



SRCC Workshop on Statistics and Mediation Analysis

Date : 02 May 2023 (Tuesday)

Time : 14:00–17:20

Venue : FSC1217, Fong Shu Chuen Library
Ho Sin Hang Campus
Hong Kong Baptist University

Yan ZHOU (Associate Professor, Shenzhen University), 14:00–14:30

Title: scDMV: A Zero-one Inflated Beta Mixture Modeling Strategy for Differential Methylation Analysis with Single-cell Bisulfite Sequencing Data

Abstract: The whole genome bisulfite sequencing has been the gold standard of DNA methylation detection at single-nucleotide resolution on a genome-wide scale. Traditionally, sequencing methods can only get the average expression level of many cells and therefore ignore heterogeneity among individual cells. To observe the multilayered status of single cells, single-cell bisulfite sequencing (scBS-seq) technologies have been rapidly developed and proven to be an effective and powerful tool in identification of differentially methylated region (DMR). However, DMR recognition with scBS-seq has low precision accuracy since data are often sparse and have excess zeros and ones, due to the relatively low sequencing depth and low coverage. A new differential methylation analysis approach that can well accommodate the special features of such data and enhance recognition accuracy is most desirable. A new beta mixture approach (scDMV) that incorporates excess zeros and ones and allows low-input sequencing is proposed for single-cell bisulfite sequencing data to analyze methylation differences between samples from different groups for a site or region. Compared with several alternative methods, the scDMV approach performs favorably in terms of both sensitivity and precision and also has a good control of the false positive rate as shown in our extensive simulation studies. In real data applications, we also find that scDMV method exhibits higher precision and sensitivity in identifying differentially methylation regions, even for low-input samples. Furthermore, scDMV can delineate important information that is missed by other methods for GO enrichment analysis with single cell whole genome sequencing data. Lastly, scDMV is available as an R package along with the tutorial at <https://github.com/PLX-m/scDMV>.

Zhonghua LIU (Assistant Professor, Columbia University), 14:30–15:00

Title: DeepMed: Semiparametric Causal Mediation Analysis with Debiased Deep Learning

Abstract: Causal mediation analysis can unpack the black box of causality and is therefore a powerful tool for disentangling causal pathways in biomedical and social sciences, and also for evaluating machine learning fairness. To reduce bias for estimating Natural Direct and Indirect Effects in



mediation analysis, we propose a new method called DeepMed that uses deep neural networks (DNNs) to cross-fit the infinite-dimensional nuisance functions in the efficient influence functions. We obtain novel theoretical results that our DeepMed method (1) can achieve semiparametric efficiency bound without imposing sparsity constraints on the DNN architecture and (2) can adapt to certain low-dimensional structures of the nuisance functions, significantly advancing the existing literature on DNN-based semiparametric causal inference. Extensive synthetic experiments are conducted to support our findings and also expose the gap between theory and practice. As a proof of concept, we apply DeepMed to analyze two real datasets on machine learning fairness and reach conclusions consistent with previous findings.

Rongmao ZHANG (Professor, Zhejiang University), 15:00–15:30

Title: On Semiparametrically Dynamic Functional-coefficient Autoregressive Spatio-Temporal Models with Irregular Location Wide Nonstationarity

Abstract: Nonlinear modelling of spatio-temporal data is often a challenge due to irregularly observed locations and location-wide non-stationarity. In this paper we propose a semiparametric family of Dynamic Functional-coefficient Autoregressive Spatio-Temporal (DyFAST) models to address the difficulties. First, we specify the dynamic autoregressive smooth coefficients depending on both a concerned regime and location so that the models can characterise not only the dynamic regime-switching nature but also the location-wide non-stationarity in real spatio-temporal data. Second, two semiparametric smoothing schemes are proposed to model the dynamic neighbouring-time interaction effects with irregular locations incorporated by (spatial) weight matrices. The first scheme popular in econometrics supposes that the weight matrix is pre-specified. In practice, many weight matrices can be generated differently by data location features. Model selection for an optimal one is popular, but may suffer from loss of features of different candidates. Our second scheme is thus to suggest a weight matrix fusion to let data combine or select the candidates. Accordingly, different semiparametric smoothing procedures are developed. Both theoretical properties and Monte Carlo simulations are investigated. The empirical application to an EU energy market dataset further demonstrates the usefulness of our DyFAST models.

Photo-taking and Coffee Break, 15:30–15:50

Guochang WANG (Professor, Jinan University), 15:50–16:20

Title: Adaptive Slicing for Functional Slice Inverse Regression

Abstract: In the paper, we propose a functional dimension reduction method for functional predictor and a scalar response. In the past study, the most popular functional dimension reduction method is the functional sliced inverse regression (FSIR) and people usually use a fixed slicing scheme to implement the estimate of FSIR. However in practical, there are two main questions for the fixed slicing scheme, the one is how many slices should be chosen and the other is how to divide all samples into different slices. To solve these problems, we first expand the functional predictor and functional regression parameters on the functional principal component basis or a given basis such as B-spline basis. Then to estimate the functional regression parameters by using the adaptive slicing for FSIR approach. Simulation and real data analysis are presented to show the merit of the proposed method.



Hongmei LIN (Associate Professor, Shanghai University of International Business and Economics),
16:20–16:50

Title: A New Method for Estimating Sharpe Ratio Function via Local Maximum Likelihood

Abstract: The Sharpe ratio function is a commonly used risk/return measure in financial econometrics. To estimate this function, most existing methods take a two-step procedure that first estimates the mean and volatility functions separately and then applies the plug-in method. In this paper, we propose a direct local maximum likelihood method to simultaneously estimate the Sharpe ratio function and the negative log-volatility function or their derivatives. We establish the joint limiting distribution of the proposed estimators, and we further extend the proposed method to estimate the multivariate Sharpe ratio function and establish its asymptotic normality. We evaluate the numerical performance of the proposed estimators through simulation studies, and compare them with existing methods. Finally, we apply the proposed method to analyze the three-month US Treasury bill interest rate datasets and capture a well-known covariate-dependent effect on the Sharpe ratio.

Yuejin ZHOU (Professor, Anhui University of Science and Technology), 16:50–17:20

Title: Causal Mediation Analysis for an Ordinal Outcome with Multiple Mediators

Abstract: Causal mediation analysis is a popular approach for investigating whether the effect of an exposure on an outcome is through a mediator to better understand the underlying causal mechanism. In recent literature, mediation analysis with multiple mediators has been proposed for continuous and dichotomous outcomes. In contrast, methods for mediation analysis for an ordinal outcome are still underdeveloped. In this paper, we first review mediation analysis methods with a continuous mediator for an ordinal outcome and then develop mediation analysis with a binary mediator for an ordinal outcome. We further consider multiple mediators for an ordinal outcome in the counterfactual framework and provide identification assumptions for identifying the mediation effects. Under the identification assumptions, we propose a regression-based method to estimate the mediation effects through multiple mediators while allowing the presence of exposure-mediator interactions. The closed-form expressions of mediation effects are also obtained for three scenarios: multiple continuous mediators, multiple binary mediators, and multiple mixed mediators. We conduct simulation studies to assess the finite sample performance of our new methods and present the biases, standard errors and confidence intervals to demonstrate that our proposed estimators perform well in a wide range of practical settings. Finally, we apply our proposed methods to assess the mediation effects of candidate DNA methylation CpG sites in the causal pathway from socioeconomic index to body mass index.

-- All interested are welcome (some talks will be given in Chinese). --