北京师范大学 BEIJING NORMAL UNIVERSITY | 统计与数据科学研究中心 Center for Statistics and Data Science

2022交叉学科前沿论坛

大数据科学交叉论坛

会议议程

主办：北京师范大学珠海校区
承办：人文和社会科学高等研究院、自然科学高等研究院
协办：统计与数据科学研究中心

中国.珠海
2022年11月12日

议程简表

日期	时间	事项	汇报人
11月12日 (周六)	14：00-14：05	开场主持	童行伟
	14：05-14：40	论坛邀请报告（1）	崔恒建
	14：40-15：05	论坛邀请报告（2）	王国长
	15：05-15：30	论坛邀请报告（3）	王江洲
	15：30-15：55	论坛邀请报告（4）	贺百花
	15：55-16：05	中场休息	
	16：05-16：30	论坛邀请报告（5）	陈欣
	16：30-16：55	论坛邀请报告（6）	曾靖
	16：55-17：20	论坛邀请报告（7）	张兴发

腾讯会议号：247-336-105 （密码：12600）

邀请报告专家介绍

崔恒建,现为首都师范大学教授，博士生导师，中国科协第十届全委会委员，曾任国务院学位委员会学科评议组专家。中国科学院系统科学研究所博士毕业。在大数据统计建模、高维统计及其稳健统计理论和方法、统计机器学习、金融统计、以及质量管理等领域取得过许多重要的研究成果，发表论文180余篇，其中包括发表在国际顶级的统计和计量经济学杂志JASA、AoS、JRSS(B)、Biometrika和JoE上。主持国家自然科学基金重点项目、杰青（B）项目以及多项面上项目、主要参加教育部重大科研基金项目、科技部863等项目。现担任《数学学报》和《应用数学学报》中、英文版以及《Statistical Theory and Related Fields》编委，中国现场统计研究会副理事长，全国工业统计教育研究会副理事长，北京应用统计学会会长，国际数理统计学会（中国分会）常务理事。曾获得教育部高等学校科学技术奖-自然科学奖二等奖；全国统计科学研究优秀成果奖一等奖等。

邀请报告专家介绍（按报告先后顺序）

王国长，暨南大学经济学院统计学系教授、博士生导师。2012年毕业于东北师范大学数学与统计学院统计系，并取得统计学博士学位。主要研究方向为函数型数据分析、时间序列、充分性降维等，迄今为止在Journal of Econometrics, Journal of the Business & Economic Statistics, Statistica Sinica等重要学术期刊发表论文30余篇。主持国家社科基金一般项目和国家自然科学基金面上和青年项目。 任中国现场统计研究会资源与环境统计分会常务理事；广东省现场统计协会秘书长，常务理事。

王江洲，南方科技大学统计与数据科学系（邵启满课题组和荆炳义课题组联合）博士后，本科毕业于东北大学，硕士和博士毕业于东北师范大学，师从郭建华教授。主要研究方向为大规模复杂网络数据的统计分析。目前发表SCI论文5篇，其中第一作者3篇（JASA 1篇，Stat 2篇）， 主持博士后基金面上资助1项，博士后基金特别资助（站中）1项。

邀请报告专家介绍（按报告先后顺序）

贺百花，中国科学技术大学管理学院特任副教授。本硕博毕业于武汉大学数学与统计学院，香港浸会大学博士后，博士期间先后前往香港理工大学，耶鲁大学，香港大学访问。主要研究领域包括生存分析，高维数据分析，模型平均以及分布式算法， 已在JASA,JMLR,biometrics等顶级统计学杂志发表若干论文。

陈欣，南方科技大学统计与数据科学系副教授，研究员，博士生导师。1999年本科毕业于南开大学数学系，2003年在新加坡国立大学获得硕士学位。2010年博士毕业于美国明尼苏达大学双子城分校。曾在美国雪城大学，新加坡国立大学任教。主要研究领域是处理高维数据的降维和变量选择的方法，他在稀疏型降维方法中提出了一个统一的框架，并用流形上的理论证明了变量选择的oracle性质。其他的研究领域包括大数据以及复杂数据分析。在统计学顶级刊物Annals of Statistics和Biometrika发表过若干篇文章。

邀请报告专家介绍（按报告先后顺序）

曾靖，中国科学技术大学管理学院特任副教授。2017年在中国科学技术大学获得数学与应用数学学士学位，2022年在佛罗里达州立大学获得统计学博士学位。目前主要研究方向为数据降维，高维数据分析，张量数据分析，以及稳健统计。主要研究成果已经发表在Journal of the American Statistical Association, Statistica Sinica, Journal of Statistical Software期刊上。

张兴发，广州大学经济与统计学院副教授，统计学博士，硕士生导师，统计系主任。兼任中国现场统计研究会理事、广东省现场统计学会副理事长。研究兴趣为时间序列分析。完成国家自然科学基金项目一项，在研广东省、广州市自然科学基金各一项，省市级教改项目各一项，在Journal of Econometrics, SCIENCE CHINA Mathematics, Statistics and its interface, Quality and Reliability Engineering International, Statistics and probability letter, 应用概率统计，应用数学学报等期刊发表科研论文30余篇。

论坛邀请报告的题目、作者和摘要

论坛邀请报告（1）

Model-free conditional screening for ultrahigh-dimensional survival data via conditional distance correlation

崔恒建/首都师范大学数学科学学院

摘要： How to select the active variables which have significant impact on the event of interest is a very important and meaningful problem in the statistical analysis of ultrahigh-dimensional data. In many applications, researchers often know a certain set of covariates are active variables from some previous investigations and experiences. With the knowledge of the important prior knowledge of active variables, we propose a model-free conditional screening procedure for ultrahigh dimensional survival data based on conditional distance correlation. The proposed procedure can effectively detect the hidden active variables which are jointly important but are weakly correlated with the response. Moreover, it performs well when covariates are strongly correlated with each other. We establish the sure screening property and the ranking consistency of the proposed method and conduct extensive simulation studies, which suggests that the proposed procedure works well for practical situations. Then we illustrate the new approach through a real data set from the diffuse large-B-cell lymphoma study.

论坛邀请报告的题目、作者和摘要

论坛邀请报告（2）

Testing for the martingale difference hypothesis in multivariate time series models

王国长/暨南大学经济学院统计学系

摘要：This paper proposes a general class of tests to examine whether the error term is a martingale difference sequence in a multivariate time series model with parametric conditional mean. These new tests are formed based on recently developed martingale difference divergence matrix (MDDM), and they provide formal tools to test the multivariate martingale hypothesis in the literature for the first time. Under suitable conditions, the asymptotic null distributions of these MDDM-based tests are established. Moreover, these MDDM-based tests are consistent to detect a broad class of fixed alternatives, and have nontrivial power against local alternatives of order n-1/2, where n is the sample size. Since the asymptotic null distributions depend on the data generating process and the parameter estimation, a wild bootstrap procedure is further proposed to approximate the critical values of these MDDM-based tests, and its theoretical validity is justified. Finally, the usefulness of these MDDM-based tests is illustrated by simulation studies and one real data example.

论坛邀请报告的题目、作者和摘要

论坛邀请报告（3）

大样本统计推断中一种新相合性的理论性质研究

王江洲/南方科技大学统计与数据科学系

摘要：在大样本统计推断中，导出所构造统计量的渐近性质是非常重要的（例如评估统计量的收敛速度、开展假设检验、构造区间估计等），但这往往也是比较困难的。基于此，本研究系统讨论了估计量的一种新的相合性Asymptotic Consistency，包括其在辅助推导统计量渐近性质时所起的作用；其与经典的弱相合性、强相合性的区别与联系；及其在社区探测、变量选择、子组分析和结构学习方面的应用。

论坛邀请报告的题目、作者和摘要

论坛邀请报告（4）

Rank-based greedy model averaging for high-dimensional survival data

贺百花/中国科学技术大学管理学院

摘要： Model averaging is an effective way to enhance prediction accuracy. However, most previous works focus on low-dimensional settings with completely observed responses. To attain an accurate prediction for the risk effect of survival data with high-dimensional predictors, we propose a novel method: rank-based greedy (RG) model averaging. Specifically, adopting the transformation model with splitting predictors as working models, we doubly use the smooth concordance index function to derive the candidate predictions and optimal model weights. The final prediction is achieved by weighted averaging all the candidates. Our approach is flexible, computationally efficient, and robust against model misspecification, as it neither requires the correctness of a joint model nor involves the estimation of the transformation function. We further adopt the greedy algorithm for high dimensions. Theoretically, we derive an asymptotic error bound for the optimal weights under some mild conditions. In addition, the summation of weights assigned to the correct candidate submodels is proven to approach one in probability when there are correct models included among the candidate submodels. Extensive numerical studies are carried out using both simulated and real datasets to show the proposed approach’s robust performance compared to the existing regularization approaches.

论坛邀请报告的题目、作者和摘要

论坛邀请报告（5）

High Dimensional Elliptical Sliced Inverse Regression in non-Gaussian Distributions

陈欣/南方科技大学统计与数据科学系

摘要： Sliced inverse regression (SIR) is the most widely-used sufficient dimension reduction method due to its simplicity, generality and computational efficiency. However, when the distribution of the covariates deviates from the multivariate normal distribution, the estimation efficiency of SIR is rather low. We propose a robust alternative to SIR - called elliptical sliced inverse regression (ESIR) for analysing high dimensional, elliptically distributed data. There are wide applications of the elliptically distributed data, especially in finance and economics where the distribution of the data is often heavy-tailed. To tackle the heavy-tailed elliptically distributed covariates, we novelly utilize the multivariate Kendall's tau matrix in a framework of so-called generalized eigenvector problem for sufficient dimension reduction. Methodologically, we present a practical algorithm for our method. Theoretically, we investigate the asymptotic behavior of the ESIR estimator under high dimensional setting. Quantities of simulation results show that ESIR significantly improves the estimation efficiency in heavy-tailed scenarios. Analysis of two real data sets also demonstrates the effectiveness of our method. Moreover, ESIR can be easily extended to most other sufficient dimension reduction methods and applied to non-elliptical heavy-tailed distributions.

论坛邀请报告的题目、作者和摘要

论坛邀请报告（6）

Subspace Estimation with Automatic Dimension and Variable Selection in Sufficient Dimension Reduction

曾靖/中国科学技术大学管理学院

摘要： Sufficient dimension reduction (SDR) methods target finding lower-dimensional representations of a multivariate predictor to preserve all the information about the conditional distribution of the response given the predictor. The reduction is commonly achieved by projecting the predictor onto a low-dimensional subspace. The smallest such subspace is known as the Central Subspace (CS) and is the key parameter of interest for most SDR methods. In this article, we propose a unified and flexible framework for estimating the CS in high dimensions. Our approach generalizes a wide range of model-based and model-free SDR methods to high-dimensional settings, where the CS is assumed to involve only a subset of the predictors. We formulate the problem as a quadratic convex optimization so that the global solution is feasible. The proposed estimation procedure simultaneously achieves the structural dimension selection and coordinate-independent variable selection of the CS. Theoretically, our method achieves dimension selection, variable selection, and subspace estimation consistency at a high convergence rate under mild conditions. We demonstrate the effectiveness and efficiency of our method with extensive simulation studies and real data examples.

论坛邀请报告的题目、作者和摘要

论坛邀请报告（7）

Daily semiparametric GARCH model estimation using intraday high-frequency data

张兴发/广州大学经济与统计学院

摘要： In this article, intraday high-frequency data are introduced to estimate the volatility function of a daily semiparametric GARCH(1, 1) model. A semiparametric volatility proxy model is proposed to obtain the estimation. Under reasonable assumptions, the asymptotic normality of model estimator is established. Furthermore, we also discuss the impact of different volatility proxies on estimation accuracy. Both the simulation and empirical results show that introducing high-frequency data can improve the estimation precision of the considered model.