Workshop on Goodness-of-fit, Change-point and Related Problems (GOFCP)

Book of Abstracts

Prague, August 2025

Contents

Welcome

We are delighted to welcome you to the Workshop on Goodness-of-fit, Change-point and Related Problems (GOFCP) 2025, taking place in the beautiful city of Prague, Czech Republic. The workshop focuses on key developments in goodness-of-fit testing and change-point analysis across a range of contemporary statistical models.

The scientific program will follow a single-session format, featuring invited oral presentations and two engaging poster sessions. This booklet includes abstracts for each presentation. The two poster sessions will take place during the coffee breaks. Presenters are kindly asked to place their posters in the designated areas before the session (e.g., in the morning of their presentation day) and to be present near their posters during the corresponding coffee break.

We thank the Scientific Programme Committee for their valuable efforts in assembling an outstanding lineup of speakers. We would also like to express our gratitude to North-West University, South Africa, for their financial support. Finally, our heartfelt appreciation goes to all participants for joining us. We hope you find the workshop both stimulating and enjoyable.

We wish you a productive experience and a memorable stay in Prague!

GOFCP 2025 Local Organising Committee

Committees and programme

Scientific Programme Committee

- Ayanendranath Basu, Indian Statistical Institute
- Giuseppe Cavaliere, Universitá di Bologna
- John Einmahl, Tilburg University
- Daniel Hlubinka, Univerzita Karlova
- Claudia Kirch, Otto von Guericke Universität Magdeburg
- Natalie Neumeyer, Universität Hamburg
- Richard Samworth, University of Cambridge

Organizing Committee

- Zdeněk Hlávka (Charles University, Czech Republic)
- Daniel Hlubinka (Charles University, Czech Republic)
- Šárka Hudecová (Charles University, Czech Republic)
- Marie Hušková (Charles University, Czech Republic)
- Michal Pešta (Charles University, Czech Republic)
- James Allison (North-West University, South Africa)
- Elzanie Bothma (North-West University, South Africa)
- Jaco Visagie (North-West University, South Africa)
- Nombuso Zondo (North-West University, South Africa)

PROGRAM: DAY 1

GOFCP 2025

Prague, Czech Republic

Day	From	To	Chair	Speaker	Talk
Thursday 08/28	16:00	20:00			*Registration and glass of beer or/and wine*
Friday 08/29	8:30	9:00			*Registration*
	9:00	9:10		Michal Pešta	Welcome and opening
			Session I: Change point in regression		
	9:15	9:40	Claudia Kirch	Lajos Horváth	Change point detection and confidence intervals for the times of changes in linear models with heteroscedastic errors
	9:40	10:05		Silvelyn Zwanzig	Some non-asymptotic rank tests for change points in regression
	10:05	10:30		Hušková Marie	Change point detection, development and remarks (personal view)
	10:30	11:10			*Coffee break*
			Session II: Structural tests for complex models		
	11:10	11:35	Thomas Verdebout	Natalie Neumeyer	Statistical inference for the error distribution in functional linear models
	11:35	12:00		Anne Leucht	Gaussian approximation for lag-window estimators of spectral densities
	12:00	12:25		Siegfried Hörmann	On epidemic change point detection under strong mixing conditions
	12:25	12:50		Michal Pešta	UWLLN and MFCLT with random number of summands
	12:50	14:20			*Lunch break*

PROGRAM: DAY 1

GOFCP 2025

Prague, Czech Republic

Day	From	To	Chair	Speaker	Talk
Friday 08/29			**Session III: Change point in time series**		
	14:20	14:45	Michal Pešta	Claudia Kirch	Data segmentation procedures for time series
	14:45	15:10		Rebecca Killick	Generalised mixed effects models for changepoint analysis of biomedical time series
	15:10	15:35		Yudong Chen	Inference on dynamic spatial autoregressive models with change point detection
	15:35	16:00		Yi He	Testing for spurious factor analysis on high dimensional nonstationary time series
	16:00	16:40			*Coffee break and poster session I.*
			Session IV: Estimation problems		
	16:40	17:05	James Allison	Noël Veraverbeke	New asymptotic results for Bernstein estimators for conditional copulas
	17:05	17:30		Aurore Delaigle	Nonparametric density estimation from streaming data
	19:00	22:00			*Welcome reception*

PROGRAM: DAY 2

GOFCP 2025

Prague, Czech Republic

Day	From	To	Chair	Speaker	Talk
Saturday 08/30			**Session V: Session dedicated to Simos G. Meintanis**		
	9:00	9:10	Marie Hušková	James Allison	Memories of Simos G. Meintanis
	9:10	9:35		María Dolores Jiménez-Gamero	Testing for the homogeneity of many populations using Gini covariance
	9:35	10:00		Bojana Milošević	Homogeneity testing in the presence of missing data
	10:00	10:25		Thomas Verdebout	Detection thresholds of tests for uniformity on hyperspheres
	10:25	10:50		Bruno Ebner	New uniformity tests on the hypersphere based on the ECF and Stein characterizations
	10:50	11:30			*Coffee break and poster session II.*
			Session VI: Inference based on distances		
	11:30	11:55	Richard Samworth	Ayanendranath Basu	Parametric estimation versus goodness-of-fit tests: The conflicts and the reconciliations
	11:55	12:20		Abhik Ghosh	Robust tests of hypotheses based on a general family of S-divergences
	12:20	12:45		Wolfgang Stummer	High-dimensional model search by means of scaled Bregman distances
	12:45	13:10		Stanislav Volgushev	On a surprising behavior of the likelihood ratio test in non-parametric mixture models
	13:10	14:40			*Lunch break*
	14:45	17:30			*Exploring Prague city center: Vintage tram and walking tour*

PROGRAM: DAY 3

GOFCP 2025

Prague, Czech Republic

Day	From	To	Chair	Speaker	Talk
Sunday 08/31			**Session VII: Problems of statistical learning**		
	9:00	9:25	Aurore Delaigle	Richard Samworth	Deep learning with missing data
	9:25	9:50		Rajen Shah	Hunt and test for assessing the fit of semiparametric regression models
	9:50	10:15		Hao Chen	Change-point detection for modern complex data
	10:15	10:40		John Einmahl	Testing for omitted variables in binary single-index models
	10:40	11:05			*Coffee break*
			Session VIII: Change point in high dimensional data		
	11:05	11:30	Rebecca Killick	Marie-Christine Düker	Multitask learning for high-dimensional time series
	11:30	11:55		Tengyao Wang	Robust mean change point testing in high-dimensional data with heavy tails
	11:55	12:10		Haeran Cho	Covariance scanning for adaptively optimal change point detection in high-dimensional linear models
	12:10	12:35		Daniel Hlubinka	*Closing*

Social program

Registration and welcome drink (Thursday evening)

Registration for the conference starts on Thursday, August 28 at 16:00 at the conference venue, Sokolovská 83, located in the Prague district of Karlín. We invite you to raise a **glass of local Prague beer or wine** as we kick off the conference. You can come **anytime between 16:00 and 20:00**, and the event is also open to accompanying persons.

If you are not able to come on Thursday, registration will also be available on Friday morning.

Welcome reception (Friday evening)

All registered conference participants are kindly invited to the welcome reception, which will take place on Friday, August 29 in the historical early Baroque building, Profesní dům (The House for Professed), located in Malá Strana. The address is Malostranské nám. 2/25.

The reception **starts at 19:00**. If you arrive earlier, we recommend exploring the beautiful neighbourhood of Malá Strana.

How to get to Profesní dům from the conference venue:

- **By tram**: Take tram number 12 from the Křižíkova stop (100 m from the conference venue) or the Karlínské náměstí stop (100 m from Hotel Fitzgerald) to the stop **Malostranské náměstí** (in the direction of Florenc; the final stop is Sídliště Barrandov).

  The ride takes approximately 20 minutes, with trams running every 10 minutes. We recommend taking the tram that departs from Křižíkova at 18:20 or 18:30. The building of Profesní dům is located directly in front of the tram stop.

- **By a walk through the Prague city center**: You can take tram 8 or metro line B from Křižíkova to Náměstí Republiky, and then enjoy a scenic walk. Start by passing the Art Nouveau building of Municipal house (Obecní dům), continue along Celetná Street to the historic Old Town Square(Staroměstské náměstí). Then cross the iconic stone Charles bridge (Karlův most) to reach **Malostranské náměstí**. A map of this recommended route is available online at this link.

  The walk is approximately 2.5 km and takes about 40 minutes, so we recommend leaving around 18:00 from Karlín (conference venue or Hotel Fitzgerald).

Tour in Prague city center (Saturday afternoon)

They say the most beautiful view of the world is from horseback. But the most beautiful view of Prague is definitely from a tram.

Saturday afternoon will be devoted to a trip to Prague city center. A vintage tram will take us on a sightseeing ride past the National Theatre and Prague Castle to the district of Střešovice. From there, we will walk to Prague Castle and then to Malá Strana.

Important information about the tram tour. We will board the tram at Těšnov stop at approximately 15:20. No delay is possible, as this is a regular tram stop – the tram must depart on time and cannot wait for anyone. The trip is available to **all registered conference participants**, so you do not need any ticket.

If you would like to join us for this trip, please:

- Meet us in front of the conference venue building **at 14:45**. From there, one group will walk (1.2 km), and the other will take a tram together.
- Alternatively, you may go directly to the Těšnov stop on your own — but please arrive by 15:10 at the latest (it can happen that the tram comes earlier and needs to leave).

Information about the walk. A map of the approximate route is available here. If you do not feel up to the whole trip, you are welcome to leave at any time – to stop by a café/pub or return to your hotel. On the other hand, if you still feel fresh and motivated after the tour, feel free to continue on your own and explore other parts of the city center.

Information about the vintage tram. The vintage tram is a beautifully preserved vehicle that represents a cherished part of Prague's public transport heritage. Today, it is used for special events, nostalgic rides, and sightseeing tours (as ours).

It was produced by the Ringhoffer factory in Prague, one of the largest tram manufacturers in the Austro-Hungarian Empire.

The tram was introduced in 1908, at a time when Prague was still part of the Austro-Hungarian Empire and the electrification of tram lines was nearing completion, replacing horse-drawn trams. These vehicles represented cutting-edge urban transport and soon became a symbol of Prague's modernization. By joining the trip, you can experience the spirit of the Austro-Hungarian Empire's final era.

Invited talks

Parametric estimation versus goodness-of-fit tests: The conflicts and the reconciliations

Ayanendranath Basu
Indian Statistical Institute, Kolkata, India

Statistical distances – or divergences – have an important role in many areas of statistical inference including parametric estimation and goodness-of-fit testing. It is in many ways natural to do parametric estimation by choosing the model element that minimizes the "separation" of the empirical data to the assumed model structure; indeed the maximum likelihood estimator is a part of this philosophy, which minimizes the Kullback-Leibler divergence between the empirical and the model. One may, more generally, consider larger classes of divergences for the same purpose, and in this talk we will fix our attention on the class of ϕ-divergences, of which the Kullback-Leibler divergence is a member. Depending on the particular member of this family chosen for inference, the small sample properties and off-the-model properties of the corresponding estimator can be very different, although under suitable assumptions their asymptotic model distributions are identical. All of these divergences have legitimate applications in the area of goodness-of-fit testing. It is of interest, however, to note that the divergences which are desirable for their off-the-model properties, are not necessarily the ones that are desirable for goodness-of-fit purposes. In this lecture we will make a distinction between some of the philosophies of parametric estimation and goodness-of-fit testing, and give a geometrical perspective of what puts the desirability of these two philosophies in conflict, and which divergences are suitable for which purposes.

Change-point detection for modern complex data

Hao Chen
Department of Statistics, University of California, Davis

Change-point analysis is thriving in this big data era, addressing problems that arise across many fields where massive data sequences are collected to study complex phenomena over time. It plays a crucial role in processing these data by segmenting long sequences into homogeneous parts for subsequent studies. Observations could be high-dimensional or not lie in the Euclidean space, such as network data, which are challenging to characterize using parametric models. We utilize the inter-point information of the observations and propose a series of nonparametric methods to address the issue. In particular, we take into account a pattern caused by the curse of dimensionality so that the proposed methods can accommodate a broad range of alternatives. Additionally, we work out ways to analytically approximate the p-values of the test statistics, enabling rapid type I error control. The methods are applied to Neuropixels data in the analysis of thousands of neurons' activities.

Inference on dynamic spatial autoregressive models with change point detection

Zetai Cen*, **Yudong Chen**†, Clifford Lam*
* Department of Statistics, London School of Economics,
† Department of Statistics, University of Warwick

We analyze a varying-coefficient dynamic spatial autoregressive model with spatial fixed effects. One salient feature of the model is the incorporation of multiple spatial weight matrices through their linear combinations with varying coefficients, which help solve the problem of choosing the most "correct" one for applied econometricians who often face the availability of multiple expert spatial weight matrices. We estimate and make inferences on the model coefficients and coefficients in basis expansions of the varying coefficients through penalized estimations, establishing the oracle properties of the estimators and the consistency of the overall estimated spatial weight matrix, which can be time-dependent. We further consider two applications of our model in change point detections in dynamic spatial autoregressive models, providing theoretical justifications in consistent change point locations estimation and practical implementations. Simulation experiments demonstrate the performance of our proposed methodology, and real data analyses are also carried out.

Covariance scanning for adaptively optimal change point detection in high-dimensional linear models

Haeran Cho
University of Bristol

This paper studies the detection of a change in high-dimensional linear models. We derive the minimax lower bounds on the detection boundary and the rate of estimation which exhibit a phase transition with respect to the sparsity of the covariance-weighted differential parameter, revealing a delicate interaction between the covariance of regressors and the change in regression parameters.

We complement these results by proposing methods that achieve minimax (near-)optimality in the sparse and the dense regimes, respectively. Referred to as McScan and QcScan, they scan the maximum and the quadratic aggregations of the local covariances at strategically selected locations for change point detection; in particular, QcScan is the first method shown to achieve consistency in the dense regime.

Further, a combined method is proposed which is adaptively optimal even when the sparsity is unknown, and we complete the study of the change point problem by considering post-detection estimation of the differential parameter and the refinement of the change point estimators. Numerical experiments confirm the new findings as well as demonstrating the computational and statistical efficiency of the covariance-scanning based methods.

This is joint work with Tobias Kley and Housen Li (University of Göttingen).

Nonparametric density estimation from streaming data

Aurore Delaigle
University of Melbourne, Australia

We consider nonparametric density estimation from streaming data such as observations collected from sensor networks. Those data are characterized by their continuous collection over

time in a high-velocity and often nonstationary environment, requiring near-real-time low-storage processing methods. We study the properties of an iterative estimator, which does not require storing data for long periods of time nor accessing them repeatedly. Then we suggest a procedure for implementing it in practice.

Multitask learning for high-dimensional time series

Marie-Christine Düker
Friedrich-Alexander-Universität Erlangen–Nürnberg (FAU), Germany

This talk introduces a novel framework for multitask learning in high-dimensional time series through hypothesis testing and data integration. The proposed procedure tests for shared structures across multiple high-dimensional factor models, determining whether they are driven by the same loading vectors up to a linear transformation. Leveraging repeated applications of singular value decomposition, the framework achieves consistent estimation of shared and non-shared loading vectors and introduces a sequential testing procedure to estimate the number of shared components. Theoretical results establish the asymptotic behavior of the test statistics and consistency. The method applies to multitask frameworks to uncover inter-individual relationships between datasets and to detecting structural changes over time in a single factor model. Applications to macroeconomic data demonstrate its practical utility.

New uniformity tests on the hypersphere based on the ECF and Stein characterizations

Bruno Ebner
Karlsruher Institut für Technologie (KIT), Institut für Stochastik, Germany

In this talk, we present two new goodness-of-fit tests of weighted L^2-type for uniformity on the d-dimensional hypersphere, based on the empirical characteristic function and Stein characterizations. We provide explicit formulas for calculation of the test statistics, and examine the asymptotic properties of these tests, including the limiting null distribution, the limit distribution under contiguous and fixed alternatives, and show their omnibus consistency. A comprehensive Monte Carlo simulation study demonstrates that our approach is competitive with the vast number of existing procedures in the literature. The results demonstrate that our tests not only provide competitive power across a range of alternatives but also offer robust performance across different distributional settings, outperforming existing methods in several cases.

This is a joint work with Paul Axmann and Eduardo García-Portugués.

Testing for omitted variables in binary single-index models

John H.J. Einmahl
Tilburg University

We wish to test whether a real-valued variable Z has explanatory power, in addition to a multivariate variable X, for a binary variable Y. Thus, we are interested in testing the hypothesis $P(Y = 1 \mid X, Z) = P(Y = 1 \mid X)$, based on n i.i.d. copies of (X, Y, Z). In order to avoid the curse of dimensionality, we follow the common approach of assuming that the dependence of both Y and Z on X is through a single-index vector only.

Based on the Y-values, we then construct a two-sample empirical process of the transformed Z-variables, after splitting the X-space into a number of strips. Studying this two-sample empirical process is challenging: it does not converge weakly to a standard Brownian bridge, but after an appropriate normalization it does. That result is, of course, the basis for obtaining the required tests.

This is joint work with Denis Kojevnikov and Bas J.M. Werker, both also at Tilburg University.

Robust tests of hypotheses based on a general family of S-divergences

Abhik Ghosh and Ayanendranath Basu
Indian Statistical Institute, Kolkata, India

Robust inference based on the minimization of statistical divergences has proved to be a useful alternative to the classical techniques based on maximum likelihood and related methods. Two such popular (one-parameter) families of divergence measures are the power divergences (PDs) of Cressie and Read (1984) and the density power divergences (DPDs) of Basu et al. (1998), which are seen to be extremely useful in generating highly efficient and robust parametric inference procedures under possible data contamination (e.g., outliers). A general two-parameter family of divergence measures, termed as the S-divergences, is later proposed by Ghosh et al. (2017) in the context of robust parameter estimation. Interestingly, this larger family of S-divergences contains both the power divergences and the density power divergences as its subclasses; but it also introduces some new divergence measures offering better trade-offs between efficiency and robustness compared to both the PD and DPD families. In this paper, we will explore the usefulness of this general family of S-divergences in developing robust tests for different hypothesis testing problems including, but not limited to, the goodness-of-fit tests. We will derive asymptotic properties of the proposed tests and also assess their robustness theoretically through the influence function analyses. The practical relevance and effectiveness of these tests will be further illustrated using some interesting data applications.

References

Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, **85**, 549–559.

Cressie, N. and Read, T. R. C. (1984). Multinomial goodness-of-fit tests. *Journal of the Royal Statistical Society B*, **46**, 440–464.

Ghosh, A., Harris, I. R., Maji, A., Basu, A. and Pardo, L. (2017). A generalized divergence for statistical inference. *Bernoulli*, **23(4A)**, 2746–2783.

Testing for spurious factor analysis on high dimensional nonstationary time series

Yi He
University of Amsterdam

Spurious factor behaviors arise in large random matrices with high-rank random signal components. This paper establishes analytical probabilistic limits and a distribution theory for these spurious behaviors in high-dimensional non-stationary time series.

We transform scree plots into Hill plots to detect spectral patterns in these spurious factor

models and develop tests to distinguish between spurious and genuine factor models. Simulations confirm the excellent size and power performance of our test in finite samples. Applying the tests to three real-life datasets, we detected spurious factors in both economic and climate data, and genuine factors in finance data.

This is joint work with Bo Zhang (University of Science and Technology of China).

On epidemic change point detection under strong mixing conditions

Siegfried Hörmann
Graz University of Technology

We analyse epidemic shifts in the mean of a time series. Here, an epidemic refers to a contiguous segment of observations within the sample where the mean shift occurs.

We derive the limiting law of our test statistic by a novel almost sure approximation for α-mixing processes. The precision of the remainder term in our approximation ensures that our test remains consistent even when the epidemic's duration scales logarithmically with the sample size.

This talk is based on joint work with István Berkes.

Change point detection and confidence intervals for the times of changes in linear models with heteroscedastic errors

Lajos Horváth
University of Utah, USA

First we test the no change in the parameter of a linear model against the alternative if there are R changes in the parameter. Heteroscedasticity of the errors is allowed. We use a weighted CUSUM process and establish the limit distribution of its supremum functional. We discuss how the heteroscedasticity of the errors affects the limit. Simulations show that neglecting possible heteroscedasticity can lead to incorrect conclusions.

In the second part of the talk, we consider the problem of estimating and deriving confidence intervals for change points in linear models with heteroscedastic errors. A CUSUM process-based estimator is proposed that is derived from weighted model residuals, and we establish its asymptotic properties when the linear regression model exhibits change points in both the regression parameters and the distribution of the errors. This theory motivates the construction of confidence intervals for multiple change points by starting with preliminary change point estimators that are refined and approximated with a distribution that is robust to heteroscedasticity. We also provide some applications.

The talk is based on joint work with Gregory Rice (University of Waterloo, Canada) and Yuqian Zhao (University of Sussex, United Kingdom).

Change point detection, development and remarks (personal view)

Marie Hušková
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague

The talk concerns (some) achievements in the area of Change Point Detection since late eighties of the last century. The particular attention is payed to the results where someone from Prague has participated. Both off-line and on-line procedures are covered.

Testing for the homogeneity of many populations using Gini covariance

María Dolores Jiménez-Gamero and María Remedios Sillero Denamiel
Universidad de Sevilla and Instituto de Matemáticas de la Universidad de Sevilla, Spain

Given k populations and assuming that independent samples are available from each of them, the problem of testing for the equality of the k populations is addressed. With this aim, an unbiased estimator of the Gini covariance is taken as test statistic. In contrast to the classical setting, where k is kept fixed and the sample size from each population increases without bound, here k is assumed to be large and the size of each sample is small in comparison to k. The asymptotic distribution of the test statistic is stated under the null hypothesis as well as under alternatives, which allows us to study the consistency of the test. Specifically, it is shown that the test statistic is asymptotically free distributed under the null hypothesis. The finite sample performance of the test based on the asymptotic null distribution is studied via simulation. The proposal is applied to a real data set.

Generalised mixed effects models for change-point analysis of biomedical time series

Rebecca Killick
Lancaster University and University of California Santa Cruz

Motivated by two distinct types of biomedical time series data, digital health monitoring and neuroimaging, we develop a novel approach for changepoint analysis that uses a generalised linear mixed model framework. The generalised linear mixed model framework lets us incorporate structure that is usually present in biomedical time series data.

We embed the mixed model in a dynamic programming algorithm for detecting multiple changepoints in biomedical data. We evaluate the performance of our proposed method across several scenarios using simulations. Finally, we show the utility of our proposed method on our two distinct motivating applications: fMRI and passive monitoring of older adults.

Data segmentation procedures for time series

Claudia Kirch
Otto-von-Guericke University Magdeburg

The segmentation of a time series into piecewise stationary segments, a.k.a. multiple change point analysis, is an important problem both in time series analysis and signal processing. In the presence of multiscale change points with both large jumps over short intervals and small

changes over long stationary intervals, multiscale methods achieve good adaptivity in their localisation but at the same time require the removal of false positives and duplicate estimators via some kind of model selection step.

In this talk, we discuss methodology based on moving sum procedures with multiple window lengths, including general methodology based on estimating functions that can be used in a variety of situations, from robust methodology for the detection of changes in the mean to the detection of changes in linear and non-linear time series including count time series.

The talk will be based on joint work with various co-authors, including Haeran Cho, Euan McGonigle, Birte Muhsal, Kerstin Reckruehm and Niclas Stoffregen.

Gaussian approximation for lag-window estimators of spectral densities

Anne Leucht
Universität Bamberg, Germany

We develop a Gaussian approximation for the maximum a suitably normalized lag-window estimator of the spectral density evaluated at all positive Fourier frequencies. This max-statistic plays an essential role to construct goodness-of-fit tests for the second-order structure of stationary time series and to derive simultaneous confidence bands for the spectral density. The Gaussian approximation opens up the possibility to verify asymptotic validity of a multiplier bootstrap procedure and, even further, to derive the corresponding rate of convergence. A small simulation study sheds light on the properties of this bootstrap proposal for finite samples.

This talk is based on a joint work with Jens-Peter Kreiss (Technische Universität Braunschweig) and Efstathios Paparoditis (Cyprus Academy of Sciences, Letters and Arts).

Homogeneity testing in the presence of missing data

Danijel Aleksić and **Bojana Milošević**
University of Belgrade, Faculty of Mathematics and Faculty of Organizational Sciences, Serbia

Here, we explore the problem of homogeneity testing when data are subject to missingness under various missing data mechanisms. We focus on energy distance-based tests and their generalizations. Beyond the standard complete-case approach, we propose a novel adaptation of the energy test statistic that takes advantage of all available information. Appropriate resampling-based approaches are developed for p-value approximation in this setting.

Additionally, we introduce a tailored bootstrap procedure designed for settings where the test statistic is evaluated on datasets that have been imputed using common imputation techniques. An extensive simulation study examines the performance of the proposed methods across various sample sizes, dimensions, underlying distributions, missingness mechanisms, and proportions of missing data. Based on the findings, we provide practical recommendations to guide the use of homogeneity tests in scenarios with incomplete data.

Statistical inference for the error distribution in functional linear models

Natalie Neumeyer, Melanie Birke, and Leonie Selk
University of Hamburg and University of Bayreuth, Germany

Some recent results on functional linear models $Y = \alpha + \langle X, \beta \rangle + \varepsilon$ with scalar response Y and functional covariate X are presented. In those models it is more challenging to deal with residual-based procedures than in regression models with vector-valued covariates. Typically, estimating the functional parameter β will dominate the asymptotic distribution of test statistics based on residuals $\hat{\varepsilon} = Y - \hat{\alpha} - \langle X, \hat{\beta} \rangle$, with a slower rate of convergence. As one example, we consider a test for independence of errors and covariates based on joint empirical characteristic functionals of residuals and functional covariates. Yet under simple assumptions there are change-point and goodness-of-fit tests for the error distribution with the same asymptotic distribution as for classical linear models. The suggested tests are based on residual empirical distribution functions and residual empirical characteristic functions.

UWLLN and MFCLT with random number of summands

Michal Pešta
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague

Within the ambit of multivariate counting processes, we generalize Korolyook's theorem and Wald's identity. Consequently, a uniform weak law of large numbers and a multivariate functional central limit theorem will be established. An application to change point analysis of counting processes is sketched.

This is joint work with Marie Hušková.

Deep learning with missing data

Richard Samworth
University of Cambridge

In the context of multivariate nonparametric regression with missing covariates, we propose Pattern Embedded Neural Networks (PENNs), which can be applied in conjunction with any existing imputation technique. In addition to a neural network trained on the imputed data, PENNs pass the vectors of observation indicators through a second neural network to provide a compact representation. The outputs are then combined in a third neural network to produce final predictions.

Our main theoretical result exploits an assumption that the observation patterns can be partitioned into cells on which the Bayes regression function behaves similarly and belongs to a compositional Hölder class. It provides a finite-sample excess risk bound that holds for an arbitrary missingness mechanism, and in combination with a complementary minimax lower bound, demonstrates that our PENN estimator attains in typical cases the minimax rate of convergence as if the cells of the partition were known in advance, up to a poly-logarithmic factor in the sample size.

Numerical experiments on simulated, semi-synthetic, and real data confirm that the PENN estimator consistently improves, often dramatically, on standard neural networks without pattern embedding.

Hunt and test for assessing the fit of semiparametric regression models

Rajen Shah
University of Cambridge

We consider testing the goodness of fit of semiparametric regression models, such as generalised additive models, partially linear models, or quantile additive regression models. We propose an approach that involves first splitting the data in two parts. On one part, we "hunt" for any signal that may be present in the score-type residuals following a fit of the model. On the remaining data, we test for the presence of the potential signal thus found.

For the first hunting stage of the procedure, our framework allows for carrying this out based on a user-chosen flexible regression method, such as a random forest. The method is thus able to leverage the power of modern machine learning methods to detect complex alternatives. A challenge with the testing step is that any first-order bias in the residuals may lead to rejection under the null. We therefore employ a debiasing step which we show amounts to performing a particular weighted least squares regression.

We show that the type I error may be controlled under relatively mild conditions, and that we have power under alternatives where with high probability the hunted signal is correlated with the true signal present in the score residuals.

High-dimensional model search by means of scaled Bregman distances

Wolfgang Stummer
Friedrich-Alexander-Universität Erlangen–Nürnberg (FAU), Germany

In statistics — as well as in the adjacent fields of machine learning, artificial intelligence, information theory, signal processing and pattern recognition — it is well known that goodness-of-fit, model-search and model-adequacy tasks can be performed by means of appropriate constrained minimization of directed distances (i.e. divergences), where the latter basically measure dissimilarities between the underlying quantities.

In this talk, we employ the *Scaled Bregman Distances* (cf. Stummer (2007), Stummer & Vajda (2012); see also Kißlinger & Stummer (2018) for change-detection applications) which *cover both* the omnipresent class of f-divergences/disparities (e.g. Kullback-Leibler divergence, Pearson chi-square divergence, (squared) Hellinger distance, GNED) *as well as* the omnipresent class of Bregman distances (e.g. density power divergences of Basu et al. (1998), (squared) L_2-distance, Bregman exponential divergence).

We present how to solve high-dimensional constrained optimization problems of Scaled Bregman Distances, in terms of the dimension-free precise bare-simulation method developed in Broniatowski & Stummer (2023, 2024); for this, almost no assumptions (like convexity) on the set of constraints are needed, which makes this approach e.g. applicable to non-parametric or semi-parametric modelling frameworks.

References

A. Basu, I. R. Harris, N. L. Hjort, and M.C. Jones. Robust and efficient estimation by minimizing a density power divergence. *Biometrika*, 85(3): 549–559, 1998.

M. Broniatowski and W. Stummer. A precise bare simulation approach to the minimization of some distances. I. Foundations. *IEEE Trans. Inf. Theory*, 69(5): 3062–3120, 2023.

M. Broniatowski and W. Stummer. A precise bare simulation approach to the minimization of some distances. II. Further foundations. *arXiv:2402.08478*, pages 1–77, 2024.

A.-L. Kißlinger and W. Stummer. A new toolkit for robust distributional change detection. *Appl. Stochastic Models Bus. Ind.*, 34: 682–699, 2018.

W. Stummer. Some Bregman distances between financial diffusion processes. *Proc. Appl. Math. Mech.*, 7(1): 1050503–1050504, 2007.

W. Stummer and I. Vajda. On Bregman distances and divergences of probability measures. *IEEE Trans. Inf. Theory*, 58(3): 1277–1288, 2012.

New asymptotic results for Bernstein estimators for conditional copulas

Noël Veraverbeke
University of Hasselt, Belgium and North-West University, Potchefstroom, South Africa

Conditional copulas are very essential in the modeling of dependence in multivariate data in the presence of a random covariate. Several authors studied the asymptotics for the conditional empirical copula function. Bernstein polynomials provide an interesting tool for obtaining smooth versions of these non-parametric estimators. Here we provide new asymptotic results for Bernstein based versions of estimators for a conditional copula, its first order partial derivative and its density function. As an application we deal with the estimation of a risk ratio for bivariate data in the presence of a covariate.

Detection thresholds of tests for uniformity on hyperspheres

Thomas Verdebout
Department of Mathematics and ECARES, ULB, Belgium

One of the most classical problems in multivariate statistics is considered, namely, the problem of testing isotropy, or equivalently, the problem of testing uniformity on the unit hypersphere. We consider various tests for the problem including the broad class of Sobolev tests. While some of these tests are known to allow for omnibus testing of uniformity, their non-null behavior and consistency rates, unexpectedly, remain largely unexplored both in the low and in the high-dimensional case. We thoroughly study the asymptotic powers of several tests including Sobolev tests. We particularly focus on consistency rates and detection thresholds.

On a surprising behavior of the likelihood ratio test in non-parametric mixture models

Stanislav Volgushev
University of Toronto

We study the likelihood ratio test in general mixture models where the base density is parametric, the null is a fixed mixing distribution, and the alternative is a general mixing distribution supported on a bounded parameter space. For Gaussian location mixtures and Poisson mixtures, we show a surprising result: the non-parametric likelihood ratio test statistic converges to a tight limit if and only if the null distribution is a finite mixture, and diverges to infinity otherwise. We further demonstrate that the likelihood ratio test diverges for a fairly general

class of distributions when the null mixing distribution is not finitely discrete.

This is joint work with Yan Zhang from the University of Toronto.

Robust mean change point testing in high-dimensional data with heavy tails

Mengchu Li, Yudong Chen, **Tengyao Wang**, Yi Yu
University of Birmingham; University of Warwick; London School of Economics and Political Science, UK

We study mean change point testing problems for high-dimensional data, with exponentially- or polynomially-decaying tails. In each case, depending on the ℓ_0-norm of the mean change vector, we separately consider dense and sparse regimes. We characterise the boundary between the dense and sparse regimes under the above two tail conditions for the first time in the change point literature and propose novel testing procedures that attain optimal rates in each of the four regimes up to a poly-iterated logarithmic factor. By comparing with previous results under Gaussian assumptions, our results quantify the costs of heavy-tailedness on the fundamental difficulty of change point testing problems for high-dimensional data.

To be specific, when the error distributions possess exponentially-decaying tails, a CUSUM-type statistic is shown to achieve a minimax testing rate up to $\sqrt{\log\log(8n)}$. As for polynomially-decaying tails, admitting bounded α-th moments for some $\alpha \geq 4$, we introduce a median-of-means-type test statistic that achieves a near-optimal testing rate in both dense and sparse regimes. In the sparse regime, we further propose a computationally-efficient test to achieve optimality. Our investigation in the even more challenging case of $2 \leq \alpha < 4$, unveils a new phenomenon that the minimax testing rate has no sparse regime, i.e. testing sparse changes is information-theoretically as hard as testing dense changes. Finally, we consider various extensions where we also obtain near-optimal performances, including testing against multiple change points, allowing temporal dependence as well as fewer than two finite moments in the data generating mechanisms. We also show how sub-Gaussian rates can be achieved when an additional minimal spacing condition is imposed under the alternative hypothesis.

Some non-asymptotic rank tests for change points in regression

Silvelyn Zwanzig and Rauf Ahmad
Uppsala University, Sweden

Rank-based tests for detecting the presence of change point in a linear regression model are considered. The main trick is to reduce a test problem in linear regression to a test problem for two i.i.d. samples. This makes an exact rank test possible. In simple linear regression, Theil's idea of data reduction is to consider the slope of all connecting lines between the data points instead of the original observations. This eliminates the regression parameters. We use this idea to generate convenient subsets of slopes which deliver i.i.d. samples, so that we can reduce the problem of testing for a known change point in regression to testing for a change of location of two i.i.d. samples. Extensions to multivariate and errors-in-variables models are also discussed.

Poster session I (Friday)

Independence testing: A new approach for mixed-type multivariate data

Dana Bucalo Jelić, **Marija Cuparić**, Bojana Milošević
University of Belgrade, Faculty of Mathematics, Serbia

We address the problem of testing independence in mixed-type data settings, where components may be discrete or positive and absolutely continuous, considering both the independence between these vectors, and total independence. The tests are based on an integral and an L^2 transformation of a special function of the distribution function, introduced by Barnighaus and Gaigall, which effectively characterizes the joint distribution in such complex, multivariate scenarios. The asymptotic properties of the proposed tests are established, and their practical relevance in higher-dimensional contexts is demonstrated through an extensive power study.

Heavy-tailed robust estimation of factor-adjusted vector autoregressive models for high-dimensional time series

Dylan Dijk
University of Bristol, UK

We study factor-adjusted vector autoregressive (VAR) models in the presence of heavy-tailed data. That is, after accounting for pervasive co-movements of the variables with a static factor model, we model the remaining idiosyncratic process as a sparse VAR process, where we implement heavy-tailed robust autocovariance estimators for the estimation of both the common component and Lasso estimation of the sparse VAR process. In particular, we utilise the elementwise truncated autocovariance estimator, and tune the truncation parameter with cross-validation. The estimation convergence rates are then provided, in the high-dimensional setting, under the assumption that for both the factor and idiosyncratic process there exists a bounded $(2 + 2\epsilon)$-th moment for some $\epsilon \in (0, 1)$. Simulation studies demonstrate the performance of the robust procedure, and how the truncation parameter varies with the dimension, sample size, and distribution of the data.

Modelling dynamics in soil moisture time series with an extended Bayesian changepoint detection algorithm

Mengyi Gong, Christopher Nemeth, Rebecca Killick, John Quinton
Lancaster University, UK

Continuous monitoring of soil moisture content via modern sensors has opened up a unique opportunity to investigate the dynamics of soil health. The abundance of data and the com-

plexity in the time series data requires the development of statistical methodologies to analyse the data appropriately. For example, the soil moisture time series can display an exponential decay pattern when the soil is relatively dry and a fluctuating pattern when the soil is close to saturation.

In this study, we proposed an extension to the Bayesian online changepoint detection (BOCPD) method to model soil moisture time series with complex patterns. In particular, the proposed method segments the time series data and assigns each segment with potentially different types of models in order to better describe the structure of the segment. It combines sequential Monte Carlo or online gradient ascent with the classic BOCPD to estimate the key unknown parameters in the models. Simulation study was carried out to compare the performance of the proposed method to the classic BOCPD method. The possibilities of reducing computational cost has been investigated. The method was applied to the soil moisture time series from the US National Ecological Observatory Network and the Alpine regions of Austria, where the soil experienced distinctively different climatic patterns over the year.

Goodness-of-fit tests with bivariate Monte Carlo critical regions

Zdeněk Hlávka, Muhammad Nauman Khan
Charles University, Faculty of Mathematics and Physics, Department of Statistics, Prague, Czech Republic

Using optimal transport theory, we propose a Monte Carlo critical region for a bivariate test statistic consisting of transformed skewness and kurtosis underlying the D'Agostino test of normality. We note that the p-value of the resulting normality test corresponds to a naturally defined nonnormality score admitting a decomposition in terms of skewness and kurtosis deviations from normality. In a simulation study, we show that the proposed test has almost the same power as the Shapiro-Wilk test in all simulation setups and that it clearly outperforms the Shapiro-Wilk test against platykurtic alternatives. We conclude by observing that the proposed method, with a bivariate "Monte Carlo mosaic" critical region, can be used as a general goodness-of-fit test also for other uni- and multivariate distributions.

Testing of time reversibility in functional data

Hedvika Ranošová
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague

Different notions of time reversibility of stochastic processes are introduced. Based on a random sample of continuous-time stochastic processes, we construct a statistical test for each version of time reversibility without any distributional or moment assumptions. The characteristic functional is utilized in the Cramér-von Mises test statistic with a pre-chosen weighting measure. Several simulations illustrate the proposed test's performance.

Testing the equality of estimable parameters across many populations

Marcos Romero Madroñal
University of Seville

The comparison of a parameter in k populations is a classical problem in statistics. Testing for

the equality of means or variances are typical examples. Most procedures designed to deal with this problem assume that k is fixed and that samples with increasing sample sizes are available from each population.

This paper introduces and studies a test for the comparison of an estimable parameter across k populations, when k is large and the sample sizes from each population are small compared with k. The proposed test statistic is asymptotically distribution-free under the null hypothesis of parameter homogeneity, enabling asymptotically exact inference without parametric assumptions.

Additionally, the behaviour of the proposal is studied under alternatives. Simulations are conducted to evaluate its finite-sample performance, and a linear bootstrap method is implemented to improve its behavior for small k. We further illustrate its practical application to a real dataset.

Sliced-Wasserstein distance based change detection with sequential empirical processes

Florian Scholze
University of Bamberg & RWTH Aachen University

We study the problem of detecting changes in the marginal distributions of a multivariate time series with a CUSUM-type detector statistic based on the (maximum-) sliced-Wasserstein distance. From a theoretical point of view, this projection-based approach has two appealing properties. Firstly, unlike the family of Wasserstein distances, it does not suffer from the curse of dimensionality, and secondly, by means of the Kantorovich duality, asymptotic properties of the detector statistic can be derived from results for function-indexed sequential empirical processes for nonstationary time series. The poster presents a new (bootstrap-) functional central limit theorem for sequential empirical processes and its application to the given testing problem.

On a new test for the two sample problem in the presence of censoring

Anke Steyn
North-West University, South Africa

A classical problem in survival analysis is testing whether two independent samples are realised from the same distribution. In practice, this problem is often complicated by the presence of random right censoring.

We introduce a new test statistic based on empirical characteristic functions. The proposed test statistic is a weighted L^2-type distance between the empirical characteristic functions of the observed samples. The counterpart of the empirical characteristic function in the presence of censoring is based on the Kaplan–Meier estimator.

We investigate the finite sample properties of the newly proposed test via simulated as well as observed data.

This is joint work with J.S. Allison and I.H.J. Visagie.

Poster session II (Saturday)

Goodness-of-fit and distribution tests for functional data

Daniel Hlubinka
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague

We present a method for constructing tests using empirical characteristic functionals to test distribution of functional random variables. In particular, we present a goodness-of-fit test and a test of symmetry and time reversibility for continuous stochastic processes.

Characteristic functionals can be used to construct distance-based tests for distribution or distributional properties of functional data. This approach is particularly useful because the characteristic functional can be consistently estimated, and test statistics are derived as if the functions were fully observable. Once constructed, we adjust the test statistics because the underlying functions are only observed at a discrete grid.

This is joint work with Hedvika Ranošová and Zdeněk Hlávka.

Optimal–transport based multivariate goodness-of-fit test

Šárka Hudecová
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague

We propose goodness-of-fit tests for multivariate observations. The test statistics are defined as two–sample criteria measuring discrepancy between multivariate ranks of the original observations and the corresponding ranks obtained from an artificial sample generated from the reference distribution under the test. Multivariate ranks are constructed using the optimal measure transport theory, rendering the tests of a simple null hypothesis distribution–free, while bootstrap approximations are necessary for testing composite null hypotheses.

This contribution is based on joint work with Zdeněk Hlávka and Simos G. Meintanis.

Local polynomial estimation of quantile density functions

Niclas Jacobsen and Natalie Neumeyer
University of Hamburg, Germany

A new approach for nonparametric estimation of quantile density functions based on local polynomial estimation is presented. Estimation of the quantile density is important because it appears, for example, in the expression for the asymptotic variance of empirical and kernel-type estimators of the quantile function.

The new approach uses a local polynomial regression on $(F_n(X_i), Q_n(F_n(X_i)))$, where F_n and Q_n represent the empirical distribution function and the empirical quantile function respectively. The new approach has more advantageous properties at the boundary than classical quantile density estimators. We present a result on the asymptotic normality of the proposed estimator. In this context, we also get the leading bias term and the asymptotic variance, which we use to compare the new estimator to classical estimators, especially focusing on the boundary.

Bagging and regression trees in individual claims reserving

Jan Janoušek
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University

This poster presents a new methodology for individual claims reserving using machine learning techniques. Claims reserving involves stochastically predicting future loss reserves to cover potential claims. Our approach utilizes regression trees combined with bootstrap aggregating (bagging) to enhance prediction accuracy.

Unlike traditional methods that focus only on claim counts, this model simultaneously addresses both claim frequency and severity. Model validation is improved through the use of out-of-bag error diagnostics. An illustrative data analysis demonstrates the method's potential to provide more precise reserve estimates, highlighting its practical value in claims reserving.

Specification tests for integer-valued time series

Aneta Kostárová
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague

This contribution addresses models for time series of integer-valued variables. Such series arise in various applications, often as increment series for counts of interest. A model with a GARCH-type structure with the Skellam conditional distribution is considered. We propose a novel testing procedure to assess the null hypothesis that a set of integer-valued observations follows such model.

Changepoint detection in tensor data with complex dependence structures

Martin Romaňák
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague

Tensor data consisting of multivariate outcomes over the items and across the subjects with longitudinal and cross-sectional dependence are considered. A completely distribution-free and tweaking-parameter-free detection procedure for changepoints at different locations is designed, which does not require training data.

A CUSUM type test statistic is employed and its asymptotic properties are derived for a large number of available individual profiles. We introduce an eigenbootstrap superstructure to address the computational challenges of high dimensionality without information loss, while preserving all dependencies within and between panels. The validity of this novel and efficient resampling technique is established in this general setting.

The empirical performance of the detection algorithm is evaluated through a simulation study. Finally, the fully data-driven test is applied to real-world data from EEG and psychometrics.

The use of the optimal–transport methods in prediction

Willy Svoboda
Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics, Charles University, Prague

Optimal transport methods can be used to construct a nonparametric prediction interval for a new observation based on the training data. The main idea is to estimate the multivariate quantile contours from the set of observed training vectors and validate them using the test vectors. Our aim is to extend this construction to stationary time series in order to develop a change-point detection tool.

Multiple change point detection in time series with non-stationary dynamics

Yuhan Tian
Department of Data Science, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany

Change point detection (CPD) identifies points within data sequences where statistical properties undergo significant changes, playing a vital role in domains such as finance, climate science, and quality control. Traditional CPD algorithms often assume piecewise stationarity, where model parameters remain constant between consecutive change points. However, this assumption is frequently violated in real-world data, which commonly exhibit continuous non-stationary dynamics. Such dynamics can obscure abrupt changes or lead to false detections, complicating the detection process.

To address these challenges, I present a model-based CPD algorithm capable of handling non-stationary dynamics while accounting for temporal and cross-correlations. The model decomposes time series into two components: a random walk to represent non-stationary dynamics and a vector autoregressive process to capture temporal and cross-correlations. Change points are identified by measuring error reduction within a moving window when abrupt changes are introduced into selected components of the time series.

Key features of the algorithm include its flexibility to incorporate a broad range of component selection procedures that satisfy the sure screening property and parameter estimation techniques under mild conditions. The consistency of the change point estimator, including both the number and locations of change points, is established for general choices. A specific implementation that employs ℓ_1-regularization for component selection and introduces a novel procedure for parameter estimation is proposed. This algorithm iteratively estimates model parameters, non-stationary dynamics, and multiple change point locations simultaneously.

The algorithm's effectiveness is demonstrated through applications to both simulated and real-world data. For instance, in images from the steel rolling process, the algorithm identifies surface defects amid gradual background variations. Similarly, in solar surface data, it detects solar flares against dynamic backgrounds. These applications highlight the algorithm's robustness and versatility across diverse scenarios.

This is joint work with Dr. Abolfazl Safikhani (Department of Statistics, George Mason University).

Multiscale detection of multiple change points in high-dimensional factor models

Yuqi Zhang and Haeran Cho
University of Bristol, UK

We introduce a multiscale, bandwidth-free procedure for detecting multiple change points in large approximate factor models, implemented via the Narrowest-over-Threshold (NOT) principle combined with Seeded Binary Segmentation (SBS), whose deterministic interval system provides full multiscale coverage at near-linear computational cost. We reparametrise the model and perform a single principal-component rotation that keeps the loading matrix globally invariant, so that only the pseudo factor covariances vary across regimes, eliminating repeated eigen-decompositions and reducing numerical error. Allowing for serial dependence and possibly heavy-tailed innovations, we show that our method consistently recovers both the number and the locations of change points; we also provide a finite-sample guarantee showing that the maximal localisation error decreases logarithmically with sample size and is modulated by the detection boundary. Simulations demonstrate high detection accuracy and localisation properties of the proposed method even when changes are closely spaced in dependent, high-dimensional data. An empirical analysis of the daily S&P 500 dataset uncovers structural breaks in volatility aligned with major market events. An R package accompanying the method is available at `https://github.com/YuqiZhangSA/NotSBS`.