南方科技大学
SOUTHERN UNIVERSITY OF SCIENCE AND TECHNOLOGY

统计学前沿理论与应用研讨会

南方科技大学 数学系

广东 深圳

2017.12.23-24

会议手册目录

Organizing Committee 组织委员会

Min Cao	曹敏	南方科技大学数学系
Xuejun Jiang	蒋学军	南方科技大学数学系
Guoliang Tian	田国梁	南方科技大学数学系
Jie Xiong	熊捷	南方科技大学数学系
Zhihong Xia	夏志宏	南方科技大学数学系
Zhaojun Yang	杨招军	南方科技大学金融系

Secretaries 秘书

Xianghui Yu	余湘辉	南方科技大学数学系
Shuang Yu	于双	南方科技大学数学系

General Information 会议指南

Registration 会议报到:

12月22日（周五）： 15:00-23:00

- 专家公寓2栋大堂。

各日会议地点:

12月 23日（周六）:

- 开幕式及全体与会者合影，地点：图书馆110报告厅；
- 会场一， 慧园三栋4楼415报告厅；
- 会场二， 慧园三栋 316 报告厅。

12月 24日（周日）:

- 会场一， 慧园三栋4楼415 报告厅；
- 会场二， 慧园三栋 316 报告厅。

就餐安排

Date	Breakfast 7:00-8:30	Lunch 12:00-13:00	Dinner
Dec. 22		Staff Dining Hall 教工餐厅自助餐	麒麟山庄贵兵楼紫云阁(18:40以前报到) 专家公寓一楼中餐厅自助餐(18:40以后报到)
Dec. 23	Staff Dining Hall 教工餐厅自助餐	Staff Dining Hall 专家公寓一楼中餐厅自助餐	麒麟山庄麒麟苑 鹏城厅(18:30)
Dec. 24	Staff Dining Hall 教工餐厅自助餐	Staff Dining Hall 专家公寓一楼中餐厅自助餐	Staff Dining Hall 专家公寓一楼中餐厅围餐(18:30)
Dec. 25	Staff Dining Hall 教工餐厅自助餐		

Contacts 联系人

Xuejun Jiang, 蒋学军 ，13924670682，jiangxj@sustc.edu.cn;
Shuang Yu, 于双，13717546998，yus@sustc.edu.cn.

学校地址：广东省深圳市南山区学苑大道1088 号

周边交通

地铁站：5 号线（环中线）塘朗站

- 塘朗地铁站D 出口出，步行约700 米到南方科技大学正门(1号门)。

公交车站：南方科技大学站（1号门，3号门）

- 线路：43，74，81，M369，M459.

来校路线:以下是从几大客运枢纽来校的推荐路线

1. 深圳北站(高铁站)

- 方式一，乘地铁5 号线（环中线）前海湾方向，深圳北—塘朗站, D 出口出站，步行700米左右至南方科技大学正门(1号门)，5-8分钟步行至专家公寓二栋；

- 方式二，高铁B2出口（西广场，维也纳方向），横穿广场至马路公交车站，81/M459路公交到南方科技大学（3号门），3-5分钟步行至专家公寓二栋；
- 方式三，乘坐出租车至南方科技大学专家公寓2栋，大约需花费22 元。

2. 深圳站(罗湖火车站)

- 方式一，乘地铁1号线（罗宝线）机场东方向，罗湖—会展中心； 转4号线(龙华线)，会展中心—深圳北站；再转5号线(环中线）前海湾方向，深圳北—塘朗站, D 出口出站，步行700米左右至南方科技大学正门(1 号门)，5-8 分钟步行至专家公寓二栋；
- 方式二，乘坐出租车至南方科技大学专家公寓2栋，大约需花费70 元。

3. 宝安国际机场

- 方式一，乘坐地铁11 号线，宝安机场—前海湾站，然后换乘地铁5号线（环中线）黄贝岭方向，黄贝岭—塘朗站，从D 出口出站，步行700 米左右到达学校正门(1号门)，5-8分钟步行至专家公寓二栋；
- 方式二，乘坐出租车至南方科技大学专家公寓2栋，从宝安机场打车来校约需90-100元。

Program 会议日程

Saturday morning, December 23, 2017

时间	报告人	报告题目	主持人
8:20-9:00		开幕式及会议合影 图书馆110	汤涛 夏志宏
会场一：慧园三栋415			
9:20-9:50	朱仲义 复旦大学	Robust Subgroup Identification	宋心远
9:50-10:20	林路 山东大学	Group-Average and Convex Clustering for Partially Heterogeneous Linear Regression	
10:20-10:40	Tea break, 茶歇		
10:40-11:10	陶剑 东北师范大学	Bayesian Joint Modeling for Item Responses and Response Times	张日权
11:10-11:40	王熙逵 University of Manitoba	反应自适应第三阶段临床试验的统计设计和分析	
11:40-12:10	张荣茂 浙江大学	Testing for Structural Breaks in Spatial Trends	
会场二：慧园三栋 316			
9:20-9:50	王启华 中科院数学与系统科学研究院/深圳大学	Sufficient Dimension Reduction under Dimension-reduction-based Imputation with Predictors Missing at Random	唐年胜
9:50-10:20	薛留根 北京工业大学	A New Nested Cholesky Decomposition and Estimation for the Covariance Matrix of Bivariate Longitudinal Data	
10:20-10:40	Tea break, 茶歇		
10:40-11:10	冯兴东 上海财经大学	Distributed Variable Selection in Quantile Regression	蒋学军
11:10-11:40	孔新兵 南京审计大学	Test on stochastic block model: local smoothing and extreme value theory	
11:40-12:10	夏寅 复旦大学	Joint Testing and False Discovery Rate Control in High-Dimensional Multivariate Response Regression Model	
12:10-13:00	Lunch: Staff Dining Hall (教工餐厅)		

Saturday afternoon, December 23, 2017

时间	报告人	报告题目	主持人
会场一：慧园三栋415			
14:30-15:00	刘卫东 上海交通大学	Quantile Regression under Memory Constraint	何书元
15:00-15:30	钟威 厦门大学	Conditional Test for Ultrahigh Dimensional Linear Regression Coefficients	
15:30-16:00	周建军 云南大学	Generalized Partial Functional Linear Models	
16:00-16:20	Tea break, 茶歇		
16:20-16:50	高巍 东北师范大学	Estimation of Treatment Effects for Heterogeneous Matched Pairs Data with Probit Models	崔恒建
16:50-17:20	赵俊龙 北京师范大学	Assessing robustness of classification using angular breakdown point	
17:20-17:50	晏挺 华中师范大学	Estimation of the degree parameters in directed network models and its applications	
会场二：慧园三栋 316			
14:30-15:00	赵鹏 江苏师范大学	Dynamics of Simultaneous Cyber Attacks over Networks	邹国华
15:00-15:30	於州 华东师范大学	Sparse SIR: Optimal Rates and Adaptive Estimation	
15:30-16:00	唐安民 云南大学	Influence analysis for skew-normal semiparametric joint models of multivariate longitudinal and multivariate	
16:00-16:20	Tea break, 茶歇		
16:20-16:50	朱复康 吉林大学	A new bivariate integer-valued GARCH model allowing for negative cross-correlation	杨瑛
16:50-17:20	丁洁丽 武汉大学	Additive Mixed Effect Model for Recurrent Gap Time Data	
17:20-17:50	张庆昭 厦门大学	Inference for Low-Dimensional Covariates in a High-Dimensional Accelerated Failure Time Model	
18:30-20:30	麒麟山庄麒麟苑鹏城厅		

Sunday morning, December 24, 2017

时间	报告人	报告题目	主持人
会场一：慧园三栋415			
8:30-9:00	李启寨 中科院数学与系统科学研究院	Order-Restricted Inference for ROC Using Correlated and Clustered Data	刘卫东
9:00-9:30	郑泽敏 中国科学技术大学	Scalable Inference for Massive Data	
9:30-10:00	李东 清华大学	Renorming volatilities in a family of GARCH models	
10:00-10:20	Tea break, 茶歇		
10:20-10:50	常晋源 西南财经大学	On the statistical inference for large precision matrices with dependent data	钟威
10:50-11:20	宋晓军 北京大学	A nonparametric test of heteroskedasticity in stock returns in the presence of measurement errors	
11:20-11:50	周季蕾 北京大学	直播打赏行为的影响因素分析	
会场二：慧园三栋 316			
8:30-9:00	王海斌 厦门大学	Rank Aggregation in Ordinal Peer Evaluation: Latent Variable Method	朱利平
9:00-9:30	李建波 江苏师范大学	Estimation and Testing for Time-varying Quantile Single-index Models with Longitudinal Data	
9:30-10:00	郭旭 北京师范大学	Enhancements of Non-parametric Generalized Likelihood Ratio Test: Bias Correction and Dimension Reduction	
10:00-10:20	Tea break, 茶歇		
10:20-10:50	赵世舜 吉林大学	Source enumeration via GBIC with a statistic for sphericity test in white Gaussian and non-Gaussian noise	李启寨
10:50-11:20	蔡敬衡 中山大学	An Additive-Multiplicative Mean Residual Life Model for Right Censored Data	
11:20-11:50	陈昱 北京大学	强化学习打麻将	
11:50-13:00	Lunch: Staff Dining Hall (教工餐厅)		

Sunday afternoon, December 24, 2017

时间	报告人	报告题目	主持人
会场一：慧园三栋415			
14:30-15:00	刘鹏飞 江苏师范大学	A New Regression	蔡敬衡
15:00-15:30	任图南 北京大学	RTB广告出价模式	
会场二：慧园三栋316			
14:30-15:00	周彦 深圳大学	Classifying next-generation sequencing data using a zero-inflated Poisson model	李建波
15:00-15:30	许可 北京大学	PM2.5 Recovering with Image Information	

Abstracts 报告摘要

An Additive-Multiplicative Mean Residual Life Model for Right Censored Data

Jingheng Cai, 蔡敬衡
中山大学
E-mail: caijheng@mail.sysu.edu.cn

Abstract: Many studies have focused on determining the effect of body mass index (BMI) on the mortality in different cohorts. In this article, we propose an additive-multiplicative mean residual life (MRL) model to assess the effects of BMI and other risk factors on the MRL function of survival time in a cohort of Chinese type 2 diabetic patients. The proposed model can simultaneously manage additive and multiplicative risk factors and provide a comprehensible interpretation of their effects on the MRL function of interest. We develop an estimation procedure through pseudo partial score equations to obtain parameter estimates. We establish the asymptotic properties of the proposed estimators and conduct simulations to demonstrate the performance of the proposed method. The application of the procedure to a study on the life expectancy of type 2 diabetic patients reveals new insights into the extension of the life expectancy of such patients.

On the statistical inference for large precision matrices with dependent data

Jinyuan Chang, 常晋源
西南财经大学
E-mail: changjinyuan@swufe.edu.cn

Abstract: Precision matrices play important roles in many practical applications. Motivated by widely observed temporally dependent multivariate data in modern social and scientific studies, we consider the statistical inference of precision matrices for high-dimensional time dependent observations. Specifically, we propose a data-driven procedure to construct a class of simultaneous confidence regions for the precision coefficients within an index set of interest. The confidence regions can be applied to test for specific structures of a precision matrix and to recover its nonzero components. We first construct an estimator of the underlying precision matrix via penalized node-wise regressions, and then develop Gaussian approximation results to approximate the distribution of the maximal difference between the estimated and true precision coefficients over the given index set of interest. A computationally feasible parametric bootstrap algorithm is developed to implement the proposed procedure. Theoretical results indicate that the proposed procedure works well without the second order cross-time stationary assumption on the data and sparse structure conditions on the long-run covariance of the estimates. Simulation studies and a real example on S&P 500 stock return data confirm the performance of the proposed approach.

强化学习打麻将

Yu Chen, 陈昱
北京大学
E-mail: Yu.chen@pku.edu.cn

Abstract: 我们应用强化学习以及深度学习的技术训练AI打麻将。本报告包含以下几个方面内容：(1) 数学规范打麻将问题 (2) 通过70万局人类麻将记录使用监督学习方法预训练策略模型和价值模型。(3) 使用Q-learning 和A3C等传统强化学习方法提升麻将AI。(4) 使用AlphaGo-Zero的MCTS强化器强化策略。(5) 强化学习训练心得，包括Explore-Exploit Balance 以及 CPU-GPU Balance。

Additive Mixed Effect Model for Recurrent Gap Time Data

Jieli Ding, 丁洁丽
武汉大学
E-mail: jlding.math@whu.edu.cn

Abstract: Gap times between recurrent events are often of primary interest in medical and observational studies. The additive hazards model, focusing on risk differences rather than risk ratios, has been widely used in practice. However, the marginal additive hazards model does not take the dependence among gap times into account. In this paper, we propose an additive mixed effect model to analyze gap time data, and the proposed model includes a subject-specific random effect to account for the dependence among the gap times. Estimating equation approaches are developed for parameter estimation, and the asymptotic properties of the resulting estimators are established. In addition, some graphical and numerical procedures are presented for model checking. The finite sample behavior of the proposed methods is evaluated through simulation studies, and an application to a data set from a clinic study on chronic granulomatous disease (CGD) is provided.

Distributed Variable Selection in Quantile Regression

Xingdong Feng, 冯兴东
上海财经大学
E-mail: feng.xingdong@mail.shufe.edu.cn

Abstract: Recently, large scale datasets appear frequently due to the development of techniques. Distributed computation has attracted attentions from statistician. Since quantile regression has been an effective alternative to the classic mean regression in many fields. However, computationally efficient quantile regression estimates for large scale datasets are less developed. In this paper, we consider the penalized quantile regression estimate obtained via a new efficient ADMM algorithm that could be implemented in a distributed manner, which has been shown to be able to carry out variable selection for massive data.

Estimation of Treatment Effects for Heterogeneous Matched Pairs Data with Probit Models

Wei Gao, 高巍
东北师范大学
E-mail: gaow@nenu.edu.cn

Abstract: Estimating the effect of medical treatments on subject responses is one of the crucial problems in medical research. Matched-pairs designs are commonly implemented in the field of medical research to eliminate confounding and improve efficiency. In this article, new estimators of treatment effects for heterogeneous matched pairs data are proposed. Asymptotic properties of the proposed estimators are derived. Simulation studies show that the proposed estimators have some advantages over the famous Heckman ' s estimator, inverse probability weighted (IPW) estimator and ATE. We apply the proposed methodologies to a blood lead level data set and an acute leukaemia data set.

Enhancements of Non-parametric Generalized Likelihood Ratio Test: Bias Correction and Dimension Reduction

Xu Guo, 郭旭
北京师范大学
E-mail: xguo12@bnu.edu.cn

Abstract: Non-parametric generalized likelihood ratio test is a popular method of model checking for regressions. However, there are two issues that may be the barriers for its powerfulness: existing bias term and curse of dimensionality. The purpose of this paper is thus twofold: a bias reduction is suggested and a dimension reduction-based adaptive-to-model enhancement is recommended to promote the power performance. The proposed test statistic still possesses the Wilks phenomenon and behaves like a test with only one covariate. Thus, it converges to its limit at a much faster rate and is much more sensitive to alternative models than the classical non-parametric generalized likelihood ratio test. As a by-product, we also prove that the bias-corrected test is more efficient than the one without bias reduction in the sense that its asymptotic variance is smaller. Simulation studies and a real data analysis are conducted to evaluate of proposed tests.

Test on stochastic block model: local smoothing and extreme value theory

Xinbing Kong, 孔新兵
南京审计大学
E-mail: kongxb@fudan.edu.cn

Abstract: The stochastic block model is widely used in modeling the community structures in network data. In this paper, to obtain a consistent estimate of the number of communities, we present a new sequential testing procedure, based on the locally smoothed adjacency matrix and the extreme value theory. Under the null hypothesis that the community number is equal to a predetermined number, the test statistic converges to the type I extreme value distribution, and otherwise, it explodes fast and the divergence rate could even reach n in the strong signal case where n is the size of the network, guaranteeing high detection power. This method is simple to use and serves as an alternative approach to the novel one in Lei (2016) using random matrix theory. To detect the change of the community structure, we also propose a two-sample test for the stochastic block model with two observed adjacency matrices. Simulation studies justify the theory. We applied our method to the political blog data set and find reasonable group structures.

Renorming volatilities in a family of GARCH models

Dong Li, 李东
清华大学
E-mail: malidong@tsinghua.edu.cn

Abstract: This paper studies the weak convergence of renorming volatilities in a family of GARCH(1,1) models from a functional point of view. After suitable renormalization, it is shown that the limiting distribution is a geometric Brownian motion when the associated top Lyapunov exponent γ ¿0 and is an exponential functional of the maximum process of a Brownian motion when γ =0. This indicates that the volatility of the GARCH (1,1)-type model has a completely different random structure according to the sign of γ . The obtained results further strengthen our understanding of volatilities in GARCH-type models. Simulation studies are conducted to assess our findings.

Order-Restricted Inference for ROC Using Correlated and Clustered Data

Qizhai Li, 李启寨
中国科学院
E-mail: liqz@amss.ac.cn

Abstract: Estimating the receiver operating characteristic (ROC) curve, which is commonly used to evaluate and compare the accuracy of diagnostic models and biometric systems, has been an important problem in diagnostic medicine, biometric recognition, signal detection, and others. In a variety of applications, the data are collected under two or more naturally ordered experimental conditions. So in these situations, it is natural to assume a stochastic ordering for the observations under different experimental conditions. More importantly, statistical inference incorporating such a stochastic ordering condition is expected to improve estimation efficiency. Clustered and correlated data occur when multiple measurements are gleaned from the same subject which makes estimation of ROC curves more complicated due to the unknown within-subject correlations. Although methods are available for the estimation of ROC curves from clustered data, to the best of our knowledge, how to impose natural ordering on the estimation of ROC curves has not been studied yet. In this article, we propose an ordered-restricted estimator for the ROC curve, as well as the area under the curve and the partial area under the curve to ccommodate the clustered and correlated data structure. We derive asymptotic properties of the proposed order-restricted estimators and theoretically show that they possess lower mean-squared errors than the existing estimators. Simulation studies demonstrate better performance of the newly proposed estimators over existing methods for finite samples. The proposed method is further illustrated using the fingerprint matching data from the National Institute of Standards and Technology Special Database 4.

Estimation and Testing for Time-varying Quantile Single-index Models with Longitudinal Data

Jianbo Li, 李建波
江苏师范大学
E-mail: Lijianbo66@163.com

Abstract: Regarding semiparametric quantile regression, the existing literature is largely focused on independent observations. We propose a time-varying quantile single-index model suitable for complex data in which the responses and covariates are longitudinal/functional, with measurements taken at discrete time points. We also develop a statistic for testing whether the time effect is signicant. The proposed methodology is illustrated using Monte Carlo simulation and empirical data analysis.

Group-Average and Convex Clustering for Partially Heterogeneous Linear Regression

Lu Lin, 林路
山东大学
E-mail: linlu@sdu.edu.cn

Abstract: In this paper, a subgroup least squares and a convex clustering are introduced for inferring a partially heterogenous linear regression that has potential application in the areas of precision marketing and precision medicine. The homogenous parameter and the subgroup-average of the heterogenous parameters can be consistently estimated by the subgroup least squares, without need of the sparsity assumption on the heterogenous parameters. The heterogenous parameters can be consistently clustered via the convex clustering. Unlike the existing methods for regression clustering, our clustering procedure is a standard mean clustering, although the model under study is a type of regression, and the corresponding algorithm only involves low dimensional parameters. Thus, it is simple and stable even if the sample size is large. The advantage of the method is further illustrated via simulation studies and the analysis of car sales data.

Quantile Regression under Memory Constraint

Weidong Liu, 刘卫东
上海交通大学
E-mail: weidongl@sjtu.edu.cn

Abstract: This paper studies the inference problem in quantile regression for a large sample size n but under a limited memory constraint, where the memory can only store a small batch of data of size m. A natural method is the naive divide-and-conquer approach, which splits data into batches of size m, computes the local quantile regression estimator for each batch, and then aggregate the estimators via averaging. However, this method only works when $n = o(m^2)$ and is computationally expensive. This paper proposes a novel inference approach and establishes the asymptotic normality result that achieves the same efficiency as the quantile regression estimator computed on all the data. Our method only requires an initial quantile regression estimator on a small batch of data. Then, all the remaining computations are simple matrix operations, and thus the method is computationally very efficient. Theoretically, our method can allow $n = o(m^{2^q})$ with q times of aggregations. Our method can also be applied to address the quantile regression under distributed computing environment (e.g,. sensor network applications) or for real-time streaming data.

RTB广告出价模式

Tunan Ren, 任图南
北京大学
E-mail: 1601211784@pku.edu.cn

Abstract: 发现RTB广告的出价模式的现存问题，建模拟合并预测得到RTB广告竞价成功的概率，并改善出价模式以降低购买广告展现的成本。

A nonparametric test of heteroskedasticity in stock returns in the presence of measurement errors

Xiaojun Song, 宋晓军
北京大学
E-mail: sxj@gsm.pku.edu.cn

Abstract: In this paper, we introduce a nonparametric test for heteroscedasticity in stock returns in the presence of measurement errors. We focus our attention on a nonparametric ARCH-type effect and we show that there is a convenient nonparametric testing procedure for testing heteroscedasticity based on the observed returns that are measured with the errors. Our test statistic is a Cramer-von-Mises-type statistic which we construct after expressing the null hypothesis of homoscedasticity in terms of unconditional moment conditions. We then use the empirical process theory to derive the asymptotic distribution of our statistic. The latter distribution, however, is a non-standard asymptotic distribution with many nuisance parameters. To tackle this issue, we propose a simple multiplier-type bootstrap procedure to approximate this asymptotic distribution and establish its asymptotic validity. Our testing procedure does not require researchers to choose nuisance parameters such as bandwidth or kernel, the wrong selection of which may lead to unreliable results as well as conflicting conclusions. Thereafter, we conduct a Monte Carlo simulation study to investigate the performance of the bootstrap-based test using a variety of data generating processes and different sample sizes. Simulation results reveal that our test controls its size whatever the sample size (frequency of the data) and it has a good power. Finally, we consider an empirical application to illustrate the usefulness of our test for testing heteroscedasticity in financial data.

Influence analysis for skew-normal semiparametric joint models of multivariate longitudinal and multivariate

Anmin Tang, 唐安民
云南大学
E-mail: tam13as@sina.com

Abstract: Normality of measurement errors is a common assumption in joint models of longitudinal and survival data, but it may lead to unreasonable or even misleading results when the longitudinal data reveal skewness. This paper proposes a new joint model for multivariate longitudinal and multivariate survival data by incorporating a nonparametric function into the trajectory function and hazard function and assuming that measurement errors of longitudinal data follow a skew-normal distribution. A Monte Carlo EM algorithm together with the penalized-splines technique is developed to estimate parameters and nonparametric functions in our considered joint models. Diagnostic measures are proposed to identify influential observations and to assess local influence of various perturbations. Simulation studies and a real example from a clinical trial are presented to illustrate the proposed methodologies.

Bayesian Joint Modeling for Item Responses and Response Times

Jian Tao, 陶剑
东北师范大学
E-mail: taoj@nenu.edu.cn

Abstract: The assumption of conditional independence between the responses and the response times (RTs) for a given person is common in RT modeling. However, when the speed of a test taker is not constant, this assumption will be violated. In this talk, we construct a conditional joint model for item responses and RTs, which incorporates a covariance structure to explain the local dependency between speed and accuracy. To obtain information about the population of test takers, the new model was embedded in the hierarchical framework proposed by van der Linden (2007). A fully Bayesian approach using a straightforward Markov chain Monte Carlo (MCMC) sampler is developed to estimate all parameters in the model. The deviance information criterion (DIC) and the Bayes factor (BF) are employed to compare the goodness of fit between the models with two different parameter structures. The Bayesian residual analysis method is also employed to evaluate the fit of the RT model. Based on the simulations, we conclude that (1) the new model noticeably improves the parameter recovery for both the item parameters and the examinees ' latent traits when the assumptions of conditional independence between the item responses and the RTs are relaxed and (2) the proposed MCMC sampler adequately estimates the model parameters. The applicability of our approach is illustrated with an empirical example, and the model fit indices indicated a preference for the new model.

Rank Aggregation in Ordinal Peer Evaluation: Latent Variable Method

Haibin Wang, 王海斌
厦门大学
E-mail: whb@xmu.edu.cn

Abstract: Students in universities and colleges benefit from ordinal peer evaluation in improving their learning outcomes, metacognition and critical thinking. Unluckily, not only lack of expertise and the presence of biases results in significant noise in their evaluations, but also limits on the number of assignments that a student is able to evaluate lends to scarcity of resources. To aggregate a reliably global ranking from these partial rankings with weak information, we propose a latent asymmetric Laplace model under the latent utility framework. To make our model more general, we further treat the quantile in asymmetric Laplace distribution as an unknown parameter and estimate it from the data. Simulation results show that even if the latent random utility departs from the asymmetric Laplace distribution, the proposed methodology can recover the true ranking efficiently. We also illustrate the proposed methodology with a real data example.

Sufficient Dimension Reduction under Dimension-reduction-based Imputation with Predictors Missing at Random

Qihua Wang, 王启华
中科院数学与系统科学研究院/深圳大学
E-mail: whb@xmu.edu.cn

Abstract: In some practical problems, a subset of predictors are frequently subject to missingness, especially when the dimension of the predictor vector is high. For this case, the standard sufficient dimension reduction (SDR) methods cannot be applied directly to avoid the curse of dimensionality. A dimension-reduction-based imputation method is developed in this article such that any of spectral-decomposition-based SDR methods for full data is applicable to the case of predictors missing at random. The sliced inverse regression (SIR) is used to illustrate this procedure. The proposed dimension-reduction-based imputation estimator of the candidate matrix for SIR, termed as DRI-SIR estimator, is asymptotically normal under some mild conditions and hence the resulting estimator of the central subspace is $\sqrt{n}$-consistent. The finite sample performance of the proposed method is evaluated through comprehensive simulations and a real data set is analyzed for illustration. It is also shown that how other two popular SDR methods, namely sliced average variance estimation (SAVE) and principal Hessian direction (PHD), are extended to the case of missing predictors with the aid of the proposed imputation procedure.

反应自适应第三阶段临床试验的统计设计和分析

Xikui Wang, 王熙逵,
University of Manitoba, Canada
E-mail: xikui.wang@umanitoba.ca

Abstract: Clinical trials are regarded as the most reliable and efficient way to evaluate the efficacy of new medical interventions. This practice has taken a prominent role in modern clinical research. Clinical experimentation on human subjects requires a careful balancing act between the benefits of the collective and the benefits of the individual. Response adaptive designs represent a major advancement in clinical trial methodology that helps balance these ethical issues and improve efficiency without undermining the validity and integrity of the clinical research. Such designs are particularly desirable in desperate medical situations in which individual ethics is often jeopardized for the collective good. Response adaptive designs for Phase III clinical trials use information so far accumulated from the trial to modify the randomization procedure and deliberately bias treatment allocation in order to assign more patients to the potentially better treatment. The talk is based on joint work with my students.

Joint Testing and False Discovery Rate Control in High-Dimensional Multivariate Response Regression Model

Xia Yin, 夏寅
复旦大学
E-mail: xiayin@fudan.edu.cn

Abstract: Multivariate response regression model with high-dimensional covariates has many applications in genomic and genetic research, in which some covariates are expected to be associated with multiple responses. This talk considers joint testing for regression coefficients over multiple responses and develops simultaneous testing methods with false discovery rate control. The test statistic is based on inverse regression and bias-corrected group Lasso estimates of the regression coefficients and is shown to have a chi-square asymptotic null distribution. A row-wise multiple testing procedure is also developed in order to identify the covariates that are associated with the responses. The procedure is shown to control the false discovery proportion and false discovery rate at the pre-specified level asymptotically. Simulations demonstrate the gain in power in detecting the covariates associated with the responses as compared to the entry-wise testing. The test is applied to an ovarian cancer data set in order to identify the miRNA regulators that regulate the expressions of ovarian cancer related proteins.

PM2.5 Recovering with Image Information

Ke Xu, 许可
北京大学
E-mail: xk0566@163.com

Abstract: Air pollution is a severe environmental problem worldwide. This is particularly true in China for the past many years. Among all the air pollutants, the PM2.5 is one of the most hazardous for human health. Therefore, monitoring and reducing PM2.5 pollution becomes a problem of fundamental importance. Despite a comprehensive air quality monitoring system has been established by Chinese government, the air quality monitoring system in China is far less than sufficient. Then, how to develop a more comprehensive air quality monitoring system, but at an extremely low cost, becomes a problem of importance. As a promising solution, we explore in this article the possibility to recover PM2.5 by using imaging data.

A New Nested Cholesky Decomposition and Estimation for the Covariance Matrix of Bivariate Longitudinal Data

Liugen Xue, 薛留根
北京工业大学
E-mail: lgxue@bjut.edu.cn

Abstract: We propose a nested modified Cholesky decomposition for modeling the covariance structure in multivariate longitudinal data analysis. The entries of this decomposition have simple structures and can be interpreted as the generalized moving average coefficient matrices and innovation covariance matrices. We model the elements of these matrices by a class of unconstrained linear models, and develop a Fisher scoring algorithm to compute the maximum likelihood estimator of the regression parameters. The consistency and asymptotic normality of the estimators are established. Furthermore, we employ the smoothly clipped absolute deviation (SCAD) penalty to select the relevant variables in the models. The resulting SCAD estimators are shown to be asymptotically normal and have the oracle property. Some simulations are conducted to examine the finite sample performance of the proposed method. A real dataset is analyzed for illustration.

Estimation of the degree parameters in directed network models and its applications

Ting Yan, 晏挺
华中师范大学
E-mail: tingyanty@mail.ccnu.edu.cn

Abstract: The heterogeneity of degrees are common in network data. To characterize this phenomenon, network models often contain the degree parameter. Since each node is assigned at least one degree parameter, the dimension of the space of parameters increases as the number of nodes grows. It makes inference difficult. In this talk, I will present a unified theoretical framework on how to estimate the degree parameter in directed network models and establish the asymptotic theories of the estimator. The results are illustrated by three important applications to the probit model with reciprocity parameter, differential private bi-degree based models and the degree-corrected stochastic block model.

Sparse SIR: Optimal Rates and Adaptive Estimation

Zhou Yu, 於州
华东师范大学
E-mail: zyu@stat.ecnu.edu.cn

Abstract: In this paper, we study the minimax estimation rates for sparse SIR under both the subspace estimation loss and prediction loss. In addition, we propose a three-stage adaptive estimation scheme to achieve such optimal estimation rates. Such three stage estimation procedures can be realized through ADMM and group lasso. And the final estimator is guaranteed to be rate optimal.

Testing for Structural Breaks in Spatial Trends

Rongmao Zhang, 张荣茂
浙江大学
E-mail: rmzhang@zju.edu.cn

Abstract: Non-stationary spatial models are widely applicable in diverse disciplines, ranging from bio-medical sciences to geophysical studies. In many of theses applications, testing for structural changes in the trend and testing the specific form of the trend are highly relevant. A novel statistics based on a discrepancy measure over small regions is proposed in this paper. Such a measure can be used to construct tests for structural trends and to identify change boundaries of the trends. By virtue of the m-dependence approximation of a stationary random field, asymptotic properties and limit distributions of these tests are established. The method is illustrated by simulations and data analysis.

Assessing robustness of classification using angular breakdown point

Junlong Zhao, 赵俊龙
北京师范大学
E-mail: zhaojunlong928@126.com

Abstract: Robustness is a desirable property for many statistical techniques. As an important measure of robustness, breakdown point has been widely used for regression problems and many other settings. Despite the existing development, we observe that the standard breakdown point criterion is not directly applicable for many classification problems. In this paper, we propose a novel breakdown point criterion for classification. To better quantify the robustness of different classification methods, we propose a new criterion, namely the angle breakdown point. Using this new breakdown point criterion, we study the robustness of binary large margin classification techniques in this paper, although the idea is applicable to general classification methods. Both bounded and unbounded loss functions with linear and kernel learning are considered. These studies provide useful insights on the robustness of different classification methods. Numerical results further confirm our theoretical findings.

Dynamics of Simultaneous Cyber Attacks over Networks

Peng Zhao, 赵鹏
江苏师范大学
E-mail: zhaop@jsnu.edu.cn

Abstract: Modeling cyber attacks is a very attractive area of research due to its practical importance. However, most of the related research in the literature does not consider the scenario of simultaneous (or coordinated) attacks which in fact is an important attack instrument in practice. This is mainly because of the complicated evolution of cyber attacks over networks. In this talk, we propose a novel model which can accommodate different types of simultaneous attacks with possible heterogenous compromise probabilities.The theoretical results are further validated by the simulation evidence.

Source enumeration via GBIC with a statistic for sphericity test in white Gaussian and non-Gaussian noise

Shishun Zhao, 赵世舜
吉林大学
E-mail: zhaoss@jlu.edu.cn

Abstract: We propose a source enumeration method via the generalised Bayesian information criterion (GBIC) based on a statistic for sphericity test in the white Gaussian and non-Gaussian noise under a large array with few samples. Instead of joint probability of observations or sample eigenvalue distribution, the proposed method is based on a statistic for testing the sphericity of a positive definite covariance matrix, to overcome the limitation of the Gaussian observations assumption. Under the white noise assumption, the covariance matrix of the noise subspace components of the observations is proportional to an identity matrix, and this identity structure can be tested by a statistic for sphericity test. The observations are decomposed into signal and noise subspace components under a presumptive number of sources. When the presumptive noise subspace components do not contain signals, the corresponding statistic for sphericity test will have a certain Gaussian distribution, and the number of sources can be estimated via the GBIC with the test statistic. Simulation results demonstrate that the proposed method provides high detection probability in both the Gaussian and the non-Gaussian noise, and performs better when the number of samples is less than the number of array sensors compared with other methods.

Scalable Inference for Massive Data

Zemin Zheng, 郑泽敏
中国科学技术大学
E-mail: zhengzm@ustc.edu.cn

Abstract: With the availability of massive data sets, how to make accurate inference with lower computational cost is the key to improving scalability. One important scenario is when both the sample size and the number of covariates are large, which is in contrast to the typical high-dimensional setting with relatively low sample size. In such cases, naive application of the existing inference procedures can be computationally inefficient or infeasible. To ameliorate the scalability, in this paper we suggest the method of inference with partitioned data (IPAD) that divides the entire sample set into subsamples for correcting the bias and constructs confidence intervals by aggregating the estimates based on subsamples. Compared to inference with the whole sample set, such an approach can substantially reduce the computational cost. Furthermore, we establish confidence intervals of the bagging estimator for aggregation, which remain largely unexplored in the literature due to the communication barriers between subsamples. Both computational advantage and theoretical guarantee of our new method are evidenced by numerical examples.

Conditional Test for Ultrahigh Dimensional Linear Regression Coefficients

Wei Zhong, 钟威
厦门大学
E-mail: wzhong@xmu.edu.cn

Abstract: This paper is concerned with a conditional test for regression coefficients in ultrahigh dimensional linear models. Conditioning on a subset of important predictors in the model, we test the significance of regression coefficients of the remaining ultrahigh dimensional predictors. We first propose a conditional test based on an estimated U-statistics of order two for a high dimensional linear regression model and prove that its null asymptotic distribution is normal under some mild assumptions. Then, we demonstrate that it performs well when the dimension is moderately high via simulations. However, the empirical power of the proposed test is inversely affected by the dimensionality of predictors. To this end, we further propose a two-stage procedure to reduce the dimensionality under the sparsity assumption and enhance the empirical power. In the first stage, we divide data randomly into two halves and apply a sure independence screening to the first half to reduce the dimensionality; In the second stage, we apply the proposed conditional test to the second half of the data. We examine the finite-sample performances of the proposed conditional test via Monte Carlo simulations and a real data analysis.

直播打赏行为的影响因素分析

Jilei Zhou, 周季蕾
北京大学
E-mail: zhoujilei@pku.edu.cn

Abstract: 近年来，直播行业发展迅猛。作为一种新兴的支付手段，打赏成为主播和直播平台重要的收入来源。因此，探究打赏行为的影响因素成为提升收入的关键问题。本研究使用斗鱼平台的真实数据探索了直播间打赏数量的影响因素。研究结论显示，用户的进入行为、弹幕行为和送礼物行为都将影响直播间的打赏数量。本研究丰富了打赏行为的相关理论研究，为直播平台或主播提升打赏数量提供理论指导。

Generalized Partial Functional Linear Models

Jianjun Zhou, 周建军
云南大学
E-mail: jjzhou@ynu.edu.cn

Abstract: In this paper, by assuming the scaler response in partial functional linear regression models belong to an exponential family distribution, we propose a generalized partial functional linear model. Based on the polynomial spline smoothing technique, we derive the quasi-likelihood estimator for the linear parameters. We also establish asymptotic normality for the estimators of the parametric components and the global convergence rate of the functional parameter. Simulation studies are conducted to investigate the proposed methodologies.

Classifying next-generation sequencing data using a zero-inflated Poisson model

Yan Zhou, 周彦
深圳大学
E-mail: Zhouy1016@szu.edu.cn

Abstract: In this paper, we propose a Zero-Inflated Poisson Logistic Discriminant Analysis (ZIPLDA) for RNA-seq data with an excess of zeros. The new method assumes that the data are from a mixture of two distributions: one is a point mass at zero, and the other follows a Poisson distribution. We then consider a logistic relation between the probability of observing zeros and the mean of the genes and the sequencing depth in the model. Simulation studies show that the proposed method performs better than, or at least as well as, the existing methods in a wide range of settings. Two real datasets including a breast cancer RNA-seq dataset and a microRNA-seq dataset are also analyzed, and they coincide with the simulation results that our proposed method outperforms the existing competitors.

A new bivariate integer-valued GARCH model allowing for negative cross-correlation

Fukang Zhu, 朱复康
吉林大学
E-mail: zfk8010@163.com

Abstract: Univariate integer-valued time series models, including integer-valued autoregressive (INAR) models and integer-valued generalized autoregressive conditional heteroscedastic (INGARCH) models, have been well studied in the literature, but there is little progress in multivariate models. Although some multivariate INAR models were proposed, they do not provide enough flexibility in modeling count data, such as volatility of numbers of stock transactions. Then a bivariate Poisson INGARCH model was suggested by Liu (2012), but it can only deal with positive cross-correlation between two components. To remedy this defect we propose a new bivariate Poisson INGARCH model, which is more flexible and allows for positive or negative cross-correlation. Stationarity and ergodicity of the new process are established. The maximum likelihood method is used to estimate the unknown parameters, and consistency and asymptotic normality for estimators are given. A simulation study is given to evaluate the estimators for parameters of interest. Real and artificial data examples are illustrated to demonstrate good performances of the proposed model relative to the existing model.

Robust Subgroup Identification

Zhongyi Zhu, 朱仲义
复旦大学
E-mail: Zhuzy@fudan.edu.cn

Abstract: We consider a more robust subgroup analysis based on median regression and model the heterogeneity by subject-specific intercepts. This model have two appealing featurs: (1) By considering median regression, we allow the errors to have heavy tails or correlation with covariates. (2) The characteristic of quantile loss function enables us to develop augmented dataset to implement Local Linear Approximation Algorithm. In this paper, we investigate nonconvex pairwise penalized median regression for simultaneous subgroup analysis and parameter estimation. To derive the oracle property, we employ a recently developed convex-differencing method to tackle the challenges of nonsmooth loss function and nonconvex penalty functions. Simulation results and an application to the Cleveland heart disease dataset demonstrate the effectiveness of the proposed method.

List of Participants 参会人员

姓名	单位	邮箱
艾明要	北京大学	myai@pku.edu.cn
蔡敬衡	中山大学	caijheng@mail.sysu.edu.cn
常晋源	西南财经大学	changjinyuan@swufe.edu.cn
陈安岳	南方科技大学	chenay@sustc.edu.cn
陈敏	中国科学院	mchen@amss.ac.cn
陈昱	北京大学	Yu.chen@pku.edu.cn
崔恒建	首都师范大学	hjcui@bnu.edu.cn
崔霞	广州大学	cuixia@gzhu.edu.cn
丁洁丽	武汉大学	jlding.math@whu.edu.cn
房祥忠	北京大学	xzfang@pku.edu.cn
冯兴东	上海财经大学	feng.xingdong@mail.shufe.edu.cn
冯艳钦	武汉大学	yqfeng.math@whu.edu.cn
高巍	东北师范大学	gaow@nenu.edu.cn
郭旭	北京师范大学	xguo12@bnu.edu.cn
何书元	首都师范大学	syhe@cnu.edu.cn
蒋文江	云南师范大学	tennisca@163.com
蒋学军	南方科技大学	jiangxj@sustc.edu.cn
孔新兵	南京审计大学	kongxb@fudan.edu.cn
李东	清华大学	malidong@tsinghua.edu.cn
李启寨	中国科学院	liqz@amss.ac.cn
李建波	江苏师范大学	Lijianbo66@163.com
李景治	南方科技大学	lijz@sustc.edu.cn
李元	广州大学	mathly@gzhu.edu.cn
林路	山东大学	linlu@sdu.edu.cn
刘卫东	上海交通大学	weidongl@sjtu.edu.cn
刘妍岩	武汉大学	liuyy@whu.edu.cn
任图南	北京大学	1601211784@pku.edu.cn
沈馨	上海交通大学	shujin-shenxin@sjtu.edu.cn
宋心远	香港中文大学	xysong@sta.cuhk.edu.hk
宋晓军	北京大学	sxj@gsm.pku.edu.cn
孙便霞	南方科技大学	sunbx@sustc.edu.cn
孙六全	中国科学院	slq@amt.ac.cn
唐安民	云南大学	tam13as@sina.com
唐年胜	云南大学	nstang@ynu.edu.cn

List of Participants 参会人员

陶剑	东北师范大学	taoj@nenu.edu.cn
田国梁	南方科技大学	tiangl@sustc.edu.cn
王军辉	香港城市大学	j.h.wang@cityu.edu.hk
王熙逵 Xikui Wang	University of Manitoba	xikui.wang@umanitoba.ca
王德辉	吉林大学	wangdh@jlu.edu.cn
王海斌	厦门大学	whb@xmu.edu.cn
王启华	中国科学院 深圳大学	qhwang@amss.ac.cn
王小舟	上海交通大学	wangxiaozhou@sjtu.edu.cn
王兆军	南开大学	zjwangnk@126.com
吴耀华	中国科学技术大学	wuyh@ustc.edu.cn
吴远山	武汉大学	shan@whu.edu.cn
夏寅	复旦大学	xiayin@fudan.edu.cn
熊捷	南方科技大学	xiongj@sustc.edu.cn
许可	北京大学	xk0566@163.com
薛留根	北京工业大学	lgxue@bjut.edu.cn
晏挺	华中师范大学	tingyanty@mail.ccnu.edu.cn
杨瑛	清华大学	yyang@math.tsinghua.edu.cn
杨招军	南方科技大学	yangzj@sustc.edu.cn
於州	华东师范大学	zyu@stat.ecnu.edu.cn
邹国华	首都师范大学	ghzou@amss.ac.cn
张宝学	首都经贸大学	zhangbaoxue@cueb.edu.cn
张庆昭	厦门大学	qzzhang@xmu.edu.cn
张日权	华东师范大学	rqzhang@stat.ecnu.edu.cn
张荣茂	浙江大学	rmzhang@zju.edu.cn
张崇岐	广州大学	cqzhang@gzhu.edu.cn
赵俊龙	北京师范大学	zhaojunlong928@126.com
赵鹏	江苏师范大学	zhaop@jsnu.edu.cn
赵世舜	吉林大学	zhaoss@jlu.edu.cn
郑泽敏	中国科学技术大学	zhengzm@ustc.edu.cn
钟威	厦门大学	wzhong@xmu.edu.cn
周季蕾	北京大学	zhoujilei@pku.edu.cn
周建军	云南大学	jjzhou@ynu.edu.cn
周彦	深圳大学	Zhouy1016@szu.edu.cn
朱复康	吉林大学	zfk8010@163.com
朱利平	中国人民大学	zhu.liping@ruc.edu.cn
朱仲义	复旦大学	Zhuzy@fudan.edu.cn

Map 南方科技大学地图