第十二届全国概率统计会议

会议手册

主办单位：中国数学会概率统计分会

承办单位：山东大学数学与交叉科学研究中心

山东大学金融研究院

山东大学数学学院

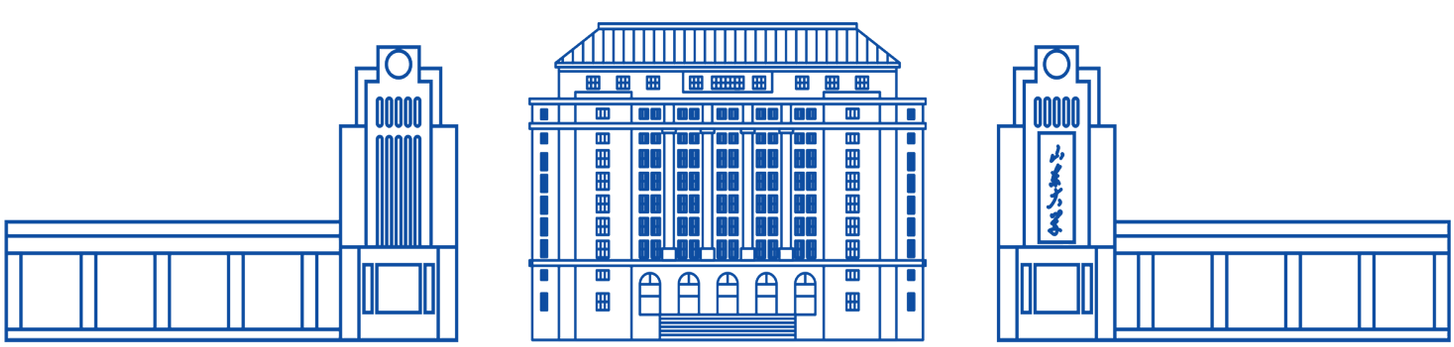
山东大学数学与统计学院

中国·青岛

2023年4月

目 录

content



会议信息

为了更好地推动概率统计及相关学科的发展，促进国内概率统计学者之间的交流和人才培养，由中国数学会概率统计分会主办，山东大学数学与交叉科学研究中心、山东大学金融研究院、山东大学数学学院和山东大学数学与统计学院承办的“中国数学会概率统计分会第十二次全国代表大会”于 2023 年 4 月 14 日至 4 月 16 日在山东省青岛市举行。这是我国概率统计学者的一次盛会，本次会议邀请了国内概率统计领域知名专家学者及教学名师，就会议相关主题作精彩报告和深入交流。

组委会

学术委员会

主任：李增沪　陈松蹊

委员：陈大岳　郭建华　郭先平　杨　瑛　王兆军　张希承　艾明要

程序委员会

主任：陈增敬　艾明要

委员：贾金柱　蒋达权　李　东　邵井海　张余辉　朱文圣　朱湘禅

地方组织委员会

主任：陈增敬　吴　臻

委员：赵永新　李　娟　張健恒　王汉超　尹　上　杨　媛

会议指南

竭诚欢迎各位专家学者莅临青岛参加第十二届全国概率统计会议。为保证会议顺利进行以及参会人员安全愉快参会，现将会议有关事宜说明如下：

报到时间 2023年4月13日14:00-22:00

报到地点 青岛黄海饭店大堂（青岛市市南区延安一路75号）

住宿酒店 青岛黄海饭店（青岛市市南区延安一路75号）

青岛汇泉王朝大饭店（青岛市市南区南海路9号）

青岛武胜关度假酒店（青岛市市南区武昌路3号）

青岛海大学术交流中心（青岛市市南区红岛路8号）

会议时间 2023年4月14日-16日

会议地点 青岛黄海饭店会议中心（地点详见日程安排）

交通说明 会议不安排接送站，请参会人员自行前往会场。

线路	出租车	乘坐公交/地铁
胶东机场——青岛黄海饭店	距离 58 公里	地铁 8 号线到青岛北站换乘地铁 3 号线到汇泉广场站（A 口）出即可抵达。
青岛站——青岛黄海饭店	距离 5 公里	地铁 3 号线到汇泉广场站（A 口）出或公交 26、316、304、312、202 路到汇泉广场公交站即可抵达。
青岛北站——青岛黄海饭店	距离 18 公里	地铁 3 号线到汇泉广场站（A 口）即可抵达。
青岛黄海饭店——青岛汇泉王朝大饭店	距离 0.9 公里	步行（8 分钟）：延安一路→南海路。
青岛黄海饭店——青岛武胜关度假酒店	距离 1.1 公里	步行（15 分钟）： 延安一路→南海路→武昌路 / 延安一路→岷山路→武昌路。
青岛黄海饭店——青岛海大学术交流中心	距离 1 公里	步行（12 分钟）： 延安一路→福山路→红岛路。

报告地点

报告厅名称	地点	备注
会议中心前区	会议中心二楼	会议中心路线：面向酒店大堂前台向右走到尽头右转进门
会议中心后区	会议中心二楼	
会议中心第二	会议中心三楼	
会议中心第三	会议中心二楼	
二楼会议厅	主楼二楼	酒店二楼

用　餐

用餐	日期	时间	地点	凭据
早餐	4日14日-17日	7:00-9:30	所住酒店	凭房卡用餐
午餐	4月14日	11:30-14:00	一楼明苑厅 二楼中华厅	凭餐券在黄海饭店用餐
	4月15日	11:30-14:00	会议中心后区 四楼锦绣厅	
	4月16日	11:30-14:00	二楼中华厅 21楼云帆阁	
晚餐	4月13日-16日	18:00-20:30	二楼中华厅	凭餐券在黄海饭店用餐

天　气

气温约为9℃-16℃，青岛春季多风，早晚温差较大，偶尔有雨，请您注意及时增减衣物，备好雨具。

温馨提示

1.请各参会人员提前到达会场，并自觉遵守会场秩序，手机等通讯工具调成静音或者振动模式。

2. 会场请勿吸烟，吸烟请移步到吸烟室。

3. 到海边游玩请务必注意安全。

4.春季传染病高发，请做好个人防护，如有发热等不适情况，请及时与会务组联系。

5.请大家自觉遵守疫情防控规定。如有发烧必须参加会议的情况，请自带核酸检测阴性证明，经会务组允许后再参会。

会务联系

会议期间，各位参会人员如有问题，可随时与会务组联系。

王汉超　18753189360

李蔚郁　13356716036

杨　媛　13465312256

郝　蕾　13687667527

孔垂柳　13573103652

傅宗奕　13969101207

祝各位来宾工作顺利，身体健康，万事如意！感谢您的支持与合作！

会议日程

2023 年 4 月 14 日（周五）

报告厅	会议中心				
8:30-9:30	开幕式　合影				
9:30-10:30	大会报告：邵启满　　主持人：陈增敬				
10:30-11:00	**茶歇**				
11:00-12:00	大会报告：姚　方　　主持人：王兆军				
12:00-14:00	**午餐**				
报告厅	**会议中心前区　主持人：刘卫东**		**报告厅**	**会议中心后区　主持人：郭先平**	
14:00-14:40	中会报告：苗　旺		14:00-14:40	中会报告：王　健	
14:40-15:20	中会报告：李丹宁		14:40-15:20	中会报告：宋　健	
15:20-15:30	**茶歇**				
报告厅	**会议中心前区**	**会议中心后区**	**会议中心第二**	**会议中心第三**	**二楼会议厅**
15:30-16:45 主持人	分组报告 01 于志勇	分组报告 02 尹传存	分组报告 03 宋永生	分组报告 04 李蔚郁	分组报告 05 王晓光
15:30-15:55	罗德军 中国科学院	戴万阳 南京大学	胡明尚 山东大学	赵世舜 吉林大学	唐炎林 华东师范大学
15:55-16:20	刘伟 武汉大学	范胜君 中国矿业大学	严晓东 山东大学	张洪 中国科学技术大学	陈大川 南开大学
16:20-16:45	鲁大伟 大连理工大学	吴盼玉 山东大学	张国栋 山东大学	温勇 南京邮电大学	潘小青 上海师范大学
16:45-17:00	**茶歇**				
17:00-18:15 主持人	分组报告 06 胡明尚	分组报告 07 戴万阳	分组报告 08 严晓东	分组报告 09 赵世舜	分组报告 10 唐炎林
17:00-17:25	胡亦钧 武汉大学	尹传存 曲阜师范大学	朱复康 吉林大学	刘玉坤 华东师范大学	王晓光 大连理工大学
17:25-17:50	柳向东 暨南大学	朱全新 湖南师范大学	谭发龙 湖南大学	王文武 曲阜师范大学	成青 西南财经大学
17:50-18:15	李欣鹏 山东大学	冯新伟 山东大学	陈泽 中国人民大学	李蔚郁 山东大学	杨文志 安徽大学
18:30	**晚餐**				

2023 年 4 月 15 日（周六）

报告厅	会议中心 前区				
9:00-10:00	大会报告：丁　剑　　主持人：张希承				
10:00-11:00	大会报告：常晋源（线上）　主持人：杨　瑛				
11:00-11:20	**茶歇**				
报告厅	**会议中心前区　主持人：林华珍**		**会议中心第二　主持人：任艳霞**		
11:20-12:00	中会报告：王学钦（线上）		中会报告：许惟钧		
12:00-14:00	**午餐**				
报告厅	**会议中心前区　主持人：林华珍**		**报告厅**	**会议中心第二　主持人：任艳霞**	
14:00-14:40	中会报告：邓　柯		14:30-15:10	中会报告：朱湘禅	
14:40-15:20	中会报告：郭　旭				
15:20-15:30	**茶歇**		15:10-15:30	**茶歇**	
报告厅	**会议中心前区**	**会议中心后区**	**会议中心第二**	**会议中心第三**	**二楼会议厅**
15:30-16:45 主持人	分组报告 11 冯新伟	分组报告 12 胡亦钧	分组报告 13 孔翠娟	分组报告 14 句媛媛	分组报告 15 王潇逸
15:30-15:55	蒋辉 南京航空航天大学	于志勇 山东大学	何勇 山东大学	张世斌 上海师范大学	刘广应 南京审计大学
15:55-16:20	戴洪帅 山东财经大学	杜恺 复旦大学	刘彬 复旦大学	刘鹏飞 江苏师范大学	谢田法 北京工业大学
16:20-16:45	宋玉平 上海师范大学	魏庆萌 东北师范大学	虞龙 上海财经大学	张兴发 广州大学	赵玮 山东大学
16:45-17:00	茶歇				
17:00-18:15 主持人	分组报告 16 蒋辉	分组报告 17 杜恺	分组报告 18 罗德军	分组报告 19 张世斌	分组报告 20 赵玮
17:00-17:25	宋永生 中国科学院	薛晓峰 北京交通大学	刘源远 中南大学	潘文亮 中国科学院	王潇逸 北师大珠海校区
17:25-17:50	杨淑振 山东大学	郭精军 兰州财经大学	赵国焕 中国科学院	句媛媛 昆明理工大学	曹雅琦 中央民族大学
17:50-18:15	李邯武 山东大学	徐平峰 东北师范大学	石权 中国科学院	孔翠娟 山东大学	刘思晟 湖南师范大学
18:30	**晚餐**				

2023 年 4 月 16 日（周日）

报告厅	会议中心 前区					
9:00-10:00	大会报告：施　展　　主持人：李增沪					
10:00-11:00	大会报告：孔新兵　　主持人：艾明要					
11:00-11:20	**茶歇**					
报告厅	**会议中心前区　主持人：王启华**		**会议中心后区　主持人：董　昭**			
11:20-12:00	中会报告：何　煦		中会报告：刘　伟			
12:00-14:00	**午餐**					
报告厅	**会议中心前区 主持人：王启华**		**报告厅**	**会议中心后区 主持人：董　昭**		
14:00-14:40	中会报告：胡　涛		14:30-15:10	中会报告：翟建梁		
14:40-15:20	中会报告：夏　寅（线上）					
15:20-15:30	**茶歇**		15:10-15:30	**茶歇**		
报告厅	**会议中心前区**	**会议中心后区**	**报告厅**	**会议中心第二**	**会议中心第三**	**二楼会议厅**
15:30-16:45 主持人	分组报告 21 杨淑振	分组报告 22 李欣鹏	15:30-16:45 主持人	分组报告 23 陈一鸣	分组报告 24 童培峰	分组报告 25 李新宇
15:30-15:55	乔会杰 东南大学	罗鹏 上海交通大学	15:30-15:45	屈直 武汉大学	高子文 中国科学院大学	李新宇 北京大学
			15:45-16:00	戴国政 浙江大学	温梦涛 南开大学	李澎涛 山东大学
15:55-16:20	李瑞囡 上海对外贸易大学	李文强 烟台大学	16:00-16:15	邹振烽 中国科学技术大学	童培峰 北京大学	张晨琳 西南财经大学
			16:15-16:30	陈一鸣 山东大学	唐科杰 上海交通大学	任怡萌 复旦大学
16:20-16:45	洪一平 北京理工大学	姜恋姿 山东科技大学	16:30-16:45	杨阳 苏州大学	毕建鑫 厦门大学	刘宇扬 上海交通大学
16:45-17:00	**茶歇**					
17:00-18:00 主持人	分组报告 26 赵林杰	分组报告 27 苏东	17:00-18:00 主持人	分组报告 28 蒲雁南	分组报告 29 梁春辉	分组报告 30 王惠远
17:00-17:15	赵林杰 华中科技大学	钱智勇 中南财经政法大学	17:00-17:15	蒲雁南 云南大学	梁春辉 东北师范大学	王惠远 北京大学
17:15-17:30	易清心 西交利物浦大学	张晴晴 南京师范大学	17:15-17:30	许诺 长春工业大学	马沄 清华大学	闫晗 北京大学
17:30-17:45	左百帅 曲阜师范大学	马雅静 首都师范大学	17:30-17:45	周智翔 南京审计大学	吉晓婷 山东大学	李杨 中国科学院大学
17:45-18:00	赵欣瑶 苏州大学	苏东 南京大学	17:45-18:00	周厚林 安徽大学	李哲 复旦大学	杨伟超 北京师范大学
18:30	**晚餐**					

大会特邀报告及报告人简介 (按姓氏拼音排序)

PT01：Statistical inferences for complex dependence of multimodal imaging data

常晋源

西南财经大学/中国科学院数学与系统科学研究院

摘要：Statistical analysis of multimodal imaging data is a challenging task, since the data involves high-dimensionality, strong spatial correlations and complex data structures. In this paper, we propose rigorous statistical testing procedures for making inferences on the complex dependence of multimodal imaging data. Motivated by the analysis of multi-task fMRI data in the Human Connectome Project (HCP) study, we particularly address three hypothesis testing problems: (a) testing independence among imaging modalities over brain regions, (b) testing independence between brain regions within imaging modalities, and (c) testing independence between brain regions across different modalities. Considering a general form for all the three tests, we develop a global testing procedure and a multiple testing procedure controlling the false discovery rate. We study theoretical properties of the proposed tests and develop a computationally efficient distributed algorithm. The proposed methods and theory are general and relevant for many statistical problems of testing independence structure among the components of high-dimensional random vectors with arbitrary dependence structures. We also illustrate our proposed methods via extensive simulations and analysis of five task fMRI contrast maps in the HCP study.

报告人简介：常晋源，西南财经大学光华特聘教授、中国科学院数学与系统科学研究院研究员、博士生导师、数据科学与商业智能联合实验室执行主任、国家杰出青年科学基金获得者、四川省特聘专家、四川省统计专家咨询委员会委员。主要从事“超高维数据分析”和“高频金融数据分析”两个领域的研究。曾担任Journal of the Royal Statistical Society Series B副主编，现担任Journal of the American Statistical Association、Journal of Business & Economic Statistics以及Statistica Sinica的副主编。

PT02：Combinatorial statistics: a common theme and a few examples

丁剑

北京大学

摘 要： In this talk, I will overview the general flavor for the topic of "combinatorial statistics", and I will describe some recent progress on shotgun assembling problems as well as random graph matching problems. Through these examples, I wish to convey the flavor of the topic of combinatorial statistics, and why it calls for joint efforts from statisticians, probabilists, computer scientists and researchers from applied sciences. The talk is based on recent joint works in various combinations with undergraduate students Hang Du, Haoyu Liu and Zhangsong Li, as well as graduate students Shuyang Gong, Heng Ma and Yiyang Jiang from Peking University.

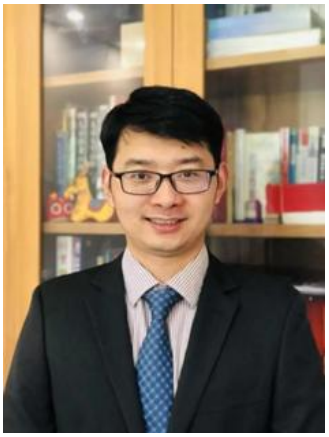
报告人简介：丁剑，北京大学数学学院讲席教授。研究兴趣为概率论，尤其着重于其与统计物理，理论计算机和理论统计的交叉方向。丁剑于 2011 年在加州大学伯克利分校获得博士学位，随后在斯坦福大学担任博士后。在回国工作以前，丁剑曾在芝加哥大学与宾西法尼亚大学担任教职，并于 2021 年获聘宾西法尼亚大学 Gilbert-Helman 讲席教授。所获得的荣誉包括斯隆奖(2015)，洛勒戴维逊奖(2017)，国际数学家大会 45 分钟邀请报告(2022)和华人数学家大会金奖(2022)。目前担任 Journal 0f American Mathematical Society, Communications in Mathematical Physics, Annals of Probability, Annals of Applied Probability 等期刊的副主编。

PT03：高维因子分析—稳健性、时变性与多截面交互

孔新兵

南京审计大学

摘要：高维因子分析作为一种基础性空间降维方法和工具，在统计学、数据科学、经济学等领域具有广泛的应用。然而，实际高维数据的分布复杂性和相依结构的复杂性为因子分析造成困难，例如分布厚尾性、概率结构的时变性、矩阵型数据复杂相依性中的双截面交互效应等，均为确定因子空间大小和估计因子空间的统计精度带来挑战。本报告将重点介绍：1）如何基于高维向量时间序列在没有任何矩条件下重构因子空间并给出统计精度；2）如何基于高维高频数据和高维伊藤半鞅识别因子空间的时变性并恢复因子空间；3）如何基于高维矩阵序列搜寻多截面交互的因子空间并给出估计精度。在给出相关统计推断理论的同时给出一些实际数据分析证据。



报告人简介：孔新兵，现为南京审计大学研究生院主持工作副院长，统计与数据科学学院教授；主要研究兴趣为高频、高维数据统计推断与机器学习；主持国家自然科学基金 3 项，教育部人文社会科学基金项目 1 项，在统计学顶级期刊 AoS、JASA、Biometrika、JoE、JBES 发表论文 20 篇，入选国家高层次青年人才和国际统计协会(ISI)当选会士，获第一届统计科学技术进步奖一等奖，香港数学会“最佳博士论文奖”。

PT04：概率统计渐近理论之随想

邵启满

南方科技大学

摘要：本报告将回顾概率统计渐近理论最新进展，特别对自正则化极限理论、正态与非正态逼近之Stein方法做些总结,同时对今后概率统计渐近理论方面的研究谈些个人感想。

报告人简介：邵启满，南方科技大学统计与数据科学系创系系主任，讲席教授。先后任教于杭州大学、新加坡国立大学、美国 Oregon 大学、香港科技大学、香港中文大学。

邵启满主要从事概率统计极限理论的研究。他系统深入地发展了自正则化极限理论，建立了自正则化大偏差、中偏差定理；发展完善了正态与非正态逼近的斯坦因方法，建立了随机浓度不等式和确定极限分布的基本方法；深入研究了相依变量极限理论，发展出了一系列重要的矩和概率不等式，建立了强逼近弱收敛等基础性工作。 在国内外的主要概率统计期刊发表论文 180 余篇，同时由 Springer 出版社出版专著三部。

邵启满于 2015 年获国家自然科学二等奖（第一完成人），2010 年在国际数学家大会作 45 分钟邀请报告，2001 年当选国际数理统计学会（IMS）会士；曾任 IMS 会士选拔委员会主席，IMS 理事会常务理事；曾任概率统计顶级国际期刊 Annals of Statistics、Annals of Applied Probability 编委；作为首位华人学者，目前出任 Annals of Applied Probability 联合主编。

PT05：随机分析的几个应用

施展

中国科学院数学与系统科学研究院

摘要：我将简单讨论随机分析的几个应用，包括 Wright-Fisher 模型、Bertoin-Le Gall 随机流、Warren-Yor 布朗贼等。这些应用的共同特点是原本技术性很强的分枝过程和扩散流，用一维布朗运动的轨道表述以后，显得特别直观。跟 Elie Aïdékon（复旦大学）和胡跃云（巴黎第十三大学）合作。

报告人简介：施展，中国科学院数学与系统科学研究院研究员。在此之前，他是巴黎第六大学的教授。他的博士导师是 Marc Yor 教授和 Paul Deheuvels 教授。施展的研究领域包括概率论、随机分析和统计物理。

PT06：Theory of Functional PCA for noisy and discretely observed data

姚方

北京大学

摘要：Functional data analysis is an important research field in statistics which treats data as random functions drawn from some infinite-dimensional functional space, and functional principal component analysis (FPCA) plays a central role for data reduction and representation. After nearly three decades of research, there remains a key problem unsolved, namely, the perturbation analysis of covariance operator for diverging number of eigencomponents obtained from noisy and discretely observed data. This is fundamental for studying models and methods based on FPCA, while there has not been much progress since the result obtained by Hall et al. (2006) for a fixed number of eigenfunction estimates. In this work, we establish a unified theory for this problem, deriving the moment bounds of eigenfunctions and asymptotic distributions of eigenvalues for a wide range of sampling schemes. We also exploit double truncation to derive the uniform convergence of such estimated eigenfunctions. The technical arguments in this work are useful for handling the perturbation series of discretely observed functional data and can be applied in models and methods involving inverse using FPCA as regularization, such as functional linear regression.

报告人简介：姚方，国家高层次人才，北京大学讲席教授，北大统计科学中心主任、概率统计系主任，数理统计学会与美国统计学会会士。2000年本科毕业于中国科学技术大学，2003获得加利福尼亚大学戴维斯分校统计学博士学位，曾任职于多伦多大学统计科学系长聘正教授。至今担任9个国际统计学核心期刊主编或编委，包括《加拿大统计学期刊》主编、顶级期刊《北美统计学会会刊》和《统计年刊》的编委。

中会 40 分钟邀请报告摘要 (按姓氏拼音排序)

01：Bayesian criteria for re-randomization

***Ke Deng,* 邓柯**

Tsinghua University kdeng@tsinghua.edu.cn

Imbalance of covariates between treatment groups is to be avoided even in randomized experiments, where this is achieved in expectation. Re-randomization based on Mahalanobis distance, referred to ReM, has been advocated as a method to achieve this. This basic method, as well as some extensions, allocates equal importance to each orthogonalized covariate. However, investigators often know a prior that some covariates are more important than the others for predicting the outcomes. Formulating such prior knowledge into a formal prior distribution and utilizing this to guide the design of re-randomization procedures can be used to establish a general Bayesian framework for re-randomization. Theoretical analyses show that the re-randomization procedure induced by the Bayesian criterion, referred to as ReB, can enjoy attractive properties, and takes ReM, and many of its extensions, as special cases. The advantages of ReB are demonstrated via simulation studies.

02：Tests for ultrahigh-dimensional partially linear regression models

***Xu Guo,* 郭旭**

Beijing Normal University xustat12@bnu.edu.cn

In this paper, we consider tests for ultrahigh-dimensional partially linear regression models. The presence of ultrahigh-dimensional nuisance covariates and unknown nuisance function makes the inference problem very challenging. We adopt machine learning methods to estimate the unknown nuisance function and introduce quadratic-form test statistics. Interestingly, though the machine learning methods can be very complex, under suitable conditions, we establish the asymptotic normality of our introduced test statistics under the null hypothesis and local alternative hypotheses. We further propose a power-enhanced procedure to improve the test statistics' performance. Two thresholding determination methods are provided for the power-enhanced procedure. We show that the power-enhanced procedure is powerful to detect signals under either sparse or dense alternatives and it can still control the type-I error asymptotically under null hypothesis. Numerical studies are carried out to illustrate the empirical performance of our introduced procedures.

03：计算机实验、实验设计

***Xu He,* 何煦**

AMSS, Chinese Academy of Sciences hexu@amss.ac.cn

计算机实验指借助数学模型对真实物理、化学等过程进行模拟的计算机程序。许多模拟复杂系统的仿真程序输入参数多、单次运行时间长，只能依据有限次计算机实验的结果建立代理模型，根据代理模型进行后续的统计分析。为了确保代理模型的精确程度，一种常用的方法是使用具有低填充距离、高分离距离以及优良的投影距离的空间填充设计。生成具有距离性质的设计并不是一个容易的问题。当前最流行的思路，是以距离准则（或其组合）作为目标函数，使用某种全局最优化方法（如模拟退火）计算最优解。然而受限于全局最优化问题的困难性，除非实验次数或变量数目很少，求得接近最优的解非常困难。

我们提出使用格生成空间填充设计。格（lattice），指的是具有群结构的点集。考虑基于格的设计，一是通过限定设计的几何结构减小最优化搜索空间；二是利用群结构简化距离性质的计算提升速度；三是有利于进行大样本分析。本报告的内容涵盖了报告人近年若干相关工作。

04：Goodness-of-fit test for semiparametric copula models with bivariate interval-censored data

***Tao Hu,* 胡涛**

Capital Normal University hutaomath@foxmail.com

This paper discusses the goodness-of-fit test of semiparametric copula models when one faces bivariate interval-censored failure time data, which often occur in many areas including epidemiological and medical studies as well as social science experiments. For the problem, three test statistics or procedures are proposed: two basedon the pseudo in-and-out-sample approach and one based on the information ratiocriterion. The asymptotic properties of the proposed test procedures are established, and in particular, the three methods are shown to be asymptotically equivalent. To assess the empirical performance of the proposed methods, an extensive simulation study is conducted and indicates that they work well in practical situations. An application to a Signal–Tandmobiel study is provided.

05：Power enhancement in high-dimensional hypothesis testing

***Danning Li,* 李丹宁**

The Northeast Normal University lidn040@nenu.edu.cn

Power-enhanced tests with high-dimensional data have received growing attention in theoretical and applied statistics in recent years. Existing tests possess their respective high-power regions, and we may lack prior knowledge about the alternatives when testing for a problem of interest in practice. There is a critical need of developing powerful testing procedures against more general alternatives. This talk is about studying the joint test of two-sample mean vectors or covariance matrices for high-dimensional data. We expand the high-power region of high-dimensional mean tests or covariance tests to a wider alternative space and then combine their strengths together in the simultaneous test. We develop a new Power-enhanced test without size distortion.

06：Monotone vs Locally monotone

***Wei Liu,* 刘伟**

Jiangsu Normal University weiliu@jsnu.edu.cn

In this talk we first recall the classical variational framework, and we will briefly review some progress of well-posedness and asymptotic properties in this direction. Then we present some recent results concerning the Mckean-Vlasov SPDEs and multi-scale stochastic systems.

07：Introducing the specificity score: measuring the credibility of causal associations

***Wang Miao*，苗旺**

Peking University mwfy@pku.edu.cn

P 值作为统计显著性度量被广泛地应用于在科学研究中。然而，近年来应用领域针对P值的质疑日益增加，Political Analysis 等知名期刊宣布禁用P值，Nature 发表评论文章反对统计显著性，在科学领域引起一场“统计危机”。这次危机的背后固然有 P 值被误用的原因，但 P 值或统计显著性有自身的局限性—很多科学研究的目的是挖掘因果关系，P 值刻画的是某种统计关联性而不一定具有因果意义，当研究中存在混杂因素时，统计关联性不代表因果关系。工具变量和代理推断等方法可以调整混杂因素，但这些使用辅助变量的方法依赖精确的因果先验或专业领域知识。在本文，作者提出基于因果特异性的因果推断理论，建立严格的因果特异性概念，提出特异性得分度量存在混杂因素时因果关系的显著性，利用特异性得分检验因果关系是否存在。本文提出的因果特异性理论只需要关于因果关系范围的粗略信息，不依赖精确的因果先验知识或辅助变量，在一定条件下，因果显著性检验可以控制住检验因果关系是否存在的第一类错误并具有趋于一的功效，可以识别多个处理和多个结果之间的因果关系。这些新理论和新方法在社会调查、生物样本数据、电子病历数据分析等领域有潜在应用。本文通过分析一个关于老鼠肥胖症的数据研究基因表达和代谢产物之间的因果关系。

08：Some recent results on stochastic wave equations

***Jian Song*，宋健**

Shandong University txjsong@hotmail.com

In the talk, some recent results on stochastic wave equations with multiplicative noise will be presented: (a) a sufficient and necessary condition for the existence and uniqueness of mild Skorohod solution; (b) a study on the existence and uniqueness of mild Stratonovich solution based on a Strichartz type estimate for the wave kernel in weighted Besov spaces; (c) moment bounds for Skorohod solutions.

09：Quantitative periodic homogenization for symmetric stable-like jumps

***Jian Wang,* 王健**

Fujian Normal University jianwang@fjnu.edu.cn

We obtain a quantitative version of homogenization for symmetric α-stable-like processes on $\mathbb{R}^d$ with periodic coefficients. In particular, the results indicate that the decay behavior of the solution near the boundary will make the homogenization slower. This is based on an on-going paper with Xin Chen, Zhen-Qing Chen and Takashi Kumagai.

10：Metric distribution function

***Xueqin Wang,* 王学钦**

University of Science and Technology of China wangxq20@ustc.edu.cn

Statistical inference aims to use observed samples to learn the unknown properties of a population. It has become an integral step in scientific reasoning. A building block of nonparametric statistical inference is distribution function. The distribution function and samples are connected to form a directed closed loop by the correspondence theorem in measure theory and the Glivenko-Cantelli and Donsker properties in statistics, and this connection creates a paradigm for statistical inference. However, existing distribution functions are defined in Euclidean spaces. Those distribution functions are no longer convenient to use or applicable in characterizing the rapidly evolving data objects of complex nature. Thus, it is imperative to develop the concept of the distribution function in a more general space to meet emerging needs. Note that the linearity allows us to use hypercubes to define the distribution function in an Euclidean space, but without the linearity in a metric space, we must work with balls as the basis of the metric topology in defining a probability measure. We introduce a class of novel quasi-distribution functions, or ball functions, for metric space-valued random objects. A ball requires a center and a radius. The center depends on the random point of interest, and the radius is determined by the distance between the center and another random point. Working with balls in defining a probability measure is particularly challenging because unlike hypercubes, the intersection of two balls may not be a ball. We overcome this challenge to prove the correspondence theorem and the Glivenko-Cantelli theorem in metric spaces that lie the foundation for conducting rational statistical inference for metric space-valued data. some relative concepts are also developed.

11：Locally adaptive transfer learning algorithms for large-scale multiple testing

***Yin Xia,* 夏寅**

Fudan University xiayin@fudan.edu.cn

Transfer learning has enjoyed increasing popularity in a range of big data applications. In the context of large-scale multiple testing, the goal is to extract and transfer knowledge learned from related source domains to improve the accuracy of simultaneously testing of a large number of hypotheses in the target domain. This talk develops a locally adaptive transfer learning algorithm (LATLA) for transfer learning for multiple testing. In contrast with existing covariate-assisted multiple testing methods that require the auxiliary covariates to be collected alongside the primary data on the same testing units, LATLA provides a principled and generic transfer learning framework that is capable of incorporating multiple samples of auxiliary data from related source domains, possibly in different dimensions/structures and from diverse populations. Both the theoretical and numerical results show that LATLA controls the false discovery rate and outperforms existing methods in power. LATLA is illustrated through an application to genome-wide association studies for the identification of disease-associated SNPs by cross-utilizing the auxiliary data from a related linkage analysis.

12：一类随机薛定谔方程的长时间表现

***Weijun Xu,* 许惟钧**

Peking University weijunxu@bicmr.pku.edu.cn

这个报告将介绍我们在一类质量临界随机薛定谔方程方面的工作，主要目标是试图理解随机性对于色散方程长时间表现的影响。我们将从经典的确定性质量临界薛定谔方程出发，衍生到随机方程的局部适定性与全局适定性，并在最后介绍我们最近在其长时间行为方面的工作。报告基于与范晨捷（中科院）和赵泽华（北理工）的合作。

13：Irreducibility of SPDEs driven by pure jump noise

***Jianliang Zhai,* 翟建梁**

University of Science and Technology of China zhaijl@ustc.edu.cn

The irreducibility is fundamental for the study of ergodicity of stochastic dynamical systems. In the literature, there are very few results on the irreducibility of stochastic partial differential equations (SPDEs) and stochastic differential equations (SDEs) driven by pure jump noise. The existing methods on this topic are basically along the same lines as that for the Gaussian case. They heavily rely on the fact that the driving noises are additive type and more or less in the class of stable processes. The use of such methods to deal with the case of other types of additive pure jump noises appears to be unclear, let alone the case of multiplicative noises.

We develop a new, effective method to obtain the irreducibility of SPDEs and SDEs driven by multiplicative pure jump noise. The conditions placed on the coefficients and the driving noise are very mild, and in some sense they are necessary and sufficient. This leads to not only significantly improving all of the results in the literature, but also to new irreducibility results of a much larger class of equations driven by pure jump noise with much weaker requirements than those treatable by the known methods. As a result, we are able to apply the main results to SPDEs with locally monotone coefficients, SPDEs/SDEs with singular coefficients, nonlinear Schrodinger equations, Euler equations etc. We emphasize that under our setting the driving noises could be compound Poisson processes, even allowed to be infinite dimensional. It is somehow surprising.

14：Recent results on Φ_d^4 model via stochastic quantization

***Xiangchan Zhu,* 朱湘禅**

AMSS, Chinese Academy of Sciences zhuxiangchan@amss.ac.cn

In this talk I will recall our recent results on large N limit of the O(N)-invariant linear sigma model, which is a vector-valued generalization of the Φ^4 quantum field theory. We study the problem via its stochastic quantization. I will also talk about perturbation theory of Φ_2^4 model by using this method.

分组报告摘要

分组报告 01

主持人：于志勇，山东大学

Eddy viscosity emerges in scaling limit of stochastic fluid equations with transport noise

***Dejun Luo*，罗德军**

AMSS, Chinese Academy of Sciences luodj@amss.ac.cn

Abstract

Transport noise models the effects of fluid small scales – eddies – on large scale fluid components. For the vorticity form of stochastic 2D Euler equations, we show that eddy viscosity appears in a suitable scaling limit of the transport noise. We also show that the 2D Smagorinsky model, which is well known in Large Eddy Simulations, is a scaling limit of suitable stochastic fluid equations driven by transport noise.

Long time behaviors on mean field interacting particle systems and McKean-Vlasov equations

***Wei Liu*，刘伟**

Wuhan University wliu.math@whu.edu.cn

Abstract

In this talk, we will present our recent studies about the long time behaviors on mean-field interacting particle systems and the McKean-Vlasov equation, by using two different methods: coupling method and functional inequalities. This talk is based on the joint works with Arnaud Guillin, Liming Wu and Chaoen Zhang.

The first exit time of fractional Brownian motion from an unbounded convex domain

***Dawei Lu*，鲁大伟**

Dalian University of Technology dwlu@dlut.edu.cn

Abstract

Consider a pair of fractional Brownian motions starting at the interior point $(\boldsymbol{x}_0, a) \in \mathbb{R}^{d+1}$, for some fixed $\boldsymbol{x}_0 \in \mathbb{R}^d$ and constant $\mathrm{a} > \mathrm{h}(\boldsymbol{x}_0)$, of an unbounded domain $\mathrm{D} = \{(\boldsymbol{x}, y) \in \mathbb{R}^{d+1}: \|\boldsymbol{x}\| < h^{-1}(y)\}$, where we take $h^{-1}(x) = Ax^{\alpha}(\log x)^{\beta}$, and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$. Let τ_D denote the first time that the fractional Brownian motion exits from D. In most cases, we give the asymptotically equivalent estimate of $\log \mathbb{P}(\tau_D > t)$. The proof methods are based on the earlier works of Li, Shi, Lifshits, and Aurzada.

主持人：尹传存，曲阜师范大学

随机微分博弈与带 Levy 跳归一化正倒向耦合随机偏微分方程

***Wanyang Dai*, 戴万阳**

Nanjing University nan5lu8@nju.edu.cn

Abstract

我们在随机微分博弈与带 Levy 跳归一化正倒向偶合随机偏微分方程之间建立起一个关系。所研究随机微分博弈系统具有 q 个博弈者组成且该系统由一般维数的向量型 Levy 过程驱驶。通过建立向量型 Ito-Ventzell 公式及求解归一化随机偏微分方程的 4 元组随机向量场解，我们得到了非零和博弈 Pareto 最优 Nash 均衡策略过程与零和博弈鞍点策略过程。所涉及的归一化随机偏微分方程是具有一般维数向量型的且是正倒向偶合的，其飘移系数与扩散系数及跳跃系数均可包含不同类型的高阶非线性偏微分算子。通过构造合适的支撑拓扑空间，我们在广义局线性增长与广义局部李普希兹条件下并利用我们最近发表的有关布朗运动情形下的相关成果证明了有关 4 元组 Levy 向量随机场强解的存在唯一性。

Multi-dimensional backward stochastic differential equations of diagonally quadratic generators: the general result

***Shengjun Fan*, 范胜君**

China University of Mining and Technology f_s_j@126.com

Abstract

The present talk is devoted to a general solvability of a multi-dimensional backward stochastic differential equation (BSDE) of a diagonally quadratic generator, by relaxing the assumptions of Hu and Tang (2016, SPA) on the generator and terminal value. Three new results are established on the local and global solutions when the terminal value is bounded and the generator is subject to some general assumptions. When the terminal value is unbounded but is of exponential moments of arbitrary order, an existence and uniqueness result is given, which seems to be the first general solvability of system of quadratic BSDEs with unbounded terminal values. This generalizes and strengthens some existing results via some new ideas. This a joint work with Prof. Ying Hu and Shanjian Tang.

On distribution dependent stochastic differential equations driven by *G*-Brownian motion

***Panyu Wu*, 吴盼玉**
Shandong University wupanyu@sdu.edu.cn

Abstract

Distribution dependent stochastic differential equations have been a very hot subject with extensive studies. On the other hand, under the *G*-expectation framework, stochastic differential equations driven by *G*-Brownian motion (in short form, *G*-SDEs) have received in creasing attentions, and the existence and uniqueness of solutions to *G*-SDEs under Lipschitz and non-Lipschitz conditions have been obtained. Based on these studies, it is very natural and also important to investigate the *G*-SDEs which are also distribution dependent. In this talk, we are concerned with the well-posedness of the distribution dependent *G*-SDEs. To this end, we first introduce a proper distance of the involved distribution functions and propose a new formulation of the distribution dependent *G*-SDEs. Then, by utilizing fix point argument, we establish existence and uniqueness of the solutions of distributed dependent *G*-SDEs under Lipschitz condition and a non-Lipschitz condition. Finally, we derive certain estimates for the solutions of the distribution dependent *G*-SDEs.

BSDEs driven by G-Brownian motion under degenerate case and its application to the regularity of fully nonlinear PDEs

***Mingshang Hu*，胡明尚**
Shandong University humingshang@sdu.edu.cn

Abstract

We obtain the existence and uniqueness theorem for backward stochastic differential equation driven by G-Brownian motion (G-BSDE) under degenerate case. Moreover, we propose a new probabilistic method based on the representation theorem of G-expectation and weak convergence to obtain the regularity of fully nonlinear PDE associated to G-BSDE. This is a joint work with Shaolin Ji and Xiaojuan Li.

非线性期望、机器学习与统计推断

***Xiaodong Yan*，严晓东**
Shandong University yanxiaodong@sdu.edu.cn

Abstract

非线性期望理论（彭实戈院士原创性成果）是结合概率测度不确定性、BSDE 的巧妙构造方法，将概率空间与非线性期望空间结合，从而跳出 Kolmogorov 的框架，得到的一系列全新理论成果。特别地，最近山东大学非线性概率团队基于多臂老虎机发展了一系列策略极限理论，作为非线性期望理论的重要研究成果。本报告考虑这一原创理论成果的统计科学问题以及提出新的统计机器学习方法，其中包括（1）双臂老虎机下的最优分布和悖论；（2）针对大数据下的双样本检验；（3）提出序贯 bootstrap 方法进行相似性假设检验和单样本检验的一般框架；（4）最后为了考虑一般化的模型，我们提出了基于随机梯度下降法的参数检验。值得注意的是，新的统计机器学习方法充分结合“知识”属于知识推理与数据驱动的统计学习方法，并且新的方法解决的假设检验问题与传统方法相比获得了更高的势。

赌博机问题的非线性极限定理

***Guodong Zhang*, 张国栋**

Shandong University zhang_gd@sdu.edu.cn

Abstract

Motivated by the study of asymptotic behavior of the bandit problems, we obtain several strategy-driven central limit theorems. To describe the fluctuations around averages, we obtain strategy-driven central limit theorems under optimal strategies. The limits in these theorems are identified explicitly, and depend heavily on the structure of the events or the integrating functions and strategies. This demonstrates the key signature of the learning structure. Our results lay the theoretical foundation for statistical inference in determining the arm that offers the higher mean reward.

LazAE: learning the latent features for the classifications of single-cell RNA-seq samples with zero-inflated negative binomial distribution as the autoencoder loss

***Shishun Zhao*, 赵世舜**

Jilin University zhaoss@jlu.edu.cn

Abstract

Single-cell RNA sequencing (scRNA-seq) is a revolutionary update to the high-throughput sequencing technology and provides the single-cell resolution map of a transcriptome. The conventional bulk RNA-seq profiles the average transcriptomes of a cell population and may not precisely represent the expression patterns of a complicated tissue system consisting multiple cell types. The scRNA-seq technology has been rapidly developed for the investigations of complicated life systems, and becomes a powerful method to answer many biomedical questions at the single-cell resolution. Many bioinformatics tools have been developed to determine the cell types of the scRNA-seq transcriptomic samples or cell type-specific differentially expressed genes. The recent innovation in biological sequencing technology has facilitated the single-cell RNA-seq (scRNA-seq) to answer many biomedical questions. One of the studies is to classify the cell types and the lesion locations of the scRNA-seq data. The high-dimensional and noisy inherent nature of the scRNA-seq data make it computationally challenging for such classification tasks. Deep learning has achieved many successful prediction questions, and this study utilized the statistical zero-inflated negative binomial (ZINB) distribution as the loss function of the autoencoder network (LazAE). Considering the excessive amount of missing data and noise within the scRNA-seq data, we hypothesized that the ZINB distribution may achieve a better capture of the inter-feature correlations than the conventional cross entropy loss. The experimental data on the simulation and real datasets supported that the proposed LazAE model outperformed the existing methods on the scRNA-seq datasets, particularly for the datasets with large numbers of samples.

Analysis of parent-of-origin effects for secondary phenotypes using case-control mother-child pair data

***Hong Zhang*, 张洪**

University of Science and Technology of China zhangh@ustc.edu.cn

Abstract

The detection of parent-of-origin effects (POEs) has become a research focus in genetic association studies since POEs play an important role in explaining the heritability of many complex human disorders. Genetic association studies are commonly conducted based on case–control designs. Case–control genetic association studies often collect additional information on secondary phenotypes other than the case–control status. Various statistical methods have been proposed to analyze the secondary phenotypes, but no methods are specifically tailored for identifying POEs of offspring genes on the secondary phenotypes. The parental origin information may not be determined unambiguously using the genotypes of the test locus for some families, and ignoring such families would lose considerable information. In this talk, we will focus on case–control mother–child pair design that has been widely used for studying human early life growth and development, and propose a robust and efficient retrospective likelihood method to detect POEs for the secondary phenotypes using multilocus genotypes. The proposed method fully utilizes the information from multilocus genotypes, Hardy–Weinberg equilibrium (HWE), Mendelian inheritance law, and conditional independence between child genotype and maternal covariate given maternal genotype. Large sample properties, including consistency and asymptotic normality, are established for our proposed statistical method. The finite sample performance of our method are demonstrated through extensive simulation studies and application to the Danish National Birth Cohort data.

基于 IADL 标准的老年人健康预期寿命研究

***Yong Wen*, 温勇**

Nanjing University of Posts and Telecommunications ywen1108@163.com

Abstract

本文基于 2013、2015 和 2018 年中国健康与养老追踪调查数据（CHARLS），通过多状态生命表法，分别计算 ADL 和 IADL 标准下我国老年人的健康预期寿命；从人口、社会支持、健康水平、经济水平和子女支持等五大特征中选取了 16 个相关变量，通过 logistic 回归，分析影响老年人独立生活能力健康预期寿命的主要因素；分析我国老年人独立生活能力健康预期寿命存在的性别差异和城乡差异，在此基础上进一步探究了影响老年人独立生活能力健康预期寿命的影响因素，为制定养老政策和分配养老资源提供了理论与政策依据。

Paired or partially paired two-sample tests with unordered samples

***Yanlin Tang*，唐炎林**

East China Normal University yltang@fem.ecnu.edu.cn

Abstract

In paired two-sample tests for mean equality, it is common to encounter unordered samples in which subject identities are not observed or unobservable, and it is impossible to link the measurements before and after treatment. The absence of subject identities masks the correspondence between the two samples, rendering existing methods inapplicable. In this paper, we propose two novel testing approaches. The first splits one of the two unordered samples into blocks and approximates the population mean using the average of the other sample. The second method is a variant of the first, in which subsampling is used to construct an incomplete U-statistic. Both methods are affine invariant and can readily be extended to partially paired two-sample tests with unordered samples. Asymptotic null distributions of the proposed test statistics are derived and the local powers of the tests are studied. Comprehensive simulations show that the proposed testing methods are able to maintain the correct size, and their powers are comparable to those of the oracle tests with perfect pair information. Four real examples are used to illustrate the proposed methods, in which we demonstrate that naive methods can yield misleading conclusions.

High dimensional beta test with high frequency data

***Dachuan Chen*，陈大川**

Nankai University dchen@nankai.edu.cn

Abstract

This is the first paper about the high dimensional beta tests with high frequency financial data, which allow the number of regressors be larger than the number of observations within each estimation block and can grow to infinity in asymptotics. In this paper, the sum-type test and max-type test have been proposed, where the sum-type test is suitable for the dense alternative and the max-type test is suitable for the sparse alternative. By showing the asymptotic independence between the sum-type test and max-type test, a Fisher's combination test is proposed, which is robust to both dense and sparse alternatives. The limiting null distributions of the three proposed tests are derived and the asymptotic behavior

of their powers are also analyzed. Monte Carlo simulations demonstrate the validity of the theoretical results developed in this paper. Empirical study with real high frequency financial data shows the robustness of the proposed Fisher's combination test under both dense and sparse alternatives. This is the joint work with Long Feng, Per Mykland and Lan Zhang.

PMAT: A tool for identifying differentially methylated regions for monozygotic twins

***Xiaoqing Pan*, 潘小青**

Shanghai Normal University xpan@shnu.edu.cn

Abstract

DNA methylation plays a vital role in gene transcriptional regulation. With the advent of next-generation sequencing technologies, reduced representation bisulfite sequencing (RRBS) is becoming increasingly common for analyzing genome-wide methylation profiles at the single nucleotide level. A major goal of RRBS studies is to detect differentially methylated regions (DMRs) between different biological conditions. Monozygotic twins, treated as unordered pairs, are classical epidemiological designs to examine the genetic and environmental influence in complex diseases. However, no DMR identification tool for paired samples is currently available. In this study, we present an innovative computational tool, PMat, combing folded normal test with a binary segmentation algorithm, to identify DMRs in twin samples.

主持人：胡明尚，山东大学

Risk measurement of joint risk of portfolios: a liquidity shortfall aspect

***Yijun Hu*，胡亦钧**

Wuhan University yjhu.math@whu.edu.cn

Abstract

In this talk, I will present a novel axiomatic framework of measuring the joint risk of a portfolio consisting of several financial positions. Precisely, from the liquidity shortfall aspect, we first construct a distortion-type risk measure to measure the joint risk of portfolios, which we refer to as multivariate distortion joint risk measure, representing the liquidity shortfall caused by the joint risk of portfolios. After its fundamental properties have been studied, then we axiomatically characterize it by proposing a novel set of axioms. Furthermore, we also propose a new class of vector-valued multivariate distortion joint risk measures, as well as with sensible financial interpretation. It turns out that this new class is rich enough, as it can not only induce new vector-valued multivariate risk measures, but also recover some popular vector-valued multivariate risk measures known in the literature with alternative financial interpretation. This talk mainly gives some theoretical results, helping one to have an insight look at the measurement of joint risk of portfolios.

This talk is based on a joint work with Shuo Gong and Linxiao Wei.

自激发跳跃粗糙仿射机波动率模型的研究 -基于快速 Fourier-SINC 方法

***Xiangdong Liu*，柳向东**

Jinan University tliuxd@jnu.edu.cn

Abstract

基于资产价格跳跃存在自激性与股票类金融资产的对数波动率存在粗糙性质的典型特征，本文提出了自激发跳跃粗糙 Heston 模型（rHeston-Hawkes）。首先证明了模型是无套利的，随后在风险中性测度下推导出模型的特征函数。针对特征函数中出现的分数 Riccati 方程和一组常微分方程，使用高效的数值算法获得数值解。通过快速 Fourier-SINC 算法(FFT-SINC)获得了欧式期权价格。实证研究表明，我国 6 个具有代表性的指数的对数波动率 Hurst 指数均远小于 1/2，即波动率存在粗糙性，并基于 FFT-SINC 对上证 50ETF 期权进行定价实验发现 rHeston-Hawkes 模型在样本内外均有较好的定价精度。研究的结果对国内外期权产品定价与精确风险管理具有重要的现实意义和应用价值。

Maximum likelihood estimation for maximal distribution under sublinear expectation

***Xinpeng Li,* 李欣鹏**

Shandong University lixinpeng@sdu.edu.cn

Abstract

Maximum likelihood estimation is a common method of estimating the parameters of the probability distribution from a given sample. This talk aims to introduce the maximum likelihood estimation in the framework of sublinear expectation. We find the maximum likelihood estimator for the parameters of the maximal distribution via the solution of the associated minimax problem, which coincides with the optimal unbiased estimation given by Jin and Peng (2021). A general estimation method for samples with dependent structure is also provided. This result provides a theoretical foundation for the estimator of upper and lower variances, which is widely used in the G-VaR prediction model in finance.

.

主持人：戴万阳，南京大学

Hessian and increasing-Hessian orderings of multivariate skew-elliptical random vectors

***Chuancun Yin*, 尹传存**

Qufu Normal University ccyin@qfnu.edu.cn

Abstract

In this talk, we establish some stochastic comparison results for multivariate skew-elliptical random vectors. These multivariate stochastic comparisons involve Hessian and increasing-Hessian orderings as well as many of their special cases. Necessary and/or sufficient conditions of the orderings are provided simply based on a comparison of the underlying model parameters.

Event-triggered control problem of stochastic nonlinear delay systems

***Quanxin Zhu*, 朱全新**

Hunan Normal University zqx22@126.com

Abstract

In this report, we introduce the the event-triggered feedback control problem of stochastic nonlinear delay systems with exogenous disturbances. By introducing the notation of input-to-state practical stability and an event-triggered strategy, we establish the input-to-state practically exponential mean-square stability of the suggested system. Moreover, we investigate the stabilization result by designing the feedback gain matrix and the eventtriggered feedback controller, which is expressed in terms of linear matrix inequalities. Also, the lower bounds of inter-execution times by the proposed event-triggered control method are obtained. Finally, an example is given to show the effectiveness of the proposed method. Compared with large number of results for discrete-time stochastic systems, only a few results have appeared on the event-triggered control for continuous-time stochastic systems. In particular, there has been no published papers on the event-triggered control for continuous-time stochastic delay systems. Our work is a first try to fill the gap on the topic.

Large deviation principles of realized Laplace transform of volatility

***Xinwei Feng*, 冯新伟**

Shandong University xwfeng@sdu.edu.cn

Abstract

In this paper, we investigate the asymptotic tail behaviour of the empirical realized Laplace transform of volatility (ERLTV) under the scenario of high-frequency data. We establish both a large deviation principle and a moderate deviation principle for the ERLTV. The good rate function for the large deviation principle is well defined in the whole real space, which indicates a limit for the normalized logarithmic tail probability of the ERLTV. Moreover, we also derive the function-level large and moderate deviation principles for ERLTV.

主持人：严晓东，山东大学

Testing for structural change of predictive regression model to threshold predictive regression model

***Fukang Zhu*, 朱复康**

Jilin University zfk8010@163.com

Abstract

We investigate two test statistics for testing structural changes and thresholds in predictive regression models. The generalized likelihood ratio (GLR) test is proposed for the stationary predictor and the generalized F test is suggested for the persistent predictor. Under the null hypothesis of no structural change and threshold, it is shown that the GLR test statistic converges to a function of a centered Gaussian process, and the generalized F test statistic converges to a function of Brownian motions. A Bootstrap method is proposed to obtain the critical values of test statistics. Simulation studies and a real example are given to assess the performances of the proposed tests.

Weighted residual empirical processes, martingale transformations, and model checking for regressions

***Falong Tan*, 谭发龙**

Hunan University falongtan@hnu.edu.cn

Abstract

This paper proposes a new methodology for testing the parametric forms of the mean and variance functions based on weighted residual empirical processes and their martingale transformations in regression models. The dimensions of the parameter vectors can be divergent as the sample size goes to infinity. We then study the convergence of weighted residual empirical processes and their martingale transformation under the null and alternative hypotheses in the diverging dimension setting. The proposed tests based on weighted residual empirical processes can detect local alternatives distinct from the null at the fastest possible rate of order $n^{-\frac{1}{2}}$ but are not asymptotically distribution-free. While the tests based on martingale transformed weighted residual empirical processes can be asymptotically distribution-free, yet, unexpectedly, can only detect the local alternatives converging to the null at a much slower rate of order $n^{-\frac{1}{4}}$, which is somewhat different from existing asymptotically distribution-free tests based on martingale transformations. As

the tests based on the residual empirical process are not distribution-free, we propose a smooth residual bootstrap and verify the validity of its approximation in diverging dimension settings. Simulation studies and a real data example are conducted to illustrate the effectiveness of our tests.

Consistency of BIC model averaging with application to importance learning

***Ze Chen*, 陈泽**

Renmin University of China chze96@163.com

Abstract

BIC weighting has been frequently applied to high-dimensional linear regression when model averaging is considered to address model selection uncertainty. It also plays a central role in model selection diagnostics. However, little research has been done on its consistency or weak consistency, which are crucial properties for a model averaging method to perform well for various purposes. In addition, previous limited work on model averaging consistency excludes the consideration of categorical covariates. In this paper, with both continuous covariates and categorical predictors (with possibly diverging numbers of levels) allowed, we establish both consistency and weak consistency for BIC weighting. On the basis of the BIC weights, we perform a sparsity oriented importance learning (SOIL) to measure importances of the predictors, which is shown to be able to asymptotically identify the variables in the true model under consistency of the weights and almost all the variables in the true model under weak consistency. A COVID-19 epidemic data analysis illustrates the superiority of the proposed method in contrast to some other variable importance measures. Furthermore, it leads to models that correct an apparently puzzling wrong sign of a key predictor reported in a Science paper.

主持人：赵世舜，吉林大学

Instability of inverse probability weighting methods and a remedy for non-ignorable missing data

Yukun Liu，刘玉坤

East China Normal University ykliu@sfs.ecnu.edu.cn

Abstract

Inverse probability weighting (IPW) methods are commonly used to analyze non-ignorable missing data under the assumption of a logistic model for the missingness probability. However, solving IPW equations numerically may involve non-convergence problems when the sample size is moderate and the missingness probability is high. Moreover, those equations often have multiple roots, and identifying the best root is challenging. Therefore, IPW methods may have low e_ciency or even produce biased results. We identify the pitfall in these methods pathologically: they involve the estimation of a moment-generating function, and such functions are notoriously unstable in general. As a remedy, we model the outcome distribution given the covariates of the completely observed individuals semiparametrically. After forming an induced logistic regression model for the missingness status of the outcome and covariate, we develop a maximum conditional likelihood method to estimate the underlying parameters. The proposed method circumvents the estimation of a moment-generating function and hence overcomes the instability of IPW methods. Our theoretical and simulation results show that the proposed method outperforms existing competitors greatly. Two real data examples are analyzed to illustrate the advantages of our method. We conclude that if only a parametric logistic regression is assumed but the outcome regression model is left arbitrary, then one has to be cautious in using any of the existing statistical methods in problems involving non-ignorable missing data.

Nonequivalence of two least-absolute-deviation estimators for mediation effects

Wenwu Wang，王文武

Qufu Normal University wangwenwu@qfnu.edu.cn

Abstract

This paper provides two groups of conditions of model consistency in least-absolute-deviation mediation models. Under model consistency, we establish the asymptotic theory of the difference estimator and the product estimator, and show that the

two estimators are not only numerically nonequivalent but asymptotically nonequivalent, which is dramatically different from the situation in the least squares mediation analysis where these two estimators are numerically equivalent. In all three possible scenarios of model parameters, both the asymptotic theories and simulation studies show that the product estimator is more efficient than the difference estimator.

Quantization: Is it possible to improve classification?

***Weiyu Li*, 李蔚郁**

Shandong University liweiyu@sdu.edu.cn

Abstract

Large-scale data classification poses great challenges to computation and storage. There are two major solutions to address the problem: the data dimension reduction and quantization. In the paper, we study the method of first reducing data dimension by random projection and then quantizing the projections to ternary and binary codes, which has been widely applied in practice. Often, the extreme quantization would degrade the accuracy of classification due to high quantization errors. Interestingly, however, we observe that the quantization could result in performance improvement, rather than degradation, if the data for quantization are preprocessed by sparse transforms. Also, the quantization gain could be obtained with the random projections of the data, if both the data and random projection matrices are sparse enough, such that the resulting projections remain sparse. The intriguing performance is verified and analyzed with extensive experiments.

Goodness-of-fit test for a parametric mixture cure model with partly interval-censored data

***Xiaoguang Wang*，王晓光**

Dalian University of Technology wangxg@dlut.edu.cn

Abstract

Partly interval-censored event time data arise naturally in medical, biological, sociological and demographic studies. In practice, some patients may be immune from the event of interest, invoking a cure model for survival analysis. Choosing an appropriate parametric distribution for the failure time of susceptible patients is an important step to fully structure the mixture cure model. In the literature, goodness-of-fit tests for survival models are usually restricted to uncensored or right-censored data. We fill in this gap by proposing a new goodness-of-fit test dealing with partly interval-censored data under mixture cure models. Specifically, we investigate whether a parametric distribution can fit the susceptible part by using a Cramér-von Mises type of test, and establish the asymptotic distribution of the test. Empirically, the critical value is determined from the bootstrap resamples. The proposed test, compared to the traditional leveraged bootstrap approach, yields superior practical results under various settings in extensive simulation studies. Two clinical data sets are analyzed to illustrate our method.

Mendelian randomization accounting for complex correlated horizontal pleiotropy while elucidating shared genetic etiology

***Qing Cheng*，成青**

Southwestern University of Finance and Economics chengqing@swufe.edu.cn

Abstract

Mendelian randomization (MR) harnesses genetic variants as instrumental variables (IVs) to study the causal effect of exposure on outcome using summary statistics from genome-wide association studies. Classic MR assumptions are violated when IVs are associated with unmeasured confounders, i.e., when correlated horizontal pleiotropy (CHP) arises. Such confounders could be a shared gene or inter-connected pathways underlying exposure and outcome. We propose MR-CUE (MR with Correlated horizontal pleiotropy Unraveling shared Etiology and confounding), for estimating causal effect while identifying IVs with CHP and accounting for estimation uncertainty. For those IVs, we map their cis-

associated genes and enriched pathways to inform shared genetic etiology underlying exposure and outcome. We apply MR-CUE to study the effects of interleukin 6 on multiple traits/diseases and identify several S100 genes involved in shared genetic etiology. We assess the effects of multiple exposures on type 2 diabetes across European and East Asian populations.

The asymptotic normality of residual density estimator in stationary and explosive auto-regressions

***Wenzhi Yang*, 杨文志**

Anhui University wzyang@ahu.edu.cn

Abstract

We consider the error density estimator in the first-order autoregressive model based on strong-mixing errors. Since the errors are not observed, the residual kernel density estimator is suggested. Then, the asymptotic normality of the residual estimator is obtained when the autoregressive model is a stationary process or an explosive process. In addition, some simulations such as the fitted curves, mean integrated square errors and histograms are illustrated to the residual kernel estimator and residual histogram estimator. It is shown that the residual kernel estimator with smooth kernel is smoother than the residual histogram estimator.

主持人：杜恺，复旦大学

Asymptotic properties of the unit root model

***Hui Jiang*，蒋辉**

Nanjing University of Aeronautics and Astronautics huijiang@nuaa.edu.cn

Abstract

In this talk, we consider asymptotic properties for the quadratic functionals and associated ordinary least squares estimator in the unit root model with Gaussian noise. Deviation inequalities, nonuniform Berry-Esseen bound and Cramer-type moderate deviations are achieved. The main methods used in this talk consist of the change of measure, deviation inequalities for multiple Wiener-Ito integrals, as well as asymptotic analysis techniques.

Heavy-traffic limit for a fluid queue with ON/OFF sources

***Hongshuai Dai*，戴洪帅**

Shandong University of Finance and Economics mathdsh@gmail.com

Abstract

Fractional Brownian motion approximation of queueing networks has been studied extensively. In the existing results related to this topic, the Hurst parameter of multidimensional fractional Brownian motion is only a constant H ($0 < H < 1$). However, just as pointed out by many scholars and practitioners, various Hurst parameters may be more appropriate. On the other hand, as a multivariate extension of fractional Brownian motion, operator fractional Brownian motion (ofBm) has operator self-similarity, and the dependence structure across the components of ofBm is determined by the Hurst matrix. Inspired by these facts, we consider a fluid queueing network with ON/OFF sources, and show that the workload process can be approximated by a reflected operator fractional Brownian motion under a heavy traffic condition.

Asymptotic normality for nonstationary high frequency data

***Yuping Song*, 宋玉平**

Shanghai Normal University songyuping@shnu.edu.cn

Abstract

In this paper, we investigate the estimation for integrated jump-diffusion models with nonstationary high frequency data. As using the original data directly to estimate the unknown parameters in the models makes theoretical analysis almost not possible, we use a local approximation to derive desired estimation based on the original data. Moreover, as extreme jumps easily produce outliers, M-estimation procedure is considered for infinitesimal coefficients associated with integrated jump-diffusion models. However, the classical M-estimation can not possess the estimation consistency, the asymptotic normality and the rate of convergence, so we propose a lag-1-based local M-estimation procedure to derive these important results. As a by-product, the designed conditions for these asymptotic results can even be weaker than those for classical independent identically distributed settings in the literature. The numerical studies suggest the advantages of our method in bias reduction and computational efficiency. A real data example about the returns for stock index under five-minute high sampling frequency is analyzed for illustration.

主持人：胡亦钧，武汉大学

Infinite horizon FBSDEs and open-loop optimal controls for stochastic LQ problems with random coefficients

Zhiyong Yu，于志勇

Shandong University yuzhiyong@sdu.edu.cn

Abstract

In this paper, we introduce a new infinite horizon domination-monotonicity framework. In this framework, by the method of continuation and some subtle techniques, we obtain an existence and uniqueness result and a pair of estimates for the solutions to a kind of infinite horizon coupled forward-backward stochastic differential equations (FBSDEs, for short). Then, the theoretical result of FBSDEs is applied to solve a stochastic linear-quadratic (LQ, for short) optimal control problem with random time-varying coefficients on infinite horizon. The unique open-loop optimal control is characterized by the solution of an infinite horizon FBSDE. Moreover, we find and illustrate a different phenomenon between the LQ problems on infinite horizon and finite horizon.

Sequential propagation of chaos

Kai Du，杜恺

Fudan University kdu@fudan.edu.cn

Abstract

A new class of particle systems with sequential interaction is proposed to approximate the McKean-Vlasov process that originally arises as the limit of the mean-field interacting particle system. The weighted empirical measure of this particle system is proved to converge to the law of the McKean-Vlasov process as the system grows. Based on the Wasserstein metric, quantitative propagation of chaos results are obtained for two cases: the finite time estimates under the monotonicity condition and the uniform in time estimates under the dissipation and the non-degenerate conditions. Numerical experiments are implemented to demonstrate the theoretical results.

Stochastic verification theorem for stochastic control problems of reflected FBSDEs

***Qingmeng Wei*, 魏庆萌**

Northeast Normal University weiqm100@nenu.edu.cn

Abstract

We study the stochastic verification theorem for stochastic control problems of reflected FBSDEs. The feedback optimal control is constructed from the solution of the obstacle problem of Hamilton-Jacobi-Bellman (HJB, for short) equation. Our work is carried out within the frameworks of classical solutions and viscosity solutions.

主持人：孔翠娟，山东大学

Huber principal component analysis for large-dimensional factor models

Yong He, **何勇**

Shandong University heyong@sdu.edu.cn

Abstract

Factor models have found widespread applications in economics and finance, but the heavy-tailed character of macroeconomic and financial data is often overlooked in the existing literature. To address this issue and achieve robustness, we propose an approach to estimate factor loadings and scores by minimizing the Huber loss function, motivated by the equivalence of conventional Principal Component Analysis (PCA) and the constrained least squares method in the factor model. We provide two algorithms based on different penalty forms. The first minimizes the ℓ_2 -norm-type Huber loss, performing PCA on the weighted sample covariance matrix and is named Huber PCA. The second version minimizes the element-wise type Huber loss and can be solved by an iterative Huber regression algorithm. We investigate the theoretical minimizer of the element-wise Huber loss function and show that it has the same convergence rate as conventional PCA under finite second-moment conditions on idiosyncratic errors. Additionally, we propose a consistent model selection criterion based on rank minimization to determine the number of factors robustly. We demonstrate the advantages of Huber PCA using a real financial portfolio selection example, and an R package called ``HDRFA" is available on CRAN to conduct robust factor analysis. This is joint work with Lingxiao Li, Dong Liu and Wen-Xin Zhou.

Change point detection for high-dimensional linear models: a general tail-adaptive approach

Bin Liu**, 刘彬**

Fudan University bin_liu@fudan.edu.cn

Abstract

We study the change point detection problem for high-dimensional linear regression models. In this work, we propose a novel tail-adaptive approach for simultaneous change point testing and estimation. The method is built on a new loss function which is a weighted combination between the composite quantile and least squared losses, allowing us to borrow information of the possible change points from both the conditional mean and quantiles. For the change point testing, based on the adjusted L_2-norm aggregation of a

weighted score CUSUM process, we pro-pose a family of individual testing statistics with different weights to account for the unknown tail structures. Through a combination of the individual tests, a tail-adaptive test is further constructed that is powerful for sparse alternatives of regression coefficients' changes under various tail structures. For the change point estimation, a family of argmax-based individual estimators is proposed once a change point is detected. In theory, for both individual and tail-adaptive tests, bootstrap procedures are proposed to approximate their limiting null distributions. Under some mild conditions, we justify the validity of the new tests in terms of size and power under the high-dimensional setup. The corresponding change point estimators are shown to be rate optimal up to a logarithm factor.

Testing the number of common factors by bootstrapped sample covariance matrix in high-dimensional factor models

***Long Yu,* 虞龙**

Shanghai University of Finance and Economics yulong@mail.shufe.edu.cn

Abstract

This paper studies the impact of bootstrap procedure on the eigenvalue distributions of the sample covariance matrix under the high-dimensional factor structure. We provide asymp- totic distributions for the top eigenvalues of bootstrapped sample covariance matrix under mild conditions. After bootstrap, the spiked eigenvalues which are driven by common factors will converge weakly to Gaussian limits via proper scaling and centralization. However, the largest non-spiked eigenvalue is mainly determined by order statistics of bootstrap resam- pling weights, and follows extreme value distribution. Based on the disparate behavior of the spiked and non-spiked eigenvalues, we propose innovative methods to test the number of common factors. According to the simulations and a real data example, the proposed meth- ods are the only ones performing reliably and convincingly under the existence of both weak factors and cross-sectionally correlated errors. Our technical details contribute to random matrix theory on spiked covariance model with convexly decaying density and unbounded support, or with general elliptical distributions.

Stochastic integral bootstrap for statistics of irregularly spaced spatial data

Shibin Zhang, **张世斌**

Shanghai Normal University zhang_shibin@shnu.edu.cn

Abstract

Resampling approaches to approximate the sampling distribution of the statistic constructed from irregularly spaced data are still far from well-developed. We propose a novel bootstrap method, the stochastic integral bootstrap (SIB), as a complement to the existing approaches. It uses a stochastic integral to approximate the sampling distribution of a statistic rather than generating a resample from the observation. Meanwhile, its variance is consistent to that of the statistic. The stochastic integral is constructed by integrating a block-constrained version of the statistic with respect to the standard Gaussian white noise, which provides a general class of resampling estimators based on irregularly spaced spatial data. Moreover, it is also proved that the SIB is applicable to various important statistics such as the sample mean, the sample variance, the auto-covariance estimator, the discrete Fourier transform and some test statistics for spatial white noise as well. In addition, the SIB can imitate the second-order dependence of multiple statistics. Simulation studies illustrate the finite sample performance of the SIB in comparison with some competitive counterparts for irregularly spaced spatial data.

Robust estimation and test based on median-of-means method

Pengfei Liu, **刘鹏飞**

Jiangsu Normal University liupengfei@jsnu.edu.cn

Abstract

Using the idea of grouping under moderate data framework, we propose the median-of-means (MoM) type nonparametric estimator for parameters of statistical model. Under certain condition on the growing rate of the number of subgroups, the consistency and asymptotic normality of the proposed estimator are investigated. Furthermore, we construct a new method to test the parameters based on the empirical likelihood method for median. Extensively numerical simulations are designed to demonstrate the superiorities of our estimator. It is shown that the new proposed estimator is quite robust with respect to outliers. We also apply the MoM method to analyze some real data sets.

Linear regression estimation using intraday high frequency data

***Xingfa Zhang,* 张兴发**

Guangzhou University xingfazhagn@gzhu.edu.cn

Abstract

Intraday high frequency data have shown important values in econometric modeling and have been extensively studied. Following this point, in this paper, we study the linear regression model for variables which have intraday high frequency data. In order to overcome the nonstationarity of the intraday data, intraday sequences are aggregated to the daily series by weighted mean. A lower bound for the trace of the asymptotic variance of model estimator is given, and a data-driven method for choosing the weight is also proposed, with the aim to obtain a smaller sum of asymptotic variance for parameter estimators. The simulation results show that the estimation accuracy of the regression coefficient can be significantly improved by using the intraday high frequency data. Empirical studies show that introducing intraday high frequency data to estimate CAPM can have a better model fitting effect.

主持人：王潇逸，北京师范大学

A variation-ratio test for volatility jumps using noisy high frequency data

***Guangying Liu*, 刘广应**

Nanjing Audit University liugying@nau.edu.cn

Abstract

This paper proposes a novel variation-ratio test for the presence of volatility jumps using high frequency data with microstructure noise. Under the null hypothesis that the volatility process is a continuous semimartingale, the test statistic is asymptotically normal. Under the alternative hypothesis that the volatility process jumps, the test statistic diverges to infinity {at rate $n^{1/4-\iota}$ for arbitrarily small $\iota > 0$, which is faster than the best rate (close to $n^{1/8}$ in the semimartingale setup) available in the literature, where n is the sampling frequency.} Simulation results corroborate our theoretical findings. Empirical results show that our test fails to reject the null hypothesis for most of the ninety US stocks under study except for quite a small portion of them. (Based on joint work with Yang Li (NAU) and Zhiyuan Zhang (SHUFE)).

Optimal confidence intervals for the relative risk and odds ratio

***Tianfa Xie*, 谢田法**

Beijing University of Technology xietf@bjut.edu.cn

Abstract

The relative risk and odds ratio are widely used in many fields, including biomedical research, to compare two treatments. Extensive research has been done to infer the two parameters through approximate or exact confidence intervals. However, these intervals may be liberal or conservative. A natural question is whether the intervals can be further improved in maintaining the correct confidence coefficient of an approximate interval or shortening an exact but conservative interval. In this paper, when two independent binomials are observed we offer an effort to improve some of the existing intervals by applying the h-function method. In particular, if the given interval is approximate, then the improved interval is exact; if the given interval is exact, then the improved interval is a subset of the given interval. This method is also applied multiple times to the improved intervals until the final resultant interval cannot be shortened any further. To demonstrate the effectiveness of the method, we use two real datasets to illustrate in detail how several good intervals in practice are improved.

Two exact intervals are then recommended for estimating each of the two parameters in different scenarios.

Person-centered semiparametric multiplicative modeling on recurrent event data

***Wei Zhao,* 赵玮**

Shandong University wzhao92@sdu.edu.cn

Abstract

In recent years, the concept of person-centered approach has gained increased attention. A person-centered focus is also useful when dealing with repeated measurements to represent heterogeneity in development trajectories. However, repeated measurement studies can be challenging due to the magnitude of data provided by long-term studies. In this work, we investigate latent class analysis of recurrent events data based on flexible semiparametric multiplicative modeling. We derive a robust estimation procedure through novelly adapting the conditional score technique and utilizing the special characteristics of multiplicative intensity modeling. The proposed estimation procedure can be stably and efficiently implemented based on existing computational routines. We provide solid theoretical underpinnings for the proposed method, and demonstrate its satisfactory finite sample performance via extensive simulation studies. An application to a dataset from research participants at Goizueta Alzheimer's Disease Research Center illustrates the practical utility of our proposals.

Continuous ergodic capacities

Yongsheng Song, **宋永生**

AMSS, Chinese Academy of Sciences yssong@amss.ac.cn

Abstract

We give a characterization of the structure of the set Θ for a continuous ergodic upper probability $V = \mathrm{Sup}_{P\in\Theta}P$.

- Θ contains a finite number of ergodic probabilities;
- Any invariant probability in Θ is a convex combination of those ergodic ones in Θ;
- Any probability in Θ Θ coincides with an invariant one in Θ on the invariant σ-algebra.

The last property has already been obtained in Cerreia-Vioglio, Maccheroni, and Marinacci [1], which firstly studied the ergodicity of such capacities.

As an application of the characterization, we prove an ergodicity result, which improves the result in [1] in the sense that the limit of the time mean is bounded by the upper expectation, instead of the Choquet integral. Generally, the former is strictly smaller.

Extended dynamic programming principle and applications to time-inconsistent control

Shuzhen Yang**, 杨淑振**

Shandong University yangsz@sdu.edu.cn

Abstract

This paper investigates an extended dynamic programming principle (DPP) for a general stochastic control problem in which the state processes are described by a forward-backward stochastic differential equation (FBSDE). A multidimensional DPP is established with auxiliary dimensions defined by a BSDE. Consequently, an extended Hamilton-Jacobi-Bellman (HJB) equation is derived. The existence and uniqueness of smooth solution and a new type of viscosity solution are investigated for this extended HJB equation. Compared to extant research on the stochastic maximum principle, the present paper is the first normal work on the partial differential equation (PDE) method for the controlled FBSDE system. Interestingly, our model provides time-consistent solution for general time-inconsistent control problems associated with the traditional mean-variance model, risk-sensitive control and utility optimization for narrow framing investors, among others.

Stochastic representation under filtration consistent nonlinear expectations

***Hanwu Li,* 李邯武**

Shandong University lihanwu@sdu.edu.cn

Abstract

In this talk, we investigate the stochastic representation problem under filtration consistent nonlinear expectations. We establish existence and uniqueness of the solutions for the discrete time case. Besides, we provide a characterization of the solution, which is helpful to derive the uniqueness of solutions for the continuous time case. The stochastic representation can be applied to a variant Skorokhod problem with the increasing process behaving in a nonlinear fashion. Besides, it provides another method to construct the optimal stopping times for optimal stopping problems under filtration nonlinear expectations without calculating the Snell envelope.

主持人：冯新伟，山东大学

Hydrodynamics and fluctuations of N-urn SIR epidemics

Xiaofeng Xue, **薛晓峰**
Beijing Jiaotong University xfxue@bjtu.edu.cn

Abstract

In this talk we are concerned with N-urn susceptible-infected-removed epidemics, where each urn is in one of three states, namely susceptible, infected and removed. We assume that recovery rates of infected urns and infection rates between infected and susceptible urns are all coordinate-dependent. We show that the hydrodynamic limit of our model is driven by a deterministic measure-valued process with density which is the solution to a nonlinear ordinary differential equation consistent with a mean-field analysis. We further show that the fluctuation of our process is driven by a generalized Ornstein-Uhlenbeck process. A key step in proofs of above main results is to show that states of different urns are approximately independent as N grows to infinity.

Multi-perspective crude oil price forecasting with a new decomposition-ensemble framework

Jingjun Guo, **郭精军**
Lanzhou University of Finance and Economics guojj@lzufe.edu.cn

Abstract

Crude oil is an important global commodity, and its price fluctuation affects the political and economic security of a country. Therefore, it is necessary to conduct crude oil price forecasting. Based on the forecasting research of multi-source information and decomposition-ensemble, we combine the two into a model and propose a multiperspective crude oil price forecasting model under a new decomposition-ensemble framework. Specifically, the crude oil price series is decomposed and reconstructed into several modes through variational mode decomposition (VMD) and fuzzy entropy (FE). Further, we screen the effective predictors from structured and unstructured multi-source data using the Granger causality test, and select the optimal input features through random forest - recursive feature elimination (RF-RFE). Finally, each reconstruction mode is individually forecasted on the basis of the selected different input features and the forecasting values obtained are combined and integrated; the final result is obtained from the integrating

prediction results through the error evaluation criterion. The West Texas Intermediate (WTI) daily spot price is adopted to validate the performance of our proposed model. The empirical results show that compared with the benchmark models, the proposed model can significantly improve forecasting accuracy.

A unified framework for nonconvex nonsmooth sparse and low-rank decomposition by Majorization-Minimization algorithm

***Pingfeng Xu*, 徐平峰**

Northeast Normal University xupf900@nenu.edu.cn

Abstract

Recovering a low-rank matrix and a sparse matrix from an observed matrix, known as sparse and low-rank decomposition (SLRD), is becoming a hot topic in recent years. The most popular model for SLRD is to use the L_1 norm and nuclear norm for the sparse and low-rank approximation. Since this convex model has certain limitations, various nonconvex models have been explored and found to be very promising. In this paper, we introduce a generalized nonconvex nonsmooth model for SLRD which covers a wide range of nonconvex surrogate functions that are continuous, concave and monotonically increasing on $[0,\infty)$ to approximate both the L_0 norm and the rank function, such as L_p norm ($0<p<1$), Logarithm, Geman, SCAD and MCP functions. The choice of the nonconvex surrogates for the sparse and low-rank components can be different. Due to the nonconvexity and extensive options of the surrogates, the optimization problem is untractable. Based on the majorization-minimization (MM) algorithm, we propose a unified framework named MM-ADMM algorithm to solve this problem, which can be applied to all eligible surrogates as long as their supergradients are available. The constrained majorizing problems established under the MM framework can be easily solved by the alternating direction method of multipliers (ADMM). The theoretical convergence properties are investigated and proved, including the convergence of the sequence of objective function values generated by the designed algorithm and a weak convergence result related to the inner ADMM-iterations. We compare MM-ADMM with BCD and GoDec to validate the efficiency of MM-ADMM in different applications of SLRD including image denoising and video background and foreground separation.

主持人：罗德军，中国科学院数学与系统科学研究院

Numerical approximation for Poisson equation

Yuanyuan Liu, **刘源远**

Central South University liuyy@csu.edu.cn

Abstract

Poisson's equation has a lot of applications in various areas. Usually it is hard to derive the explicit expression of the solution of Poisson's equation for a Markov chain on an infinitely many state space. We will present a computational framework for the solution for both discrete-time Markov chains and continuous-time Markov chains, by developing the technique of augmented truncation approximations. The convergence is established for two types of truncation approximation schemes: the censored chain and the linear augmented truncation. Moreover, truncation approximations to the variance constant in central limit theorems are also considered. The convergence error analysis is also discussed.

Martingale problem for a class of Lévy-type operators with low singularity kernels

Guohuan Zhao, **赵国焕**

AMSS, Chinese Academy of Sciences gzhao@amss.ac.cn

Abstract

In this talk, I will focus on the martingale problem for a class of Lévy-type operators with critically low singularity kernels. After showing the audience how to formulate some new regularity estimates for pseudodifferential operators with slow-growth symbols in generalized Orlicz-Besov spaces, I will explain how to use these estimates to prove the uniqueness of solutions to corresponding martingale problems. Moreover, a new Krylov-type estimate for the associated Markov processes will be introduced in this talk.

Integer compositions, random trees, and interval partition evolutions

***Quan Shi*, 石权**

AMSS, Chinese Academy of Sciences quan.shi@amss.ac.cn

Abstract

A composition of a positive integer n is a sequence of positive integers that sum to n. In this talk, I will study the scaling limits of a family of Markov chains on integer compositions. The limiting processes are interval-partition-valued diffusions with Poisson--Dirichlet (pseudo-)stationary distributions. This model is closely related to Pitman--Dubins Chinese restaurant processes and Markov chains on integer partitions studied by Borodin--Olshanski and Petrov. I will also talk about some applications of our model in population genetics and continuum-tree-valued dynamics.

This talk is based on joint work with Noah Forman, Douglas Rizzolo, and Matthias Winkel.

主持人：张世斌，上海师范大学

Nonparametric statistical inference via metric distribution function in metric spaces

***Wenliang Pan*，潘文亮**

AMSS, Chinese Academy of Sciences panwliang@amss.ac.cn

Abstract

Distribution function is essential in statistical inference, and connected with samples to form a directed closed loop by the correspondence theorem in measure theory and the Glivenko-Cantelli and Donsker properties. This connection creates a paradigm for statistical inference. However, existing distribution functions are defined in Euclidean spaces and no longer convenient to use in rapidly evolving data objects of complex nature. It is imperative to develop the concept of the distribution function in a more general space to meet emerging needs. Note that the linearity allows us to use hypercubes to define the distribution function in a Euclidean space, but without the linearity in a metric space, we must work with the metric to investigate the probability measure. We introduce a class of metric distribution functions through the metric between random objects and a fixed location in metric spaces. We overcome this challenging step by proving the correspondence theorem and the Glivenko- Cantelli theorem for metric distribution functions in metric spaces that lie the foundation for conducting rational statistical inference for metric space-valued data. Then, we develop a homogeneity test and a mutual independence test for non-Euclidean random objects and present comprehensive empirical evidence to support the performance of our proposed methods.

Research on image segmentation based on statistical method and deep learning

***Yuanyuan Ju*，句媛媛**

Kunming University of Science and Technology jundeyy@126.com

Abstract

In recent years, deep learning has developed rapidly and achieved excellent results. The breakthrough of deep learning has not only had a profound impact in the computer field, but also affected many fields. In the field of metallurgy, gas-liquid two-phase flow image segmentation is the key to accurately measure bubble parameters. However, deep learning method is directly used to study gas-liquid two-phase flow image segmentation,

which is difficult to obtain label data and has many parameters. Therefore, this paper proposes a self-supervised learning method based on Dirichlet process to solve the problem of label data. Secondly, based on the full convolutional neural network, a spatial transformers atrous spatial pyramid pooling module is constructed to improve the precision of image segmentation by using auxiliary information. Finally, factor analysis is introduced to reduce the number of convolution kernels for achieving the reduction of the parameters in the model.

Quantile difference estimation with censoring indicators missing at random

Cuijuan Kong, **孔翠娟**

Shandong University cuijuankong@sdu.edu.cn

Abstract

In this paper, we define estimators of distribution functions when the data are right-censored and the censoring indicators are missing at random, and establish their strong representations and asymptotic normality. Besides, based on empirical likelihood method, we define maximum empirical likelihood estimators and smoothed log-empirical likelihood ratios of two-sample quantile difference in the presence and absence of auxiliary information, respectively, and prove their asymptotic distributions. Simulation study and real data analysis are conducted to investigate the finite sample behavior of the proposed methods.

主持人：赵玮，山东大学

Adaptive tests for bandedness of high-dimensional covariance matrices

Xiaoyi Wang, **王潇逸**

Beijing Normal University wangxy059@bnu.edu.cn

Abstract

Estimation of the high-dimensional banded covariance matrix is widely used in multivariate statistical analysis. To ensure the validity of estimation, we aim to test the hypothesis that the covariance matrix is banded with a certain bandwidth under the high-dimensional framework. Though several testing methods have been proposed in the literature, the existing tests are only powerful for some alternatives with certain sparsity levels, whereas they may not be powerful for alternatives with other sparsity structures. The goal of this paper is to propose a new test for the bandedness of high-dimensional covariance matrix, which is powerful for alternatives with various sparsity levels. The proposed new test also be used for testing the banded structure of covariance matrices of error vectors in high-dimensional factor models. Based on these statistics, a consistent bandwidth estimator is also introduced for a banded high dimensional covariance matrix. Extensive simulation studies and an application to a prostate cancer dataset from protein mass spectroscopy are conducted for evaluating the effectiveness of the proposed adaptive tests and bandwidth estimator for the banded covariance matrix.

A constrained maximum likelihood approach to developing well-calibrated models for predicting binary outcomes

Yaqi Cao, **曹雅琦**

Minzu University of China yaqicaostats@163.com

Abstract

The added value of candidate predictors for risk modeling is routinely evaluated by comparing the performance of models with or without including candidate predictors. Such comparison is most meaningful when the estimated risk by the two models are both unbiased in the target population. Oftentimes, data for standard predictors in the base model is richly available from the target population, but data for candidate predictors are available only from nonrepresentative convenience samples. While the base model can be naively updated using the study data without recognizing the discrepancy between the underlying

distribution of the study data and that in the target population, the resultant risk estimates as well as evaluation of the candidate predictors are biased. Towards building a well-calibrated updated model, we propose a semiparametric method for model fitting that enforces calibration against a well-calibrated base model. Our method allows unbiased assessment of model improvement by candidate predictors without requiring a representative sample from the target population, thereby overcoming a major bottleneck in practice. We study theoretical properties for model parameter estimates, and demonstrate improvement in model calibration via extensive simulation studies. Finally, we apply the proposed method to data extracted from Penn Medicine Biobank to inform the added value of breast density for breast cancer risk assessment in the Caucasian woman population.

A framework to select tuning parameters for nonparametric derivative estimation

Sisheng Liu, **刘思晟**
Hunan Normal University ssl1989@hunnu.edu.cn

Abstract

In this paper, we propose a general framework to select tuning parameters for the nonparametric derivative estimation. The new framework enlarges the scope of the generalized Cp criterion Charnigo et al. (2011) by replacing the empirical derivative via any other linear nonparametric smoother. We provide the theoretical support of the proposed derivative estimation in a random design and justify it through simulation studies. The practical application of the proposed framework is demonstrated in the study of the age effect on hippocampal gray matter volume in healthy adults from the IXI dataset.

主持人：杨淑振，山东大学

The central limit theorem for stochastic Volterra equations with singular kernels

***Huijie Qiao*, 乔会杰**

Southeast University hjqiaogean@seu.edu.cn

Abstract

This work concerns stochastic Volterra equations with singular kernels. Under the suitable conditions, we prove the central limit theorem for them. Moreover, we apply our result to stochastic Volterra equations with the kernels of fractional Brownian motions with the Hurst parameter $H \in (0, 1)$.

A large deviation principle for the stochastic heat equation with general rough noise

***Ruinan Li*, 李瑞囡**

Shanghai University of International Business and Economics liruinan@suibe.edu.cn

Abstract

In this talk, we focus on the Freidlin-Wentzell's large deviation principle for one dimensional nonlinear stochastic heat equation driven by a Gaussian noise which is white in time and fractional in space with Hurst parameter $1/4 < \mathrm{H} < 1/2$. Weak convergence approach will be taken. To that end, we first prove the uniqueness of the solution of the skeleton equation. The large deviation principle follows by verifying the new sufficient condition proposed by Matoussi, Sabbagh, Zhang (2021) for the weak convergence criterion. This talk is based on a paper joint with Ran Wang and Beibei Zhang.

A nonstationary autologistic model for space-time data

***Yiping Hong,* 洪一平**

Beijing Institute of Technology yiping.hong@bit.edu.cn

Abstract

In many research fields such as meteorology, ecology, and epidemiology, the spatio-temporal datasets are binary, describing the existence of particular species or events. The spatial dependence of the binary observations often exhibits nonstationarity. However, most nonstationary models in the literature are based on Gaussian random fields, which may be computational demanding. We propose a nonstationary spatio-temporal autologistic regression model, which allows the spatial covariances to vary in space. We investigate the spatial and temporal correlation of autologistic models with different coding and centering settings and recommend the non-centered $\{-1,1\}$ coding model. We then develop the maximum pseudolikelihood method for parameter estimation, prediction, and variable selection. The simulation studies show the superior performance of the proposed methods compared to the commonly-used stationary model. We apply our model to analyze the fine particulate matter (PM2.5) concentrations from the Community Multiscale Air Quality (CMAQ) data for the continental US, thresholding at the health standard. Our analyses characterize the nonstationarity in spatial dependence over the continental US and predict the probability that the PM2.5 concentration is over the threshold.

主持人：李欣鹏，山东大学

A unified approach to global solvability for FBSDEs with diagonal generators

***Peng Luo,* 罗鹏**

Shanghai Jiao Tong University peng.luo@sjtu.edu.cn

Abstract

In this paper, we study the global solvability of multidimensional forward–backward stochastic differential equations (FBSDEs) with diagonally Lipschitz, quadratic or super-quadratic generators. Under a certain "monotonicity" condition, we provide a unified approach which shows that there exists a decoupling field that is uniformly Lipschitz in its spatial variable. This decoupling field is closely related to bounded solution to an associated characteristic BSDE. Applications to stochastic optimal controls and stochastic differential games are investigated.

Backward stochastic differential equations with conditional reflection and related recursive optimal control problems

***Wenqiang Li,* 李文强**

Yantai University wenqiangli@ytu.edu.cn

Abstract

In this talk, we introduce a new type of reflected backward stochastic differential equations (BSDEs) for which the reflection constraint is imposed on its main solution component, denoted as Y by convention, but in terms of its conditional expectation $E[Y_t|G_t]$ on a general sub-filtration G_t. We thus term such equation as conditionally reflected BSDE (for short, conditional RBSDE). Conditional RBSDE subsumes classical RBSDE with a pointwise reflection barrier, and the recent developed BSDE with a mean reflection constraint, as its two special and extreme cases: they exactly correspond to G_t being the full filtration to represent complete information, and the degenerated filtration to deterministic scenario, respectively. For conditional RBSDE, we obtain its existence and uniqueness under mild conditions by combining the Snell envelope method with Skorokhod lemma. We also discuss its connection, in the case of linear driver, to a class of optimal stopping problems in presence of partial information. As a by-product, a new version of comparison theorem is obtained. With the help of this connection, we study weak formulations of a class of optimal control problems with reflected recursive functionals by

characterizing the related optimal solution and value. Moreover, in the special case of recursive functionals being RBSDE with pointwise reflections, we study the strong formulations of related stochastic backward recursive control and zero-sum games, both in non-Markovian framework, that are of their own interests and have not been fully explored by existing literatures yet.

This talk is based on the joint work with Professors Ying Hu (Université de Rennes 1, France) and Jianhui Huang (The Hong Kong Polytechnic University, China).

A universal robust limit theorem for nonlinear Lévy processes under sublinear expectation

***Lianzi Jiang,* 姜恋姿**

Shandong University of Science and Technology jianglianzi95@163.com

Abstract

This talk is devoted to the study of a universal robust limit theorem under a sublinear expectation framework. Under moment and consistency conditions, we show that, for $\alpha \in (1,2)$, the i.i.d. sequence

$$\left[\left\{\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i, \frac{1}{n}\sum_{i=1}^{n} Y_i, \frac{1}{\sqrt[\alpha]{n}}\sum_{i=1}^{n} Z_i\right)\right\}_{n=1}^{\infty}\right]$$

converges in distribution to $\widetilde{L_1}$, $where\ \widetilde{L_t}, t \in [0,1]$, is a multidimensional nonlinear Lévy process. This nonlinear Lévy process is characterized by a fully nonlinear and possibly degenerate partial integro-differential equation (PIDE). To construct the limit process, we develop a weak convergence approach based on the notions of tightness and weak compactness on a sublinear expectation space. We further prove a new type of Lévy-Khintchine representation formula to characterize the nonlinear Lévy processes. As a byproduct, we also provide a probabilistic approach to prove the existence of the above fully nonlinear degenerate PIDE.

This talk is based on a recent joint work with Shige Peng, Mingshang Hu, and Gechun Liang.

主持人：陈一鸣，山东大学

Berry-Esseen bounds and precise deviations for a jump-type CIR process

***Zhi Qu,* 屈直**

Wuhan University zqu.math@whu.edu.cn

Abstract

In this paper, we consider a jump-type CIR process driven by a standard Wiener process and a subordinator. We derive some Berry-Esseen bounds and precise deviations for the sample mean of the CIR process and the MLE of the growth rate. This is a joint work with Fuqing Gao.

Tail bounds on the spectral norm of sub-exponential random matrices

***Guozheng Dai,* 戴国政**

Zhejiang University 11935022@zju.edu.cn

Abstract

Let X be an $n \times n$ symmetric random matrix with independent but non-identically distributed entries. The deviation inequalities of the spectral norm of X with Gaussian entries have been obtained by using the standard concentration of measure results in Gauss space which is unfortunately not suitable for the sub-Exponential case. This paper establishes an upper tail bound of $\|X\|$ with sub-Exponential entries. Our general method relies upon a crucial ingredient of a novel chaining argument that essentially depends on the distribution of coordinates of a point on the unit sphere. What makes this approach work is the particular structure of the sets used for the chaining.

Risk management

***Zhenfeng Zou,* 邹振烽**

University of Science and Technology of China newzzf@mail.ustc.edu.cn

Abstract

This paper studies the portfolio diversification of iid ultra heavy-tailed (i.e., infinite mean) Pareto losses. With the notion of majorization order, we show that a more diversified portfolio of iid ultra heavy-tailed Pareto losses is larger in the sense of the first-order stochastic dominance. This result is generalized for ultra heavy-tailed Pareto losses which are triggered by catastrophic events, and random losses of which the tails follow an ultra

heavy-tailed Pareto distribution. Finally, we discuss the implications of our results in optimal decisions for agents.

The generic chainning method in φ-Sub-Gaussian space

Yiming Chen, **陈一鸣**

Shandong University chenyiming960212@mail.sdu.edu.cn

Abstract

We calculated the tail bound of φ-Sub-Gaussian process by truncated generic chaining method, which is a novel method in φ-Gaussian space, the tail bound obtained in which space is obviously better than the classical Dudley entropy bound, we can consider more improvements of classical problems in this space with new method. As an application, we provided the bound of independent random variable, derivation of compressed sensing problem, JL lemma and concentration inequality.

Time-consistent reinsurance-investment games for multiple mean-variance insurers with mispricing and default risks

Yang Yang, **杨阳**

Soochow University yangy_1217@126.com

Abstract

This paper studies a non-zero-sum stochastic differential game for multiple mean-variance insurers. Insurers can purchase proportional reinsurance and invest in a financial market consisting of a risk-free asset, a market index, a defaultable bond and a pair of mispriced stocks. The dynamics of the mispriced stocks satisfy a "cointegrated system" where the expected returns follow the mean reverting processes, and the bond is defaultable with a recovering proportional value at default. Each insurer's objective is maximizing a function of her terminal wealth and competitors' relative wealth under the mean-variance criterion. Using techniques in stochastic control theory, we establish the extended Hamilton-Jacobi-Bellman equations and obtain the equilibrium strategies. Particularly, the derived solutions are analytical and time-consistent. We represent our results in terms of the M-matrices, which help us prove the existence and uniqueness of the solutions and further explicitly analyze how the crucial arguments in the model affect the equilibrium strategies. Numerical examples with detailed sensitivity analyses are presented to support our conclusions.

主持人：童培峰，山东大学

Reliability estimation of supply chain system with model uncertainty

Ziwen Gao, **高子文**

AMSS, Chinese Academy of Sciences ziwen@amss.ac.cn

Abstract

Estimating the reliability of supply chains provides support for their proper operation, and the more accurate the estimate, the better. Using the copula function to measure the dependence structure among suppliers, this paper proposes a model averaging method based on Kullback–Leibler (KL) loss to estimate the reliability of the supply chain with a k-out-of-n: G system. We prove the asymptotic optimality of the proposed estimator and the consistency of weights. Additional simulation studies and a real dataset demonstrate the proposed method's effectiveness.

Semi-supervised distribution estimation with applications

Mengtao Wen, **温梦涛**

Nankai University mtwen97@gmail.com

Abstract

Distribution estimation is one of the most general problems in statistical theory and machine learning. Traditional statistical theory guarantees that the empirical cumulative distribution function is an efficient estimator of the target distribution. However, in semi-supervised learning, with small-sized labeled data and large-sized unlabeled data, a natural question arises: how to improve the empirical cumulative distribution function of the response based on the labeled data with the help of unlabeled data? In this article, we develop a general semi-supervised method for distribution estimation of the response variable by using both labeled and unlabeled data. The proposed distribution estimator is root-n consistent in l_∞ norm and asymptotically normal under mild conditions. Furthermore, we explore the benefits of the proposed distribution estimator in quantile estimation and conformal prediction. A real data analysis on gene expression exhibits the superiority of the proposed method.

Multivariate calibration with supporting covariates: A new perspective on inverse regression and causality

Peifeng Tong, **童培峰**
Peking University tongpf@pku.edu.cn

Abstract

We consider a multivariate calibration problem in the presence of covariates. Conventional calibration analysis aims at inferring covariates $\boldsymbol{X}$ based on the observed response variable $\boldsymbol{Y}$. In this study, we extended the problem with the involvement of additional covariates $\boldsymbol{Z}$. We will discuss and compare the properties of the classical Generalized Least Square (GLS) estimator, inverse regression estimator and a newly proposed shrinkage estimator, with interpretations from the Bayesian and Frequentist perspectives. When we have multiple measurements in the controlled calibration condition, it is showed that the shrinkage estimator converges to the GLS estimator as the number of repeated measurements tends to infinity. We will also discuss how to deal with heterogeneity using a mixed random effects model. The shrinkage estimator had the best performance compared with the inverse regression and GLS estimators, and it outperformed the random forest both in simulation and case studies.

Multi-consensus decentralized primal-dual fixed point algorithm for distributed learning

Kejie Tang, **唐科杰**
Shanghai Jiao Tong University tangkj00@sjtu.edu.cn

Abstract

Decentralized distributed learning has recently attracted significant attention in many applications in machine learning and signal processing. To solve a decentralized optimization with regularization, we propose a Multi-consensus Decentralized Primal-Dual Fixed Point (MD-PDFP) algorithm. We apply multiple consensus steps with the gradient tracking technique to extend the primal-dual fixed point method over a network. The communication complexities of our procedure are given under certain conditions. Moreover, we show that our algorithm is consistent under general conditions and it enjoys global linear convergence under strong convexity. With some particular choices of regularizations, our algorithm can be applied to decentralized machine learning applications. Finally, several numerical experiments and real data analyses are conducted to demonstrate the effectiveness of the proposed algorithm.

Measuring and testing conditional dependence with high-dimensional conditioning variables

***Jianxin Bi*, 毕建鑫**

Xiamen University bjxstatistics@stu.xmu.edu.cn

Abstract

Measuring conditional dependence is a pivotal step of many statistical procedures, such as variable selection, network analysis and causal inference. Nevertheless, relevant works are still limited under the setting of high-dimensional conditioning variables, which is commonly encountered in the big data era. To address the high dimensionality, most existing literature impose certain model structures. However, the high-dimensional conditioning variables often inevitably bring about spurious correlation in these models. In this paper, we investigate estimation biases of commonly used conditional dependence measures at presence of spurious variables. We discuss the estimation inconsistency both intuitively and theoretically and show that such dependence can be either overestimated or underestimated under different scenarios. We then propose a new measure for high-dimensionally conditional dependence, based on data-splitting and refitting, that can eliminate the estimation biases and achieve consistency. A conditional independence test is also developed using the newly advocated measure, with a tuning-free asymptotic null distribution. Furthermore, the proposed test is applied to generating high dimensional network graphs in graphical modeling. The superior performances of newly proposed methods are illustrated both theoretically and through simulation studies. We also utilize the method to construct the gene-gene networks using a dataset of breast invasive carcinoma, which contain interesting discoveries that are worth further scientific exploration.

Correcting for nonignorable nonresponse bias in overestimation of turnout using callabck data

***Xinyu Li*, 李新宇**

Peking University xinyu.li@pku.edu.cn

Abstract

For decades, overestimation of turnout has been a classic problem in election surveys, and researchers have struggled with how to adjust for turnout bias. From 2000 to 2016, the gap between ANES self-reported and official turnouts in U.S. presidential elections has never been less than 15 percentage points, and other public opinion surveys also suffer from sizable overestimation bias. Nonresponse bias is regarded as one of the main sources of turnout bias. Adjusting for nonignorable nonresponse bias is substantially challenging, since identification generally fails to hold in the absence of additional information. Recently, more and more researchers have recognized the potential of callback data in adjusting for nonresponse bias. We aim to offer in-depth insight into whether and how callbacks in follow-up survey studies can help lessen nonignorable nonresponse bias of voter turnout. We characterize a stableness of resistance assumption, and establish the nonparametric identification under this assumption. The stableness of resistance assumption states that the impact of the missing outcome on the response propensity is stable across the first two call attempts, which does not impose parametric functional restrictions on propensity scores. Semiparametric efficiency theory under this assumption is established, which comprises the characterization of the tangent space, efficient influence function, and semiparametric efficiency bound for estimating a general full-data functional. Then a variety of semiparametric methods are proposed for estimation and inference, including a doubly robust and locally efficient one. We apply these semiparametric estimation approaches to compensate for nonresponse bias in a real ANES mailed post-election survey concerning the 2020 U.S. presidential election. Our proposed method successfully captures the trend of declining willingness to vote as response reluctance or contact difficulty increases, and gives a closer estimate to the true value.

On Catoni's M-estimation

***Pengtao Li*, 李澎涛**
Shandong University pengtaoli@mail.sdu.edu.cn

Abstract

Catoni proposed a robust M-estimator and gave the deviation inequality for one test function. In this talk, we focus on the uniform concentration inequality for a family of test functions. As an application, we consider empirical risk minimization with heavy-tailed losses.

Functional data analysis with covariate-dependent mean and covariance structures

***Chenlin Zhang*, 张晨琳**
Southwestern University of Finance and Economics 928147754@qq.com

Abstract

Functional data analysis has emerged as a powerful tool in response to the ever-increasing resources and efforts devoted to collecting information about response curves or anything that varies over a continuum. However, limited progress has been made with regard to linking the covariance structures of response curves to external covariates, as most functional models assume a common covariance structure. We propose a new functional regression model with covariate-dependent mean and covariance structures. Particularly, by allowing variances of random scores to be covariate-dependent, we identify eigenfunctions for each individual from the set of eigenfunctions that govern the variation patterns across all individuals, resulting in high interpretability and prediction power. We further propose a new penalized quasi-likelihood procedure that combines regularization and B-spline smoothing for model selection and estimation and establish the convergence rate and asymptotic normality of the proposed estimators. The utility of the developed method is demonstrated via simulations, as well as an analysis of the Avon Longitudinal Study of Parents and Children concerning parental effects on the growth curves of their offspring, which yields biologically interesting results.

Distributed estimation and inference for spatial autoregression model with large scale networks

***Yimeng Ren*, 任怡萌**
Fudan University ymren22@m.fudan.edu.cn

Abstract

The rapid growth of online network platforms generates large-scale network data and it poses great challenges for statistical analysis using the spatial autoregression (SAR) model. In this work, we develop a novel distributed estimation and statistical inference framework for the SAR model on a distributed system. We first propose a distributed network least squares approximation (DNLSA) method. This enables us to obtain a one-step estimator by taking a weighted average of local estimators on each worker. Afterwards, a refined two-step estimation is designed to further reduce the estimation bias. For statistical inference, we utilize a random projection method to reduce the expensive communication cost. Theoretically, we show the consistency and asymptotic normality of both the one-step and two-step estimators. In addition, we provide theoretical guarantee of the distributed statistical inference procedure. The theoretical findings and computational advantages are validated by several numerical simulations implemented on the Spark system. Lastly, an experiment on the Yelp dataset further illustrates the usefulness of the proposed methodology.

Feature selection in ultrahigh-dimensional additive models with heterogeneous frequency component functions

***Yuyang Liu*, 刘宇扬**
Shanghai Jiaotong University d0408x@sjtu.edu.cn

Abstract

In this paper, we consider feature selection in ultrahigh-dimensional additive models with heterogeneous frequency component functions. Firstly, we introduce a new concept, weighted sum of squared conditional correlations (WSSCC), which measures the correlation between a random variable and its function. Afterwards, we propose a sure independence screening procedure based on WSSCC (WSSCCSIS), whose sure screening property is established. Furthermore, a sequential feature selection procedure based on WSSCC (WSSCCFR) is proposed. Numerical studies including comprehensive simulation and a real data analysis are carried out to demonstrate the advantage of our method over other existing approaches.

主持人：赵林杰，华中科技大学

Hydrodynamic limits for facilitated exclusion processes

***Linjie Zhao*，赵林杰**

Huazhong University of Science and Technology linjie_zhao@hust.edu.cn

Abstract

The theory of hydrodynamic limits is concerned with the macroscopic properties of interacting systems, which are usually described by PDEs. In this talk, we discuss hydrodynamic limits for a degenerate interacting particle system, i.e., the facilitated exclusion process. The hydrodynamic equation turns out to satisfy the Stefan problems. If time permitted, we shall also discuss stationary fluctuations for the facilitated exclusion process. Joint work with Clement Erignoux and Marielle Simon.

Asymptotics for the conditional higher moment coherent risk measure with weak contagion

***Qingxin Yi*，易清心**

Xi'an Jiaotong-Liverpool University Qingxin.Yi18@student.xjtlu.edu.cn

Abstract

Various measures have been proposed in the existing literature to evaluate the extreme risk exposure of a risk under the effect of an "observable" factor. Due to the nice properties of the higher-moment (HM) coherent risk measures, we propose a conditional version of the HM risk measure (CoHM) by taking into account the information of an "observable" factor. We conduct the asymptotics for this measure of extreme risks at high confidence levels under the weak contagion of risk, which are further applied to the special case of Conditional Haezendonck-Goovaerts risk measure (CoHG). Numerical illustrations are also provided to examine the accuracy of the asymptotic formulas and to analyze the sensitivity of the risk contribution based on the Haezendonck-Goovaerts risk measure.

Multivariate doubly truncated moments for generalized skew-elliptical distributions with applications

***Baishuai Zuo,* 左百帅**

Qufu Normal University bszuo124@163.com

Abstract

In this paper, we focus on multivariate doubly truncated first two moments of generalized skew-elliptical distributions. This class of distributions includes many useful distributions, such as skew-normal, skew Student-t, skew-logistic and skew-Laplace-normal distributions, as special cases. The formulas of multivariate doubly truncated covariance (MDTCov) for generalized skew-elliptical distributions is also given.Further, we compute multivariate doubly truncated expectation (MDTE) and MDTCov for 2-variate skew-normal distribution, and use Monte Carlo method to simulate and compare with the above results. As applications, the results of multivariate tail conditional expectation (MTCE) and multivariate tail covariance (MTCov) for generalized skew-elliptical distributions are derived. In addition, an optimal problem involving MDTE and MDTCov risk measures is proposed. Finally, we use real data to fit normal and skew-normal distributions, the best distribution is selected using the Akaike information criterion (AIC) and Bayesian information criterion (BIC) methods, respectively, and discuss MTCEs and MTCovs of logarithm of opening prices for two portfolios consisting of 3 companies from S&P (Standard & Poor's) sectors.

How does node centrality in a complex network affect prediction?

***Xinyao Zhao,* 赵欣瑶**

Soochow University xyzhao1996@stu.suda.edu.cn

Abstract

In complex financial networks, systemically important nodes usually play vital roles. We consider networks consisting of major global assets and explore how node centrality affects price forecasting by applying a hybrid random forest algorithm. We find two counterintuitive phenomena: (i) factors with low centrality have better forecasting ability; (ii) nodes with low centrality can be predicted more accurately in direction. Using the notion of entropy, which measures the quantity of information, we show that factors with low centrality have more useful information and less noise for the forecast asset price than those with high centrality do. In addition, while predicting a systemically unimportant node, we demonstrate that the other nodes within the network have a higher information rate to it. Finally, we verify the robustness of our results using an alternative deep learning method.

主持人：苏东，南京大学

The generalization ability of logistic regression with Markov sampling

Zhiyong Qian, **钱智勇**

Zhongnan University of Economics and Law 202111190165@stu.zuel.edu.cn

Abstract

In this paper, we propose a new algorithm for Logistic regression model based on uniformly ergodic Markov samples and study the generalization ability, weak learning rate and convergence of the algorithm. It is found that the performance of the new algorithm is more effective than the Logistic regression model based on random samples, and its performance is better than the classical machine model algorithms such as random forest and Adaboost.

A Stackelberg-Nash equilibria with investment and reinsurance in mixed leadership games

Qingqing Zhang, **张晴晴**

Nanjing Normal University

Abstract

In this paper, we investigate the optimal reinsurance and investment problem from joint interests of the insurer and the reinsurer under the framework of the mixed leadership game, which also termed the Stochastic Stackelberg-Nash game. More specifically, the reinsurer is the leader to decide on an optimal reinsurance premium to charge, while the insurer acts as the leader to determine the amount he/she invests into the risky assets to hedge the liabilities retained. Besides, the insurer acts as a follower for the retention level, and the reinsurer is the follower to determine the amount he/she invests into the risky assets. A correlation between insurer's liabilities and the risky assets is introduced. Both the insurer and the reinsurer aim to maximize the expected utility on the terminal wealth and the explicit optimal strategies are derived by solving the Hamilton-Jacobi-Bellman(HJB) dynamic programming. Compare to the traditional Stackelberg differential game, we find that both the insurer and the reinsurer make better strategies under the mixed leadership game. Numerical examples are provided to analyze the economic intuition and insights.

Minimal joint entropy and order-preserving couplings

***Yajing Ma,* 马雅静**

Capital Normal University mayajing121@126.com

Abstract

This paper focuses on the extreme-value problem for Shannon entropy of the joint distribution with given marginals. It is proved that the minimum-entropy coupling must be of order-preserving, while the maximum-entropy coupling coincides with the independent one. Note that in this sense, we interpret entropy as a measure of system disorder.

Approximation for a generalized langevin equation with high oscillation in time and space

***Dong Su,* 苏东**

Nanjing University 1961548346@qq.com

Abstract

This paper derives an approximation for a generalized Langevin equation driven by a force with random oscillation in time and periodic oscillation in space. By a diffusion approximation and the weak convergence of periodic oscillation function, the solution of the generalized Langevin equation is shown to converge in distribution to the solution of a stochastic partial differential equations (SPDEs) driven by time white noise.

Self-supervised distributional and contrastive learning model for image anomaly detection

Yannan Pu, **蒲雁南**

Yunnan University puyannan123@mail.ynu.edu.cn

Abstract

Anomaly detection aims to detect anomalies that are away from the distribution of normal data. It has been widely investigated in computer vision and machine learning. In this paper, we propose a self-supervised distributional and contrastive learning model for anomaly detection. Specifically, in the feature space, based on the augmented dataset with given rotated images, we propose to model latent features of normal sample set under each rotation transformation via a Gaussian Mixture Model (GMM). By the proposed GMM clustering loss, we maximize the data probability to the GMM model with rotation that applied to the data, and minimize its probability to the GMMs of other rotations. We also discriminate augmented features via rotation-based contrastive learning. Our approach jointly conducts distributional modeling of image features considering feature clustering properties, and discriminative learning of features in a self-supervised way. Extensive experiments on image datasets demonstrate that our method achieves favorable performance compared with several state-of-the-art methods on unsupervised image anomaly detection tasks.

Multivariate threshold integer-valued autoregressive processes with explanatory variables

Nuo Xu, **许诺**

Changchun University of Technology

Abstract

To capture the multivariate count time series showing piecewise phenomena, we introduce a class of first-order multivariate threshold integer-valued autoregressive process. The component of autoregressive coefficient is driven by explanatory variables via a logistic link function. Basic probabilistic and statistical properties of the model are discussed. The model parameters are estimated by means of conditional least squares and conditional maximum likelihood methods. A new algorithm to estimate the threshold parameter of the model is also provided. Moreover, the nonlinearity test of the model and existence test of explanatory variables are well addressed using a wald-type test. Finally, some numerical results of the estimates and a real data example are presented.

On estimating the covariance matrix for large-dimensional matrix factor model

***Zhixiang Zhou,* 周智翔**

Nanjing Audit University xiaohufeixiang@hotmail.com

Abstract

The projection procedure proposed in Yu et al. (2021) is popular in the inference of the factor spaces for large-dimensional matrix factor models. However, via the projection approach, estimating the covariance matrix for large-dimensional matrix factor models with matrix time series is not yet considered. In this paper, we introduce a factor-projection-based covariance (precision) matrix estimators for matrix time series. The convergence rates of the idiosyncratic covariance matrix, total covariance matrix as well as its inverse are obtained under the relative norm and the spectral norm. Our estimators converge faster than that of the vectorized POET estimators. Extensive numerical studies justify the empirical performance of the estimators. An application into the linear discriminant analysis (LDA) demonstrates the superiority of our estimators.

Change-point models with applications to feedforward neural networks

***Houlin Zhou,* 周厚林**

Anhui University 474455137@qq.com

Abstract

Compared with traditional parametric or nonparametric models, the change-point models can describe data changes more flexibly. While, the sophistication of these models leads to the complexity of the estimator expression. In order to avoid overly complicated form of estimator, this paper considers to build a neural network framework and fit parameters of the estimator by data. It is not easy to prove the theoretical properties directly. To obtain the limiting properties of the estimator generated by feedforward neural network (FNN, for short), we need to rely on a widely studied cumulative sum (CUSUM, for short) estimator by establishing the approximate relationship between the two estimators.

In the case of single change point, we derive the complete f-moment consistency for the CUSUM estimator. By the universal approximation theorem, similar theoretical property of the estimator generated by FNN is obtained. In the simulation experiment, we find that FNN estimator has less influence on signal strength and sample size compared with CUSUM estimator. Therefore, in the case of small sample size or weak data change, FNN estimator has obvious advantage. Finally, we give a case of stock data for analysis.

主持人：梁春辉，东北师范大学

Adaptive block banding precision matrix estimation for multivariate longitudinal data

***Chunhui Liang,* 梁春辉**

Northeast Normal University liangch801@nenu.edu.cn

Abstract

We propose an estimator for precision matrices with the structure of Banded Kronecker Sparse forms (BKS). BKS takes advantage of the special feature of a precision matrix, which has the form of the kronecker product of an adaptively banded matrix and a sparse matrix, both are positive definite. Such precision matrix arises frequently in practice in finance data, medical data and time series data. We achieve the adaptive bandedness via a specially designed penalty, and enforce the sparsity via lasso. We design a computationally efficient procedure named Alternative Convex Search (ACS) algorithm to implement BKS. We establish the computational convergence and show the statistical guarantee through establishing the asymptotic rate. Our extensive simulation studies indicate the superior finite sample performance of BKS in comparison to existing methods. We apply BKS to an EEG data to improve the performance of the discriminant analysis result.

On the best approximation of finite Gaussian mixtures

***Yun Ma,* 马沄**

Tsinghua University mayun21@mails.tsinghua.edu.cn

Abstract

We consider the problem of approximating a general Gaussian location mixture by finite mixtures. The minimum order of finite mixtures that achieve a prescribed accuracy (measured by various f-divergences) are determined within constant factors for families of mixing distributions with constraints on support, tail probability, or moment conditions. While the upper bound is achieved using the technique of moment matching, the lower bound is established by relating the best approximation error to the low-rank approximation of certain trigonometric moment matrices, followed by a refined spectral analysis of the minimum eigenvalue. In the case of Gaussian mixing distributions, this result corrects a previous lower bound in [Wu and Verdú, 2010].

Opposite online learning via sequentially integrated stochastic gradient descent estimators

***Xiaoting Ji*, 吉晓婷**
Shandong University jixiaoting@mail.sdu.edu.cn

Abstract

The stochastic gradient descent algorithm, often depicted as SGD, has been widely employed in various fields of artificial intelligence and is a prototype of online learning algorithms. In the article, we propose a novel and general framework of one-sided testing for streaming data based on SGD. The proposed method constructs an online-updated test statistic sequentially by integrating the selected batch-specific estimators or its opposite, which is referred to as opposite online learning. Notably, the batch-specific online estimators are chosen strategically according to the proposed sequential tactics designed by the two-armed bandit process. Theoretical results prove the strategy's advantage, ensuring that the test statistic distribution is optimal under the null hypothesis. We also supply the theoretical evidence of power enhancement compared with classical test statistics. In application, the proposed method is appealing for statistical inference of two-sided testing and it is scalable and adaptable for any model. Finally, the superior finite-sample performance is evaluated by simulation studies.

Consistent selection of the number of groups in panel models via sample-splitting

***Zhe Li*, 李哲**
Fudan University zheli20@fudan.edu.cn

Abstract

Group number selection is a key question for group panel data modelling. In this work, we develop a cross validation method to tackle this problem. Specifically, we split the panel data into a training dataset and a testing dataset on the time span. We first use the training dataset to estimate the parameters and group memberships. Then we apply the fitted model to the testing dataset and then the group number is estimated by minimizing certain loss function values on the testing dataset. We design the loss functions for panel data models either with or without fixed effects. The proposed method has two advantages. First, the method is totally data-driven thus no further tuning parameters are involved. Second, the method can be flexibly applied to a wide range of panel data models. Theoretically, we establish the estimation consistency by taking advantage of the optimization property of the estimation algorithm. Experiments on a variety of synthetic and empirical datasets are carried out to further illustrate the advantages of the proposed method.

主持人：王惠远，北京大学

Heterogeneous federated learning

Huiyuan Wang, **王惠远**

Peking University huiyuan.wang@pku.edu.cn

Abstract

Federated learning, where algorithms are trained across multiple decentralized devices without sharing local data, is increasingly popular in distributed machine learning practice. Typically, a graph structure G exists behind local devices for communication. In this work, we consider parameter estimation in federated learning with heterogeneity in communication and data distribution, coupled with a limited computational capacity of local devices. We encode the distribution heterogeneity by parametrizing distributions on local devices with a set of distinct p-dimensional vectors. We then propose to jointly estimate parameters of all devices under the M-estimation framework with the fused Lasso regularization, encouraging an equal estimate of parameters on connected devices in G. We provide a general statistical guarantee for our estimator, which can be further calibrated to obtain convergence rates for various specific problem setups. Surprisingly, our estimator attains the optimal rate under certain graph fidelity condition on G, as if we could aggregate all samples sharing the same distribution. If the graph fidelity condition is not met, we propose an edge selection procedure via multiple testing to ensure the optimality. To ease the burden of local computation, a decentralized stochastic version of ADMM is provided, with convergence rate $O(T^{\{-1\}}\log T)$ where T denotes the number of iterations. We highlight that our algorithm transmits only parameters along edges of G at each iteration, without requiring a central machine, which preserves privacy. To address communication heterogeneity, we further extend it to the case where devices are randomly inaccessible during the training process, with a similar algorithmic convergence guarantee. The computational and statistical efficiency of our method is evidenced by simulation experiments and the 2020 US presidential election data set.

Statistical inference for four-regime segmented regression models

Han Yan, **闫晗**

Peking University hanyan@stu.pku.edu.cn

Abstract

Segmented regression models are attractive for their flexibility and interpretability as compared to the global parametric and the nonparametric models, and yet are challenging in both estimation and inference. We consider a four-regime segmented model for

temporally dependent data with two segmenting boundaries depending on multivariate covariates with nondiminishing boundary effects. A mixed integer quadratic programming algorithm is formulated to facilitate the least square estimation to both the regression and the boundary coefficients. The rates of convergence and the asymptotic distributions of the least square estimators are obtained, which shows differential convergence rates and limiting distributions between the regression and the boundary coefficients. Estimation and testing for degenerated segmented models with less than four segments are also considered with a testing procedure to decide if neighboring segments can be merged. Numerical simulations and a case study on air pollution in Beijing are conducted to demonstrate the proposed model and the inference results. In particular, it shows that the segmented models with three or four regimes are suitable for the modeling of the meteorological effects on the PM2.5 concentration.

Physical Parameter Calibration

***Yang Li,* 李杨**

AMSS, Chinese Academy of Sciences 1051337458@qq.com

Abstract

Computer simulation models are widely used to study complex physical systems. A related fundamental topic is the inverse problem, also called calibration, which aims at learning about the values of parameters in the model based on observations. In most real applications, the parameters have specific physical meanings, and we call them physical parameters. To recognize the true underlying physical system, we need to effectively estimate such parameters. However, existing calibration methods cannot do this well due to the model identifiability problem. This paper proposes a semi-parametric model, called the discrepancy decomposition model, to describe the discrepancy between the physical system and the computer model. The proposed model possesses a clear interpretation, and more importantly, it is identifiable under mild conditions. Under this model, we present estimators of the physical parameters and the discrepancy, and then establish their asymptotic properties. Numerical examples show that the proposed method can better estimate the physical parameters than existing methods.

Score function-based tests for ultrahigh-dimensional linear models

***Weichao Yang,* 杨伟超**

Beijing Normal University yangweichao@mail.bnu.edu.cn

Abstract

To sufficiently exploit the model structure under the null hypothesis such that the conditions on the whole model can be mild, this paper investigates score function-based tests to check the significance of an ultrahigh-dimensional sub-vector of the model coefficients when the nuisance parameter vector is also ultrahigh-dimensional in linear models. We first reanalyze and extend a recently proposed score function-based test to derive, under weaker conditions, its limiting distributions under the null and local alternative hypotheses. As it may fail to work when the correlation between testing covariates and nuisance covariates is high, we propose an orthogonalized score function-based test with two merits: debiasing to make the non-degenerate error term degenerate and reducing the asymptotic variance to enhance the power performance. Simulations evaluate the finite-sample performances of the proposed tests, and a real data analysis illustrates its application.