MSM 2023

MSM 2023

Third Symposium on Modern Statistical Methods:
From Data to Knowledge

Krakow, Poland – December 13-14, 2023

Conference Program and Abstracts

WMS
WYDZIAŁ MATEMATYKI STOSOWANEJ AGH

AGH
AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA
W KRAKOWIE

Timetable

T: Tutorial, IS: Invited Speaker.

Wednesday, 13 of December, 2pm–6pm

2:00–3:15pm	T	**Cristian Jimenez** KAUST, Saudi Arabia	An introduction to quantile spectral analysis
3:15–3:40pm	**Coffee break**		
3:40–4:05pm	IS	**John Kornak**	Practical modeling of longitudinal neuropsychological and neuroimaging brain change
4:05–4:30pm	IS	**Timothy Johnson**	A Bayesian non-parametric Potts model for fMRI for presurgical planning
4:30–4:55pm	IS	**Carlos Fernandez**	Harris recurrent Markov chains and nonlinear monotone cointegrated models
4:55–5:10pm	**Short break**		
5:10–5:35pm	IS	**Iris Gauran**	High-dimensional testing based on exhaustive nested cross-validation
5:35–6:00pm	IS	**Hernando Ombao**	Overview of functional dependence in brain networks

Thursday, 14 of December, 9am–1pm

9:00–10:15am	T	**Paolo Redondo** KAUST, Saudi Arabia	A short tutorial on general dependence measures
10:15–10:40am	**Coffee break**		
10:40–11:05am	IS	**Małgorzata Bogdan**	Asymptotic distribution of low-dimensional patterns by regularizors with convex non-differentiable penalties
11:05–11:30am	IS	**Bartosz Majewski**	Inference for signals exhibiting irregular statistical cyclicity with applications to electrocardiograms
11:30–11:55am	IS	**Damian Brzyski**	Unit commitment in electrical power system: how to decompose the problem by using lagrangian relaxation
11:55–12:10pm	**Short break**		
12:10–12:35pm	IS	**Jaroslaw Harezlak**	Novel penalized regression method applied to study the association of brain functional connectivity and alcohol drinking
12:35–1:00pm	IS	**Constantin Yiannoutsos**	Small area estimation in the Indiana COVID-19 prevalence study

List of Abstracts

Wednesday, 13 of December

Tutorial: An introduction to quantile spectral analysis

Cristian Jimenez

Statistics program — CEMSE Division, King Abdullah University of Science and Technology (KAUST)

The aim of this tutorial is to introduce the basics of quantile spectral analysis for both univariate and multivariate time series. The tutorial commences by presenting the concept of quantile regression, which, unlike traditional regression methods that focus on modeling the conditional mean, explores the conditional quantiles. In the second part of this tutorial, we delve into spectral analysis by introducing the ordinary periodogram as a nonparametric estimator for the spectral density function. Furthermore, we present the quantile periodogram, which extends the ordinary periodogram to offer a robust frequency-domain representation of serial dependency in random processes by estimating the spectrum at a set of quantile levels. The last part of the tutorial is dedicated to multivariate time series analysis and the introduction of coherence as a measure of linear dependency between two series at different frequencies. Since real data applications often exhibit nonlinear dependency in the frequency domain, quantile coherence characterizes nonlinear dependency by defining coherence at a set of quantile levels. Although quantile coherence is a more powerful tool, its estimation remains challenging due to the high level of noise. We present a new semi-parametric estimation technique that approximates the quantile spectral matrix using the parametric form of the spectrum of the vector autoregressive (VAR) model, along with nonparametric smoothing across quantiles. The session closes with some data examples in bivariate financial time series.

References

[1] Jiménez-Varón, C. F., Y. Sun, and T.-H. Li (2023). A semi-parametric estimation method for quantile coherence with an application to bivariate financial time series clustering.

Practical modeling of longitudinal neuropsychological and neuroimaging brain change

John Kornak

University of California San Francisco

Understanding this temporal ordering of effects in dementia and other neurological diseases/illnesses would have major benefits for both individual level prediction and clinical trial design. I will discuss nonlinear trajectory modeling approaches aimed at estimating normalized cognitive test scores for individuals that appropriately account for demographic and other factors. These normalized scores are subsequently used in Bayesian disease progression modeling of brain biomarker trajectories (cognition, imaging and otherwise) in dementia; the goal being to determine potential differences in disease progression across patients. This is collaborative work with the UCSF Memory and Aging Center along with the Berry Consultants group.

A Bayesian non-parametric Potts model for fMRI for presurgical planning

Timothy Johnson

Department of Biostatistics, University of Michigan

There is a growing interest in using fMRI data in clinical practice. I present a fully Bayesian model for fMRI that may be more suitable for clinical applications than standard fMRI tools. A time-varying autoregressive model is used to capture any non-stationary behavior over time. Low frequency drift is modeled using adaptive B-spline bases. A conditional autoregressive-type prior is placed on the model variances. Hyperpriors are specified on the HRF hyperparameters allowing greater modeling flexibility. A non-parametric Potts model is used to partition the parameters of interest into deactivated, activated, and null classes. We compare our modeling approach to the standard mass-univariate approach on a presurgical fMRI dataset where the patient has a temporal lobe glioblastoma.

Harris recurrent Markov chains and nonlinear monotone cointegrated models

Carlos Fernandez

LTCI, Telecom Paris, Institut Polytechnique de Paris

In this talk, we study a nonlinear cointegration-type model of the form $Z_t = f_0(X_t) + W_t$ where f_0 is a monotone function and X_t is a Harris recurrent Markov chain. Using a localization argument, we develop a nonparametric Least Square Estimator to locally estimate f_0, and under mild conditions, we show its strong consistency and obtain its rate of convergence. New results (of the Glivenko Cantelli type) for localized null recurrent Markov chains are also presented.

High-dimensional testing based on exhaustive nested cross-validation

Iris Gauran

King Abdullah University of Science and Technology (KAUST)

Cross-validation is a widely utilized algorithmic technique for estimating prediction error, tuning regularization parameters, and selecting the best predictive models. Nevertheless, its behavior is intricate due to various complex factors at play. This study introduces a novel high-dimensional test based on the exhaustive nested cross-validation procedure. This method is straightforward to apply and operates almost automatically in numerous scenarios, with minimal assumptions about the underlying data distribution. Furthermore, our proposed approach can establish valid confidence intervals for comparing prediction error differences between two model-fitting algorithms. To address concerns about computational complexity, we have derived a highly efficient expression for the cross-validation estimator. Our research also delves into strategies for enhancing statistical power in high-dimensional scenarios while preserving the Type I error rate. Lastly, we showcase the application of our method to an RNA sequencing study and biological data, illustrating its practical utility in real-world scenarios.

Overview of functional dependence in brain networks

Hernando Ombao

Statistics Program, King Abdullah University of Science and Technology

Brain activity is complex. A full understanding of brain activity requires careful study of its multi-scale spatial-temporal organization (from neurons to regions of interest; and from transient events to long-term temporal dynamics). Motivated by these challenges, we will explore some characterizations of dependence between components of a brain network. This is potentially interesting because alterations in functional brain connectivity are associated with mental and neurological diseases.

In this talk, we provide an overview of functional dependence measures. We present a general framework for exploring dependence through the oscillatory activities derived from each component of the tine series. The talk will draw connections of this framework to some of the classical notions of spectral dependence such as coherence, partial coherence, and dual-frequency coherence. Moreover, this framework provides a starting point for exploring potential non-linear cross-frequency interactions. These interactions include the impact of phase of one oscillatory activity in one component on the amplitude of another oscillation. The proposed approach captures lead-lag relationships and hence can be used as a general framework for spectral causality. Under this framework, we will also present some recent work on inference using spectral mutual information and entropy measures.

This is joint work with Marco Pinto (UC Irvine), Paolo Redondo (KAUST) and Raphael Huser (KAUST).

Thursday, 14 of December

A short tutorial on general dependence measures

Paolo Redondo

Statistics Program, CEMSE Division, King Abdullah University of Science and Technology

General dependence measures capture the relationship (beyond linear) between two variables. There are several frameworks on how these measures are formulated. In this tutorial, two frameworks will be discussed; copula theory and information theory. In the former, the relationship between the variables is obtained by separating the dependence function from the marginal characteristics of the variables. For the latter, quantities, such as entropy and mutual information, are used to summarize the dependence. The link between these two frameworks will also be discussed. Suppose the variables represent time series, for example, recorded brain signals. Rather than the relationship between the time series at their observed scale, another interest is to determine at which frequency oscillations the dependence is most prominent. Coherence analysis is a common strategy to answer such query and will be covered in this tutorial. Lastly, an information-theoretic counterpart of coherence analysis, the spectral transfer entropy, will be presented.

Asymptotic distribution of low-dimensional patterns by regularizors with convex non-differentiable penalties

Małgorzata Bogdan

University of Wrocław

We explore the asymptotic distribution of the patterns generated by regularizers with non-differentiable penalties. These patterns depend on the penalty through its subdifferential and can take various forms, such as the sign vector of regression coefficients for LASSO, or the more refined SLOPE pattern, which also identifies clusters of coefficients with the same absolute values. We focus on the classical asymptotics, where the sample size approaches infinity while the number of regressors remains fixed. We derive the asymptotic distribution of the $\sqrt{n}$ scaled estimation error and its pattern for a broad class of regularizers. To achieve the pattern convergence, we utilize the Hausdorff distance, which provides a suitable notion of convergence for the penalty subdifferentials. Our framework encompasses various regularizers, including Generalized LASSO, SLOPE or Elastic net. Importantly, it extends beyond ordinary least squares to the robust Huber and Quantile loss functions. For SLOPE, we also establish asymptotic control of the false discovery rate in the context of an asymptotically orthogonal design of the regressors. Additionally, sampling from the asymptotic error distribution facilitates comparisons between different regularizers. We provide a short simulation study showcasing an illustrative comparison between the asymptotic properties of LASSO, fused LASSO and SLOPE. This is a joint work with Ivan Hejny and Jonas Wallin from Lund University.

Inference for signals exhibiting irregular statistical cyclicity with applications to electrocardiograms

Bartosz Majewski

AGH University of Krakow

We present a novel statistical method for analyzing electrocardiogram (ECG) signals. We consider an amplitude-modulated time-warped cyclostationary model which can capture the quasi-periodic behavior of ECG signals. To conduct statistical inference for ECGs, we develop two bootstrap procedures based on the well-known Circular Block Bootstrap. Additionally, we propose a bootstrap test to identify significant frequencies of the mean and autocovariance functions. The proposed methods are used to discover irregular patterns in ECG signals and improve inference for the considered model.

Unit commitment in electrical power system: how to decompose the problem by using lagrangian relaxation

Damian Brzyski

Hitachi Energy

Unit commitment (UC) problem is based on optimizing the schedule of power generators so that the operating cost is minimized. In the full version, it is an NP-hard, mixed-integer, nonlinear, non-convex optimization problem. The common approach for finding the solution is to consider the iterative process involving the linearization of some of the constraints. In each iteration, the mixed-integer linear program (MILP) is solved, which can still be very time consuming. In my talk, I will present the idea of using the special structure of such a MILP problem to relax it and then decompose it into a collection of independent, much simpler subproblems, which can be solved in parallel. This direction is based on the Lagrangian relaxation technique, which will be introduced. I will also focus on the semi-automatic procedure that defines the balanced subproblems using the tools of graph theory.

Novel penalized regression method applied to study the association of brain functional connectivity and alcohol drinking

Jaroslaw Harezlak

Indiana University School of Public Health

The intricate associations between brain functional connectivity and clinical outcomes are difficult to estimate. Common approaches used do not account for the interrelated connectivity patterns in the functional connectivity (FC) matrix, which can jointly and/or synergistically affect the outcomes. In our application of a novel penalized regression approach called SpINNEr (Sparsity Inducing Nuclear Norm Estimator), we identify brain FC patterns that predict drinking outcomes. Results dynamically summarized in the R shiny app indicate that this scalar-on-matrix regression framework via the SpINNEr approach uncovers numerous reproducible FC associations with alcohol consumption.

Small area estimation in the Indiana COVID-19 prevalence study

Constantin Yiannoutsos

Richard M. Fairbanks School of Public Health

From 25-29 April, 2020, the state of Indiana undertook testing of randomly chosen state residents for the novel SARS-CoV-2 virus, the agent causing COVID-19 disease. This was the first statewide randomized study of COVID-19 testing in the United States. RT-PCR and serological tests were administered to all study participants. We describe the analysis of these data in order to generate county-level estimates of disease prevalence, leading to a nuanced and detailed picture of the COVID-19 epidemic in the state of Indiana at the time. We use a Bayesian approach, which offers a unified platform for addressing non-response among various demographic groups and testing errors to reduce bias in the estimates of the disease prevalence. Our model incorporates all available information on disease prevalence obtained from data reported by the Indiana State Department of Health, along with diagnostic test performance, and data on the size of various demographic groups in the state, obtained from census of the Indiana population. Both adjustments appear to have substantial impact on the unadjusted estimates, due to high levels of non-response by participants of non-white race and Hispanic ethnicity and false-positive and false-negative test results among both the PCR and antibody tests utilized in the study. The resulting estimate provide a rich, nuanced picture of the differing impact of the epidemic among racial and ethnic groups and shows how the epidemic spread into the state from urban areas within Indiana and concentrated epidemics in adjacent states, as well as focused epidemics in meat processing plans and other vulnerable areas. Our approach, while not novel from a methodological perspective, provides a way to incorporate all available evidence and develop a coherent picture at the local level, which shows how unequal the effects of the early COVID-19 epidemic has been in the state.