risk scores combined explain 27% of the variance of pigment spots and 24% of wrinkles. The highest contribution to both phenotypes is achieved by the genetic risk score comprising the considered SNPs, while about 15% are attributable to environmental and lifestyle factors.
Our approach enables to quantify contributions of various risk factors, which have been assessed by several correlated variables, to a certain outcome. For skin aging, we are able to explain one fourth of the outcomes' variances. However, further studies are needed to improve the selection of relevant predictors.

Machine Learning II

304 Low-rank estimation with Missing Non At Random data

Aude Sportisse[1,3], Claire Boyer[1,2], Julie Josse[3]

[1]Laboratoire de Probabilités Statistique et Modélisation, Sorbonne Université, France
[2]Département de Mathématiques et applications, Ecole Normale Supérieure, Paris, France
[3]Centre de Mathématiques Appliquées, Ecole Polytechnique, France

Missing values challenge data analysis because many supervised and unsupervised learning methods cannot be directly applied to incomplete data. Matrix completion based on low-rank models has recently become very popular since it ensures algorithmic, methodological and theoretical guarantees to deal with missing values. However, most of these existing methods only consider the case of missing completely at random or missing at random values, which is not appropriate for many cases.
We propose therefore matrix completion methods based on the low-rank assumption with informative missing values, namely Missing Not At Random (MNAR) data. A classical example is when the patient's health indicator cannot be measured because of his or her condition.
Our first contribution is to suggest a model-based estimation strategy by modeling the missing mechanism distribution. An EM algorithm is then implemented: we propose to solve the M-step including a nuclear-norm penalization by using a Fast Iterative Soft-Thresholding Algorithm (FISTA).
Our second contribution is to suggest an efficient surrogate estimation by implicitly considering the joint distribution of the data and the missing mechanism: the data matrix is concatenated with the indicator matrix coding for the missing values; a low-rank structure is assumed on this new matrix, for implicitly encoding links between variables and missing mechanisms. Finally, this problem is proposed to be numerically solved using FISTA algorithm.
This work is motivated by a collaboration with the Traumabase group of the Paris Hospital, with the analysis of a large incomplete register for polytrauma patients.

305 Scaled Expected Improvement for Bayesian Optimization

Umberto Noè[1], Dirk Husmeier[2]

[1]Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), Bonn, Germany
[2]School of Mathematics and Statistics, University of Glasgow, UK

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).

306 Infill Criterion for Multimodal Model-Based Optimisation

Dirk Surmann[1], Claus Weihs[2], Uwe Ligges[2]

[1]camLine, Germany
[2]Faculty of Statistics, TU Dortmund, Germany

Physical systems are modelled and investigated within simulation software in an increasing range of applications. In reality an investigation of the system is often performed by empirical test scenarios which are related to typical situations. Our aim is to derive a method which generates diverse test scenarios each representing a challenging situation for the corresponding physical system. From a mathematical point of view challenging test scenarios correspond to local optima. Hence, we focus to identify all local optima within mathematical functions. Due to the fact that simulation runs are usually expensive we use the model-based optimisation approach with its well-known representative efficient global optimisation. We derive an infill criterion which focuses on the identification of local optima. The criterion is checked via fifteen different artificial functions in a computer experiment. Our new infill criterion performs better in identifying local optima compared to the expected improvement infill criterion and Latin Hypercube Samples.

307 Stable Feature Selection

Andrea Bommert, Jörg Rahnenführer, Michel Lang
TU Dortmund, Germany

Feature selection is a key part of data analysis. Especially for high-dimensional datasets, it is necessary to filter out the irrelevant and redundant features by choosing a suitable subset. The feature selection should be stable, i.e. robust with respect to small changes in the data. An unstable feature selection would question the reliability of the interpretability of the results.
We consider statistical learning methods which perform embedded feature selection. The tuning of the hyperparameters of such statistical learning methods is almost always done only with respect to predictive accuracy. We investigate the influence of performing the tuning with respect to both predictive accuracy and stability of the feature selection, both on simulated and on real data. We compare stability measures that take into account similarities between the features (e.g. correlations) and measures that do not.

Network Analysis I

308 Inference for Social Network Models from Egocentrically-Sampled Data

Pavel N. Krivitsky[1], Martina Morris[2]
[1]University of Wollongong, Australia
[2]University of Washington, USA

Egocentric network sampling observes the network of interest from the point of view of a set of sampled actors, who provide information about themselves and anonymised information on their network neighbours. In survey research, this is often the most practical, and sometimes the only, way to observe certain classes of networks, with the sexual networks that underlie HIV transmission being the archetypal case. Although methods exist for recovering some descriptive network features, there is no rigorous and practical statistical foundation for estimation and inference for network models from such data. We identify a subclass of exponential-family random graph models (ERGMs) amenable to being estimated from egocentrically sampled network data, and apply pseudo-maximum-likelihood estimation to do so and to rigorously quantify the uncertainty of the estimates. For ERGMs parametrised to be invariant to network size, we describe a computationally tractable approach to this problem. We use this methodology to help understand persistent racial disparities in HIV prevalence in the US. Lastly, we discuss how questionnaire design affects what questions can and cannot be answered with this analysis.

309 Nonparametric inference in the dynamic stochastic block model

Ann-Kristin Becker[1], **Hajo Holzmann**[2]
[1]University Medicine Greifswald
[2]Uni Marburg, Germany

We show nonparametric identification of the parameters in the dynamic stochastic block model as recently introduced in Matias and Miele (2017, JRSS B) in case of binary, finitely weighted and general edge states. We formulate conditions on the true parameters which guarantee actual point identification instead of mere generic identification, and which also lead to novel conclusions in the static case. In particular, our results justify in terms of identification the applications in Matias and Miele (2017, JRSS B) to finitely weighted edges with three edge states. We also deal with extensions to inhomogeneous, nonstationary Markov chains. Further, we make the required changes to the variational EM algorithm for the inhomogeneous case, and give numerical illustrations in simulation settings covered by our identification analysis.

310 Bayesian and Spline based Approaches for (EM based) Graphon Estimation

Goeran Kauermann, **Benjamin Sischka**

Ludwig-Maximilians-Universität München, Germany

We propose a new method for graphon estimation using the principles of the EM algorithm. The approach considers both, variability with respect to the latent variables within [0,1] relating to the nodes and estimation of the unique representation of a graphon. To do so (linear) B-splines are used, which allows to easily accommodate constraints in the estimation routine so that essential features of the estimated graphon are achieved. The graphon estimate itself allows to apply Bayesian ideas to explore the plausibility of repositioning the latent variables. Variability and uncertainty is taken into account using MCMC techniques. Combining both steps yields an EM based approach for graphon estimation. Additionally, some quantile adjustments can be incorporated into the graphon estimation step in the form of compressing and stretching with respect to the latent variables. This allows additional flexibility in the gradients and thus enables also the estimation of community structures and even mixed structures, meaning a fluent shifting within the community structure.

Robust and Nonparametric Statistics III

311 Real-time detection of sudden location changes in time series with a time-varying trend

Sermad Abbas, Roland Fried
TU Dortmund, Germany

The real-time detection of sudden level shifts in the signal of a time series is a frequently arising problem in monitoring applications. When a restriction of the false-alarm frequency is required, control charts are typically used for this task.
Ordinary control charts are not suitable for signals which exhibit natural systematic variations in the form of periodic fluctuations or trends. Then, using a single target value and fixed control limits may not be reasonable. Moreover, outlier patches caused by measurement errors or irrelevant singular events may be confused with relevant location changes.
We investigate control charts that are based on applying nonparametric two-sample location tests in a moving time window. The most recent observations are compared to their immediate predecessors in the time series. Local regression allows for the elimination of disturbing in-control structures, e.g. trends, so that we can search for shifts in the sequence of one-step-ahead forecast errors.
We present results on the in- and out-of-control properties of these charts. The Wilcoxon-Mann-Whitney test leads to a control chart with a distribution-free in-control average run length for locally constant signals. A test based on the two-sample Hodges-Lehmann estimator provides a chart that is approximately distribution free and shows a more robust behaviour for locally constant and locally linear signals. Additionally, we also study the ability to use this principle on autoregressive processes with a slowly varying mean and show some results on real-world data.

312 Paradoxical Results with Ranks for Unequal Sample Sizes

Edgar Brunner
Universität Göttingen / UMG, Germany

If rank methods are used for $d > 2$ samples $X_{ik} \sim F_i,\ i = 1, \ldots, d;\ k = 1, \ldots, n_i$, then paradoxical results may be obtained in case of unequal sample sizes since the quantities $p_i = \int H dF_i$, on which rank procedures are based depend on sample sizes through the definition of the weighted mean distribution function $H = \frac{1}{N} \sum_{i=1}^{d} n_i F_i$ in the experiment. This undesirable property applies to the well-known Kruskal-Wallis (1952) test, the tests for trend by Hettmansperger-Norton (1987) and by Jonckheere-Terpstra (1954) test, as well as for the Akritas-Arnold-Brunner procedures (Akritas et al, 1997) in factorial designs. For example, in two-way layouts, interactions may appear or disappear just by changing the ratios of the samples sizes while keeping fixed the underlying distributions. Such designs appear in the so-called sub-group analysis in clinical trials or in clinical epidemiology.
This problem can be solved by using the unweighted mean distribution $G = \frac{1}{d} \sum_{i=1}^{d} F_i$ of the distributions $F_1, \ldots, F_d$ in the experiment. Thus, the quantities $\psi_i = \int G dF_i$ are fixed model quantities. They can be estimated by the simple plug-in estimator $\widehat{\psi}_i = \int \widehat{G} d\widehat{F}_i$ which can be represented by the so-called pseudo-ranks R_{ik}^{ψ} of the observations X_{ik} as

$$\widehat{\psi}_i = \frac{1}{N} \left(\overline{R}_{i\cdot}^{\psi} - \frac{1}{2} \right)$$

, where $\overline{R}_{i\cdot}^{\psi} = \frac{1}{n_i} \sum_{k=1}^{n_i} R_{ik}^{\psi}$ denotes the mean of the pseudo-ranks of R_{ik}^{ψ} (Kulle, 1999), Gao et al., 2005), and Thangavelu et al. (2007). The pseudo-rank estimators $\widehat{\psi}_i$ have similar properties (unbiased, consistent, asymptotically normal) as the usual rank estimators $\widehat{p}_i$. The details are to be found in Brunner et al. (2018). The R-package rankFD performing the computations in general factorial designs can be downloaded from CRAN.
References
Akritas, M. G., Arnold, S. F. and Brunner, E. (1997). Nonparametric hypotheses and rank statistics for unbalanced factorial designs. Journal of the American Statistical Association 92, 258-265. Brunner, E, Bathke, A.C., and Konietschke, F. (2018). Rank- and Pseudo-Rank Procedures for Independent Observations in Factorial Designs, Springer Series in Statistics.
Gao, X. and Alvo, M. (2005). A Unified Nonparametric Approach for Unbalanced Factorial Designs. Journal of the American Statistical Association, 100, 926-941. Hettmansperger, T.P. and Norton R.M. (1987). Tests for patterned alternatives in k-sample problems. Journal of the American Statistical Association, 82, 292-299.
Jonckheere, A.R. (1954). A distribution-free k-sample test against ordered alternatives. Biometrika 41, 133-145.
Kruskal, W.H. and Wallis, W.A. (1952). The use of ranks in one-criterion variance analysis. Journal of the American Statistical Association 47, 583-621. Kulle, B. (1999). Nichtparametrisches Behrens-Fisher-Problem im Mehrstichprobenfall. Diploma Thesis, Inst. of Math. Stochastics, University of Göttingen.
Thangavelu, K. and Brunner, E. (2007). Wilcoxon Mann-Whitney test for stratified samples and Efron's paradox dice. Journal of Statistical Planning and Inference 137, 720-737.

313 On robust two-way MANOVA tests with applications

Bernhard Spangl

Universität für Bodenkultur, Wien, Austria

We propose robust tests as alternative to the classical Wilk's lambda test in two-way MANOVA with interactions. Based on work of Todorov and Filzmoser (Todorov, V., and Filzmoser, P., 2010. Robust statistic for the one-way MANOVA. Computational Statistics & Data Analysis, 54, 37-48.) and Van Aelst and Willems (Van Aelst, S., and Willems, G., 2011. Robust and Efficient One-Way MANOVA Tests. Journal of the American Statistical Association, 106, 706-718.) we extend the proposed test statistics for the one-way MANOVA to the two-way case. Monte Carlo simulations are used to investigate the power of the new tests, as well as their robustness against outliers. We compare their performance with the classical MANOVA and other nonparametric test procedures proposed by Konietschke et al. (Konietschke, F., Bathke, A.C., Harrar, S.W., and Pauly, M., 2015. Parametric and nonparametric bootstrap methods for general ANOVA. Journal of Multivariate Analysis, 140, 291-301.) and Bathke and Harrar (Bathke, A.C., and Harrar, S.W., 2016. Rank-Based Inference for Multivariate Data in Factorial Designs. In: Liu, R.Y., and McKean, J.W. (Eds.). Robust Rank-Based and Nonparametric Methods, Springer, New York, 121-139.). Finally, we illustrate the use of these robust test statistics on a real data example.

314 Generalized sign tests: From asymptotics to efficient computation

Kevin Leckey, Christine H. Müller, Dennis Malcherczyk

TU Dortmund, Germany

The K-sign-depth, $K \geq 2$, is a family of robust residual based test statistics. While the 2-sign-depth is essentially equivalent to the classical sign test, other choices for K lead to more powerful test statistics. Two major drawbacks of the K-sign-depth are the lack of limit laws (and thus asymptotic quantiles) for $K \geq 4$ and the fairly high computational effort of $\Theta(n^K)$ for residuals of length n.
We will discuss a solution to both problems for $K = 3$ and $K = 4$. The proof strategy for the limit theorems is based on an approach by Kustosz, Leucht and Müller (2016) dealing with the case $K = 3$. However, a significant part of the proof has been simplified by an application of Donsker's invariance principle which made a generalization to other values for K easier. Fortunately, the calculations in the proof also lead to equivalent formulae for the 3- and 4-sign-depth that can be computed in linear time.

Survival and Event History Analysis V (Competing Risks and Multistate Models II)

315 State transition modeling of complex monitored health data

Jörn Schulz

University of Stavanger, Norway

An innovative application of a non-parametric state intensity regression method to monitored health data is presented. The collection of such data is increasing constantly due to new and the growing availability of sensors. In clinical settings the health status is often tightly monitored as for example in an intensive care unit. In everyday life, smart watches and other devices are becoming increasingly popular. Often, one or several signals are reflecting the current health status that can be represented by a finite number of states, in addition to a set of covariates. State intensity regression allows to study the time dependent effects of covariates on the state transition intensities. Given a state space S and states $i, j \in S$, the conditional state transition intensity $\lambda^{ij}(t)$ of a transition $i \to j$ for an individual at time t can be written by

$$\lambda^{ij}(t) = \beta_0^{ij}(t) + \beta_1^{ij}(t)X_1(t) + \ldots + \beta_K^{ij}(t)X_K(t)$$

where $X_k(t), k = 1, \ldots, K$ are the covariates that might influence the intensity, and $\beta_k^{ij}(t)$ are the regression parameters indicating the effect of the covariates $X_k(t)$ and $\beta_0^{ij}(t)$ is the baseline intensity. The method can handle baseline, time varying as well as dynamic covariates. The covariates can be aggregated by a suitable functional form over a time history window. Inspired by scale spaces, we propose to study the estimated cumulative regression parameters for different lengths of the time history window simultaneously instead of one. The proposed framework is applied to resuscitation data of newborns collected in Tanzania as a part of the Safer Births project.

316 Resampling complex time-to-event data without individual patient data, with a view towards time-dependent exposures

Tobias Bluhmki[1], Hein Putter[2], Arthur Allignol[3], Jan Beyersmann on behalf of the COMBACTE-MAGNET consortium[1]

[1]Ulm University, Germany
[2]Leiden University Medical Center, The Netherlands
[3]Merck KGaA, Darmstadt, Germany

A renewed interest in flexible algorithms to simulate biologically plausible time-to-event data in the presence of time-dependent exposures has recently been observed. We propose non- and semi-parametric resampling of multistate event histories by generating multistate trajectories from an empirical multivariate hazard measure. This technique has been described in detail for competing risks [1] and briefly for general multistate models in the context of prediction in reduced rank Cox models [2]. The advantages of an empirical perspective are diverse: (i) It allows for mimicking complex real-world time-to-event data and (ii) provides for a convenient resampling technique that may be based on published information not necessarily requiring individual patient data. This attractive for, e.g., study planning. (iii) The concept extends to left-truncation and right-censoring mechanisms, non-degenerate initial distributions, non-proportional and non-Markov settings.

A special focus of the talk is on its connection to survival data generation in the presence of time-dependent exposures. Frequently used techniques simulate the covariate processes a priori and subsequently draw the event time from conditional survival distributions. Even though we demonstrate that this procedure leads to the correct data structure, a multistate perspective gives a more natural interpretation of how such data evolve over the course of time by avoiding hypothetical transition times such as disease progression after death and leading to population-level quantities. A simulation study investigating the effect of liver functionality on survival in patients with liver cirrhosis serves as a proof of concept.

References

[1] A. Allignol, M. Schumacher, C. Wanner, C. Drechsler, J. Beyersmann, Understanding competing risks: a simulation point of view, BMC Medical Research Methodology 11 (2011).

[2] M. Fiocco, H. Putter, H. C. van Houwelingen, Reduced-rank proportional hazards regression and simulation-based prediction for multi-state models, Statistics in Medicine 27 (2008) 4340–4358.

317 Recurrent neural networks for time to event predictions with competing risks

Marvin N Wright[1,2,4], Laust H. Mortensen[2], Sasmita Kusumastuti[3], Rudi G.J. Westendorp[3], Thomas A. Gerds[4]

[1]Leibniz Institute for Prevention Research & Epidemiology – BIPS, Bremen, Germany
[2]Statistics Denmark, Copenhagen, Denmark
[3]Section of Epidemiology, Department of Public Health and Center for Healthy Aging, University of Copenhagen, Denmark
[4]Section of Biostatistics, Department of Public Health, University of Copenhagen, Denmark

Everyone who reaches the age of 80 in Denmark is offered annual preventive home visits from their municipality of residence. For residents aged 65-79, preventive home visits are by law to be offered to persons "particularly at risk" of declining health, but there is no established strategy yet on how to accurately identify the high risk group. We aim at identifying older persons at high risk of declining health by predicting who will eventually use chronic care services provided by the municipality. We build machine learning models on register data provided by Statistics Denmark to predict the time to the first use of these care services. This data includes personal information such as age, sex, education and family status, income and pension data as well as the person's medical history, obtained over several years. The challenges for the prediction modeling are the competing risk of death and how to incorporate the longitudinal information on the registered life course data.

We use recurrent neural networks to build models that predict the t-year risk of outcome in the presence of competing risks. The networks are built from layers shared between causes and cause-specific sub-networks. The network outputs cause-specific non-conditional event probabilities, which allows simple calculation of cumulative incidence functions. As loss function, we use the likelihood of parametric inference for the cumulative incidence functions. We implemented our proposed method in the R package survnet. The implementation is based on the deep learning framework keras and thus capable of efficiently handling very large datasets. As a proof of concept, we compare our proposed method with the cause-specific Cox model, random survival forests for competing risks and standard artificial neural networks in a simulation study. The new method performs favorably to the alternatives in terms of calibration and discrimination. On the Danish register data, we show that the method is able to accurately predict the use of chronic care services. Thus, we can identify persons at high risk of declining health, who might benefit from preventive home visits.

318 Methodological aspects in the analysis of adverse events in time-to-event data

Regina Stegherr[1], Jan Beyersmann[1], Claudia Schmoor[2], Michael Luebbert[3], Tim Friede[4]
[1]Institute of Statistics Ulm University, Germany
[2]Clinical Trials Unit, Faculty of Medicine and Medical Center, University of Freiburg, Germany
[3]Hematology, Oncology, and Stem-Cell Transplantation, Faculty of Medicine and Medical Center, University of Freiburg, Germany
[4]Department of Medical Statistics, University Medical Center Göttingen, Germany

Safety analyses in terms of adverse events (AEs) are a very relevant aspect of benefit-risk assessment of therapies. As compared to efficacy analyses AE analyses are often rather simplistic [1]. The probability of an AE of a specific type is most often estimated by the incidence proportion given by the number of patients experiencing the AE out of all patients in the respective population. The incidence proportion underestimates the AE probability in case of censoring. The incidence density estimates the AE hazard under the parametric constant hazard assumption and accounts for censoring, but it does not estimate the AE probability [2]. Additionally to censoring, competing events such as death or premature discontinuation of participation in the study can occur preventing the observation of the AE. In the case of competing events, a parametric estimation based on the incidence density or a non-parametric Kaplan-Meier estimation leads to an overestimation of the AE probability. In the presence of censoring and competing events, an unbiased estimator of the AE probability is the non-parametric Aalen-Johansen estimator. Under the assumption of constant hazards for the AE and for the competing event, a parametric estimator of the AE probability can be constructed from these two hazards.
The aim of this work is to investigate the relative importance of these three sources of bias, namely censoring, competing events, and model misspecifications. The different methods for estimation of the AE probability will be compared to the gold standard Aalen-Johansen estimator. As the estimators will have large variances at the end of follow-up, we compare them not only at the maximal event time but also at the 60 and 90 percent quantiles of the observed times.
Furthermore, the impact of using different estimators on group comparisons is unclear. Therefore, we also compute estimators of the difference and ratio of AE probabilities, and estimators of the hazard ratio based on the Cox proportional hazards model, based on the Nelson-Aalen estimators of the cumulative AE hazard, and based on incidence densities. All of these comparisons are conducted in a real data example and in a simulation study considering constant and non-constant hazards, different censoring mechanisms and different frequencies of observed AEs and competing events.
References:
1. Unkel, S., Amiri, M., Benda, N. et al. (2018) On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. Pharmaceutical Statistics; https://doi.org/10.1002/pst.1915; early view.
2. Allignol, A., Beyersmann, J., Schmoor, C. (2016) Statistical issues in the analysis of adverse events in time-to-event data. Pharmaceutical Statistics, 15(4):297–305.

Time Series Analysis V

319 Analyzing Different Facets of Forecast Quality through Decompositions of Loss Functions

Marc-Oliver Pohle
Goethe University Frankfurt, Germany

Sound decision-making in a wide variety of situations needs to be based on high-quality forecasts of economic variables. But what does quality mean with regard to economic forecasts? In the econometric forecast evaluation literature, the focus has been placed on measuring the distance between forecasts and realizations via loss functions. Besides that, a large literature on forecast optimality testing has emerged, which aims to assess the optimal use of the given information.
We propose to augment this approach by using some measures of forecast quality and the so-called Murphy decomposition that are well known in the meteorolocical literature. The Murphy decomposition, which we discuss for a rather general type of loss function, namely consistent scoring functions, expresses the scoring function as the sum of an uncertainty component of the observations - which can be regarded as a generalized variance or entropy, the (mis-)calibration and the resolution. Miscalibration is closely related to the optimality of the forecasts and measures a generalized conditional bias, while resolution measures the ability of the forecasts to discriminate between outcome values, in other words their informational content. The decomposition illuminates the relationship between the distance-measure and the optimality-testing approach to forecast evaluation by adding a measure of the informational content to the equation. It shows that both the informational content and the optimal use of information determine the size of the forecast errors and reveals which of the two is the driving force. Thus, it can be crucial in understanding the sources of forecast errors and consequently lead to improved forecasts.
We discuss estimation of the decomposition terms employing kernel regressions to estimate the occuring conditional functionals and provide limit theory, enabling for example tests of forecast optimality and informational content. Then we apply the decomposition to evaluate mean forecasts of US inflation and GDP growth from the Survey of Professional Forecasters and quantile forecasts of inflation derived from the Bank of England fan charts. For the former we find that miscalibration plays a negligible role in terms of determining forecast accuracy and that resolution is what drives the mean squared forecast error. Resolution drops to zero already for two-quarter ahead forecasts, putting into question the usefulness of the forecasts from then on, even though they are optimal and do not seem to contain systematic mistakes. For the latter, a quite different picture arises with a gradual rise in miscalibration as well as a gradual fall in resolution. Here, even though the informational content of course falls with rising forecast horizon, it would suffice to provide useful forecasts for several quarters ahead if the forecasts were closer to optimal. The applications show the potential of the proposed methodology in complementing the existing approaches to forecast evaluation in econometrics, providing valuable insights and at the same time being sufficiently simple to implement and interpret.

320 Coupled state-switching models with applications in ecology and medicine

Jennifer Pohle[1], Roland Langrock[1], Ruth King[2], Mihaela van der Schaar[3]
[1]Bielefeld University, Germany
[2]University of Edinburgh, UK
[3]University of Oxford, UK

Hidden Markov models (HMMs) are time series models which assume the observations to depend on an underlying unobserved Markov chain with finitely many states. They have been applied in many different areas, e.g. speech recognition, finance, medicine, and ecology. In the case of multivariate time series, within a basic HMM formulation, the variables would be expected to evolve synchronously in the sense that they are driven by the same underlying state sequence. However, in some applications, for example in medicine, the observed variables, although correlated, do not necessarily evolve in lockstep. For instance, a substantial change in a patient's rate of breathing may or may not be accompanied by immediate visible changes in other vital signs. Analogously, in ecology, two interacting animals may or may not at a given time be in the same behavioral mode, such as resting or foraging. Coupled hidden Markov models (CHMMs) overcome this limitation by assuming separate but correlated state sequences to underlie the different variables observed, hence "couplingthe state processes of multiple HMMs.
In this talk, we first present some simulation studies that highlight the importance of accounting for both the asynchronous evolution and the interactions between the underlying latent processes. This will then be further illustrated using two case studies, a) on the interactions between a dolphin mother and her calf as inferred from movement data, and b) considering electronic health record data collected for 702 patients within the medical intensive care unit at the University of California in Los Angeles. In the latter case study, the observed variables, such as blood pressure, depend not only on the underlying states, but also on external factors, including medical treatments. The corresponding model represents an extension of CHMMs to allow for covariates in the observation processes, which lead to the flexible class of coupled Markov-switching regression models.

321 Approximate leave-future-out cross-validation for time series models

Paul Bürkner[1], Jonah Gabry[2], Aki Vehtari[3]

[1]University of Münster, Germany
[2]Columbia University, USA
[3]Aalto University, Finland

One of the most common goals of a time series analysis is to use the observed series to inform predictions for future observations. In the absence of any actual new data to predict, cross-validation can be used to measure a model's predictive accuracy for instance for the purpose of model comparison or selection. As exact cross-validation is often practically infeasible for Bayesian models because it requires too much time, approximate cross-validation methods have been developed; most notably methods for leave-one-out cross-validation (LOO-CV). However, for time-series models, it does not make sense to leave out observations one at a time because then we are allowing information from the future to influence predictions of the past. To apply the idea of cross-validation to time-series models, we thus need some form of leave-future-out cross-validation (LFO-CV). Like exact LOO-CV, exact LFO-CV requires refitting the model many times to different subsets of the data, which is computationally very costly for most nontrivial examples, in particular for Bayesian models. Using Pareto-smoothed importance sampling, we propose a method for approximating exact LFO-CV that drastically reduces the computational burden while also providing informative diagnostics about the quality of the approximation.

Visualisation and Exploratory Data Analysis

322 Visual Inference: leveraging the power of our eyes

Heike Hofmann
Iowa State University, Ames IA, United States of America

When we are 'looking' at data in plots, we often find a set of interesting points or identify a pattern that is interesting. What is the significance of a finding like that? Visual inference gives us protocols that help us to quantify the strength of a visual finding in a framework similar to to confirmatory statistical hypothesis testing. Visual inference helps analysts determine if structure is real or spurious. In this framework, plots take on the role of test statistics, and human cognition the role of statistical tests. Statistical significance of visual "discoveries" is measured by having the human viewer compare the plot of the real dataset with collections of plots of simulated or null datasets. We will discuss some applications of visual inference and aspects of visual tests, such as determining the power of competing designs.

323 HJ-Biplot as a data visualization tool in Social Sciences

Asbel Bohigues, Cristina Rivas
University of Salamanca, Spain

The HJ-Biplot is a multivariate analysis technique that allows the approach and visualization of three or more variables as well as the cases of a given data matrix in the same plane of reduced dimension; to this purpose, it is based on the factor analysis and the main components analysis logic. This technique presents a series of advantages over the limitations of bivariate correlations and scatter plots, meaning a great simplification of descriptive analyzes. It does not require advanced statistical knowledge, it is easy to interpret, and its use opens many possibilities in the exploration and treatment of data in Social Sciences.

324 A Toolbox for Manipulating and Assessing Color Palettes for Statistical Graphics

Achim Zeileis[1], Jason C. Fisher[2], Kurt Hornik[3], Ross Ihaka[4], Claire D. McWhite[5], Paul Murrell[4], Reto Stauffer[1], Claus O. Wilke[5]

[1]Universität Innsbruck, Austria
[2]U.S. Geological Survey, United States of America
[3]WU Wirtschaftsuniversität Wien, Austria
[4]University of Auckland, New Zealand
[5]The University of Texas at Austin, United States of America

The software package colorspaceprovides a broad toolbox for selecting individual colors or color palettes, manipulating these colors, and employing them in various kinds of statistical graphics. The package is available for the statistical software systems R (http://colorspace.R-Forge.R-project.org/) and, more recently, also Python (https://python-colorspace.readthedocs.io/).
Three types of palettes are provided based on the HCL (hue-chroma-luminance) color space whose axes match those of the human visual system very well: (1) Qualitative for coding categorical information. (2) Sequential for coding ordered/numeric information, where colors go from high to low (or vice versa). (3) Diverging for coding numeric information around a central neutral value, i.e., where colors diverge from neutral to two extremes.
To aid choice and application of these palettes the package provides scales for use with ggplot2; shiny (and tcltk) apps for interactive exploration (also online at http://hclwizard.org/); visualizations of palette properties; accompanying manipulation utilities (like desaturation, lighten/darken, and emulation of color vision deficiencies).

Computational Statistics and Statistical Software V (invited)

325 Identifying Mixtures of Mixtures Using Bayesian Estimation

Gertraud Malsiner-Wallli[2], Sylvia Frühwirth-Schnatter[2], **Bettina Grün**[1]

[1]Johannes Kepler Universität Linz, Austria
[2]Wirtschaftsuniversität Wien, Austria

Finite mixture models enable the approximation of data distributions in a semi-parametric way and the identification of groups in data. The mixture of mixtures approach uses a two-level model exploiting both. On the lower level mixtures approximate the cluster distribution and on the upper level mixtures group the data. Identifying the clusters in this setting is challenging and in general either achieved by imposing constraints on the model or by using post-processing procedures. Within the Bayesian framework, we propose a different approach based on sparse finite mixtures. We specify a hierarchical prior with carefully selected hyperparameters to reflect the cluster structure aimed at and which enables model estimation using standard MCMC sampling methods. In combination with a post-processing approach to resolve the label switching issue, our approach allows us to simultaneously (1) determine the number of clusters, (2) flexibly approximate the cluster distributions in a semi-parametric way and (3) identify cluster-specific parameters and classify observations. The proposed approach is illustrated in simulation studies and applications.

326 Robust outcome prediction across data sources through altered tuning parameter value selection

Nicole Schüller[1], Anne-Laure Boulesteix[1], Bernd Bischl[2], Roman Hornung[1]

[1]Institute for Medical Information Processing, Biometry and Epidemiology, University of Munich, Munich, Germany

[2]Department of Statistics, University of Munich, Munich, Germany

In many application areas of high-dimensional data, covariate-based prediction rules that have been trained using data from a particular source do not perform well when applied to data from other sources. This is because data from different sources on the same subject matter can feature slightly differing distributions. Many prediction methods involve one or several tuning parameters, the values of which are commonly chosen by maximizing the cross-validated prediction performance. The latter procedure, however, presumes that the data to which the prediction rule is applied follow the same distribution as the training data. In situations in which this is not the case, it could be worthwhile to consider more robust prediction rules that slightly underfit the training data in order to increase prediction performance. The value of a tuning parameter does not only control the degree to which the prediction rule is adjusted to the observations in the training data, but also, more generally, the degree to which it is adjusted to the distribution of the training data. On the basis of this idea, we consider various approaches for choosing tuning parameter values that lead to more robust prediction rules than those obtained using tuning parameter value optimization based on cross-validation. For most of these approaches an external data set is used for choosing the tuning parameter values. Consulting a large collection of real data sets we compare these approaches and based on this comparison determine two strategies that can lead to better generalizing prediction rules in applications.

327 Measuring Stability of Replicated LDA Runs

Jonas Rieger, Lars Koppers, Carsten Jentsch, Jörg Rahnenführer
Technische Universität Dortmund, Germany

Topic modelling provides tools for organizing large text corpora. A widely used method is Latent Dirichlet Allocation (LDA), a generative probabilistic model, which models single texts of a text collection as mixtures of latent topics. Each topic is characterized by its word distribution. Parameter estimation for generating topic distributions relies on stochastic elements leading to different results in replicated runs of modelling the same text data. This is an often neglected fact in everyday practice.
We study the stability of LDA by comparing estimated models from different runs. Therefore, the similarity of two topics is quantified by a modified Jaccard coefficient. Building on this, topics can be clustered. The stability of replicated LDA runs is then calculated by a new algorithm based on the idea that two runs lead to a number of pairs of similar topics. In addition, the measure is suitable for comparing LDA runs modelled on different text data. In this regard, comparisons of different media and the same publication form (e.g. newspapers) as well as comparisons of different publication forms (e.g. newspaper vs. twitter) are of interest.
The interpretation of results based on LDA modelling contains two types of uncertainty. Besides the variance caused by stochastic elements in the estimation procedure, estimation of the variance of the generative process itself is also of interest. For this purpose, we propose a bootstrap technique based on random sampling of sentences with replacement within texts.
Preprocessing of text data, and generating and visualizing LDA models is done with our R package called tosca (tools for statistical content analysis) which additionally offers tools for validation.

Design of Experiments and Clinical Trials VIII (Clinical Trials I)

328 Sample size considerations for paired experimental design with incomplete outcomes

Chul Ahn

UT Southwestern Medical Center, United States of America

Paired experimental design is widely used in clinical and health behavioral studies, where each study unit contributes a pair of observations. Investigators often encounter incomplete observations of paired outcomes in the data collected. Statistical inference for paired experimental design with incomplete observations of continuous outcomes has been extensively studied in literature. However, sample size method for such study design is sparsely available. We derive a closed-form sample size formula for continuous and binary outcomes based on the generalized estimating equation approach by treating the incomplete observations as missing data in a linear model. The proposed method properly accounts for the impact of mixed structure of observed data: a combination of paired and unpaired outcomes. We derive a closed-form sample size formula accommodates different missing patterns, magnitude of missingness, and correlation parameter values. Simulation studies are conducted to evaluate the finite-sample performance of the sample size formula. Simulation results suggest that the calculated sample size maintains the empirical power and type I error under various design configurations. We demonstrate the proposed method using a real application example.

329 A two-level matching algorithm for a multi-center case-control study using registry data

Benjamin Mayer
Ulm University, Germany

Background: Lacking structural equality is a major issue to be addressed in observational studies. On the contrary, their advantages against randomized controlled trials are often reduced efforts in data collection as well as more realistic effect estimates due to an increased external validity. Numerous approaches have been developed to account for covariates which may be unequally distributed in comparison groups, including mutiple regression, subgroup analysis, and matched case-control designs. The latter has been often described as a useful tool if extensive control data sets are available, i.e. the pool of possible controls for each case is comparatively large.
Methods: A two-level matching algorithm is presented which enables to conduct a multi-centric case-control study. In particular, the algorithm includes the possibility to define the matching strategy as a combination of an exact matching approach and a subsequent consideration of further matching variables to be controlled by means of any distance measure, e.g. Euclidean distance or propensity score. The presented algorithm is applied to a case-control-based study on the treatment effect of an anti-leukemic drug using different registries as source data. Furthermore, a concept is presented to evaluate the quality of the applied matching.
Results: Applying the presented matching algorithm to the demographic and clinical data revealed well-balanced comparison groups in a 1:2 ratio (cases-controls), whereas most important covariates associated with an acute myeloid leukemia were considered as first level (exact) matching criteria, and the distribution of further covariates was controlled by means of a propensity score matching (second level matching). In general, the quality of the matching and thus the interval validity were found satisfactory regarding all matching variables used. No statistically significant treatment effect could be demonstrated.
Discussion: This two-level matching algorithm is a very flexible and useful approach to deal with the aim of finding comparable cases and controls in observational data. It is able to increase structural equality by means of balancing the most important covariates which might be of different importance for the matching process. It has been implemented in an object-oriented manner in the statistical software SAS (version 9.4) which offers therefore high flexibility regarding its application to various data analysis projects. The application of the presented algorithm in the course of a clinical evaluation of the therapeutic effect of an anti-leukemic drug was approved by the European Medicines Agency. This demonstrates the acceptance of the use of observational data in clinical studies assuming that respective measures are taken towards a maximal structural equality of comparison groups.

330 Equivalence testing with dependent data and unequal variances: Simulation of power and type 1 error for modifications of the TOST procedure

Christina Loley[2], Beate Krüger[2], Michael Matiu[1], Jürgen von Frese[4], Hannes Buchner[1], Armin Boehrer[2], Erich Bluhmki[2,3]

[1]Staburo GmbH, Germany
[2]Boehringer Ingelheim Pharma GmbH & Co. KG, Biberach a.d. Riss, Germany
[3]Hochschule Biberach, University of Applied Sciences, 88400 Biberach, Germany
[4]Data Analysis Solutions DA-SOL GmbH

Equivalence tests are frequently used in pharmacology to draw inference about the comparability of two samples. Statistically, equivalence is mostly demonstrated by means of the two one-sample t-tests (TOST) approach. However, the conventional t-test assumes independence of each observation and equality of variances in the two compared samples; an assumption that is not always to be assumed. Especially in process development in scale-down-models, where dependence between fermentations can be expected, and variances are likely to differ. Herewith, distinctive simulations of power and type 1 errors were performed for the standard TOST and its variants that account for inequality of variances (Welch's adaptation to the standard t-test) or dependence (random effects model) or both (random effects model with variance function). The impact in the following settings were investigated: number of observations in each of the two samples, their grouping within and between samples, the variance ratio between samples and the variance between groups within the samples. Our results show that power increases if the appropriate method is used, while at the same time the type 1 error is kept below the required level. Moreover, for some settings, an inappropriate method has even substantially higher type 1 error rates. Consequently, costs and risks in the pre-clinical development can be reduced by first identifying the structure of the data and then choosing the appropriate TOST variant.

331 Optimal Decisions in the Portfolio Problem

Robert Richmond Peck
University of Bath, United Kingdom

Suppose a trial sponsor has a portfolio of various drugs in phase II trials which if successful will become available for phase III trials in the near future. We consider the portfolio problem, which deals with the optimal allocation of a R&D budget to phase III confirmatory trials within some planning horizon. We require a design strategy to allocate the overall budget to the available drugs for phase III trials, in order to maximise the expected net present value of the portfolio. This design strategy must specify the optimal phase III sample sizes, given the remaining budget and drug parameters.
Previous approaches to this problem have used integer programming formulations or simulation models. We use a dynamic programming approach to derive the optimal decisions, which may scale up more efficiently in computational workload as the portfolio considered becomes more complex.
The use of group sequential designs which allow early stopping can benefit the portfolio by increasing the time one can market the drug until patent expiry, in addition to allowing the reinvestment of saved resources back into the portfolio. We analyse the impact on the portfolio when using group sequential methods.

Epidemiology III (Chronic and infectious disease methodology)

332 Prevalence of chronic diseases: Comparison between an analytical relationship and a micro-simulation.

Tim Filla[1], Annika Hoyer[2], Thaddäus Tönnies[2,4], Ralph Brinks[2,3]

[1]Institute for Medical Statistics, Heinrich-Heine Universität Düsseldorf, Germany
[2]German Diabetes Center, Institute for Biometrics and Epidemiology, Duesseldorf, Germany
[3]Hiller Research Unit for Rheumatology, University Duesseldorf, Germany
[4]Department of Statistics, Ludwig-Maximilians-University Munich

Prevalence of chronic diseases: Comparison between an analytical relationship and a micro-simulation.
Filla T, Hoyer A, Tönnies T, Brinks R
Chronic diseases, as for example, type 2 diabetes mellitus, impose a threat to humanity which has gained attention by highest international panels. To describe burden of such a chronic disease, temporal changes in prevalence, incidence and mortality are commonly used. These measures can be estimated from illness-death models that are completely governed by a differential equation (DE) [1]. Up to now, it has not been examined if this relationship is valid for every population size. The aim of this project is to evaluate the correspondence between estimation accuracy from the DE and different population sizes. We simulated two different population dynamics: first, a theoretical disease with a high prevalence and second, a hypothetical disease with a low prevalence. For each population dynamics scenario, we estimated the prevalence of the disease using the DE. Then, we performed a micro-simulation with different population sizes ranging from 100 to 1500 [2]. Afterwards, we calculated the empirical 95% confidence interval (CI) to evaluate whether the DE is an adequate representation of these population dynamics. For this, we additionally assume that the numbers of diseased participants follow a binomial distribution with prevalence as success probability and corresponding (asymptotic) 95% CIs. Therefor we compared the width of the theoretical 95% CI from this binomial distribution with prevalence calculated from the DE with the width of the CI of the empirical prevalence. We found that for population sizes less than 100 the width of the theoretical binomial CIs differ from the width of the empirical CI. For larger population size the difference between the widths of the confidence intervals gets very small. This indicates that for population sizes higher than 100, results from the DE are reliable and can be used for practical implications.
References
[1] Brinks R, Landwehr S (2015) A new relation between prevalence and incidence of a chronic disease, Math Med Biology 32 (4): 425-35, DOI 10.1093/imammb/dqu024
[2] Brinks R, Landwehr S, Fischer-Betz R, Schneider M, Giani G (2014) Lexis Diagram and Illness-Death Model: Simulating Populations in Chronic Disease Epidemiology, PLoS ONE 9 (9), DOI 10.1371/journal.pone.0106043

333 Compression of morbidity due to chronic diseases in Germany? Results from the Survey of Health, Ageing and Retirement in Europe (SHARE) 2004-2015

Ralph Brinks[1,2], Annika Hoyer[1,3]

[1]German Diabetes Center, Institute for Biometrics and Epidemiology, Duesseldorf, Germany
[2]Clinic, Department an Hiller Research Unit for Rheumatology, University Duesseldorf, Germany
[3]Department of Statistics, Ludwig-Maximilians-University Munich, Germany

Introduction: In many countries of the world, the life expectancy (LE) at birth is increasing. In Germany, for instance, LE at birth has more than doubled in the past 140 years. James Fries stated in the 1980ies that the decrease in mortality is accompanied with a decreased life expectancy with morbidity (LEM) due to chronic diseases, which he termed "hypothesis of compression of morbidity"(HCOM) [1]. There is an ongoing debate, if the HCOM is true or not.
Methods: We use the illness-death model to provide a mathematical formulation of the HCOM and derive a necessary and sufficient condition for the HCOM to hold true. Data from the SHARE study from 2004 to 2015 are analysed to examine the HCOM in the population aged 50 years and older of Germany during this period.
Results: After a formal treatment the HCOM using the illness-death model, the concepts are illustrated by empirical findings from SHARE. Coarsely speaking, the prevalence of morbidity in any of the age groups during 2004 to 2015 remains unchanged. The LE for 50y old men and women increased from 32.1 to 33.6 and 37.1 to 38.3, respectively. During the same period, LEM for 50y old men increased from 4.7 (95% CI: 3.9 to 5.7) to 4.9 (95% CI: 4.0 to 6.4) years. For 50y old women the increase in LEM was from 7.8 (95% CI: 6.8 to 8.8) to 8.7 (7.2 to 10.2) years. From 2004 to 2015, the increase of LEM at age 50 was 0.24 years for men (95% CI: -0.04 to 0.71) and 0.84 years for women (95% CI: 0.39 to 1.37). In relative terms, the percentage of life expectancy with morbidity from overall life expectancy (i.e., LEM/LE) increased 0.1 (95% CI: -0.01 to 1.4) percentage points in men and 1.5 (95% CI: 0.4 to 2.8) percentage points in women.
Conclusion: Burden of chronic diseases remains high over the study period 2004 to 2015. We have no evidence for an absolute or relative decrease of LEM in men, but have evidence for an absolute and relative increase of LEM in women.

334 Evaluating forecasts of infectious disease spread

Sebastian Meyer

Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Forecasting the future course of epidemics has always been one of the main goals of epidemic modelling. We review statistical methods to quantify the accuracy of epidemic forecasts. We distinguish point and probabilistic forecasts and describe different methods to evaluate and compare the predictive performance across models. Two case studies demonstrate how to apply proper scoring rules to uni- and multivariate forecasts of count time series from routine public health surveillance: weekly counts of influenza-like illness in Switzerland, and age-stratified counts of norovirus gastroenteritis in Berlin. Data and code to reproduce the results is provided at https://hidda.github.io/forecasting/.
Reference:
Held and Meyer (2018). Forecasting Based on Surveillance Data. arXiv:1809.03735, to appear in the Handbook of Infectious Disease Data Analysis, Chapman and Hall/CRC.

335 Estimation of multivariate hidden population sizes from register data

Birgit Debrabant

University of Southern Denmark, Denmark

Prevalence estimates of infectious diseases are important in many contexts. However, it is not always possible to simply count the infected individuals when incubation times are long or diagnosis is complicated or lengthy. Consequently there is a need for methods that can estimate the size of the corresponding hidden populations. One possible approach uses registration frequencies and estimates the population size based on an underlying probability model. Such an approach was recently presented by [LW14], who on the basis of binomial removal sampling propose a maximum likelihood estimate for the size of the hidden population. Our aim is the simultaneous estimation of age-cohort specific population sizes for several age groups. We discuss shortcomings of the ML approach and present a moment-based approach for estimation.
[LW14] A. Ledberg and P. Wennberg. “Estimating the size of hidden populations from register data”. In: BMC Medical Research Methodology 14.1 (2014).

Machine Learning III

336 Calculating Optimal Subgroup Weights for Survival Analysis using Model-Based Optimization

Jakob Richter, Katrin Madjar, Jörg Rahnenführer
TU Dortmund, Germany

A frequent problem in survival analysis is that patient cohorts consist of different subgroups, e.g., patients from different clinical centres. We consider the case where the goal is to obtain a predictive model with high accuracy for a specific subgroup, also called target subgroup. This is also a common problem in machine learning when predictions for a specific target group of observations are of interest and data from multiple other groups with the same features are available. On the one hand, simple pooling of data from all groups can deteriorate the predictive accuracy, due to potential heterogeneity between the subgroups. On the other hand, pooling increases sample size considerably especially for subgroups with small sample size and thus can improve the predictive accuracy by variance reduction. Many machine learning methods allow the use of observation weights. This enables to balance the influence of observations from different subgroups, depending on how beneficial they are for the predictive accuracy. We propose a method that optimizes such subgroup weights using MBO (model-based optimization), a state-of-the-art optimizer for expensive black-box functions. We conducted a survival analysis benchmark study with ten subgroups (cohorts from different lung cancer studies) in a nested cross-validation setting. With MBO we obtained higher values for the Concordance index, compared to results from the two naive approaches of pooling data from all subgroups or using only the data from the target subgroup.

337 Inference for L2-Boosting

David Rügamer, Sonja Greven
LMU Munich, Germany

We propose a statistical inference framework for the component-wise functional gradient descent algorithm (CFGD) under normality assumption for model errors, also known as L_2-Boosting. The CFGD is one of the most versatile toolboxes to analyze data as it scales well to high-dimensional data sets, allows for a very flexible definition of additive regression models and incorporates inbuilt variable selection. Due to this variable selection, we build on recent proposals for post-selection
inference. However, the iterative nature of component-wise boosting, which can repeatedly select the same component to update, necessitates adaptations and extensions of existing approaches. We propose tests and confidence intervals for linear, grouped and penalized
additive model components selected by L_2-Boosting. Our concepts also transfer to slow-learning algorithms more generally and to other selection techniques which restrict the response space to more complex sets than polyhedra. We apply our framework to an additive model for sales prices of residential apartments and investigate the properties of our concepts in simulation studies.

338 An application of Statistical learning to the analysis of mortality by homicide in Mexico, 2014-2017

Eliud Silva, Miguel A. Villalobos

Anahuac University, Mexico

Between 2006 and 20012, the Mexican government intensified its efforts to fight drug trafficking. This frontal war against organized crime led to an increase in violence, having, consequently, an important increase in homicides. According to the United Nations Office on Drugs and Crime (UNODC) (2014), in America, the top homicide rates per one hundred thousand people correspond to Honduras and Venezuela (with 90.4 and 53.7 respectively). However, 2017 preliminary figures from the Mexican National Institute of Statistics and Geography (INEGI, 2018), which is the government's entity responsible for the country's official statistical data, indicate that the states of Colima and Baja California Sur have very high homicide rates (113 and 91 respectively).
The objective of this work is compare several statistical learning and statistical methods to predict the homicides and identify the key variables that influence death by homicide, considering all types of deaths considered in the International Classification of Diseases, 10th revision (ICD-10). The information employed in our analysis include, among other: (1) INEGI's vital mortality statistics for the 2014-2017 period (VMS from here on); (2) census information on sociodemographic segregation (INEGI, 2015 and CONAPO, 2016); (3) peace index at the state level (Institute of Economics and Peace, 2018), gender alert information, which indicates the states that have a very high incidence of female homicides obtained from the Nacional Women´s Institute (https://www.gob.mx/inmujeres/); (5) Mexico Travel advisory from the US government (https://bit.ly/2DjRQvF); and (6) several wellbeing, poverty and social inequality indices from the CONEVAL (Council of Evaluation of the Mexican Social Policies).

339 Statistical Analysis of Benchmark Results

Daniel Horn, Nils Jannik Schüßler

TU Dortmund, Germany

In machine learning, new algorithms and improvements to old ones are introduced frequently. In most such publications some experiments are run in order to show that the proposed methods are superior to exististing ones. In a typical experiment, k algorithms are compared to each other in n different test situations (e.g. on different data sets or with different parameters of a data generating process) and each experiment is replicated l times. The target typically is to find the best performing algorithm. Results of such experiments have to be analysed with statistical methods. For example, in the case of $k = 2$ algorithms and $l = 1$ replications this can be achieved by simply performing a paired t-tests. Otherwise, more sophisticated methods are needed.

Unfortunately, in many such publications, the statistical analysis is performed poorly because either the proper statistical methods are not known to the authors or there are no such methods available. In 2006, Janez Demšar analyzed several hundred publications [1]. Less than half of these publications used hypothesis tests to find significant differences between the algorithms and only very few publications used corrections like Bonferroni to account for multiple hypothesis testing.

Janez Demšar also gave recommendations on which statistical methods should be used: At first, mean performance values over the replications are calculated. Afterwards, a Friedman test is performed, followed up by the corresponding post-hoc test, the Nemenyi test. However, there are two major drawbacks in his procedure. At first, he only uses averaged values of the l replications and therefore does ignore the variance within the replications. At second, both the Friedman and the Nemenyi tests only give global results. However, it is very unlikely that the algorithms behave comparably in all n test situations. For example, it may be possible that algorithm A significantly outperforms algorithm B on half of the test problems, while B outperforms A on the other half. In many such situations it may be beneficial to find such groups of test situations and to analyze afterwards, why the algorithms behaviors vary in these groups.

Therefore, we propose an extension to Demšar's approach: Prior to performing significance tests, we perform a cluster analysis. Based on the performance values of the algorithms in each replication, the n test situations are grouped into multiple clusters. Afterwards, Friedman and Nemenyi tests are applied to each individual cluster. The results of the tests, including a ranking of the algorithms, are displayed using directed graphs. Moreover, we present an exemplary application, in which seven optimization algorithms are compared on 40 different parameterizations of a simple test function.

[1] Demšar, J. (2006). „Statistical Comparisons of Classifiers over Multiple Data Sets“. In: J. Mach. Learn. Res. 7, S. 1–30.

Network Analysis II

340 Iterative Estimation for Exponential Random Graph Models with Nodal Random Effects

Sevag Kevork, Goeran Kauermann
LMU, Germany

The presence of unobserved heterogeneity in Exponential Random Graph Models (ERGM) is an obvious concern. We extend the well-known Exponential Random Graph Model (ERGM) by including random effects to account for unobserved heterogeneity in the network. This leads to an ERGM with random structure on the coefficients. Estimation is carried out by combining approximate penalized pseudo-likelihood estimation for the random effects with maximum likelihood estimation for the remaining parameters in the model. This allows to fit nodal heterogeneity effects even for large scale networks.

341 Modelling Time-Varying Dependence in Dynamic Networks with Applications to Regression and Model-Checking in Survival Analysis

Alexander Günther Kreiß
University of Mannheim, Germany

We will consider stochastic processes $(Z_e)_{e \in E_n}$ indexed the edges (or alternatively vertices) of a network $G_n = (V_n, E_n)$ with vertex set V_n and edge set E_n. Suppose that there exists a time-varying, random and unobserved distance measure on G_n (e.g., the shortest path respecting random weights on the edges). We will assume that the dependence-structure of the process is linked to this distance measure in the following way: Two random variables Z_{e_1} and Z_{e_2} are non-trivially dependent at a time point t if e_1 and e_2 are close according to the distance measure at time t (e.g., if they are adjacent), while we assume that the dependence decreases as the distance of e_1 and e_2 in the network increases.
In the first part we will consider three concepts based on correlation, β-mixing and (conditional) m-dependence which make the above intuition mathematically precise. We will also prove weak-dependency results (e.g. an exponential inequality) which might be of independent interest. In the second part we intent to demonstrate the use of these concepts in applications to regression and model-checking in survival analysis in a relational event network framework. We will consider, for example, a network of actors who can interact with each other. Suppose that the intensities of the interactions depend on certain covariates which describe the relation between the involved actors. Moreover, we suppose that actors can change their interaction partners and how much other actor's interactions influence their own interactions. We will apply the methods from the first part to test the accuracy of parametric models for this situation. This is illustrated by an application to rental bike data.

342 On the Construction of Invariant Measures for Graph Partition Comparison

Andreas Geyer-Schulz, Fabian Ball
Karlsruher Institut für Technologie, Germany

In this presentation we first prove that all existing graph partition comparison measures do not always work for partitions of graphs with non-trivial automorphism groups. Next, we present three ways of building invariant graph comparison measures based on Hausdorff's and von Neumann's construction of invariant measures on a pseudo-metric space. By a combination of a pseudo-metric and a metric space we provide a measure decomposition which separates an invariant part which captures the structural difference and a part which is attributed to the action of the graph automorphism group on the partitions compared. Constructive algorithms and applications to clustering several well known graphs (Zachary's Karate graph, the Petersen Graph, ...) are provided.

Robust and Nonparametric Statistics IV

343 Halfspace depth for scatter matrices

Davy Paindaveine[1], Germain Van Bever[2]
[1]Universite Libre de Bruxelles, Belgium
[2]Universite de Namur, Belgium

We propose halfspace depth concepts for scatter, concentration and shape matrices. For scatter matrices, our concept is similar to those from Chen et al. (2018) and Zhang (2002). Rather than focusing, as in these earlier works, on deepest scatter matrices, we thoroughly investigate the properties of the proposed depth and of the corresponding depth regions. We do so under minimal assumptions and, in particular, we do not restrict to elliptical distributions nor to absolutely continuous distributions. Interestingly, fully understanding scatter halfspace depth requires considering different geometries/topologies on the space of scatter matrices. We also discuss, in the spirit of Zuo and Serfling (2000), the structural properties a scatter depth should satisfy, and investigate whether or not these are met by scatter halfspace depth. Companion concepts of depth for concentration matrices and shape matrices are also proposed and studied. We show the practical relevance of the depth concepts considered in a real-data example from finance.

344 Choosing among notions of depth for multivariate data

Karl Mosler[1], Pavlo Mozharovskyi[2]
[1]Universität zu Köln, Germany
[2]Telecom Paris Tech, France

Classical multivariate statistics measures the outlyingness of a point by its Mahalanobis distance from the mean, which is based on the mean and the covariance matrix of the data. Since the early 1990's more general depth statistics have been developed for measuring centrality and outlyingness of multivariate data in a nonparametric way. A depth function is a function which, given a point and a distribution in d-space, yields a number between 0 and 1, while satisfying certain postulates regarding invariance, monotonicity, convexity and continuity. Accordingly, numerous notions of depth have been proposed in the literature, some of which are also robust against outlying data.
The departure from classical Mahalanobis distance does not come without cost.There is a trade-off between invariance, robustness and computational feasibility. In the last few years, efficient exact algorithms as well as approximate ones have been made available in R-packages like ddalpha, DepthProc, and others, implementing specific depths and their applications. Consequently, in many applications the choice of a depth statistic is no more restricted to one or two notions due to computational limits. Rather often more notions are feasible, among which we have to decide.
We discuss aspects and general principles of this choice. The speed of exact algorithms is compared, the limitations of popular approximate approaches like the random Tukey depth are demonstrated, and guidelines are provided for the construction of depth-based statistical procedures as well as for practical applications, e.g. in classification.

345 Test based on sign depth for multiple regression

Melanie Horn, Christine Müller

Technische Universität Dortmund, Germany

Following the idea to utilize the simplicial depth for regression, it was shown in 2005 that this depth can be used to test the parameter vector of regression models. As it was proven in 2016, the simplicial depth often reduces to counting the subsets with alternating signs of the residuals in the context of regression, especially in the case of polynomial regression. We will refer to this depth as sign depth.
Due to the sign depth's dependence on the order of the data, one generally assumes that the N real values of the d-dimensional explanatory variable can be sorted with respect to an inherent order. While the one-dimensional real space possesses such a natural order, for $d > 1$ in most cases one cannot sort these values that easily because there exists no distinct order of the data.
For this scenario, we present some approaches for ordering and compare them with more naive approaches like taking the order in the dataset or ordering on the basis of a single component of the explanatory variable. To compare the introduced approaches, we will look at the computational runtime as well as at the power of the resulting tests for testing the parameter vector of different multiple regression models. We compare our results to the results obtained by the classical sign test and the F-Test.

Survival and Event History Analysis VI (Prediction)

346 Joint Modelling approaches to survival analysis via likelihood-based boosting techniques.

Colin Griesbach[1], Andreas Groll[2], Elisabeth Waldmann[1]

[1]Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[2]Technische Universität Dortmund

When analyzing data where event-times are recorded alongside a longitudinal outcome, one commonly used approach in practice is separate modelling of the two outcomes without considering any interaction effects. Especially in survival analysis one main interest is incorporating time-varying covariates into the model. This however is quite a challenge, since popular methods like the extended cox regression produce biased results. Joint modelling on the other hand combines a longitudinal and a survival submodel in one single joint likelihood and thus accounts for interactions like time-varying covariates measured with error, which can be often found in follow-up studies. Previous works proposed algorithms to fit joint models via component-wise gradient boosting techniques which focus on minimizing the predictive risk, offer advantages like variable selection and also work with high dimensional data. However, gradient boosting leads to problems in the survival part of the model, since time-varying effects can not be estimated so easily. Likelihood-based boosting approaches on the other hand are, as verified in various literature, capable of handling time-dependent covariates in survival analysis, since likelihood-based boosting directly optimizes the likelihood by using newton algorithms with a component-wise updating procedure.

347 Bayesian joint latent class models of longitudinal and time-to-event outcomes

Matthias Brueckner

Lancaster University, United Kingdom

Joint models of longitudinal and time-to-event outcomes can be used to exploit the prognostic value of longitudinal trajectories on the risk of event, to correct for bias induced by the occurrence of the event or to investigate how repeated measurements and risk of event are linked, e.g. in the evaluation of surrogate biomarkers. Joint latent class models are a useful alternative to the more commonly used joint shared random effects models, especially when modelling heterogeneous populations and for prediction problems. These models assume that the entire dependency between the longitudinal trajectory and the risk of event is captured by a latent class structure instead of individual random effects, i.e. the longitudinal and the time-to-event outcomes are conditionally independent given latent class membership. There is a rich literature on frequentist and Bayesian estimation of joint shared random effects models. Existing work on joint latent class models has focused on maximum-likelihood estimation. Bayesian estimation of joint latent class models faces similar challenges as estimation of mixture models with an unknown number of components, including identifiability issues such as the problem of label switching. We develop a Markov chain Monte Carlo method for sampling from the posterior distribution of the parameters of a joint latent class model consisting of class-specific linear mixed models and proportional hazards models. We assess the performance in a simulation study and compare our method with maximum-likelihood estimation in an application of a joint model of prostate cancer recurrence times and repeated prostate-specific antigen measurements to a real data set.

348 A simulation study comparing different approaches for detection of covariate-by-treatment interactions

Bernhard Haller

Technische Universität München, Germany

Predictive biomarkers, i.e. biomarkers that are associated with the expected treatment effect, play an important role for treatment stratification which constitutes a relevant step towards a personalized medicine. Consequently, identification of predictive biomarkers is fundamental for informed treatment decisions. In randomized clinical trials, treatment effect heterogeneity over the range of relevant variables is commonly assessed by the means of subgroup analyses, presenting estimated treatment effects within subsets of patients complemented by a test for treatment-by-covariate interaction. This is an intuitive procedure for categorical variables, but relies on categorization of continuous variables. It is well known that categorization of continuous variables leads to a loss of information and consequently - for the presented research question - to a loss of efficiency regarding the detection of true treatment-by-covariate interactions. Different methods for assessment of an interaction between treatment and a continuous covariate that do not require categorization of the continuous covariate were proposed in recent years. These procedures use e.g. overlapping subgroups (Subpopulation Treatment Effect Pattern Plots, STEPP, Bonetti & Gelber, 2000), assume linear interaction terms in a common regression model, use transformations for the covariate(s) of interest (Multivariable Fractional Polynomials for Interaction, MFPI, Royston and Sauerbrei, 2004), or perform local weighted parameter estimations (Local Partial Likelihood Estimation, LPLE, Liu, Jiang & Chen, 2015). Best to our knowledge, only a very small number of studies exist that directly compare the performance of some of these methods regarding detection of true interactions in relevant scenarios (Sauerbrei, Royston & Zapien, 2007; Royston & Sauerbrei, 2014 & 2015). Therefore, we performed a simulation study considering a time-to-event endpoint, which appears to be the most relevant type of endpoint for oncological studies, where most predictive biomarkers were identified. Aim of the study is to compare properties of various methods for investigation of covariate-by-treatment interactions that rely on categorization of continuous covariates, assume linear interaction terms, or allow modeling of non-linear treatment-by-covariate interactions. We investigated different scenarios under absence of a true treatment-by-covariate interaction, presence of linear interactions and presence of non-linear interactions. We want to present results of the simulation study regarding observed type I error probabilities and power to detect truly present interactions of different approaches in the investigated scenarios.

349 Maximum Likelihood Prediction of Record Values

Grigoriy Volovskiy, Udo Kamps

Institute of Statistics, RWTH Aachen University

The analysis of (upper) record values also refers to event times of nonhomogeneous Poisson processes or to some minimal repair scheme in reliability studies. Likelihood based predictions of the next records or event times to come are studied and compared. An alternative likelihood-based prediction principle is proposed, which is seen to be related to an information-theoretic approach. Several examples for particular distributions, such as exponential, Pareto and Weibull distributions are shown.

Statistics in Science, Technology and Industry IV

350 What we might miss: Stress-testing measurements of dark energy

Ben Moews[1], Rafael S. de Souza[3], Emille E. O. Ishida[2], Alex I. Malz[5], Caroline Heneka[4], Ricardo Vilalta[6], Joe Zuntz[1]

[1]Institute for Astronomy, University of Edinburgh, UK
[2]Laboratoire de Physique de Clermont, University of Clermont Auvergne, France
[3]Department of Physics and Astronomy, University of North Carolina at Chapel Hill, USA
[4]Cosmology Research Group, Scuola Normale Superiore di Pisa, Italy
[5]Center for Cosmology and Particle Physics, New York University, USA
[6]Department of Computer Science, University of Houston, USA

Since its inception two decades ago, the nature of dark energy has puzzled cosmologists, sparking sophisticated approaches to measure its equation of state, the ratio of pressure to energy density. One of the most prominent data sources are supernovae, the bright deaths of certain types of stars. We investigate the ability of a standard Lambda-CDM analysis pipeline to discriminate between a cosmological constant and redshift-dependent dark energy equations of states based on Type Ia supernovae data. In this paper, as part of the Cosmostatistics Initiative (COIN) under the International Astrostatistics Association, we introduce a physically motivated novel random curve generator to simulate curves from different constraint families that result in representative samples exhibiting various degrees of deviation from a constant equation of state of dark energy. Using the Kullback-Leibler divergence between posterior densities of parameters, we show that an arbitrarily large set of potential deviations from the standard model of cosmology are indistinguishable based on current data and methodology, highlighting the need for both data collection with future sky surveys and advances in astrostatistical methodology.

351 Joint and conditional dependence modelling of district heating demand and weather conditions: a copula-based approach

F. Marta L. Di Lascio[1], Andrea Menapace[2], Maurizio Righetti[2]

[1]Faculty of Economics and Management, Free University of Bozen-Bolzano, Italy
[2]Faculty of Science and Technology, Free University of Bozen-Bolzano, Italy

We analyze the dependence relationship between district heating demand and weather conditions, such as outdoor temperature and solar radiation, through copula function. Copula function makes it possible to represent complex associations among variables for many different dependence structures of the data generating process without requiring specific distributional forms for the margins. Thus, copula is a flexible tool to model complex and multivariate relationships.
Our aim is to derive the conditional copula of heat demand given weather conditions, with particular attention to extreme climatic events, in order to provide useful implications for the management and production of thermal energy. We consider the case of the city of Bozen-Bolzano (Italy) and data concern district heating demand observed from January 2014 to November 2017.
The methodology used comprises three steps. First, the univariate marginal probability distribution of the variables of interest are estimated though a seasonal autoregressive moving average model for time series using the well-known Box&Jenkins procedure. Thus, we account for the non-stationarity of each time series, and model the serial dependence structure in the time series taken separately. Next, the residuals of the estimated models are computed and, being non-autocorrelated, they enable resorting to copula theory. In the second step of the analysis, we, indeed, model the complex cross-dependence relationship between heat demand and weather conditions through several different models. The most appropriate copula is selected on the basis of the empirical quantile dependence plot, the Akaike information criterion and the goodness-of-fit copula test. Finally, we derive the conditional probability function of heat demand given weather conditions. The analysis of the percentiles of such conditional distribution shows that the proposed approach can be a potentially powerful tool to improve the management of thermal energy and its storage in district heating systems.

352 Modeling Fuel Injector Spray Characteristics of Jet Engines Using Vine Copulas

Maximilian Coblenz, Oliver Grothe, Simon Holz, Rainer Koch
Karlsruhe Institute of Technology, Germany

We model the distribution of fuel drops generated by a fuel injector in a jet engine, which can be assumed as a 5-dimensional problem in terms of drop size, position, and velocity. The data is generated by numerical simulations of the fuel atomization process using the Smoothed Particle Hydrodynamics (SPH) method. The pertinent engineering literature provides no modeling approach of the interdependencies of the variables involved. In numerical simulations of combustion chambers, the variables are usually assumed to be independent. However, our data show that this is clearly not the case. Even more, the dependence between some of the variables is non-monotone and asymmetric, which makes the modeling task more difficult. We employ vine copulas which provide a flexible way to construct a multivariate distribution function. However, for this application, we need to use non-standard copulas as bivariate building blocks which are usually not implemented in software packages. After calibration, the resulting vine construction provides interesting insights into the underlying physical system, for example by discovering certain conditional independence properties. Using this copula representation enables us to create samples of realistic droplets of the fuel spray which is required for the numerical prediction of the combustion process. This approach is significantly faster than the usual way of solving the fuel disintegration problem with the SPH method. By interpolation between copula parameters we are furthermore able to capture various engine operating conditions for which no data is available.

353 Bayesian Prediction for failure times in Fatigue Behavior of Prestressed Concrete

Sophie Tchanyou Ganme, Katja Ickstadt
Technische Universitaet Dortmund, Germany

The analysis of the fatigue behavior of prestressed concretes, bridges for example is crucial to determine their lifetime. Here we are concerned with a Bayesian modelling of the damage process under low cyclic load. First we consider the Basquin model as a link function and analyse different prior distributions to predict the failure times of wires in the concrete. Secondly, we assume unknown hyperparameters that we describe with prior distributions and we build a hierarchical model witch accumulated prior information to model the damage accumulation.

Visualisation and Exploratory Data Analysis

354 Graphics in Research and Teaching illustrated in Forschung und Lehre

Antony Unwin
University of Augsburg, Germany

The Deutsche Hochschulverband, an association for professors and researchers, publishes a monthly journal, Forschung und Lehre. It is mainly composed of short articles on topics of intellectual or professional interest to the members. The authors come from a wide range of academic subjects and institutions and the articles are generally well written and well presented.
This talk looks at the use of statistical graphics in the journal. It is assumed that professors writing for professors will apply the same high standards of graphical presentation that they use in their own research and teaching. Thus these examples should give a representative picture of the level of graphics common amongst German academics.

355 A new approach to model and visualize Airbnb listing prices by the use of a smoothing surface on spatial information

Bernhard Hrobath, Friedrich Leisch
University of Natural Resources and Life Sciences, Vienna, Austria

The sharing economy thrived in the recent years in several sectors. One of the largest examples is the peer-to-peer accommodation platform Airbnb.com. The global leader in peer-to-peer accommodation innovated the tourism sector, especially in city destinations. This lead to new research interests in the mechanisms behind the players involved. In tourism research a big part of the research focuses on Airbnb as the biggest platform on the market. Several studies investigate the relationship between certain features of a listing and its host on the demanded prices. Some of these studies include spatial information, mostly by using the distance to a certain point of interest for tourists, such as the city centre. In our work, we focus on the location of a listing as a key feature on prices in traditional and peer-to-peer tourism accommodation. We follow a new approach by including the location into the model in a more flexible way. We model the spatial information in a bivariate smooth term in an additive model. Our main goal is to detect and visualize patterns in the spatial pricing structure within a city. We apply our approach to several major European and international tourism destinations. Findings support the assumption that the simple distance to a single point is not sufficient to include spatial information in this context.

356 iSEE: RNA-sequencing data exploration made easy and reproducible

Federico Marini[1,2], Aaron Lun[3], Charlotte Soneson[4,5,7], Kevin Rue-Albrecht[6]
[1]Institute of Medical Biostatistics, Epidemiology and Informatics (IMBEI), Mainz
[2]Center for Thrombosis and Hemostasis (CTH), Mainz
[3]Cancer Research UK Cambridge Institute, University of Cambridge
[4]Institute of Molecular Life Sciences, University of Zurich
[5]SIB Swiss Institute of Bioinformatics
[6]Kennedy Institute of Rheumatology, University of Oxford
[7]Friedrich Miescher Institute for Biomedical Research

Data exploration is crucial for the comprehension of large biological datasets obtained by high-throughput assays such as sequencing.
Interactive exploration is a key element in bioinformatics, as it fosters the efficient generation of novel data-driven hypotheses prior to rigorous statistical analysis, enables diagnosis of potential problems during quality control, and facilitates interpretation of the results in the context of a specific scientific question.
Most existing tools for intuitive and interactive visualization are limited to specific assays or analyses and lack support for reproducible analysis.
As a result of a community-driven effort in the scope of the Bioconductor project, we have built a general-purpose tool, iSEE - Interactive SummarizedExperiment Explorer, designed to accommodate any experimental data (bulk RNA-seq, single-cell RNA-seq, mass cytometry, ...) and/or associated metadata, stored in an instance of a SummarizedExperiment container.
iSEE is implemented in R using the Shiny framework, and is compatible with many existing R/Bioconductor packages for analysing high-throughput biological data.
Salient features of iSEE include:

- a customizable interface with different panel types, linked among each other via user selection (brushing)
- automatic tracking, storage, and rendering of the exact R code to generate all visible plots
- interactive tours to showcase datasets and findings, with step-by-step description of relevant publication-ready plots
- extendability with the definition of custom panel types

Example applications are available online to demonstrate the interactive exploration of the TCGA RNA sequencing data (https://marionilab.cruk.cam.ac.uk/iSEE_tcga),
single-cell RNA sequencing data (https://marionilab.cruk.cam.ac.uk/iSEE_allen, https://marionilab.cruk.cam.ac.uk/iSEE_pbmc4k),
and mass cytometry data (https://marionilab.cruk.cam.ac.uk/iSEE_cytof).

357 Short ordinal patterns in time series analysis

Karsten Keller

University of Luebeck, Germany

The analysis of ordinal pattern distributions provides a relativeiy new and robust approach to nonlinear time series analysis leading, for example, to the interesting concept of permutation entropy. Whereas nice asymptotic results have been found for pattern length going to infinity, in practical data analysis very short patterns describing certain features of data and models behind them are of some special interest. Our main aim is to demonstrate this point of ordinal data analysis by examples. Here we mainly discuss statistics based on counting monotone changes and on considering asymmetries in ordinal pattern distributions.

Advanced Regression Modeling V

358 Using mixed multinomial probit models to explain daily mobility behavior in a large panel data set

Manuel Batram, Sebastian Büscher, Dietmar Bauer
Bielefeld University, Germany

Understanding the daily mobility behavior of a population is an essential prerequisite for the planning and administration of transportation infrastructure. The German Mobility Panel is a large mobility survey with a rotating sample where households are sampled and each person within the household is asked to record all trips, which are made during the day, for one week. The corresponding trip diary conforms to the KONTIV design and collects trip start times and locations, trip end times and locations, means of transport, distances covered, and trip purposes.
We preprocess the trip dairies and compute a human mobility motif for each day/participant. Motifs are a parsimonious way to encode the dependencies between the trips a person makes during the course of the day. We then proceed to model the sequence of motif choices for each person within the random utility framework.
For the purpose of this paper we focus on 10 years from 2004 to 2013. Estimating the mixed multinomial probit (MMNP) model, which is the most flexible choice model (esp. w.r.t substitution patterns) within the class of random utility models, for large data sets using standard methods leads to infeasible computation times.
As an alternative Bhat proposed the Maximum Approximate Composite Marginal Likelihood (MACML) method which combines composite marginal likelihood (CML) estimation and an analytic approximation to allow for simulation-free, and fast but approximate estimation of MMNP models. For CML the full-likelihood is substituted by a composite likelihood, where - instead of modeling repeated ($t = 1, ..., T$) observations of choices C_{nt} jointly - all pairs of choices are modeled as if they were independent,

$$cml(\theta) = \sum_{n=1}^{N} cml_n(\theta) = \sum_{n=1}^{N} \sum_{t=1}^{T-1} \sum_{t'=t+1}^{T} w_{tt'} \log P(C_{nt} = Y_{nt}, C_{nt'} = Y_{nt'} | X, \theta),$$

where $w_{tt'}$ is a weighting factor. Each cml_n still involves a multivariate normal cdf, which has a dimension equal to the number of potential choices, and this is addressed by an analytic approximation. In our implementation, which utilizes R and C++, we rely on the Solow-Joe approximation.
In the paper we show that the combination of human mobility motifs and MACML estimation enables us to employ the computationally burdensome but flexible MMNP model to a large panel data set with a large choice set.
We compare the estimates to the more restricted multinomial logit model estimation and investigate a number of different alternatives with respect to weighting schemes of the CML criterion function. Regarding the empirical results we observe that motif choice is very stable across time. To a differing degree this is also true for the estimated models which explain only a remarkably small fraction of the variation of choices.

359 Baseline-adjusted Proportional Odds Models for Quantification of Treatment Effects in Neurological Trials with Ordinal Outcomes

Muriel Buri[1], Armin Curt[2], John Steeves[3], Torsten Hothorn[1]

[1]Epidemioogy, Biostatistics and Prevention Institute, University of Zurich, Switzerland
[2]Spinal Cord Injury Center, Balgrist University Hospital, Zurich, Switzerland
[3]ICORD, University of British Columbia, Vancouer/Kelowna, Canada

Ordinal outcomes are common in neurological randomized clinical trials. The approaches routinely employed for assessing treatment effects, such as t-tests or Wilcoxon tests, are not particularly powerful in detecting changes in relevant parameters or lack the ability to incorporate baseline information. Hence, tailored statistical methods are needed for the analysis of ordinal outcomes in neurological trials.
We propose baseline-adjusted proportional odds logistic regression models to overcome previous limitations in the analysis of ordinal outcomes in neurological trials. For the validation of our method, we focus on the upper extremity motor score, the spinal cord independence measure, and the self-care subscore of the latter. We compared the statistical power of our method to other conventional approaches in a large simulation study of two-arm randomized clinical trials based on data from the European Multicenter Study about Spinal Cord Injury (EMSCI, ClinicalTrials.gov Identifier: NCT01571531).
The simulation study of all postulated trial settings demonstrated that the statistical power of the new method was greater than that of conventional methods. Baseline adjustments were more suited for the analysis of the upper extremity motor score compared to the spinal cord independence measure and its self-care subscore.
The proposed baseline-adjusted proportional odds models allow the global treatment effect to be directly interpreted. This clear interpretation, the superior statistical power compared to the conventional analysis approaches, and the availability of open-source software support the application of this novel method for the analysis of future clinical trials.

Causal Inference II (Aspects of Propensity Score Methods)

360 Measuring global covariate balance in matched propensity score analysis

Lina Glaubitz[1], Tim Filla[1], Oliver Kuß[2]

[1]Institut für Statistik in der Medizin, Universitätsklinikum Düsseldorf, Germany
[2]Deutsches Diabetes-Zentrum, Germany

Due to random allocation, treated and untreated subjects in randomized controlled trials (RCT) do not differ systematically from each other in both measured and unmeasured baseline covariates. This is not true in nonrandomized studies, but here propensity score matching (PSM) can be used to account for systematic differences at baseline, at least in the measured baseline covariates. It is obvious that the quality of results in a PSM analysis strongly depends on achieving balance of baseline covariates between groups, a quality which can be assessed using standard differences or z-differences [1]. Up to now, both measures have always been calculated and assessed separately for each baseline covariate. We here introduce a global measure for covariate balance which is based on the z-difference [1]. Expressing overall covariate balance in one single quantity simplifies fitting the propensity score model. Additionally, we show how the measure can be used to compare the achieved balance in the observational data to that from an RCT or to the perfectly matched propensity score analysis described by Rubin and Thomas [2,3].

1. Kuss, O., The z-difference can be used to measure covariate balance in matched propensity score analyses. Journal of Clinical Epidemiology, 2013. 66(11): p. 1302-1307.
2. Rubin, D.B., Thomas, N., Characterizing the effect of matching using linear propensity score methods with normal distributions. Biometrika, 1992. 79(4): p. 797-809.
3. Rubin, D.B., Thomas, N., Matching using estimated propensity scores: relating theory to practice. Biometrics, 1996. 52(1): p. 249-64.

361 Propensity Weighting in the Estimation of Direct Effects.

Christiana Drake[1], Julie Smith-Gagen[2]

[1]University of California, Davis, CA, USA
[2]University of Nevada, Reno, NV, USA

We investigate propensity weighting in assessing direct effects in a model where treatment may be mediated by another risk factor. In particular, we compare two approaches to estimating direct effects, counterfactual approaches as discussed by Vanderweele (2015) and principal stratification as suggested by Rubin (2002, 2004). We demonstrate the ideas via simulation studies and apply the method to a study of Cancer of Unknown Primary (CUP). The exposure variable is Cancer of Unknown Primary (CUP) and the mediator is treatment. A CUP is confirmed after a set of recommended tests are performed but a primary cancer is not found. A CUP is unconfirmed if the tests are not done. The outcome is survival beyond a fixed time point. A direct effect of a CUP diagnosis would suggest that CUP is directly affecting survival, irrespective of treatment. An indirect effect would be survival by modified by the treatment.

Classification and Pattern Recognition II

362 Evaluate the diagnostic accuracy for disease of longitudinal markers with missing data

Cuiling Wang, Richard Lipton, Ellen Grober
Albert Einstein College of Medicine, United States of America

Logistic model is commonly used to perform the Receiver operating characteristic (ROC) analysis to evaluate the diagnostic accuracy of a marker or markers. This requires all markers considered to be clearly defined and completely measured. When the marker is longitudinal, this requirement is often violated because of missing data and the fact that the marker might be collected at different times across subjects. By examine the mean pattern of the longitudinal marker among the disease and control group, we can deduce the accuracy of using the longitudinal marker to diagnose disease. Simulation studies were performed to evaluate the performance of our methods. The method is applied to an example of using the longitudinal free recall score of the Free and Cued Selective Reminding Test (FCSRT) with Immediate Recall to diagnose Alzheimer's Disease (AD) neuropathology in the Einstein Aging Study (EAS).

363 Measuring conditional agreement in method comparison studies by mixed-effects model trees

Alexander Hapfelmeier

Institute of Medical Informatics, Statistics and Epidemiology, Germany

Clinical measurement techniques are subject of constant redevelopment, improvement and modification. While standard techniques may be well-established these are often associated with invasive procedures, stress and risks on the patients' side or costs and further investment of time and other efforts on the applicants side. New approaches are introduced to reduce such disadvantages while they are supposed to provide measurements that are at least similar and preferably identical to those of the standard technique. The most commonly used analysis method and rightful standard to assess agreement of measurements, which are made by two techniques on a continuous scale, has been introduced by Martin Bland and Douglas Altman more than three decades ago (Bland & Altman 1986). It has ever since been known as the ''Bland-Altman plot" as a main outcome is a plot of differences against mean measurement values.

Much attention has been devoted to the question whether the agreement between techniques depends on the magnitude of the measured value, i.e. whether there is a relation of differences to mean values. Another research question, which has been untouched so far, is whether agreement is conditional upon other factors like experimental settings or patient characteristics. A new device or technique might simply not be applicable in any situation for any patient. The present work addresses this issue by introducing the concept of conditional agreement and proposes a corresponding analysis method for the case of repeated measurements per subject. The new method exploits the fact that agreement can be modeled through properly specified linear mixed-effects models (Carstensen 2011). In combination with the recently developed mixed-effects model trees (Fokkema 2018) it is possible to use model-based recursive partitioning to define patient subgroups with different agreement of measurement techniques. A clinical dataset is used to exemplify the application of the new method.

References:

Bland, J. Martin, and DouglasG Altman. SStatistical methods for assessing agreement between two methods of clinical measurement.The lancet 327.8476 (1986): 307-310.

Carstensen, Bendix. Comparing clinical measurement methods: a practical guide. Vol. 108. John Wiley & Sons, 2011.

Fokkema, Marjolein, et al. "Detecting treatment-subgroup interactions in clustered data with generalized linear mixed-effects model trees."Behavior research methods 50.5 (2018): 2016-2034.

Design of Experiments and Clinical Trials IX (Clinical Trials II)

364 Adaptive Propensity Score Procedure Improves Matching in Prospective Observational Trials

Dorothea Weber, Lorenz Uhlmann, Meinhard Kieser

Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

Objective:
Randomized controlled trials are the gold-standard for clinical trials. However, randomization is not always feasible. In this article we propose a prospective and adaptive matched case-control trial design assuming that a control group already exists.
Study Design:
We propose and discuss an interim analysis step to estimate the matching rate using a resampling step followed by a sample size recalculation. The sample size recalculation is based on the observed mean resampling matching rate. We applied our approach in a simulation study and to a real data set to evaluate the characteristics of the proposed design and to compare the results to a naive approach.
Results:
The proposed design achieves at least 10% higher matching rate than the naive approach at final analysis, thus providing a better estimation of the true matching rate. A good choice for the interim analysis seems to be a fraction of around $\frac{1}{2}$ to $\frac{2}{3}$ of the control patients.
Conclusion:
The proposed resampling step in a prospective matched case-control trial design leads to an improved estimate of the final matching rate and, thus, to a gain in power of the approach due to sensible sample size recalculation.
References:
PR Rosenbaum and DB Rubin. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician, 39(1):33-38, 1985.
AB Frakt. An observational study goes where randomized clinical trials have not. Jama, 313(11):1091-1092, 2015.

365 Diskussion der Estimand-Strategien aus Sicht der Nutzenbewertung

Ralf Bender

IQWiG, Germany

Seit einigen Jahren wird vor allem im Bereich der Arzneimittelzulassung über Estimands diskutiert. Unter Estimand wird hierbei der in einer geplanten Studie zu schätzende Effekt verstanden, der aufgrund der wissenschaftlichen Fragestellung durch Spezifikation der interessierenden Population, der Zielvariable, des Umgangs mit auftretenden Zwischenereignissen sowie des verwendeten Effektmaßes zu wählen ist. Neben zahlreichen Veranstaltungen und Publikationen in wissenschaftlichen Fachzeitschriften beschäftigt sich auch der Entwurf des am 30.08.2017 von der EMA publizieren Addendums zur ICH E9 Guideline über die statistischen Prinzipien für klinische Studien mit Estimands (EMA, 2017).

In diesem Addendum werden 5 Strategien beschrieben, die zu unterschiedlichen Estimands führen, von denen jedoch ein Teil mit traditionellen Methoden der klinischen Forschung nicht mehr ohne hohes Verzerrungspotenzial schätzbar ist. Es stellt sich die Frage, welche Estimands im Rahmen der Nutzenbewertung sinnvoll verwendet werden können und welche Estimands für HTA-Entscheidungen weniger relevant oder sogar irrelevant sind. Es besteht möglicherweise die Gefahr, dass die beschriebenen Strategien als Rechtfertigung verwendet werden, in klinischen Studien keine vollständige Datenerhebung durchzuführen. Aus Sicht der Nutzenbewertung sollte primär die Treatment-Policy-Strategie gewählt werden, und zwar sowohl für die gewünschten Effekte einer Therapie als auch die unerwünschten Ereignisse (Bender et al., 2016). Die anderen Strategien zur Wahl von Estimands können unter Umständen hilfreich sein für Sensitivitätsanalysen sowie ergänzende Auswertungen-

Literatur

Bender, R., Beckmann, L. & Lange, S. (2016): Biometrical issues in the analysis of adverse events within the benefit assessment of drugs. Pharm. Stat. 15, 292-296

EMA (2017): ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. European Medicines Agency, London, EMA/CHMP/ICH/436221/2017.

Data Fusion and Meta-Analysis IV

366 Dynamically borrowing strength from another study

Christian Röver, Tim Friede
University Medical Center Göttingen, Germany

Meta-analytic methods may be used to combine evidence from different sources of information. Quite commonly, the normal-normal hierarchical model (NNHM) including a random-effect to account for between-study heterogeneity is utilized for such analyses. The same modeling framework may also be used to not only derive a combined estimate, but also to borrow strength for a particular study from another by deriving a shrinkage estimate. For instance, a small-scale randomized controlled trial could be supported by a non-randomized study, e.g. a clinical registry. This would be particularly attractive in the context of rare diseases. We demonstrate that a meta-analysis still makes sense in this extreme case, effectively based on a synthesis of only two studies, as illustrated using a recent trial and a clinical registry in Creutzfeld-Jakob disease. Derivation of a shrinkage estimate within a Bayesian random-effects meta-analysis may substantially improve a given estimate even based on only a single additional estimate while accounting for potential effect heterogeneity between the studies. Alternatively, inference may equivalently be motivated via a model specification that does not require a common overall mean parameter but considers the treatment effect in one study, and the difference in effects between the studies. The proposed approach is quite generally applicable to combine different types of evidence originating e.g. from meta-analyses or individual studies. An application of this more general setup is provided in immunosuppression following liver transplantation in children.

367 Meta-analysis of few studies involving rare events

Burak Kürsad Günhan, Christian Röver, Tim Friede
University Medical Center Göttingen, Germany

Meta-analyses of clinical trials targeting rare events face particular challenges when the data lack adequate numbers of events for all treatment arms. Especially when the number of studies is low, standard meta-analysis methods can lead to serious distortions because of such data sparsity. To overcome this, we suggest the use of weakly informative priors (WIP) for the treatment effect parameter of a Bayesian meta-analysis model, which may also be seen as a form of penalization (Greenland and Mansournia, 2015). As a data model, we use a binomial-normal hierarchical model (BNHM) which does not require continuity corrections in case of zero counts in one or both arms (Smith et al, 1995). We suggest a normal prior for the log odds ratio with mean 0 and standard deviation 2.82, which is motivated (1) as a symmetric prior centred around unity and constraining the odds ratio to within a range from 1/250 to 250 with 95% probability, and (2) as consistent with empirically observed effect estimates from a set of 37 773 meta-analyses from the Cochrane Database of Systematic Reviews (Günhan et al, 2018). In a simulation study with rare events and few studies, our BNHM with a WIP outperformed a Bayesian method without a WIP and a maximum likelihood estimator in terms of smaller bias and shorter interval estimates with similar coverage. Furthermore, the methods are illustrated by a systematic review of rare safety events in immunosuppression following paediatric transplantation. A publicly available R package, MetaStan (https://github.com/gunhanb/MetaStan), is developed to automate the Stan implementation of meta-analysis models using WIPs.

Keywords: Random effects meta-analysis, rare events, few studies, Bayes, weakly informative priors

[1] Greenland S, Mansournia MA. Penalization, bias reduction, and default priors in logistic and related categorical and survival regressions. Stat Med. 2015;34:3133–3143.
[2] Smith TC, Spiegelhalter DJ, Thomas A. Bayesian approaches to random-effects meta-analysis: A comparative study. Stat Med. 1995;14:2685–2699.
[3] Günhan, BK and Röver, C and Friede, T (2018). Meta-analysis of few studies involving rare events. arXiv preprint https://arxiv.org/abs/1809.04407.

Epidemiology IV

368 Statistical Tools for Assessing the Exposome

Mercè Garí
Helmholtz Zentrum München, Germany

The concept of exposome was proposed in 2005, shortly after the end of the Human Genome Project, to indicate the set of environmental factors to which human beings are exposed throughout their lives, even before they are born (Wild, 2005). These factors include not only exposure to chemical substances, but also diet, lifestyle, use (and abuse) of drugs, stress, infections and, in general, all environmental factors, internal and external, to which individuals are exposed on a daily basis. This also includes those indoor - and outdoor - environments in which individuals spend much of their time, such as the workplace, home, school and means of transportation, among others.
To evaluate the exposome, a series of advanced analytical technologies have recently come into play. On the one hand, the inclusion of ömicstechnologies, such as transcriptomics, proteomics or metabolomics, are essential for the study of the exposome, offering very powerful tools to investigate the effects of environmental exposure and better understand the etiology of most common diseases (Vineis et al., 2009). On the other hand, the use of sensors or mobile apps (äpps"), capable of monitoring the state of health or monitoring individuals throughout the day or period, will provide additional and holistic information about the external factors of environmental exposures, registering parameters such as physical exercise, sedentary lifestyle, heart rate or time spent in indoor and outdoor environments, among many others (Rappaport, 2011).
The proposed paper aims to address some statistical tools that can be useful for assessing the exposome, including EWAS (environmental-wide association studies), SEM (structural equation models) and Multilevel modelling. For this purpose, three examples on environmental pollutants in humans are presented: (i) Mercury levels measured in hair samples from children (Garí et al., 2013; Junqué et al., 2018); (ii) Persistent organic pollutants (POPs) in serum from a general adult population (Porta et al., 2012; Garí and Grimalt, 2013); and (iii) Pesticides in an occupationally-exposed population (Garí et al., 2018). Each example provides a unique case of study on socio-demographic factors and on health outcomes, including neurodevelopment (for mercury exposures) and metabolic syndrome (associated to POPs). In addition to that, geografical distribution of pollutant exposure was addressed using multilevel modelling.
The aforementioned statistical tools are applied for each case study, and the obtained results are thoroughly presented. In summary, although there are many environmental factors that affect each individual every day, the exposome allows to characterize and quantify the simultaneous exposures to environmental pollutants, anticipating the possible diseases that an individual can develop, according to the characteristics of their surrounding environment. In this sense, the concept of exposome is fundamental to improve the science of disease prevention.

369 Quantile regression for the applied user – opportunities, challenges, examples

Andreas Beyerlein

Helmholtz Zentrum München, Germany

Quantile regression is a statistical technique to model quantiles (i. e. percentiles) within a regression framework. Although quantile regression has been of greater interest to statistical methodologists and is implemented in standard statistical packages, it appears to be heavily underused in medical research. This talk will address the advantages and potential drawbacks of using quantile regression in comparison to standard modelling approaches such as linear and logistic regression from an applied user's perspective. Successful applications of quantile regression in epidemiological research on risk factors of childhood overweight and type 1 diabetes will also be presented.

Machine Learning IV

370 Forecasting of high-dimensional realized covariances with reservoir computing

Lyudmila Grigoryeva[1], Oleksandra Kukharenko[1], Juan-Pablo Ortega[2,3]
[1]University of Konstanz, Germany
[2]CNRS, France
[3]University of St. Gallen, Switzerland

The problem of forecasting high-dimensional realized covariance (RV) matrices computed out of intraday returns of the components of the S&P 500 market index is considered. The study focuses on a novel machine learning paradigm known as reservoir computing (RC) for producing multistep ahead forecasts for time series of realized covariances. Various families of reservoir computers have been recently proved to have universal approximation properties when processing stochastic discrete-time semi-infinite inputs. The goal is to implement with reservoir computers the forecasting of realized covariances. We examine the empirical performance of RC in comparison with many conventional state-of-the-art econometric models for various RV estimators, periods, and dimensions. We show that universal RC families consistently demonstrate superior predictive ability for various designs of empirical exercises.

371 Gaussian-Process Approximations for Big Data

Matthias Katzfuss
Texas A&M University

Gaussian processes (GPs) are popular, flexible, and interpretable probabilistic models for functions. GPs are well suited for big data in areas such as machine learning, regression, and geospatial analysis. However, direct application of GPs is computationally infeasible for large datasets. We consider a framework for fast GP inference based on the so-called Vecchia approximation. Our framework contains many popular existing GP approximations as special cases. Representing the models by directed acyclic graphs, we determine the sparsity of the matrices necessary for inference, which leads to new insights regarding the computational properties. Based on these results, we propose novel Vecchia approaches for noisy, non-Gaussian, and massive data. We provide theoretical results, conduct numerical comparisons, and apply the methods to satellite data.

Statistics of High Dimensional Data III

372 The Growth Curve model under high dimensions with applications to profile analysis

Dietrich von Rosen

Swedish University of Agricultural Sciences, Sweden

The Growth Curve model is a bilinear model useful for studying short balanced time series. In high dimensions the mean parameter space is fixed but the size of the dispersion matrix becomes large, meaning that there are an infinite number of nuisance parameters. Estimation of mean parameters are derived under high dimensional assumptions. Moreover, we consider profile analysis in high dimensions where some of the tests follow from knowledge about the growth curve model.

Statistics in Science, Technology and Industry V

373 Inlier Detection

Undine Falkenhagen[1,2], Wolfgang Kössler[1], Hans-Joachim Lenz[3]
[1]Humboldt Universität zu Berlin
[2]Universität Potsdam
[3]Freie Universität Berlin

Inliers are values hidden in the interior of a sample which seem to be generated by a different mechanism than the rest of the sample. In the univariate case it is not unlikely that a value is extremely close to the mean. Still, a number of values very close to the mean might be suspicious. We look for inlier-contaminated samples because they could hint to data fraud or structural defects. Unlike for outliers, there is not much literature on inlier detection and in addition it seems to be a more complicated task.
We suppose a method to identify a normal inlier distribution with small variance within an otherwise normally distributed sample with higher variance using a likelihood ratio test. As Wilks' Theorem on the distribution of the likelihood ratio test statistic under the null hypothesis is not applicable in our situation, we use a simulation-based approach. The method outperforms a simple Shapiro-Wilk-Test on normality and therefore helps to further investigate the authenticity of data.

374 A new statistical index to evaluate sleep quality using sensors

Gloria Gheno
Innovative data analysis, Italy

In the last decades the studies of the analysis of life quality are becoming increasingly important. Recent estimates show that one third of life is spent sleeping, consequently the analysis of the nature of sleep becomes essential to evaluate, in the subjects examined, the conditions of health and any complications caused by the various types of insomnia. Currently, some methods have been developed to assess correctly sleep quality, such as polysomnography, a test used to determine sleep disturbances, and the examination of data collected from sensors applied to the monitored subject. The latter method is inexpensive, it does not require hospitalization in appropriate facilities, it can be done at home and it is easily interpretable using the appropriate index, which, however, has the limit to be deterministic. To solve this limitation, I propose a new stochastic index associated with specific statistical tests which evaluate its goodness and determine its possible outliers. To improve the monitoring of sleep, I also insert some additional variables related to the nocturnal movements of the body, which change depending on the 5 phases into which the sleep is subdivided. Indeed, a good quality of sleep requires that the percentage of non-REM sleep is less than 66.6%. The examinations, proposed so far, subdivide the REM phase, which can be distinguished by the random and rapid movements of the eyes, by the lower muscle tone and by the propensity to vivid dreams, from the other 4 phases, defined as non-REM. In my work, considering the 5 phases separately, I can improve the sleep monitoring and establish more precisely possible parasomnias and dyssomnias. The goodness of my index is demonstrated by its application to real cases and by its comparison with other methods present in literature.

Advanced Regression Modeling VI (Modeling Multivariate Dependence)

375 Generalised Joint Regression Modelling

Giampiero Marra[1], Rosalba Radice[2]
[1]University College London, UK
[2]Cass Business School City, University of London, UK

Regression is one of the core statistical methods and is used in a wide variety of empirical applications. It typically involves one response variable and a set of covariates. However, the importance of modelling simultaneously two or more responses conditional on some covariates has been increasingly recognised. In this talk I will provide an overview of the joint bivariate regression models that I have been co-developing for the past 10 years. The approach is flexible in that it can accommodate a number of model representations, several distributional choices and a variety of regressor effects. I will discuss the main methodological and computational aspects of the modelling framework as well as the corresponding open-source software, and illustrate the methods using some case studies.

376 Predicting matches in international football tournaments via generalised joint regression modelling

Hendrik van der Wurp, Andreas Groll
TU Dortmund University, Germany

In this talk we investigate the Generalised Joint Regression Modelling (GJRM) class proposed by Marra and Radice (2017) and illustrate how it can be used to predict the outcomes of football matches. Common approaches to predict e.g. FIFA World Cups are working without dependency structures or are assuming simple correlation when modelling goals for each team in a given match. The GJRM model class uses generalised linear regression approaches to fit marginal distributions which are then evaluated inside a given copula. The copula structure is taken into account while fitting the margins. This leads to very flexible dependency structures.
However, regarding (international) football data, such as e.g. FIFA World Cup matches, it is reasonable to assume that the covariate effects are equal for both competing teams. This is particularly problematic in FIFA World Cup matches where the first-named team is more frequently the higher ranked team and, hence, the margins are not assigned randomly. For this reason, we present a small adjustment to the methodology by Marra and Radice (2017) and incorporate a specific type of penalisation into the estimation procedure.
In both a short simulation study and a real data example on FIFA World Cup data we then compare the predictive strength of our proposed extension with both the original method and an alternative approach to deal with the imbalance in the tournament design of the FIFA. Finally, we plan to investigate more complex approaches inside GJRM which include non-linear modelling via splines. This way, we expect to improve our model step by step. We will compare the performance with the predictions by bookmakers, serving as a natural benchmark, and other published modelling approaches by Groll et al. (2015) and Schauberger and Groll (2018), who used regularised Poisson regression and Random Forests.

377 Structured additive multiple-output noncrossing Bayesian quantile regression models

Bruno Santos, Thomas Kneib
University of Goettingen, Germany

Quantile regression models are a powerful tool for studying different points of the conditional distribution of univariate response variables. Their multivariate counterpart extension though is not straightforward, starting with the definition of multivariate quantiles. Here in this work, we propose a flexible Bayesian quantile regression model when the response variable is multivariate, where we are able to define a structured additive framework for all predictor variables. We build on previous ideas considering a directional approach to define the quantiles of a response variable with multiple-outputs. We combine this approach with a proposal in the literature to define non-crossing quantiles in every directional quantile model. We define a Markov Chain Monte Carlo (MCMC) procedure for model estimation, where the noncrossing property is obtained considering a Gaussian process design to model the correlation between several quantile regression models. We illustrate the results of these models using German data from the Socio Economic Panel, where the interest lies in explaining more dimensions of inequality in the population, such as income and health, taking into account the dependence between these two variables.

378 Non-stationary spatial regression for modelling monthly precipitation in Germany

Isa Marques[1], Thomas Kneib[1], Nadja Klein[2]
[1]Georg-August-University of Göttingen, Germany
[2]Humboldt-University of Berlin, Germany

It is widely accepted that spatial dependencies have to be acknowledged appropriately in data that are spatially aligned. However, most spatial models still rely on the assumption that the dependence structure does not vary over space. While assuming stationarity considerably facilitates estimation, it is too restrictive when describing a variety of atmospheric phenomena, such as precipitation. Even so, general applicability of non-stationary models is often hindered, as their use reveals to be cumbersome and improvements over stationary models can be hard to identify. The stochastic partial differential equation approach to spatial modelling allows for flexible specification of non-stationary models where explanatory variables can easily be included in the dependence structure. Given the German orographic diversity, it makes sense to model precipitation in a non-stationary way. We suggest an SPDE based model where both the mean and dependence structure are allowed to vary with elevation. Results show that, according to the widely applicable Bayesian information criterion, a non-stationary model provides a better fit to the data. Taking German monthly precipitation as a case study, we use a simulation study to explore the model's ability to correctly identify non-stationarity.
This ability is analyzed, under a fixed domain, for increasing sample sizes and signal-to-noise ratios, as well as for different scalings of the spatially varying parameters of the SPDE. In addition, alternative, more flexible, ways of modelling spatial dependence, such as through multivariate B-Splines, are explored.

Causal Inference III (Neyman-Rubin Model and Observational Studies)

379 Estimating continuous treatment effect functions with joint sufficient dimension reduction

Ming-Yueh Huang

Academia Sinica, Taiwan

The estimation of continuous treatment effect functions using observational data often requires parametric specification of the effect curves, the conditional distributions of outcomes and treatment assignments given multi-dimensional covariates.
While nonparametric extensions are possible, they typically suffer from the curse of dimensionality.
To deal with this problem, dimension reduction is often inevitable and we propose a sufficient dimension reduction framework to balance parsimony and flexibility.
The joint central subspace can be estimated at a $n^{1/2}$-rate without fixing its dimension in advance, and the treatment effect function is estimated by averaging local estimates of a reduced dimension.
Asymptotic properties are also studied.
Unlike binary treatments, continuous treatments require multiple smoothing parameters of different asymptotic orders to borrow different facets of information, and their joint estimation is proposed by a non-standard version of the infinitesimal jackknife.

380 Causal inference in multi-state models - estimands and estimators of the population-attributable fraction

Maja von Cube[1,2], Martin Schumacher[1,2], Martin Wolkewitz[1,2]

[1]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

[2]Freiburg Center for Data Analysis and Modelling, University of Freiburg, Germany

A relevant statistical quantity in epidemiology is the population-attributable fraction (PAF). Initially, the PAF has been defined for cross-sectional studies and cohort studies of fixed length. It summarizes the population-attributable burden by taking both the relative risk (RR) of the exposure and the prevalence into account.
Provided adequate consideration of confounding, the PAF defined for these data settings is interpretable as a causal effect measure and quantifies the benefit of a preventive intervention against a harmful exposure.
However, in hospital epidemiology researchers are confronted with complex time-to-event data. To study for example the burden of hospital-acquired infections (HAIs) in terms of intensive-care mortality, one has to account for the time-dependency of an acquisition of HAIs. All patients are initially unexposed at intensive-care unit (ICU) admission. Then, an infection may occur over the course of the ICU stay of the patients. Moreover, discharge alive is a competing event to death in the ICU. Disregarding the competing risk by e.g. censoring discharged patients can lead to a strong overestimation of ICU mortality.
Literature provides two different approaches to obtain the PAF for time-dependent exposures and competing outcomes. However, an extension of the PAF to a time-dependent setting is not straightforward.
By critically revising the approaches proposed to define and estimate the PAF for time-to-event data with competing risks and time-dependent exposures, we clarify misconceptions of interpretation and estimation of the PAF in these data settings.
Statistical nomenclature allows for a clear differentiation between causal effect measures defined with counterfactuals and measures defined as associational measures based on observable quantities. By properly differentiating between these two types of definition, we show that the basic rational "in a perfect randomized trial association is causation" is not valid when the exposure depends on time. As a result, there are two distinct definitions of the PAF for data settings with time-dependent exposure. These two definitions are the same when the exposure is time-independent. However, they have different interpretations in the more complex setting we consider. This finding provides an explanation for misinterpreted results found in literature.
The interpretational difference between the two definitions of the PAF is demonstrated by a real data example of patients in intensive-care. In this setting we use the PAF to estimate the proportion of excess ICU death cases associated with hospital-acquired pneumonia (HAP). Moreover, we estimate the proportion of preventable cases if HAP could be prevented.

References:
[1] Schumacher, M., Wangler, M., Wolkewitz, M. and Beyersmann, J. (2007) Attributable mortality due to nosocomial infections. A simple and useful application of multistate models. Methods of information in medicine, 46, 595.
[2] Bekaert, M., Vansteelandt, S. and Mertens, K. (2010) Adjusting for time-varying confounding in the subdistribution analysis of a competing risk. Lifetime data analysis, 16, 45–70.
[3] Hernán, M. A. (2004) A definition of causal effect for epidemiological research. Journal of Epidemiology and Community Health, 58, 265–271.

381 Practical and Effective Estimation of Effect Heterogeneity by Modified Causal Forests

Michael Lechner
University of St. Gallen, Switzerland

Uncovering the heterogeneity of causal effects of polices and business decisions at various levels of granularity provides substantial value to decision makers. This paper develops new estimation and inference procedures for multiple treatment models in a selection-on-observables frame-work by modifying the Causal Forest approach suggested by Wager and Athey (2018). The new esti-mators have desirable theoretical and computational properties for various aggregation levels of the causal effects. An Empirical Monte Carlo study shows that they may outperform previously suggested estimators. Inference tends to be accurate for effects relating to larger groups and conservative for effects relating to fine levels of granularity. An application to the evaluation of an active labour mar-ket programme shows the value of the new methods for applied research.

382 Search for predictive factors based on observational studies

Julia Krzykalla, Axel Benner, Annette Kopp-Schneider
German Cancer Research Center, Heidelberg

One of the main endeavors in clinical research is to provide evidence for tailored treatment decisions ("stratified medicine"). In the pursuit of optimal treatment decision rules, one strives to identify biological markers ("biomarkers") that are able to distinguish patient subgroups with differential treatment outcome and therefore qualify as predictive factors. The usual setting to this end is a randomized controlled trial (RCT), the final analysis of which is followed by a set of pre-defined interaction tests or subgroup analyses "to check that the estimated overall effect is broadly applicable" (European Medicines Agency (EMA) Draft Guideline on the investigation of subgroups in confirmatory clinical trials). However, cancer diseases, even within the same entity, are very diverse, plus the mode of action of many cancer therapies is not fully understood. Hence, from a pathological or therapeutic point of view, a large number of biomarkers are to be considered, as they are representing different pathways for example. Additionally, the interplay of several biomarkers might be necessary to get the full picture. Structures as complex as these cannot be captured accurately by basic interaction tests or subgroup analyses, but ask for more advanced techniques.
We performed a review on statistical methods that are suitable for such situations. Almost all of the methods that we were able to identify have been developed for randomized data only.
Although RCTs are the gold-standard in drug development, observational studies are sometimes the only source of data available. "Naïve" application of the above-mentioned methods to non-randomized data is expected to cause biased results. In the world of causal inference, two approaches have been proposed to adjust for confounding in a setting with heterogeneous treatment effects: marginal structural models using inverse probability of treatment weighting and structural nested mean models. We investigate in which data situations these causal models are advantageous over simple covariate adjustment. In a second step, we explore how such ideas could be used in combination with the advanced methods for the search of predictive factors mentioned above.

Design of Experiments and Clinical Trials X (Clinical Trials III)

383 Blinded sample size reestimation in multi-centre randomized controlled clinical trials

Markus Harden, Tim Friede

Universitätsmedizin Göttingen, Germany

Many phase II/III trials recruit subjects at multiple study sites. This will introduce some hierarchical structure in the data that can result in a power-loss compared to a more homogeneous single-centre trial. Building on a recently proposed approach to sample size determination in multi-centre clinical trials, we suggest a blinded sample size reestimation procedure for multi-centre trials with a continuous endpoint [1]. Compared to other sample size approaches for multi-centre trials, our formula does not rely on balanced data and is therefore advantageous, especially for sample size reestimation at interim stages. We illustrate the proposed methodology by an example on diabetes management systems. A simulation study is carried out to assess the operation characteristics of the blinded sample size reestimation procedure demonstrating the influence of parameters as between-centre heterogeneity, residual variance of observations, block length and number of study sites.

The implementation of an interim pilot study design for multi-centre trials works similar to other study designs and can help to correct misspecified initial sample size calculations during a trial.

References:

[1] Harden M, Friede T (2018) Sample size calculation in multi-centre clinical trials. BMC Medical Research Methodology 18:156; https://doi.org/10.1186/s12874-018-0602-y

384 A Bayesian decision-theoretic framework for evaluation of Bayesian clinical trials performance and robustness to prior-data conflict

Silvia Calderazzo, Manuel Wiesenfarth, Annette Kopp-Schneider
Division of Biostatistics, German Cancer Research Center, Germany

Bayesian clinical trials allow to take advantage of relevant external information through the elicitation of prior distributions. The information contained in the prior eventually impacts on Bayesian posterior parameter estimates and test decisions, which are commonly evaluated in terms of frequentist operating characteristics, i.e., type I error, power and MSE. Improvements of such operating characteristics can be achieved if prior information is consistent with the true parameter value generating the current data, at the cost of type I (or type II) error and MSE inflation, otherwise. To limit harmful effects in case of prior-data conflict, approaches which adaptively discount possibly conflicting prior information, such as robust mixture (see e.g. Schmidli et al., 2014) and empirical Bayes power priors (Gravestock and Held, 2017) have been proposed. However, for a range of applications where uniformly most powerful tests exist, no power gains are possible if strict control of frequentist type I error is required (Kopp-Schneider et al., in preparation).
To weight the relative advantages of each prior specification in a fully Bayesian framework, for different degrees of agreement between prior information and current data, we take a Bayesian decision theoretic standpoint. We adopt the convention of e.g. Psioda and Ibrahim (2018), and assume that two priors are elicited: a sampling prior, which generates the observed data and represents the 'true' state of nature, and an analysis prior which is ultimately adopted to perform the trial and fit the data. This distinction allows us to perform sensitivity analyses for different degrees of discrepancy between sampling and analysis prior specification. We extend on the available literature in the field of Bayesian decision-theoretic approaches to sample size determination for hypothesis testing, by including the estimation error in the evaluation of the integrated risk and by placing particular focus on the increasingly popular robust mixture and empirical Bayes power analysis prior specifications.
Gravestock, I., Held, L., on behalf COMBACTE?Net consortium. (2017). Adaptive power priors with empirical Bayes for clinical trials. Pharmaceutical statistics, 16(5), 349-360.
Kopp-Schneider, A., Calderazzo, S., & Wiesenfarth, M. (in preparation). Use of external information in clinical trials: What can be gained in terms of frequentist power?
Psioda, M. A., & Ibrahim, J. G. (2018). Bayesian clinical trial design using historical data that inform the treatment effect. Biostatistics.
Schmidli, H., Gsteiger, S., Roychoudhury, S., O'Hagan, A., Spiegelhalter, D., & Neuenschwander, B. (2014). Robust meta-analytic-predictive priors in clinical trials with historical control information. Biometrics, 70(4),
1023-1032.

385 Planning sequential Bayesian designs: Sample size prediction and stopping boundary specification

Angelika M. Stefan[1], Felix Schönbrodt[2]
[1]University of Amsterdam
[2]Ludwig-Maximilians-Universität München

Well-designed experiments balance the needs for efficiency and informativeness. Sequential Bayesian designs address these needs as they guarantee compelling evidence with on average substantially lower sample sizes than comparable frequentist designs. In a sequential Bayesian design, data are collected until the Bayes factor in the focal hypothesis test arrives at an upper or lower boundary, that is, until compelling evidence for the null or the alternative hypothesis is reached. Thus, the sample size of the experiment is not known to the experimenter before the data collection starts. At the planning stage of an experiment, sequential Bayesian designs invoke two practical questions: How should the stopping boundaries be chosen and what sample sizes can be expected given a certain experimental setup? In this talk, we want to introduce Bayes Factor Design Analysis (BFDA), an a-priori simulation-based design analysis method that aims at providing an answer to these questions. We show how the results of a BFDA can be used to explore the effects of different stopping boundaries on the properties of the design and how a BFDA can be used to predict the expected sample size conditional on the chosen boundaries. We analyze symmetric, and non-symmetric as well as constant and collapsing stopping boundaries and derive recommendations for their usage in experimental design.

386 A new conditional performance score for evaluating sample size recalculation rules in adaptive designs

Carolin Herrmann[1], Kevin Kunzmann[2], Maximilian Pilz[2], Meinhard Kieser[2], Geraldine Rauch[1]

[1]Charité - Universitätsmedizin Berlin, Germany
[2]Institute of Medical Biometry and Informatics, University of Heidelberg

A precise sample size calculation is of major importance for a successful and efficient clinical trial. Under- or overpowering trials should be avoided for ethical and economic reasons. As calculation of the "correct" sample size in the planning stage is based on a number of parameter assumptions, which are related to a certain level of uncertainty, an adjustment of the sample size during an ongoing trial is appealing. After recruiting and evaluating a first sequence of patients, updated knowledge on the required parameters is available which can be used to adapt the sample size or to decide on an early stopping.
So far, there exist no unique standards to assess the performance of adaptive sample size recalculation rules. Consequently, a fair comparison between different recalculation rules is difficult. Single performance criteria commonly reported are given by the power and the average sample size (under the null- or alternative hypothesis) which are obviously highly correlated. Other performance measures such as the variability of the recalculated sample size and the conditional power distribution are often ignored. Liu et al. [1] were the first who presented a performance score for adaptive designs based on sample size and power criteria. This score compares the power and the average sample size of an adaptive design in relation to the "perfect" fixed design (under the true parameter setting) as a gold standard. The performance score has the potential shortcoming that it does not take into account the variability of sample size and that it is not well defined under the null hypothesis of the underlying test problem. Moreover, it is highly questionable whether the "perfect" fixed sample size design is really a valid gold standard.
Therefore, the need for an optimized performance score combining all relevant performance criteria is evident.
In this talk, we present a new conditional performance score and compare it to the one by Liu et al. [1].
[1] Liu GF, Zhu GR, Cui L. Evaluating the adaptive performance of flexible sample size designs with treatment difference in an interval. Stat. Med. 2008, 27:584-596.

Machine Learning V

387 Learning in artificial and real neural networks

Günther Palm
University of Ulm, Germany

In this talk I will compare the implementation of learning in artificial and real neural networks in terms of: the local learning rules, the overall architecture, the learning schedules. One idea, in particular, is to compare layers in multi-layer architectures to areas of the cortex.

388 Conditional sampling for exploring biological connections in single cell RNA-Seq data with Deep Boltzmann Machines

Moritz Hess, Stefan Lenz, Harald Binder
Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

The high dimensional gene expression profiles of single cells, inferred by deep RNA sequencing (RNA-Seq) should theoretically allow to precisely characterize the function and state of cells and finally to model intercellular interactions, which is crucial for better understanding pathogenesis of diseases. Deep feed forward architectures such as convolutional neural networks can learn complex dependencies but are by design confined to learn structure with respect to a specific task. In contrast generative models such as Deep Boltzmann Machines (DBM) learn the joint distribution of the training data. A trained DBM can thus be regarded as a model of the complex co-expression pattern that is linked with the physiological and biochemical processes occurring in a cell. We employ DBMs to model this pattern in different neuron types in the mouse brain. By sampling from the DBMs, conditional on the expression level of individual genes and gene sets, we in-silico study, how a change in the expression levels of a small group of genes affects the expression levels of the remaining genes. This allows us to study the connection of well known cell type-specific marker genes with potential gene networks that are linked with a differentiation of neuronal function such as the type of neurotransmitter that is employed by the cell. We validate our findings using information about known co-regulatory relationships of genes. Since to date, DBMs only work well with binary data, we employ an ordinal coding scheme which allows us to model the quantitative expression data.

389 Pattern Detection of Life Events and Daily Hassles Using Longitudinal Deep Boltzmann Machines

Göran Köber[1,2], Stefan Lenz[1], Haakon Engen[3], Kenneth S.L. Yuen[3], Anita Schick[3], Raffael Kalisch[3], Harald Binder[1]

[1]Institute of Medical Biometry and Statistics (IMBI), Faculty of Medicine and Medical Center—University of Freiburg, Germany
[2]Freiburg Center for Data Analysis and Modeling, University of Freiburg, Germany
[3]German Resilience Center (DRZ)—University Medical Center of the Johannes Gutenberg University, Mainz, Germany

Generative deep learning provides a versatile toolbox for investigating patterns in data in an unsupervised manner. We specifically, use deep Boltzmann machines (DBMs) to learn longitudinal binary representations. This is motivated by an application in psychological resilience research where the occurrence of life events and daily hassles (referred to as stressors hereafter) was assessed repeatedly every three months in the MARP and LORA study. An understanding of longitudinal stressor patterns would foster predictions of future mental health which, in turn, can substantially improve intervention decisions. We train a longitudinal deep Boltzmann machine to learn the cross-sectional co-occurrences and longitudinal sequences of stressors. Since DBMs are considered difficult to train, we discuss architectural decisions facilitating successful learning, e.g., pre-training parts of the DBM consisting of cross-sectional groups of more strongly related stressors. Furthermore, the small sample size is addressed by training examples with a sliding window $w = 2$, i.e., with stressors at T and $T + 1$. Thereby, the DBM learns both the co-occurrences and immediate sequences of the stressors. We apply heatmaps to show and discuss the uncovered patterns. To further evaluate performance, we investigate the sampled stressors at $T + 1$ conditionally on actual realization of values of T. More precisely, we fix the sampling process at the actually occurred stressors at T and repeatedly sample stressors at $T + 1$ from the DBM. This procedure allows us to compare conditionally sampled stressors at $T + 1$ with the actually realized stressors. Lastly, loglinear models together with a permutation approach are used on the generated data to extract interrelated groups of stressors. The results highlight the feasibility of deep learning for resilience research, and more generally for longitudinal binary patterns.

390 Releasing Differentially Private Synthetic Micro-Data with Bayesian GANs

Christian Arnold[1], Marcel Neunhoeffer[2], Sebastian Sternberg[2]
[1]Cardiff University
[2]University of Mannheim, Germany

This paper shows how to generate differentially private synthetic data using generative adversarial nets (GANs). We bring together insights from three literatures. First, generating artificial copies of original data is considered the gold standard in differential privacy, since any further analysis of this kind of data does not spend any extra amount of the privacy budget. Second, GANs became prominent in learning and generating the representation of visual and audio data. However, to capture the diversity of social micro-level data, we apply, third, Bayesian GANs. We show how BayesGAN can generate differentially private data when injecting the right amount of noise during training with a Stochastic Gradient Langevin Dynamics sampler. We are the first to generate differentially private data using BayesGAN. Our experiments show that we generate differentially private micro-data that are at least as useful for analysis and prediction as synthetic data generated with other, so far considered methods. Adding to this, we also incorporate the privacy loss parameters ϵ and δ into our framework which allows users to control the desired privacy loss of the synthetic data ex-ante.

Network Analysis III

391 Preservation of multivariate correlation-based networks constructed from high-dimensional data

Pascal Schlosser[1], Anna Köttgen[1], Martin Schumacher[2]

[1]Institute of Genetic Epidemiology, Faculty of Medicine and Medical Center - University of Freiburg, Germany

[2]Institute of Medical Biometry and Statistics, Faculty of Medicine and Medical Center - University of Freiburg, Germany

Background:
Correlation-based network analyses are successfully used in many high-dimensional Omics settings. We recently proposed Netboost, a weighted correlation-based network analysis based on a multivariate boosting filter (Schlosser et al. submitted). Applied to diverse Omics datasets (DNA methylation, RNA sequencing, metabolomics), aggregate measures of identified networks were shown to be predictive of overall survival in acute myeloid leukemia and kidney renal clear cell carcinoma, disease severity in Huntingtin mice, and associated with genetic determinants in chronic kidney disease patients.
Methods:
In this study we investigated preservation of networks constructed by Netboost, by Weighted Gene Co-expression Network Analysis (WGCNA, Langfelder et al. 2008), and by k-means clustering (Hartigan and Wong 1979). We compared their stability under resampling in two settings. In the first setting, we investigated 1172 metabolomic features from 5088 patients participating in the German Chronic Kidney Disease (GCKD) study. In the second setting, we studied 396,065 DNA methylation and 17,104 gene expression features from 169 acute myeloid leukemia patients from The Cancer Genome Atlas (TCGA). In both instances we compared Netboost and WGCNA on the level of their adjacency matrixes by the network preservation statistics proposed in Langfelder et al. 2011. To evaluate the reproducibility of the full approaches down to the partition of variables, we compared resulting clusterings by the Jaccard Index and the adjusted Rand Index.
Results:
DNA methylation, gene expression and metabolomics are known to exhibit strong correlation structures, which was reflected in superior performance of Netboost and WGCNA in comparison to k-means with respect to clustering similarity measured by the Jaccard Index and the adjusted Rand index. When comparing these two correlation-based network analyses directly, results depended on the application dataset with no method consistently outperforming the other. Investigation of additional preservation statistics is ongoing. In many instances, network analysis recaptured known biochemical relationships or pathways in an unsupervised manner. This also highlights the potential of strongly associated but novel clusters of variables.
Conclusion:
Network analyses can be applied for dimension reduction to stably extract aggregate measures from high-dimensional omics data. These measures can show much stronger associations with measures of interest (e.g. overall survival and genotypes) than the original variables in univariate analyses, and were stable under resampling in our applications.

392 Tensor decomposition for dynamic clustering in multi-modal social networks

Angelika Schmid, Sven Apel
University of Passau, Germany

The mesoscopic structure of many digital communities can only be fully understood when considering more than one type of interaction. For example in open-source software development, communication and coordination can be just as important for understanding developer interaction and clustering as collaborative source code editing. Uncovering community structure and their dynamics in this context can help to optimize online collaboration platforms and to predict the future development of a project. We use four-dimensional tensors to represent multi-modal interaction among developers, and tensor decompositions for clustering. Special challenges arise because of the sparsity of the networks and the count data nature of the directed edges we use. We apply the method to several large open-source projects, mining repository and different types of communication data. Our results show that real-world events that change the organizational structure of the projects manifest themselfs in the group dynamics of the projects.

393 Regression-based Network Reconstruction with Nodal and Dyadic Covariate Effects

Michael Lebacher, Göran Kauermann
Ludwig-Maximilians-Universität München, Department of Statistics, Germany

Network (or matrix) reconstruction is a general problem which occurs if the margins of a matrix are given and the matrix entries need to be predicted. In this paper we show that the predictions obtained from the Iterative Proportional Fitting Procedure (IPFP) or equivalently Maximum Entropy (ME) can be obtained by a restricted maximum likelihood estimation framework relying on Augmented Lagrangian optimization. Based on the equivalence we extend the framework of network reconstruction towards regression by allowing for exogenous covariates and random heterogeneity effects. The proposed estimation approach is compared with different competing methods for network reconstruction and traffic matrix estimation. Exemplary, we apply the approach to interbank lending data, provided by the Bank for International Settlement (BIS). This dataset provides full knowledge of the real network and is therefore suitable to evaluate the predictions of our approach. It is shown that the inclusion of exogenous information allows for superior predictions in terms of L_1 and L_2 errors. Additionally, the approach provides meaningful interpretation for the effects of the exogenous variables.

394 A smooth dynamic network model for patent collaboration data

Verena Bauer, Göran Kauermann
LMU München, Germany

The development and application of models which take the evolution of networks with a dynamical structure into account are receiving increasing attention. Our research focuses on a partial likelihood approach to model time-stamped event data for a large-scale network applied on patent collaborations. We propose a flexible non-parametric approach including covariates built from the network history.
We apply our approach to a large data set of patents submitted jointly by inventors from Germany between 2000 and 2013 on a monthly basis.

Robust and Nonparametric Statistics V

395 M-Estimation with Incomplete and Dependent Multivariate Data

Gabriel Frahm, Klaus Nordhausen, Hannu Oja
HSU, Germany

We extend the theory of M-estimation to incomplete and dependent multivariate data. ML-estimation can still be considered a special case of M-estimation in this context. In order to guarantee the consistency of an M-estimator, the unobserved data must be missing completely at random but not only missing at random, which is a typical assumption of ML-estimation. We show that the weight functions for scatter must satisfy a critical scaling condition, which is implicitly fulfilled both by the Gaussian and by Tyler's weight function. We generalize this principal result by introducing the class of power weight functions. The two aforementioned weight functions represent limiting cases of a power weight function. A simulation study confirms our theoretical findings. If the data are heavy tailed or contaminated, the M-estimators turn out to be favorable compared to the ML-estimators that are based on the normal-distribution assumption.

396 Robust estimators and tests for Gaussian graphical models

Daniel Vogel
University of Aberdeen, United Kingdom

Robust covariance matrix estimators which obey a given zero pattern in their inverse are studied. The theory for these dates back to Vogel & Fried (2011) and Vogel & Tyler (2014). In the present talk, the emphasis is on practical aspects. Some recent developments prompt me to revisit this topic: An R- package is available, and an application to music performance anxiety will be presented (Wiedemann et al. 2018). Estimating the covariance matrix subject to knowing the zero positions (i.e., knowing the graph) seems of lesser interest than estimating the positions of the zeros along with the covariance matrix, which is usually done by means of L1-penalization. However, the former provides the basis for tests. Its main benefit is for confirmatory analysis, while the latter is foremost an exploratory tool. This overview talk is based on joint work with a number of co-authors, notably Roland Fried (TU Dortmund), David Tyler (Rutgers), Stuart Watt (University of Aberdeen), and Anna Wiedemann (University of Aberdeen).
References
Vogel, D. & Fried, R. (2011). Elliptical graphical modelling. Biometrika, 98(4), 935-951.
Vogel, D. & Tyler, DE. (2014). Robust estimators for nondecomposable elliptical graphical models. Biometrika, 101(4), 865-882.
Wiedemann, A., Vogel, D., Voss, C., Nusseck, M., Hoyer, J. (2018). The role of retrospectively perceived parenting style and adult attachment behaviour in music performance anxiety. Forthcoming in Psychology of Music. Available at psyarxiv.com/k5npv

397 A simple non-parametric goodness-of-ft test for elliptical copulas

Miriam Jaser, **Stephan Haug**, Aleksey Min
Technical University of Munich, Germany

In this talk, we present a simple non-parametric goodness-of-fit test for elliptical copulas of dimension d, see also [1]. It is based on the equality of Kendall's tau and Blomqvist's beta for all bivariate margins of an elliptical copula. The test statistic is based on all $d(d-1)/2$ differences between the corresponding estimators of Kendall's tau and Blomqvist's beta. Under the assumption of known marginal distributions the test statistic has an asymptotic normal distribution. Nominal level and power of the proposed test are investigated in a Monte Carlo study. Further we suggest two approaches to deal with the case of unknown marginal distributions. An empirical application illustrates our goodness-of-fit test at work.
References:
[1] Jaser, M., Haug, S. and Min, A. (2017). A simple non-parametric goodness-of-ft test for elliptical copulas. Dependence Modeling 5, 330-353.

Statistics of High Dimensional Data IV

398 Sampling Distributions of Optimal Portfolio Weights and Characteristics

Taras Bodnar[1], Arjun K. Gupta[2], Nestor Parolya[3], **Erik Thorsén**[1]

[1]Department of Mathematics, Stockholm University
[2]Department of Mathematics and Statistics, Bowling Green State University
[3]Institute of Statistics, Leibniz University Hannover

Optimal portfolio selection problems are determined by the (unknown) parameters of the data generating process. If an investor wants to realise the position suggested by the optimal portfolios then he/she need to estimate the unknown parameters. Most often, the parameters of interest are the mean vector and the covariance matrix of the asset return distribution. In this paper we characterise the exact sampling distribution of the estimated optimal portfolios obtained as solutions of different optimization problems. Additionally, their asymptotic distributions are derived under the high-dimensional setting when the number of variables $p \to \infty$ and the sample size $n \to \infty$ such that $p/n \to c$, where c is a positive constant.

399 Best Subset Selection: The Holy Grail for Variable Selection?

Louis Dijkstra*, Moritz Hanke*, Ronja Foraita, Iris Pigeot, Vanessa Didelez
Leibniz Institute for Prevention Research and Epidemiology - BIPS

*Shared first authorship
When it comes to regression analysis, best subset selection (BSS) is often seen as the best possible approach or the 'holy grail'. The intuition is that considering all possible subsets should surely lead to the optimal selection.
Despite this widespread belief, BSS was and is still rarely used due to its computational burden; the number of subsets increases exponentially with the number of variables. Naively looping over all possible combinations becomes quickly infeasible.
Only with the recent approach of Bertsimas et al. [1] has it become possible to carry out BSS to problems with the number of observations in the 1000s and the number of variables in the 100s in minutes to provable optimality. Bertsimas et al. [1] compared the performance of BSS to forward step-wise selection (FSS) and the LASSO. Hastie et al. [2] extended the comparison by using a more varied simulation set-up. In both cases, the comparison was solely with regard to predictive performance.
However, when the main objective is to gain a better understanding of a particular process or phenomenon, being able to successfully identify the true predictors is of greater relevance than prediction accuracy. Hence, in the present work, we compare BSS, FSS, the LASSO and, additionally, the elastic net with respect to their ability to recover the true predictors in both low and high dimensional problems. Throughout, we assume a sparse model, i.e., only a handful of variables are true predictors. In [1] and [2] the focus was on regression problems with Toeplitz-structured correlation matrices. We extend the simulation set-up by considering block structures as well, since many fields deal with clusters of highly correlated variables, e.g., genetics. As in [1] and [2], we investigate the impact of the signal-to-noise ratio and the correlation strength on the performance. We use minimizing the L_2 norm as the objective function for each method. The best observed F_1 score is used to assess the methods' performance.
Interestingly, BSS does not uniformly outperform the other approaches. In fact, sparse regression methods do better in most of the considered settings. BSS and FSS perform best when the signal is strong, i.e., a high signal-to-noise ratio or low dimensionality. Noticeably, FSS exhibits a similar performance to BSS, while being computationally considerably more efficient.
In summary, our findings suggest that BSS is not the holy grail for model selection. Our simulation-based results indicate that regularized regression approaches should be preferred over BSS when dealing with noisy, correlated sparse data and/or high dimensional problems.
References
1. Bertsimas, D., King, A., & Mazumder, R. (2015). Best subset selection via a modern optimization lens. Annals of Statistics, 44(2), 813–852.
2. Hastie, T., Tibshirani, R., & Tibshirani, R. J. (2017). Extended comparisons of best subset selection, forward stepwise selection, and the Lasso, 1–52. Retrieved from http://arxiv.org/abs/1707.08692

400 Nonparametric Bayesian dependent Chained Equation Multiple Imputation for Incomplete Surveys

Humera Razzak, Christian Heumann
LMU, Germany

Using data based on health, development and nutrition indicators for children under 5 years old, from MICS 2014 Punjab, we present an easy to implement non-parametric dependent sequential regression multiple imputation technique to impute survey data with high dimensional categorical and continuous data. 3-stage Hybrid Multiple Imputation (HMI) approach can be used for missing data imputation in large scale surveys where sometimes log linear or sequential MI methods fail due to compatibility and various limitations. We estimate parameters of interest using GLM for continuous response (height for age percentiles NCHS) and various child health dependent co-variates. Contrast is made with existing MI approaches. Purposed HMI method shows better 95% confidence coverage rates and smaller root mean square errors for various settings as compared to the existing MI methods. Regardless of any combination, purposed MI method is markedly superior to the existing MI methods in terms of computational efficiency. Simulation results are supported by real data example.

401 Adaptive Discrete Smoothing for (High-Dimensional and Nonlinear) Panel Data

Martin Spindler[1], Xi Chen[2], Victor Chernozhukov[3], Ye Luo[4]
[1]University of Hamburg, Germany
[2]New York University
[3]MIT
[4]University of Hong Kong

In this paper we develop a data-driven smoothing technique for non-linear panel data models. We allow for individual specific (non-linear) functions and estimation with econometric or machine learning methods by using weighted observations from other individuals. The weights are determined by a data-driven way and depend on the similarity between the corresponding functions and are measured based on initial estimates. The key feature of such a procedure is that it clusters individuals based on the distance / similarity between them, estimated in a first stage. Our estimation method can be combined with various statistical estimation procedures, in particular modern machine learning methods which are in particular fruitful in the high-dimensional case and with complex, heterogeneous data. The methods can also be applied for estimation of Random Coefficient models and for estimation of nonparametric functions with many categorical variables ("cells"). The theoretical properties of the proposed estimators for important cases are derived which improve on many classical methods. We conduct a simulation study which shows that the prediction can be greatly improved by using our estimator. Finally, we analyze a big data set from didichuxing.com, a leading company in transportation industry, to analyze and predict the gap between supply and demand based on a large set of covariates. Our estimator clearly performs much better in out-of-sample prediction compared to existing non-linear panel data estimators.

Time Series Analysis VI (Time Series Resampling)

402 Detecting Regimes of Predictability in the U.S. Equity Premium

A.M. Robert Taylor
University of Essex, United Kingdom

We investigate the stability of predictive regression models for the U.S. equity premium. A new approach for detecting regimes of temporary predictability is proposed using sequential implementations of standard (heteroskedasticity-robust) regression t-statistics for predictability applied over relatively short time periods. Critical values for each test in the sequence are provided using subsampling methods. Our primary focus is to develop a real-time monitoring procedure for the emergence of predictive regimes using tests based on end-of-sample data in the sequential procedure, although the procedure could be used for an historical analysis of predictability. Our proposed method is robust to both the degree of persistence and endogeneity of the regressors in the predictive regression and to certain forms of heteroskedasticity in the shocks. We discuss how the detection procedure can be designed such that the false positive rate is pre-set by the practitioner at the start of the monitoring period. We use our approach to investigate for the presence of regime changes in the predictability of the U.S. equity premium at the one-month horizon by traditional macroeconomic and financial variables, and by binary technical analysis indicators. Our results suggest that the one-month ahead equity premium has temporarily been predictable (displaying so-called 'pockets of predictability'), and that these episodes of predictability could have been detected in real-time by practitioners using our proposed methodology.

403 Extending the validity of frequency domain bootstrap methods to general stationary processes

Marco Meyer[1], Efstathios Paparoditis[2], Jens-Peter Kreiß[1]
[1]TU Braunschweig, Germany
[2]University of Cyprus

Existing frequency domain methods for bootstrapping time series have a limited range. Essentially, these procedures cover the case of linear time series with independent innovations, and some even require the time series to be Gaussian. In this paper we propose a new frequency domain bootstrap method which is consistent for a much wider range of stationary, even nonlinear processes and can be applied to a large class of periodogram-based statistics. It introduces a new concept of convolved periodograms of smaller samples which uses pseudo periodograms of subsamples generated in a way that correctly imitates the weak dependence structure of the periodogram. We show consistency for this procedure for a general class of stationary time series, ranging clearly beyond linear processes, and for general spectral means and ratio statistics. Furthermore, we show how existing bootstrap methods can be corrected using this new approach. The finite sample performance of the new bootstrap procedure is illustrated via simulations.

404 Bootstrapping characteristic functions under local stationarity

Carina Beering[1], Carsten Jentsch[2], Anne Leucht[1], Marco Meyer[1]
[1]Technische Universität Braunschweig, Germany
[2]Technische Universität Dortmund, Germany

We propose a kernel-type estimator for the local characteristic function of locally stationary processes introduced by Dahlhaus in 1997. Under weak moment conditions, we provide a functional central limit theorem for the local empirical characteristic function process, which we are able to generalize further. Since in some cases asymptotic confidence intervals cannot be computed due to unknown parameters, we use the block bootstrap proposed in Dowla et al. in 2013 to generate a bootstrap estimator for the local CF. Subsequently, we show consistency for this method. Finally, we illustrate the finite sample behaviour of the procedure in a small simulation study for time-varying α-stable distributions for $\alpha \in (1, 2)$.
References:

(1) Dahlhaus, R. (1997): Fitting time series models to nonstationary processes. *The Annals of Statistics* 25, 1–37

(2) Dowla, A. et al. (2013). Local block bootstrap inference for trending time series. *Metrika* 76, 733–764

405 The impact of selecting the truncation indices on the order estimation of subspace methods—a simulation study with seasonally integrated processes.

Rainer Buschmeier

Bielefeld University, Germany

Estimation of state space systems by use of the subspace algorithm of Larimore (1983, Proceedings of the American Control Conference) is consistent for stationary, integrated, and seasonally integrated VARMA processes and is a promising initial estimator for subsequent quasi maximum likelihood estimation.
It requires the user to set truncation parameters f and p that determine to what extent present, future, and past observations are considered in the estimation. A proposed choice of f and p that is sufficient for consistent estimation of the system matrices is to let them depend on the sample size as $f = p = d\hat{k}_{AIC}$ with $d > 2$ (Bauer 2005, Econometric Theory, vol. 21). Here $\hat{k}_{AIC}$ is the lag length selected by AIC in auxiliary autoregressions.
One of the subsequent steps in the algorithm is to choose the system order n. A large enough n is crucial for the determination of the number of seasonal unit roots, since each additional common trend requires an additional state. Choosing n too large is not desirable, however, because it increases the number of parameters, the reduction of which is a major motivation behind using state space representations of VARMA processes. Fortunately, n can be estimated consistently by means of a singular value criterion (SVC) as in Bauer and Wagner (2009, Computational Statistics & Data Analysis, vol. 53) or Garcia-Hiernaux et al. (2012, Computational Statistics, vol. 27) and also other estimators exist based on information criteria or statistical tests.
For the case of a seasonally integrated process it is shown here using simulations that the range from which the system order is chosen for given f, p is very sensitive to changes in f and p in finite samples. Namely, even a slight increase in f and p may lead to a sudden increase in the estimated orders such that they mayy not include the true n anymore. An adequate choice of f and p is thus found to be crucial for the choice of n. In this regard it is shown by simulation that popular choices such as $\hat{k}_{AIC}$ lead to values of f, p that are too large, resulting in bad order estimation performances.
It is further investigated using simulation which of the order estimation procedures proposed in the literature performs best with the choice $f = p = d\hat{k}_{AIC}$. Secondly, the robustness of the various order selection criteria with respect to the choice of f and p is examined. Finally it is analyzed how sensitive the previous findings are to different persistence structures of the data generating process.

CONTENT INDEX

AUTHOR INDEX

Name, Page(s), (Abstract(s))

DAG STAT
DEUTSCHE ARBEITSGEMEINSCHAFT STATISTIK

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

www.dagstat2019.de
www.dagstat.de

Deutsche Statistische Gesellschaft • Internationale Biometrische Gesellschaft - Deutsche Region • Fachgruppe Stochastik der DMV • Gesellschaft für Klassifikation e.V. • Verband Deutscher Städtestatistiker • Fachbereich Biometrie der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V. • Verein zur Förderung des schulischen Stochastikunterrichts e.V. • Deutsche Gesellschaft für Epidemiologie e.V. • Ökonometrischer Ausschuss des Vereins für Socialpolitik • Fachgruppe Methoden und Evaluation der DGPs • Sektion Methoden der Empirischen Sozialforschung der DGS • European Network for Business and Industrial Statistics - Deutsche Sektion • Statistisches Bundesamt • Sektion Methoden der DVPW