



全国工业统计学教学研究会青年统计学家协会 2024 年年会
暨第二届中国青年统计学家论坛

会议手册

主办单位：全国工业统计学教学研究会青年统计学家协会

承办单位：江苏师范大学

会议资助：江苏高校统计学优势学科

国家自然科学基金重大项目

中国·徐州

2024 年 4 月

一、参会须知

尊敬的各位参会代表，热烈欢迎并衷心感谢您的莅临！为了保证会议的顺利召开，同时也为了方便您在会议期间的工作与生活，敬请留意以下事项：

1. 报到

- 报到时间：4月19日10:00-20:00，4月20日8:30-10:00。如您错过报到时间，请及时联系会务组。
- 报到地点：徐州喜来登酒店一楼大厅（江苏省徐州市铜山区大学路35号）

2. 食宿

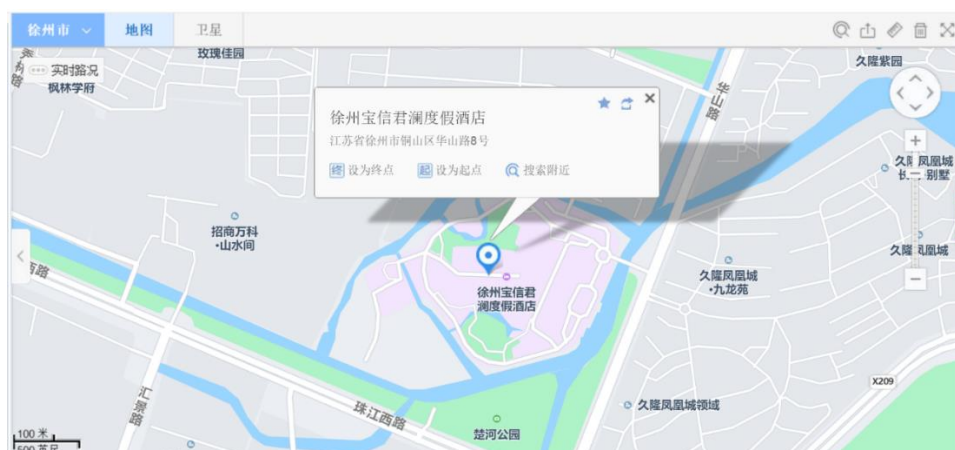
- 本次会议期间将提供江苏师范大学附近不同条件的酒店供与会者选择，请与会者自行联系酒店住宿：

- ✓ 徐州喜来登酒店（**主会场**，江苏省徐州市铜山区大学路35号）



订房时报会议名称联系刘经理，电话：13912046964（会议协议价）

- ✓ 徐州宝信君澜度假酒店（江苏省徐州市铜山区华山路8号）距主会场喜来登酒店2.6km



订房时报会议名称联系赵经理，电话：13814424775（会议协议价）酒店电话：0516-66999999

- ✓ 徐州云龙湖矿业大学亚朵酒店（江苏省徐州市铜山区大学路 202 号矿业大学东门对面）距离主会场 1.4km



订房联系人：崔经理，电话：15851409526，总机：0516-85068889

- ✓ 全季酒店（江苏省徐州市铜山区大学路 299 号怡和广场 A 座）距离主会场 2.2km



订房联系人：邱经理，电话：19951500514，总机：0516-85055988

- 报到时，请您随工作人员至酒店前台办理住宿手续，请带好身份证或护照等有效证件，住宿发票退房时请自行在前台开具。
- 本次会议食宿自理，请您按照会议日程自行安排用餐时间。

- ✓ 果紫民宿是一家遍布徐州全城城市民宿，入住寻常百姓家，体验彭城烟火气！
该民宿提供房型均为在喜来登酒店附近的三室户型，每室可多住 2 人，每户均可入住 6 人。



联系人： 小果 联系电话（微信同号）：18552806997 （江苏师大会议协议价）

- 请您提前添加小果微信在线办理住宿手续，请带好身份证或护照等有效证件，住宿发票退房时请自行在前台或在线开具。房间详情请电话在线咨询。
- 本次会议食宿自理，请您按照会议日程自行安排用餐时间。

3. 联络信息

在您参会期间，如有任何问题，请联络会务组人员：

丁维勇，18914857024

刘鹏飞，18361291761

赵琬迪，15011000390

二、组织机构

主办单位：全国工业统计学教学研究会青年统计学家协会

承办单位：江苏师范大学

会议协办：狗熊会

科学委员会：

邓柯（清华大学）

冯兴东（上海财经大学）

马秀颖（吉林财经大学）

王学钦（中国科学技术大学）

夏寅（复旦大学）

姚方（北京大学）

郑术蓉（东北师范大学）

组织委员会：

主席：赵鹏（江苏师范大学）、王汉生（北京大学）

秘书长：潘蕊（中央财经大学）

副秘书长：赵琬迪（首都经济贸易大学）

成员：

丁维勇（江苏师范大学）

李丰（中央财经大学）

李建波（江苏师范大学）

李月玲（江苏师范大学）

刘鹏飞（江苏师范大学）

王菲菲（中国人民大学）

钟威（厦门大学）

周勤（江苏师范大学）

三、日程简表

日期	时间	日程安排	地点
4月19日	全天	会议报到	
4月20日 上午	9:00-9:30	大会开幕式	淮海宴会厅
	9:30-12:00	大会主题报告	
	12:00-13:30	午餐	
4月20日 下午	13:30-15:10	分会场一：因果学习、深度学习与强化学习的最新进展	彭城宴会厅 A
		分会场二：机器学习与统计推断	彭城宴会厅 B
		分会场三：大数据计算与应用	淮海宴会厅 A
		分会场四：大规模网络及时序数据建模与计算	淮海宴会厅 B
		分会场五：生物信息与生物学大数据	淮海宴会厅 C
		分会场六：矩阵数据统计推断	会议室 1
		分会场七：高维统计推断	会议室 2
		分会场八：案例教学专场	会议室 3
	15:10-15:25	间歇	
	15:25-17:05	分会场一：深度学习的理论与应用	彭城宴会厅 A
		分会场二：机器学习前沿方法	彭城宴会厅 B
		分会场三：统计学习的最新进展	淮海宴会厅 A
		分会场四：大规模网络建模和统计推断	淮海宴会厅 B
		分会场五：复杂生物学数据的统计推断	淮海宴会厅 C
		分会场六：高维时间序列	会议室 1
分会场七：无模型假设的统计推断		会议室 2	
13:30-17:00	招聘专场&博士后论坛	淮海前厅	
4月21日	全天	离会	

四、会议日程

大会主题报告（4月20日 上午）

地点：淮海宴会厅

时间	议程内容		
大会开幕仪式			主持人
9:00-9:30	江苏师范大学校领导致欢迎辞 陈木法 中国科学院院士、发展中国家科学院院士、 江苏师范大学数学研究院院长 陈松蹊 中国科学院院士、北京大学讲席教授 陈 敏 全国工业统计学教学研究会会长、中国科学院数学与系统科学 学研究院研究员		王汉生
时间	报告人	报告题目	主持人
9:30-10:20	陈松蹊（北京大学）	数据同化：机理与观测相结合的 数据融合方法	孙六全
10:20-11:10	朱雪宁（复旦大学）	Distributed Estimation and Inference for Spatial Autoregression Model with Large Scale Networks	冯兴东
11:10-12:00	赵鹏（江苏师范大学）	网络风险统计建模及保 险应用	於州

分会场报告（一）（4月20日下午，彭城宴会厅A）

分会场一（上半场）：因果学习、深度学习与强化学习的最新进展			
时间	报告人	演讲题目	会议主席
召集人：於州（华东师范大学）			
13:30-13:55	崔逸凡（浙江大学）	Marginal structural Cox models for time-varying endogenous treatment	於州（华东师范大学）
13:55-14:20	刘林（上海交通大学）	On asymptotic model selection criterion in causal inference	
14:20-14:45	邱怡轩（上海财经大学）	ReHLine: Regularized Composite ReLU-ReHU Loss Minimization with Linear Computation and Linear Convergence	
14:45-15:10	周帆（上海财经大学）	Recent advances in Distributional Reinforcement Learning	
间歇（15:10-15:25）			
分会场一（下半场）：深度学习的理论与应用			
召集人：焦雨领（武汉大学）			
15:25-15:50	黄维然（上海交通大学）	Theoretical Insights into Self-Supervised Contrastive Learning	焦雨领（武汉大学）
15:50-16:15	刘勇（中国人民大学）	In-context Learning隐式更新机理研究	
16:15-16:40	毛晓军（上海交通大学）	Decentralized Reduced Rank Regression for Response Partition	
16:40-17:05	谢琦（西安交通大学）	参数化卷积方法及其应用	

分会场报告（二）（4月20日下午，彭城宴会厅B）

分会场二（上半场）：机器学习与统计推断			
时间	报告人	演讲题目	会议主席
召集人：刘旭（上海财经大学）			
13:30-13:55	申国豪（香港理工大学）	Complexity of Feed-Forward Neural Networks from the Perspective of Functional Equivalence	朱利平 （中国人民大学）
13:55-14:20	韩东啸（南开大学）	Inference for High Dimensional Proportional Hazards Model with Streaming Survival Data	
14:20-14:45	宋珊珊（香港中文大学）	A general M-estimation theory in semi-supervised framework	
14:45-15:10	张晓（香港中文大学深圳分校）	Hypothesis testing in high-dimensional censored transformation models	
间歇（15:10-15:25）			
分会场二（下半场）：机器学习前沿方法			
召集人：王菲菲（中国人民大学）			
15:25-15:50	杨朋昆（清华大学）	A Statistical Perspective of Federated Learning Algorithms and Model Personalization	常象宇 （西安交通大学）
15:50-16:15	张琼（中国人民大学）	Federated Learning with Mixed-Type Labels	
16:15-16:40	亓颢博（北京师范大学）	Mini-batch Gradient Descent with Buffer	
16:40-17:05	朱映秋（对外经济贸易大学）	A communication efficient boosting method for distributed spectral clustering	

分会场报告（三）（4月20日下午，淮海宴会厅A）

分会场三（上半场）：大数据计算与应用			
时间	报告人	演讲题目	会议主席
召集人：李丰（中央财经大学）			
13:30-13:55	熊昊一（百度研究院大数据实验室负责人、主任架构师）	基于大模型的数据工程与数据科学解决方案	张新雨（中国科学院数学与系统科学研究院）
13:55-14:20	汪澎洋（澳门大学智慧城市物联网国家重点实验室）	时空图数据表示学习	
14:20-14:45	孟祥溪（北京大学肿瘤医院）	用数据科学搭建医学图像到临床决策的桥梁	
14:45-15:10	杨亚磊（北京大学肿瘤医院）	贝叶斯网络在医学影像中的应用：参数估计与分割的不确定性研究	
间歇（15:10-15:25）			
分会场三（下半场）：统计学习的最新进展			
召集人：潘蕊（中央财经大学）			
15:25-15:50	李挺（香港理工大学）	Conditional Stochastic Interpolation for Generative Learning	郑泽敏（中国科学技术大学）
15:50-16:15	师佳鑫（北京大学）	Mixture Conditional Regression with Ultrahigh Dimensional Text Data for Estimating Extralegal Factor Effects	
16:15-16:40	吕绍高（南京审计大学）	基于贝叶斯表示的个性化联邦学习	
16:40-17:05	高原（北京大学）	An Asymptotic Analysis of Random Partition Based Minibatch Momentum Methods for Linear Regression Models	

分会场报告（四）（4月20日下午，淮海宴会厅B）

分会场四（上半场）：大规模网络及时序数据建模与计算			
时间	报告人	演讲题目	会议主席
召集人：朱雪宁（复旦大学）			
13:30-13:55	伍书缘（上海财经大学）	Quasi-Newton Updating for Large-Scale Distributed Learning	朱雪宁（复旦大学）
13:55-14:20	蒋斐宇（复旦大学）	Testing serial independence of object-valued time series	
14:20-14:45	许杏柏（厦门大学）	TranSAR: a spatial transfer learning framework	
14:45-15:10	蒋滨雁（香港理工大学）	A two-way heterogeneity model for dynamic networks	
间歇（15:10-15:25）			
分会场四（下半场）：大规模网络建模和统计推断			
召集人：兰伟（西南财经大学）			
15:25-15:50	冯龙（南开大学）	Computationally efficient and data-adaptive change-point inference in high dimension	兰伟（西南财经大学）
15:50-16:15	刘耀午（西南财经大学）	Ensemble methods for testing a global null	
16:15-16:40	刘秉辉（东北师范大学）	Joint community detection in random effects stochastic block models via the split-likelihood method	
16:40-17:05	范新妍（中国人民大学）	Projected semi-parametric network analysis	

分会场报告（五）（4月20日下午，淮海宴会厅C）

分会场五（上半场）：生物信息与生物医学大数据			
时间	报告人	演讲题目	会议主席
召集人：席瑞斌（北京大学）			
13:30-13:55	胡刚（南开大学）	基于深度学习的空间转录组数据增强方法	席瑞斌（北京大学）
13:55-14:20	侯琳（清华大学）	Inference of heterogeneous effect in Perturb-Seg experiments	
14:20-14:45	明静思（华东师范大学）	Automated and fast annotation of cell types with hierarchical structure for single-cell RNA-seq data	
14:45-15:10	金锁钦（武汉大学）	Systematic analysis of cell-cell communication from spatial transcriptomics data	
间歇（15:10-15:25）			
分会场五（下半场）：复杂生物医学数据的统计推断			
召集人：周静（中国人民大学）			
15:25-15:50	卜德亮（首都经济贸易大学）	Summary statistics-based association test for identifying the pleiotropic effects with set of genetic variants	李启寨（中国科学院数学与系统科学研究院）
15:50-16:15	梅好（中国人民大学）	Clinical Human Disease Networks with Healthcare Administrative Claims Data	
16:15-16:40	高慕鸿（中科院数学与系统科学研究院）	Learning network-structured dependence from multivariate point process data	
16:40-17:05	李晶茂（厦门大学）	Hierarchical Multi-Label Classification with Gene-Environment Interactions in Disease Modeling	

分会场报告（六）（4月20日下午，会议室1）

分会场六（上半场）：矩阵数据统计推断			
时间	报告人	演讲题目	会议主席
召集人：孔新兵（南京审计大学）			
13:30-13:55	何勇（山东大学）	Robust Statistical Inference for Large-dimensional Matrix-valued Time Series via Iterative Huber Regression	孔新兵（南京审计大学）
13:55-14:20	虞龙（上海财经大学）	Extreme eigenvalues of sample covariance matrices under generalized elliptical models with applications	
14:20-14:45	吴彬（中国科学技术大学）	Statistical Inferences for Staleness Factor Models and Robust Volatility Estimation with Large Panel High-Frequency Data	
14:45-15:10	蔡雄（南京审计大学）	Matrix-factor-augmented regression	
间歇（15:10-15:25）			
分会场六（下半场）：高维时间序列			
召集人：涂云东（北京大学）			
15:25-15:50	马辰辰（北京大学）	A Tale of Two Types of Structural Instabilities in High Dimensional Factor Models	何勇（山东大学）
15:50-16:15	吴潇然（浙江大学）	Sparse factor model for high dimensional time series	
16:15-16:40	罗祥夫（厦门大学）	Adaptive shrinkage estimation for high dimensional change point detection	
16:40-17:05	高尚（复旦大学）	Conditional Factor Models with High-dimensional Instrument Characteristics: Estimation and Variable Selection	

分会场报告（七）（4月20日下午，会议室2）

分会场七（上半场）：高维统计推断			
时间	报告人	演讲题目	会议主席
召集人：郭旭（北京师范大学）			
13:30-13:55	许凯（安徽师范大学）	Testing the Effects of High-Dimensional Covariates via Aggregating Cumulative Covariances	谭发龙（湖南大学）
13:55-14:20	杨松山（中国人民大学）	New Tests for High-Dimensional Two-sample Mean Problems with Consideration of Correlation Structure	
14:20-14:45	周叶青（同济大学）	Rank-based indices for testing independence between two high-dimensional vectors	
14:45-15:10	杨伟超（北京师范大学）	High-dimensional testing based on quadratic norm statistics	
间歇（15:10-15:25）			
分会场七（下半场）：无模型假设的统计推断			
召集人：邹长亮（南开大学）			
15:25-15:50	梁德才（南开大学）	Tests of Partial/Weak Separability for Complex Functional Data	郑术蓉（东北师范大学）
15:50-16:15	任好洁（上海交通大学）	Online multiple changepoints detection with false discovery rate control	
16:15-16:40	王光辉（华东师范大学）	Reliever: Relieving the Burden of Costly Model Fits for Changepoint Detection	
16:40-17:05	严晓东（山东大学）	策略统计学习	

分会场报告（八）（4月20日下午，会议室3）

分会场八（上半场）：案例教学专场			
时间	报告人	演讲题目	会议主席
召集人：潘蕊（中央财经大学）			
13:30-13:55	朱雪宁（复旦大学）	从《红楼梦》文本看红楼梦作者之谜	常莹（狗熊会）
13:55-14:20	王菲菲（中国人民大学）	案例教学模式探索：以时间序列分析为例	
14:20-14:45	常莹（狗熊会）	数据科学人才实践能力的培养：科阶段实践类教学活动的设计与实施	
14:45-15:10	黄丹阳（中国人民大学）	研究生商务大数据案例分析课程教学	
间歇（15:10-15:25）			
分会场八（下半场）：青年学者成长论坛			
对话嘉宾	黄丹阳（中国人民大学）		
	杨松山（中国人民大学）		
	邱怡轩（上海财经大学）		
	邓婉璐（清华大学）		

五、主题报告人简介

1. 陈松蹊

报告题目：数据同化：机理与观测相结合的数据融合方法

报告人简介：陈松蹊，中国科学院院士，现任北京大学讲席教授、数学科学学院、光华管理学院教授，中国概率统计学会理事长；美国科学促进会、美国统计学会、数理统计学会会士。他的主要研究方向为超高维大数据统计分析、计量经济，环境统计等。

报告摘要：The ensemble Kalman Filter (EnKF), as a fundamental data assimilation approach, has been widely used in many fields of earth science, engineering and beyond. However, there are several unknown theoretical aspects of the EnKF, especially when the state variable is of high dimensional accompanied with high resolution observation and physical models.

This paper first proposes several high dimensional EnKF methods which provide consistent estimators for the important forecast error covariance and the Kalman gain matrix. It then studies the theoretical properties of the EnKF under both the fixed and high dimensional state variables, which provides the mean square errors of the analysis states to the underlying oracle states offered by the Kalman filter and gives the much needed insight into the roles played by forecast error covariance on the accuracy of the EnKF. The accuracy of the data assimilation under the misspecified physical model is also considered. Numerical studies on the Lorenz-96 and the Shallow Water Equation models illustrate that the proposed high dimensional EnKF algorithms perform better than the standard EnKF methods as they provide more robust and accurate assimilated results.

2. 朱雪宁

报告题目：Distributed Estimation and Inference for Spatial Autoregression Model with Large Scale Networks

报告人简介：朱雪宁，复旦大学大数据学院副教授，博士生导师。2017 年获得北京大学光华管理学院商务统计与经济计量系博士学位，2017-2018 在美国宾夕法尼亚州立大学从事博士后研究工作。入选 2019 年度上海市青年科技英才扬帆计划，2022 年获得国家自然科学基金优秀青年基金项目资助。主要研究领域为网络数据分析、空间计量模型、高维数据建模等，研究成果发表于 Journal of Econometrics, Journal of the American Statistical Association, Annals of Statistics, 中国科学等国内外经济计量与统计学期刊，著有教材 2 本。

报告摘要：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.

3. 赵鹏

报告题目：网络风险统计建模及保险应用

报告人简介：赵鹏，江苏师范大学教授、副校长。主要从事可靠性统计和网络可靠性等领域的研究工作，先后入选基金委国家优青、杰青项目。现任 Commun. Stat.、《应用概率统计》《数理统计与管理》编委；全国工业统计教学研究会青年统计学家协会副会长、中国现场统计研究会大数据统计分会副理事长等。

报告摘要：In recent years, data breaches have become a significant concern, leading to substantial financial losses annually. However, the lack of suitable statistical approaches for assessing breach risks poses an obstacle. To address this challenge, we first propose a novel statistical model that focuses on analyzing hacking breach risks at the individual company level. We then develop a multivariate frequency-severity framework that examines breach risks at the state level. Additionally, we introduce the concept of the data breach lifecycle. By incorporating this lifecycle and utilizing a novel self-exciting marked point process model, we enhance our understanding of the temporal dynamics of data breaches. Applications in insurance industry are also presented.

分会场一（上半场）：因果学习、深度学习与强化学习的最新进展

1. 崔逸凡

报告题目: Marginal structural Cox models for time-varying endogenous treatment

报告人简介: 崔逸凡, 浙江大学研究员, 博士生导师。2018 年于北卡罗来纳大学教堂山分校获得统计与运筹专业博士学位, 曾在宾夕法尼亚大学沃顿商学院从事博士后研究工作。回国前任职于新加坡国立大学统计与数据科学系担任助理教授, 国家级青年人才计划入选者 (2021)。当选 ISI (国际统计学会) Elected Member, 入选福布斯亚洲 U30 杰出青年, 现担任 Biostatistics & Epidemiology, Biometrical Journal 的 Associate Editor 以及 Journal of Machine Learning Research 的 editorial board reviewer。

报告摘要: Robins (1998) introduced marginal structural models, a general class of counterfactual models for the joint effects of time-varying treatments in complex longitudinal studies subject to time-varying confounding. Robins (1998) established the identification of marginal structural model parameters under a sequential randomization assumption, which rules out unmeasured confounding of treatment assignment over time. The marginal structural Cox model is one of the most popular marginal structural models for evaluating the causal effect of time-varying treatments on a censored failure time outcome. In this paper, we establish sufficient conditions for identification of marginal structural Cox model parameters with the aid of a time-varying instrumental variable, in the case where sequential randomization fails to hold due to unmeasured confounding.

2. 刘林

报告题目: On asymptotic model selection criterion in causal inference

报告人简介: 刘林, 于 2020 年入职上海交大自然科学研究院与数学科学学院, 于交大-耶鲁生物统计与数据科学联合中心双聘, 任职长聘教轨副教授。主要研究方向为半参数理论 因果推断与机器学习理论。

报告摘要: Data-driven confounder selection can be an important component in practical causal inference done daily in large tech companies. In this talk, we discuss the "mechanics" of several commonly employed model selection criterion on the performance of the AIPW estimator for the counterfactual mean, through the lens of Le Cam's third lemma and local perturbing laws. In particular, we consider the following criterion: change-in-estimates (formally known as the Lepski's adaptation scheme in the math-stats literature), focused information criterion, CTMLE, and several other proposals in the literature. We do not give any practical recommendation but instead consider the following questions: what are their algebraic connections and under what conditions, a certain method should be favored under the "gold-standard" that the selected estimator is closer to the unbiased one.

3. 邱怡轩

报告题目: ReHLine: Regularized Composite ReLU-ReHU Loss Minimization with Linear Computation and Linear Convergence

报告人简介: 邱怡轩, 上海财经大学统计与管理学院副教授, 博士毕业于普渡大学统计系, 毕业后曾于卡内基梅隆大学担任博士后研究员。主要研究方向包括统计计算、深度学习、贝叶斯计算与推断等, 科研成果发表在统计学国际权威期刊及机器学习顶级会议上。长期参与建设统计学与数据科学社区“统计之都”, 是众多开源软件的开发者与维护者。

报告摘要: Empirical risk minimization (ERM) is a crucial framework that offers a general approach to handling a broad range of machine learning tasks. In this paper, we propose a novel algorithm, called ReHLine, for minimizing a set of regularized ERMs with convex piecewise linear-quadratic loss functions and optional linear constraints. The proposed algorithm can effectively handle diverse combinations of loss functions, regularizations, and constraints, making it particularly well-suited for complex domain-specific problems. Examples of such problems include FairSVM, elastic net regularized quantile regression, Huber minimization, etc. In addition, ReHLine enjoys a provable linear convergence rate and exhibits a per-iteration computational complexity that scales linearly with the sample size. The algorithm is implemented with both Python and R interfaces, and its performance is benchmarked on various tasks and datasets. Our experimental results demonstrate that ReHLine significantly surpasses generic optimization solvers in terms of computational efficiency on large-scale datasets. Moreover, it also outperforms specialized solvers such as Liblinear in SVM, hqreg in Huber minimization, and Lightning (SAGA, SAG, SDCA, SVRG) in smoothed SVM, exhibiting exceptional flexibility and efficiency.

4. 周帆

报告题目: Recent advances in Distributional Reinforcement Learning

报告人简介: 周帆, 上海财经大学统计与管理学院副教授, 博士毕业于美国北卡罗来纳大学教堂山分校。主要研究方向包括强化学习, 深度学习, 因果推断。在 Journal of American Statistical Association, Journal of Machine Learning Research, Biometrics, Nature Genetics 等统计学机器学习期刊以及 NeurIPS, ICML, KDD, IJCAI 等国际人工智能顶会接收发表一作通讯文章数十篇, 曾获得国际泛华统计协会新研究者奖, 北卡教堂山分

校 Barry H. Margolin Award, 并入选上海市人才计划 (青年)。

报告摘要: Although distributional reinforcement learning (DRL) has been widely examined in the past few years, very few studies investigate the validity of the obtained Q-function estimator in the distributional setting. We discuss some of our works in ensuring the monotonicity of the obtained quantile estimates and the theoretical necessity. Moreover, we undertake a comprehensive analysis of how the approximation errors within the Q-function impact the overall training process in DRL. We both theoretically analyze and empirically demonstrate techniques to reduce both bias and variance in these error terms, ultimately resulting in improved performance in practical applications.

分会场一（下半场）：深度学习的理论与应用

1. 黄维然

报告题目: Theoretical Insights into Self-Supervised Contrastive Learning

报告人简介: 黄维然, 上海交通大学清源研究院副教授、博士生导师, 专注于机器学习理论、自监督学习、小样本学习等研究领域, 荣获 2023 年“CCF-百度松果基金”资助。他毕业于清华大学, 分别取得了本科和博士学位。他在机器学习三大顶会 ICML、NeurIPS、ICLR 以及计算机视觉顶会 CVPR、ICCV 等国际高水平会议上发表论文二十余篇, 其研究工作曾入选 ICCV 2023 最佳论文奖的候选名单。

报告摘要: Self-supervised learning has recently attracted significant attention because it requires only unlabeled data for model training. Contrastive learning, a popular method within self-supervised learning, has demonstrated promising empirical performance. However, the theoretical understanding of its generalization ability remains limited. In this talk, I will analyze self-supervised contrastive learning from a theoretical perspective, demonstrating that its generalization ability is linked to three key factors: the alignment of positive samples, the divergence of class centers, and the concentration of augmented data. Moreover, I will show that self-supervised contrastive learning fails to learn domain-invariant features, thereby limiting its transferability. To address this issue, I will introduce Augmentation-robust Contrastive Learning (ArCL) and show how it significantly improves the transferability of self-supervised contrastive learning.

2. 刘勇

报告题目: In-context Learning 隐式更新机理研究

报告人简介: 刘勇, 中国人民大学, 副教授, 博士生导师。长期从事机器学习基础理论研究, 共发表论文 80 余篇, 其中以第一作者/通讯作者发表顶级期刊和会议论文近 50 篇, 涵盖机器学习领域顶级期刊 TIT、JMLR、TPAMI、Artificial Intelligence 和顶级会议 ICML、NeurIPS 等。获中国人民大学“杰出学者”、中国科学院“青年创新促进会”成员、中国科学院信息工程研究所“引进优青”等称号。主持国家自然科学基金面上/基金青年、北京市面上项目、中科院基础前沿科学研究计划、腾讯犀牛鸟基金、CCF-华为胡杨林基金等项目。

报告摘要: 预训练大语言模型表现出惊人的上下文学习能力 (In-context Learning, ICL)。给定少数几个示例, 模型在没有参数更新的情况下实现在新任务上表现出极好的学习性能, 然而关于 ICL 的内在学习机理仍不清楚。将 ICL 的推理过程解释为一种对比学习模式下的隐式梯度更新过程, 从对比学习的视角给出了 ICL 一种全新解释。此外, 从对比学习的角度提出了几种改进原有 ICL 方法的思路。

3. 毛晓军

报告题目: Decentralized Reduced Rank Regression for Response Partition

报告人简介: 毛晓军, 上海交通大学长聘教轨副教授。他的研究领域包括分布式统计推断, 推荐系统和高维数据分析。主要研究成果已经发表于 JASA, JMLR, IEEE TIT, IEEE TSP, ICML, 《管理世界》等顶级期刊及会议上。入选 2023 年度上海市青年科技启明星计划, 2019 年度上海市青年科技英才扬帆计划。目前是国际重要学术期刊 Journal of Multivariate Analysis 的 Early Career Advisory Board 成员。先后主持国家自然科学基金面上项目和青年基金, 参与国家自然科学基金重大研究计划等。

报告摘要: Distributed learning in decentralized networks has been extensively studied and applied in various machine-learning scenarios. However, previous research primarily focused on data partitioning based on samples. In this paper, we address the less explored scenario of response partition, where different components of the response vector are collected and stored across multiple nodes in a multi-agent network. To mitigate the information loss resulting from response partitioning, we use the Reduced Rank Regression (RRR) model to establish connections between the response components. Subsequently, we formulate an optimization problem that involves both local and global parameters within the framework of matrix factorization, capturing both inter-node and intra-node correlations. To solve this problem efficiently, we propose an algorithm based on Decentralized Gradient Descent with Gradient Tracking (DGGT), which incorporates an additional step for local estimation. The theoretical analysis yields non-asymptotic error bounds for both estimation error and consensus error. As the number of iterations tends to infinity, the statistical error rate converges to the

optimal performance achieved in the centralized case. Furthermore, we validate the effectiveness of our method through simulations and real-world applications. The numerical results not only align with our theoretical findings but also demonstrate the superiority of our approach over local reduced-rank regression methods.

4. 谢琦

报告题目: 参数化卷积方法及其应用

报告人简介: 西安交通大学数学与统计学院副教授，博导。于 2013 年 7 月和 2020 年 12 月分别获西安交通大学理学学士与理学博士学位。2017 年 8 月至 2018 年 9 月曾赴普林斯顿大学访学。目前主要从事机器学习与计算机视觉的基础问题研究。在 CCF A 类期刊与会议发表论文 18 篇，IEEE Trans. 论文 10 篇，其中以第一作者在领域顶刊 TPAMI 发表论文 3 篇；三篇论文入选 ESI 高被引论文。2015 年至今，谷歌学术被引 4160 次，H 指数为 21。入选 2022 年 CCF 优秀博士学位论文激励计，曾获“2023 年陕西省优秀博士学位论文奖”、“Valse 年度最佳学生论文提名奖”、“徐宗本应用数学论文奖”、“2021 年 ACM 中国优博提名奖”等奖项。

报告摘要: 卷积算子是卷积网络核心，在现代深度学习领域发挥了重要的作用。然而，常用的离散形式卷积核无法灵活进行变换，功能上仍具有局限性，不适用于旋转与尺度不变性刻画、卷积核动态调整和非网格卷积等操作。参数化（连续化）卷积方法是克服上述问题的一个重要技术，因逐渐引起重视。现阶段，参数化卷积方法的研究处在起步阶段，本报告将介绍一种适用于底层视觉任务的参数化卷积方法，并以旋转等变卷积算子的构造为代表，介绍一系列参数化卷积方法的应用。

分会场二（上半场）：机器学习与统计推断

1. 申国豪

报告题目: Exploring the Complexity of Deep Neural Networks through Functional Equivalence.

报告人简介: 申国豪是香港理工大学应用数学系的助理教授。2022 年毕业于香港中文大学获统计学博士学位。他的研究兴趣包括统计机器学习和非参数统计学，尤其关注深度学习的基础理论。他的研究成果已发表在 *Annals of Statistics*, *Biometrika*, *Journal of Econometrics*, and *NeurIPS* 等期刊和会议上。

报告摘要: We investigate the complexity of deep neural networks through the lens of functional equivalence, which posits that different parameterizations can yield the same network function. Leveraging the equivalence property, we present a novel bound on the covering number for deep neural networks, which reveals that the complexity of neural networks can be reduced. Additionally, we demonstrate that functional equivalence benefits optimization, as overparameterized networks tend to be easier to train since increasing network width leads to a diminishing volume of the effective parameter space. These findings can offer valuable insights into the phenomenon of overparameterization and have implications for understanding generalization and optimization in deep learning.

2. 韩东啸

报告题目: Inference for High Dimensional Proportional Hazards Model with Streaming Survival Data

报告人简介: 韩东啸博士是南开大学统计与数据科学学院副教授，在 *Joe*, *Jbes*, *biometrics*, *Scandinavian of Statistics* 等杂志上发表过多篇文章，主要研究方向为生物统计，高维数据统计推断，机器学习。

报告摘要: We propose an online inference procedure for high dimensional streaming survival data based on the proportional hazards model. We offer an online Lasso method for regression parameter estimation and establish the non-asymptotic error bounds of the corresponding Lasso estimators for the regression parameter vector. In addition, we study the pointwise and group inference for the regression parameters by utilizing a debiased Lasso method. Extensive simulations are conducted to evaluate the finite sample performance of the proposed method. An application to a colon cancer dataset is provided to demonstrate the practical utility of the proposed methodology.

3. 宋珊珊

报告题目: Semi-supervised Inference for Block-wise Missing Data without Imputation

报告人简介: 宋珊珊博士，目前在香港中文大学统计系做博士后研究员。在此之前，她在 2020 年于上海财经大学取得博士学位。宋博士的研究方向包括大数据分析，半监督学习，统计机器学习。相应成果已发表在 *JASA*, *GJS*, *Fundamental Research* 等期刊。

报告摘要: We consider statistical inference for single or low dimensional parameters in a high-dimensional linear model under a semi-supervised setting, wherein the data are a combination of a labelled block-wise missing data set of a relatively small size and a large unlabelled data set. The proposed method utilises both labelled and unlabelled data without any imputation or removal of the missing observations. The asymptotic properties of the estimator are established under

regularity conditions. Hypothesis testing for low-dimensional coefficients are also studied. Extensive simulations are conducted to examine the theoretical results. The method is evaluated on the Alzheimer's Disease Neuroimaging Initiative data.

4. 张晓

报告题目: Hypothesis testing in high-dimensional censored transformation models

报告人简介: 张晓现为香港中文大学（深圳）数据科学学院博士后，曾任新加坡国立大学博士后。他 2021 年于上海财经大学统计与管理学院取的博士学位，主要研究方向包括高维数据统计推断、单细胞/空间转录组数据分析，曾在 *nature communications*、*Journal of Computational and Graphical Statistics* 以及 *The Econometrics Journal* 等国际知名期刊发表论文。

报告摘要: As the rapid development of modern technologies, high-dimensional statistical inference of survival time is increasingly important in various fields, such as biostatistics and financial risk. Detecting whether the high-dimensional features exist significantly is crucial in practice for accurate model interpretation and effective decision-making. Herein, we propose an efficient rank-based test statistic that is asymptotically normally distributed. The proposed test statistic allows covariance dependent censoring, robust against heavy-tailed distributions and potential outliers for response of interest. Out of practical needs, furthermore, we propose a new rank-based test statistic to test existence of high-dimensional features with high-dimensional control factors. The asymptotic distributions under the null hypothesis and local alternatives are established for proposed test statistics. Extensive numerical studies evaluate the finite-sample performances of the proposed tests and its robustness across heavy-tailed error's distributions and transformation function changes. We also apply the proposed methods to two SKCM data to illustrates its practical performance.

分会场二（下半场）：机器学习前沿方法

1. 杨朋昆

报告题目: A Statistical Perspective of Federated Learning Algorithms and Model Personalization

报告人简介: Pengkun Yang is an assistant professor at the Center for Statistical Science at Tsinghua University. Prior to joining Tsinghua, he was a Postdoctoral Research Associate at the Department of Electrical Engineering at Princeton University. He received a Ph.D. degree (2018) and a master degree (2016) from the Department of Electrical and Computer Engineering at University of Illinois at Urbana-Champaign, and a B.E. degree (2013) from the Department of Electronic Engineering at Tsinghua University. His research interests include statistical inference, learning, optimization, and systems. He is a recipient of Thomas M. Cover Dissertation Award in 2020, and a recipient of Jack Keil Wolf ISIT Student Paper Award at the 2015 IEEE International Symposium on Information Theory (semi-plenary talk).

报告摘要: Abstract: Federated learning is a promising framework with immense potential in privacy preservation and reducing the computation load at the cloud. The successful deployment faces many challenges in both theory and practice such as data heterogeneity and client unavailability. In this talk, I will discuss the resolution from a statistical perspective including the statistical efficiency of FedAvg and FedProx from a nonparametric regression viewpoint, and a new algorithm achieving global convergence when the clients exhibit cluster structure. One notable innovation in our analysis is a uniform estimate on clustering errors, which we prove by bounding the VC dimension of general polynomial concept classes based on the theory of algebraic geometry.

2. 张琼

报告题目: Federated Learning with Mixed-Type Labels

报告人简介: 张琼，2015 年本科毕业于中国科学技术大学少年班学院。2022 年博士毕业于加拿大英属哥伦比亚大学统计系。2022 年 9 月起加入中国人民大学统计与大数据研究院并担任助理教授。目前的研究兴趣包括：混合模型、分布式学习、联邦学习等。她的研究论文发表在 *Journal of Machine Learning Research*, *IEEE Transactions on Information Theory*, *ICCV* 等机器学习期刊和会议上，主持国家自然科学基金青年项目。

报告摘要: In federated learning (FL), classifiers (e.g., deep networks) are trained on datasets from multiple data centers without exchanging data across them, which improves the sample efficiency. The classical setting of FL assumes that the same labeling criterion is employed across data centers being involved in training, which greatly limits the applicability of FL in practice. For example, standards for disease diagnosis are likely to be different in different clinical centers, where we cannot apply methods under the classical setting. In this work, we consider this important yet under explored setting of FL, namely FL with mixed type labels, where different labeling criteria are allowed, leading to inter-center label space differences. Our proposed method is advantageous since it is versatile and can be easily combined with various FL methods such as FedAvg. The efficiency of the proposed method is shown empirically on benchmark and medical datasets.

3. 亓颢博

报告题目: Mini-batch Gradient Descent with Buffer

报告人简介: 亓颢博, 北京师范大学统计学院师资博士后, 博士毕业于北京大学光华管理学院。他的主要研究方向包括统计优化算法、大规模数据统计建模、网络结构数据分析等。在 *Journal of Computational and Graphical Statistics*、*Neurocomputing*、*Computational Statistics & Data Analysis* 等期刊发表多篇论文。

报告摘要: In this paper, we studied a buffered mini-batch gradient descent (BMGD) algorithm for training complex model on massive datasets. The algorithm studied here is designed for fast training on a GPU-CPU system, which contains two steps: the buffering step and the computation step. In the buffering step, a large batch of data (i.e., a buffer) are loaded from the hard drive to the graphical memory of GPU. In the computation step, a standard mini-batch gradient descent (MGD) algorithm is applied to the buffered data. Compared to traditional MGD algorithm, the proposed BMGD algorithm can be more efficient for two reasons. First, the BMGD algorithm uses the buffered data for multiple rounds of gradient update, which reduces the expensive communication cost from the hard drive to GPU memory. Second, the buffering step can be executed in parallel so that the GPU does not have to stay idle when loading new data. We first investigate the theoretical properties of BMGD algorithms under a linear regression setting. The analysis is then extended to the Polyak-Lojasiewicz loss function class. The theoretical claims about the BMGD algorithm are numerically verified by simulation studies. The practical usefulness of the proposed method is demonstrated by three image-related real data analysis.

4. 朱映秋

报告题目: A communication efficient boosting method for distributed spectral clustering

报告人简介: 朱映秋, 对外经济贸易大学统计学院讲师, 主要研究方向包括电子商务数据挖掘、高频金融数据分析、深度学习、社交网络分析等。目前已在 *Journal of the Royal Statistical Society: Series C (Applied Statistics)*、*Neural Networks*、*IEEE Transactions on Knowledge and Data Engineering*、*Neurocomputing*、*Computational Statistics & Data Analysis* 等统计、计算机方向国际权威期刊发表多篇论文。

报告摘要: Spectral clustering is one of the most popular clustering techniques in statistical inference. For large-scale datasets, spectral clustering is typically implemented through distributed computing. However, existing distributed implementations face two major challenges. First, the clustering performance is negatively affected by distributed computing since the topological structure of all objects has to be divided into distributed parts. Second, the communication among computers within a distributed system results in high communication costs. To address these issues, we propose a communication-efficient algorithm for distributed spectral clustering. Our motivation stems from a theoretical comparison between the conventional spectral clustering algorithm, which operates on the entire dataset, and the local spectral clustering, which is performed on a subsample using a single computer. We identify the critical factor that leads to the difference between the performances of global spectral clustering and local spectral clustering. Based on the findings, we propose a novel approach that iteratively aggregates the intermediate results generated by local spectral clustering. In this process, only low-dimensional vectors are exchanged between computers. The results of simulations and real data analysis demonstrate that the proposed method apparently enhances the performance of distributed spectral clustering with low communication costs.

分会场三（上半场）：大数据计算与应用

1. 熊昊一

报告题目: 基于大模型的数据工程与数据科学解决方案

报告人简介: 熊昊一博士, 正高级工程师、百度主任架构师、百度研究院大数据实验室团队负责人。曾任美国密苏里科技大学终身序列助理教授、博士生导师。研究方向为普适计算与深度学习。在 *UbiComp*、*KDD*、*ICML* 等会议期刊上发文 70 余篇, 引用 5200 余次。他参与百度飞桨等深度学习技术研发, 获得中国电子学会科技进步一等奖, 入选北京市海聚工程高层次人才计划, 科技创新 2030 人工智能重大项目课题负责人, CCF 普适计算专委会常委, IEEE 高级会员。

报告摘要: 高效准确的数据操作是数据工程、数据科学的重要组成部分。而大量商业/数据分析师或不具备编写数据查询(如 SQL 或 Gremlin)代码能力。本讲座主要介绍了星致助手——基于大模型的数据工程与数据科学解决方案。星致助手是一种全新的基于自然语言交互的数据分析 AI 原生应用, 旨在降低数据使用门槛、减少研发成本, 解决传统查询流程复杂、取数流程长和现有的 NL2SQL 能力有限等问题。它由全过程无需人工操作和复杂工具使用, 以聊天的形式与助手沟通即可快速、高效、精确地完成探索性数据分析需求。

2. 汪澎洋

报告题目: 时空图数据表示学习

报告人简介: 汪澎洋, 博士, 澳门大学智慧城市物联网国家重点实验室助理教授, 澳门大学珠海研究院副研究员, 博士生导师。研究兴趣包括数据挖掘、机器学习和大数据分析, 特别是时空图数据的表征学习, 及其在智慧城市、交通、人/系统行为建模、关键基础设施防护和复杂网络分析上的应用, 在数据挖掘顶级期刊和会议上发表文章

40 余篇；获得了“AI 华人新星百强”奖，SIGKDD 2018 最佳论文提名奖，和 SIGSPATIAL 2020 最佳论文提名奖。其中，自动城市规划的工作获得了“Synced Review”和“UCF Today”等媒体的报道，城市活力系列工作获得了美国自然科学基金的报道。

报告摘要：随着物联网和信息物理系统的发展，时空图结构数据在很多领域变得越来越普遍，例如智慧城市、智慧电网、智慧农业、智慧医疗等。在这些系统里，传感器网络产生了大量的时空数据。挖掘这些时空图数据可以帮助我们感知系统和环境，并进一步提供决策支持。在今天的报告中，我将首先介绍什么是时空图数据和对时空图数据建模的关键技术难点。然后，我将集中介绍（1）对关键性质进行建模的时空图数据的表征学习；（2）以及他们在用户行为建模上的应用。最后，我将做出总结，并从宏观上对建设闭环可信的智能感知-决策系统做出展望。

3. 孟祥溪

报告题目：用数据科学搭建医学图像到临床决策的桥梁

报告人简介：孟祥溪，北京肿瘤医院核医学科助理研究员。主要研究方向是核医学影像物理学，光学分子影像技术，分子影像中的定量问题等。2019 年于北京大学、佐治亚理工学院和埃默里大学获生物医学工程博士学位。在 PET/MR 的成像方法、高能 X 射线切伦科夫成像、PET 影像在疾病诊断中的应用等方面作为第一作者和通讯作者发表学术论文十余篇。承担和参与北京高校高精尖学科建设课题、科技部重大专项等课题。获得发明专利授权 2 项，实用新型 1 项。

报告摘要：在医学影像的临床研究中，存在三个学术传统和方法学视角，即以流行病学和循证医学为基础的征象分析，以模式识别与统计学习为基础的影像生物标志物建模，和以深度学习与人工智能为基础的端到端模型应用。三者临床实践中相互补充，为诊断和治疗决策提供帮助。核医学与分子影像是一个跨学科的研究领域，同样可以采取这样的视角进行分析，且面临着一些特殊的挑战与机会。我们通过对医学图像噪声特性的总结与研究，提出了一种内容-噪声互补的学习策略，基于对抗生成网络实现了图像增强与去噪，并表现出了极佳的泛化能力，可以在 PET 等多种不同医学影像模态得到应用。在临床研究方面，我们通过传统的流行病学研究，证实了 PET/MR 在纵隔肿瘤诊断方面具有更强的诊断效能。在分类模型方面，我们利用影像组学技术构建了基于 PET/CT 的肺癌 PD-L1 诊断模型、肺癌胸膜侵犯术前评估模型、前纵隔淋巴瘤和胸腺来源肿瘤的鉴别模型等。

4. 杨亚磊

报告题目：贝叶斯网络在医学影像中的应用：参数估计与分割的不确定性研究

报告人简介：杨亚磊，格拉斯哥大学，统计学博士，北京大学肿瘤医院与联影医疗联合培养博士。擅长统计建模，统计推断，统计分析，图像处理（医学）。论文发表在统计与人工智能期刊 *Journal of the Royal Statistical Society: Series C, Artificial Intelligence in Medicine* 等。

报告摘要：贝叶斯网络，也被称为层次贝叶斯模型，是一种包含多层具有复杂关系参数的统计模型。这一族模型能够在贝叶斯统计的框架下以图模型的形式展现不同层参数之间的依赖关系。在我们的工作中，我们将这一族模型应用于不同类型的医学图像之上，例如，心肌灌注 DCE-MRI，多模态脑部 MRI 以及动态 PET 图像。这一族模型具有高度的灵活性以包含基于生理模型的方法，如 Fermi 模型，two tissues compartmental 模型，或者不基于生理模型的方法，如时空自回归模型。同时，它还能够利用马尔可夫随机场来结合邻域信息或者其它的先验信息。在我们的工作中，我们还引入了一个隐变量来表达医学图像中每一个像素的标签，亦即感兴趣区域的性质。参数的条件后验分布是可以利用它的马尔可夫毯来进行推断的，而参数的估计或者抽样则可以利用最大后验估计或者马尔可夫链蒙特卡罗来进行。因为贝叶斯网的统计推断在我们的工作中可以被明确给出，所以估计或者分割的不确定性定量分析能够自然得到。

分会场三（下半场）：统计学习的最新进展

1. 李挺

报告题目：Conditional Stochastic Interpolation for Generative Learning

报告人简介：Dr. Li is an assistant professor in the Department of Applied Mathematics at Hong Kong Polytechnic University. Prior to joining PolyU, he was a postdoctoral associate in Yale University, Biostatistics Department. He received his PhD in Hong Kong University of Science and Technology. His research focuses on data science and statistical learning on complex data, especially on network data, brain data and imaging genomics.

报告摘要：We propose a conditional stochastic interpolation (CSI) approach for learning conditional distributions. CSI learns probability flow equations or stochastic differential equations that transport a reference distribution to the target conditional distribution. This is achieved by first learning the velocity function and the conditional score function based on conditional stochastic interpolation, which are then used to construct a deterministic process governed by an ordinary differential equation or a diffusion process for conditional sampling. We establish the transport equation and derive the explicit form of the conditional score function with mild conditions. We also incorporate an adaptive diffusion term in

our proposed CSI model to address the instability issues arising during the training process. Furthermore, we establish non-asymptotic error bounds for learning the target conditional distribution via conditional stochastic interpolation in terms of KL divergence, taking into account the neural network approximation error. We illustrate the application of CSI on image generation using benchmark image data.

2. 师佳鑫

报告题目: Mixture Conditional Regression with Ultrahigh Dimensional Text Data for Estimating Extralegal Factor Effects

报告人简介: 师佳鑫, 北京大学光华管理学院商务统计与经济计量系在读博士生, 师从王汉生教授。主要研究方向为高维数据中的隐藏结构分析, 因子模型, 复杂网络数据分析等。研究论文被 *Annals of Applied Statistics* 期刊接收。

报告摘要: Testing judicial impartiality is a problem of fundamental importance in empirical legal studies, for which standard regression methods have been popularly used to estimate the extralegal factor effects. However, those methods cannot handle control variables with ultrahigh dimensionality, such as those found in judgment documents recorded in text format. To solve this problem, we develop a novel mixture conditional regression (MCR) approach, assuming that the whole sample can be classified into a number of latent classes. Within each latent class, a standard linear regression model can be used to model the relationship between the response and a key feature vector, which is assumed to be of a fixed dimension. Meanwhile, ultrahigh dimensional control variables are then used to determine the latent class membership, where a naive Bayes type model is used to describe the relationship. Hence, the dimension of control variables is allowed to be arbitrarily high. A novel expectation-maximization algorithm is developed for model estimation. Therefore, we are able to estimate the key parameters of interest as efficiently as if the true class membership were known in advance. Simulation studies are presented to demonstrate the proposed MCR method. A real dataset of Chinese burglary offenses is analyzed for illustration purposes.

3. 吕绍高

报告题目: 基于分期贝叶斯表示的个性化联邦学习

报告人简介: 吕绍高, 现为南京审计大学统计与数据科学学院教授, 博士生导师。2011年毕业于中国科大-香港城市大学联合培养项目, 获得理学博士学位。主要研究方向是统计机器学习, 当前研究兴趣包括联邦学习、统计强化学习以及几何深度学习。迄今为止在 SCI 检索的国际期刊上发表论文 20 多篇, 包括统计学类期刊《*Annals of Statistics*》、人工智能类期刊《*Journal of Machine Learning Research*》、CCF A 类国际会议“*NeurIPS*”与“*IJCAI*”以及计量经济学期刊《*Journal of Econometrics*》等。曾主持过国家自然科学基金项目 3 项。长期担任人工智能顶级会议“*NeurIPS*”、“*ICML*”、“*AAAI*”以及“*AIStat*”程序委员或审稿人。

报告摘要: 联邦学习是一种去中心化的隐私保护技术, 能让多个客户端与服务器协作学习全局模型, 而不会暴露自己的私人数据。然而客户机之间存在统计异质性会带来挑战, 尤其全局模型可能很难在每个客户的特定任务上表现出色。在异质联邦学习中, 如何有效地表示全局共享信息与局部个性化信息是异质算法是否成功的关键。为了解决这个问题, 我们引入了一个新的视角, 即通过分期贝叶斯元学习 (Amortized Bayesian Meta-Learning) 实现个性化联邦学习。具体来说, 我们提出了一种名为 FedABML 的新算法, 它采用了跨客户端的分层变异推理。全局先验旨在捕捉来自异构客户端的共同内在结构表征, 然后将其转移到各自的具体任务中, 并通过少量局部更新帮助生成精确、特定的、客户端近似后验。我们的理论分析给出了平均泛化误差的上限, 并保证了在未见数据上的泛化性能。最后, 通过几项实证结果表明, FedABML 的性能优于几种具有常用的基线方法。

4. 高原

报告题目: An Asymptotic Analysis of Random Partition Based Minibatch Momentum Methods for Linear Regression Models

报告人简介: 高原, 博士毕业于华东师范大学统计学院, 现为北京大学光华管理学院博士后。主要研究方向包括大规模数据统计计算方法、网络数据分析等, 目前在《*Journal of Computational and Graphical Statistics*》、《*Stat*》、《*Journal of Econometrics*》、《*Neural Networks*》等国际期刊上发表多篇论文。

报告摘要: Momentum methods have been shown to accelerate the convergence of the standard gradient descent algorithm in practice and theory. In particular, the random partition based minibatch gradient descent methods with momentum (MGDM) are widely used to solve large-scale optimization problems with massive datasets. Despite the great popularity of the MGDM methods in practice, their theoretical properties are still underexplored. To this end, we investigate the theoretical properties of MGDM methods based on the linear regression models. We first study the numerical convergence properties of the MGDM algorithm and derive the conditions for faster numerical convergence rate. In addition, we explore the relationship between the statistical properties of the resulting MGDM estimator and the tuning parameters. Based on these theoretical findings, we give the conditions for the resulting estimator to achieve the optimal statistical efficiency. Finally, extensive numerical experiments are conducted to verify our theoretical

results. (Joint work with Xuening Zhu, Haobo Qi, Guodong Li, Riquan Zhang, and Hansheng Wang)

分会场四（上半场）：大规模网络及时序数据建模与计算

1. 伍书缘

报告题目: Quasi-Newton Updating for Large-Scale Distributed Learning

报告人简介: 伍书缘, 上海财经大学统计与管理学院助理研究员。2023 年从北京大学获得经济学博士学位。主要研究方向为分布式计算、再抽样方法、统计优化算法、大规模数据统计建模等。目前已在 Journal of the Royal Statistical Society Series B: Statistical Methodology, Journal of Business and Economic Statistics, Statistica Sinica 等国际权威期刊上发表多篇论文。

报告摘要: Distributed computing is critically important for modern statistical analysis. Herein, we develop a distributed quasi-Newton (DQN) framework with excellent statistical, computation, and communication efficiency. In the DQN method, no Hessian matrix inversion or communication is needed. This considerably reduces the computation and communication complexity of the proposed method. Notably, related existing methods only analyze numerical convergence and require a diverging number of iterations to converge. However, we investigate the statistical properties of the DQN method and theoretically demonstrate that the resulting estimator is statistically efficient over a small number of iterations under mild conditions. Extensive numerical analyses demonstrate the finite sample performance.

2. 蒋斐宇

报告题目: Testing serial independence of object-valued time series.

报告人简介: 蒋斐宇, 复旦大学管理学院统计与数据科学系青年副研究员。2021 年从清华大学获得统计学博士学位, 现主持国家自然科学基金青年科学基金一项, 上海市扬帆计划一项。主要研究领域为时间序列分析、变点分析、金融计量经济学等, 研究成果发表在 Biometrika, JRSSB, JOE 等期刊。

报告摘要: We propose a novel method for testing serial independence of object-valued time series in metric spaces, which is more general than Euclidean or Hilbert spaces. The proposed method is fully nonparametric, free of tuning parameters and can capture all nonlinear pairwise dependence. The key concept used in this paper is the distance covariance in metric spaces, which is extended to auto-distance covariance for object-valued time series. Furthermore, we propose a generalized spectral density function to account for pairwise dependence at all lags and construct a Cramér von-Mises type test statistic. New theoretical arguments are developed to establish the asymptotic behaviour of the test statistic. A wild bootstrap is also introduced to obtain the critical values of the nonpivotal limiting null distribution. Extensive numerical simulations and two real data applications on cumulative intraday returns and human mortality data are conducted to illustrate the effectiveness and versatility of our proposed test.

3. 许杏柏

报告题目: TranSAR: a spatial transfer learning framework

报告人简介: 许杏柏, 厦门大学王亚南经济研究院和经济学院副教授, 博士生导师。2016 年从俄亥俄州立大学获得经济学博士学位。现在主持国家自然科学基金面上项目一项。主要研究领域为空间计量经济学和网络计量经济学。多份研究成果发表在 Journal of Econometrics, Econometric Theory, Regional Science and Urban Economics 等学术期刊上。

报告摘要: The spatial autoregressive (SAR) model is widely used in studying spatial dependent data, as it is tailored to capture spatial autocorrelation by integrating neighboring observations into the regression model. SAR models have found extensive utility, primarily focusing on specific target datasets. Consequently, there exists a compelling demand for the exploration of methodologies aimed at enhancing the estimation performance of SAR models through the assimilation of information derived from similar source datasets, a topic that warrants deeper investigation. In the context of this research, we endeavor to address this gap by introducing a transfer learning framework to the SAR model, designed to ameliorate the estimation performance, which has not been extensively examined from a theoretical standpoint within the realm of spatial econometrics. Furthermore, the paper introduces error bounds for the proposed two-step estimator that are relevant to any plausible first-step primary estimator within the spatial model framework. We evaluate the finite sample performance of these methods across diverse scenarios, demonstrating the outstanding performance of our proposed approach. Subsequently, we apply the proposed method to predict election outcomes for swing states. Leveraging the polling data from the 2016 US presidential election, alongside other demographic and geographical data, we aim to predict election results using the proposed technique, outperforming traditional methods in the majority of cases.

4. 蒋滨雁

报告题目: A two-way heterogeneity model for dynamic networks

报告人简介: 蒋滨雁, 香港理工大学应用数学系副教授。2007 年获中国科学技术大学统计学学士学位, 2012 年

大学获新加坡国立大学统计与应用概率学博士学位。博士毕业后在美国卡内基梅隆大学从事博士后工作，并于2015年入职香港理工大学。主要研究领域包括网络数据分析，高维数据分析和生存分析。统计理论和建模方面的代表性研究成果曾在 *Journal of Machine Learning Research*, *Journal of American Statistical Association*, *Biometrika*, *Biostatistics*, *Statistica Sinica* 等学术期刊上。跨领域的合作成果也曾发表在相关领域的顶级期刊和会议上，例如教育心理学期刊 *Child Development* 以及计算机会议 *International World Wide Web Conference*。

报告摘要: Analysis of networks that evolve dynamically requires the joint modelling of individual snapshots and time dynamics. This paper proposes a new flexible two-way heterogeneity model towards this goal. The new model equips each node of the network with two heterogeneity parameters, one to characterize the propensity to form ties with other nodes statically and the other to differentiate the tendency to retain existing ties over time. With observed networks each having nodes, we develop a new asymptotic theory for the maximum likelihood estimation of parameters when np tends to infinity. We overcome the global non-convexity of the negative log-likelihood function by the virtue of its local convexity, and propose a novel method of moment estimator as the initial value for a simple algorithm that leads to the consistent local maximum likelihood estimator (MLE). To establish the upper bounds for the estimation error of the MLE, we derive a new uniform deviation bound, which is of independent interest. The theory of the model and its usefulness are further supported by extensive simulation and a data analysis examining social interactions of ants.

分会场四（下半场）：大规模网络建模和统计推断

1. 冯龙

报告题目: Computationally efficient and data-adaptive change-point inference in high dimension

报告人简介: 冯龙现任南开大学统计与数据科学学院副教授、特聘研究员、博士生导师。入选教育部青年长江学者、南开大学百名青年学科带头人。曾获得教育部学术新人奖，南开大学优秀博士论文奖。主要从事高维数据分析方面的研究，在统计学国际顶尖杂志 *JRSSB*, *JASA*, *Biometrika*, *Annals of Statistics*, *JOE*, *JBES* 等发表多篇论文。主持一项国家自然科学基金面上项目和青年项目。

报告摘要: High-dimensional change-point inference that adapts to various change patterns has received much attention recently. We propose a simple, fast yet effective approach for adaptive change-point testing. The key observation is that two statistics based on aggregating cumulative sum statistics over all dimensions and possible change-points by taking their maximum and summation, respectively, are asymptotically independent under some mild conditions. Hence we are able to form a new test by combining the p-values of the maximum- and summation-type statistics according to their limit null distributions. To this end, we develop new tools and techniques to establish asymptotic distribution of the maximum-type statistic under a more relaxed condition on component wise correlations among all variables than that in existing literature. The proposed method is simple to use and computationally efficient. It is adaptive to different sparsity levels of change signals, and is comparable to or even outperforms existing approaches as revealed by our numerical studies.

2. 刘耀午

报告题目: Ensemble methods for testing a global null

报告人简介: 刘耀午，西南财经大学统计学院教授。他的研究兴趣包括统计遗传学，大规模假设检验，全基因组关联性分析等。他的多项研究成果发表于 *JASA* 和 *JRSSB* 等统计学顶级期刊，以及 *American Journal of Human Genetics*, *Nature Genetics* 等遗传学顶级期刊。

报告摘要: Testing a global null is a canonical problem in statistics and has a wide range of applications. In view of the fact that no uniformly most powerful test exists, prior and/or domain knowledge are commonly used to focus on a certain class of alternatives to improve the testing power. However, it is generally challenging to develop tests that are particularly powerful against a certain class of alternatives. In this paper, motivated by the success of ensemble learning methods for prediction or classification, we propose an ensemble framework for testing that mimics the spirit of random forests to deal with the challenges. Our ensemble testing framework aggregates a collection of weak base tests to form a final ensemble test that maintains strong and robust power. We apply the framework to four problems about global testing in different classes of alternatives arising from Whole Genome Sequencing (WGS) association studies. Specific ensemble tests are proposed for each of these problems, and their theoretical optimality is established in terms of Bahadur efficiency. Extensive simulations and an analysis of a real WGS dataset are conducted to demonstrate the type I error control and/or power gain of the proposed ensemble tests.

3. 刘秉辉

报告题目: Joint community detection in random effects stochastic block models via the split-likelihood method

报告人简介: 刘秉辉，东北师范大学教授，博士生导师，统计学主任。主要研究方向为统计机器学习和网络数据分析，在统计学，计算机&人工智能，计量经济学期刊发表论文 30 余篇。部分成果发表在 *JASA*, *AOS*, *JOE*, *JMLR*,

JBES 上。支持国家自然科学基金多项，入选国家级青年人才计划，国家天元数学东北中心优秀青年学者，吉林省拔尖创新人才。担任中国现场统计研究会因果推断分会副理事长，中国现场统计研究会统计交叉科学研究分会副理事长等。与中国联通公司，长春市长公开电话办公室等单位合作，主持大数据产品开发、大数据培训若干。

报告摘要: In this study, we tackle the joint community detection in multi-layer networks under a random effects stochastic block model. This model presents a unique challenge as it induces variability in the community structure across each layer of the multi-layer network. This variability is a random transformation originating from a common community structure that permeates all layers. The exact fit for this model is an NP-hard problem. We propose a solution, the ‘split-likelihood method’, which balances detection accuracy and computational efficiency. It employs an approximate likelihood maximization process by decoupling the row and column labels of community assignment. We further establish the convergence theory for our proposed method, along with the consistency theories for the estimated community labels derived from it. Extensive numerical results suggest that our method excels in both detection accuracy and computational efficiency. Finally, we conducted a resting state fMRI study on schizophrenia, to demonstrate the practical applicability of the proposed method.

4. 范新妍

报告题目: Projected semi-parametric network analysis

报告人简介: 范新妍，现任中国人民大学统计学院副教授。主要研究方向为多源数据分析、网络数据分析等，在 JoE, AoAS, JBES, 统计研究等期刊发表论文二十余篇。

报告摘要: Covariate-assisted network analyses such as community detection and link prediction have drawn much attention in recent years. It is worth noting that most of the existing methods have two typical limitations that warrant attention: (1) sensible to the mis-specification of the similarity matrix of nodal covariates; and (2) failed of link prediction for newly added network nodes. To alleviate the above two limitations, we propose the a novel projected semi-parametric network analysis model. The key idea is to model the connection probability matrix as the sum of the inner product of semi-parametric functions of nodal covariates and that of the unobservable low-rank matrix. To estimate the model with both the unknown functions and the latent matrix, we propose a novel projection method by projection the adjacency matrix of the network onto a given linear space spanned by the basis functions induced by nodal covariates to remove the information of latent matrix. We study the statistical properties of the proposed method for both community detection and link prediction, and also discuss the conditions under which the error rates of clustering converge to zero faster than the stochastic block models and the degree corrected block models. The effectiveness of the proposed method has been examined through numerous numerical studies including both simulation studies and real data analysis.

分会场五（上半场）：生物信息与生物医学大数据

1. 胡刚

报告题目: 基于深度学习的空间转录组数据增强方法

报告人简介: 胡刚，南开大学统计与数据科学学院教授，博士生导师，数据科学系主任。博士毕业于南开大学数学科学学院概率论与数理统计专业，2014年3月-2014年7月，2017年10月至2018年10月美国宾夕法尼亚大学生物统计系访问副教授；2007-2018年多次在加拿大阿尔伯塔大学 Cross 癌症研究所和计算机与电子工程系访问。主要从事生物信息学和统计基因组学研究，主持国家自然科学基金和天津市自然科学基金等项目多项。以主要作者在 Circulation, Nature Communications, Briefings in Bioinformatics, Bioinformatics 等杂志发表论文多篇。曾获得天津市自然科学一等奖。

报告摘要: 空间转录组相较于单细胞转录组，同时提供细胞在组织空间的具体位置信息，能让我们进一步理解组织结构以及细胞间相互作用。但是目前的空间转录组数据要么不能达到单细胞分辨率，要么测序深度较浅，限制了空间转录组数据的应用。为了克服这些缺点，我们开发了基于深度学习的空间转录组数据增强方法，通过自监督学习和迁移学习，对空间转录组数据进行去噪和数据增强，从而能够克服目前空间转录组数据测序深度较浅以及分辨率低等缺点。

2. 侯琳

报告题目: Inference of heterogeneous effect in Perturb-Seq experiments

报告人简介: Lin Hou, Tsinghua University.

报告摘要: The integration of CRISPR screening and single-cell RNA sequencing has arisen as a powerful tool for profiling the impact of genetic perturbations on the entire transcriptome at the single-cell scale. Computational methods have been developed to estimate average perturbation effects. Here we present a new method that disentangles perturbation effect from heterogeneous cell state, and infers perturbation effect at single-cell resolution. We demonstrate

in simulation studies and real datasets that our method facilitates genetic interaction analysis, clustering of perturbation effect, and prioritization of genes in various biological processes.

3. 明静思

报告题目: Automated and fast annotation of cell types with hierarchical structure for single-cell RNA-seq data

报告人简介: 明静思, 华东师范大学统计学院和统计交叉科学研究院助理教授。2018年博士毕业于香港浸会大学, 2018-2020年在香港科技大学从事博士后研究工作, 2020年加入华东师范大学。主要研究方向包括统计遗传学, 生物信息学, 统计机器学习等, 研究成果发表于 *Briefings in Bioinformatics*, *Bioinformatics*, *Nature Computational Science*, *Journal of Computational and Graphical Statistics* 等期刊, 入选上海市扬帆计划, 主持一项国家自然科学基金青年基金项目。

报告摘要: Cell type identification is a critical step in analyzing single-cell RNA sequencing (scRNA-seq) data, necessitating the development of fast and reliable computational methods for automatic cell annotation. Cells can be systematically categorized at different levels of granularity, from broad cell types to more specific subtypes. Leveraging the vast collection of well-annotated scRNA-seq datasets and the hierarchical structure of cell types, we present a deep neural network-based model with a progressive growing network structure, which enables rapid and accurate cell type identification. Through extensive benchmark studies, we demonstrate the superior performance of our method over existing approaches in terms of accuracy, stability, and efficiency for cell type annotation.

4. 金锁钦

报告题目: Systematic analysis of cell-cell communication from spatial transcriptomics data

报告人简介: 金锁钦, 武汉大学数学与统计学院教授, 博士生导师, 国家级青年人才计划入选者。主要从事数学、机器学习与生物医学交叉研究, 研究方向包括生物信息学、计算生物学与大数据分析等。在生物医学大数据的数学建模和智能挖掘、发展数学的理论与方法应用于解决生物医学前沿科学问题等方面开展了系列研究, 研究成果发表在 *Nature Communications*, *Nature Neuroscience*, *Genome Biology*, *Nucleic Acids Research*, *SIAM J Appl Math* 等学术期刊上。

报告摘要: Recent advances of single-cell and spatial sequencing technologies provide an unprecedented opportunity for probing underlying intercellular communications that often drive heterogeneity and cell state transitions in tissues. We developed a computational framework for systematic inference and analysis of spatially-proximal cell-cell communication by integrating gene expression, spatial locations and prior knowledge of the interactions between signaling molecules. Furthermore, we can also infer high-order interactions that capture the underlying cell-cell communication patterns and reveal how cells and signals coordinates together for function.

分会场五（下半场）：复杂生物医学数据的统计推断

1. 卜德亮

报告题目: Summary statistics-based association test for identifying the pleiotropic effects with set of genetic variants

报告人简介: 卜德亮, 首都经济贸易大学讲师, 博士毕业于中国科学院大学。主要研究方向是全基因组关联分析和高维统计推断。目前已有数篇文章发表在 *Bioinformatics* 和 *Genetic Epidemiology* 等期刊上。

报告摘要: Traditional genome-wide association study focuses on testing one-to-one relationship between genetic variants and complex human diseases or traits. While its success in the past decade, this one-to-one paradigm lacks efficiency because it does not utilize the information of intrinsic genetic structure and pleiotropic effects. Due to privacy reasons, only summary statistics of current genome-wide association study data are publicly available. Existing summary statistics-based association tests do not consider covariates for regression model, while adjusting for covariates including population stratification factors is a routine issue.

In this work, we first derive the correlation coefficients between summary Wald statistics obtained from linear regression model with covariates. Then, a new test is proposed by integrating three-level information including the intrinsic genetic structure, pleiotropy, and the potential information combinations. Extensive simulations demonstrate that the proposed test outperforms three other existing methods under most of the considered scenarios. Real data analysis of polyunsaturated fatty acids further shows that the proposed test can identify more genes than the compared existing methods.

2. 梅好

报告题目: Clinical Human Disease Networks with Healthcare Administrative Claims Data

报告人简介: 中国人民大学统计学院讲师, 中国人民大学杰出青年学者, 2021 年博士毕业于美国耶鲁大学, 曾就职于耶鲁纽黑文医院临床实效研究中心, 腾讯医疗健康事业部。主要研究方向为网络数据分析、生存分析、复杂数据建模等统计学方法及其在医疗健康、决策预测等领域的应用。研究成果发表在 *Biometrics*, *Statistics in Medicine*, *Annals of Emergency Medicine*, *BMC health services research* 等期刊, 总引用量 300 次以上。

报告摘要: Clinical treatment outcomes are the quality and cost targets that healthcare providers aim to improve. Most existing outcome analysis focuses on a single disease or all diseases combined, which ignores the complex interconnections among diseases. Motivated by the success of molecular and phenotypic human disease networks (HDNs), we develop clinical HDNs that describes the interconnections among diseases in terms of multiple clinical treatment outcomes. In this framework, one node represents one disease, and two nodes are linked with an edge if their outcomes are conditionally dependent. The data experiments validate the performance of the proposed models in identifying correct edges. Analyzing key network properties, such as connectivity, module/hub, and temporal variation, using healthcare administrative claims data, the findings are not only biomedically sensible, but also uncover information that are less/not investigated in the literature. Overall, clinical HDNs can provide additional insight into diseases' properties and their interconnections and assist more efficient disease management and health-care resources allocation.

3. 高慕鸿

报告题目: Learning network-structured dependence from multivariate point process data

报告人简介: 高慕鸿, 中科院数学与系统科学研究院博士后。博士毕业于美国威斯康星大学麦迪逊分校统计系, 本科毕业于北京大学数学科学学院。主要研究方向为神经影像数据分析, 随机过程, 统计学习等。曾在 2023 年获得 ICSA Excellent Junior Researcher Award。已有数篇研究工作发表于国际知名统计学期刊上。主持中国博士后科学基金会引进项目一项。

报告摘要: Learning the sparse network-structured dependence among nodes from multivariate point process data has wide applications in information transmission, social science, and computational neuroscience. This work develops new continuous-time stochastic models of the conditional intensity processes to learn the network structure underlying an array of non-stationary multivariate counting processes. For studying the stochastic mechanism of the proposed model, we introduce a new marked point process for intensity discontinuities, derive the compact representations of their conditional distributions, and demonstrate the cyclicity property driven by recurrence time points. These new theoretical properties further enable us to establish statistical consistency and convergence properties of the proposed penalized M-estimators for graph parameters under mild regularity conditions. Simulation evaluations demonstrate the computational simplicity of the proposed method and its increased estimation accuracy compared to existing methods. Real multiple neuron spike train recordings are analyzed to infer connectivity in neuronal networks.

4. 李晶茂

报告题目: Hierarchical Multi-Label Classification with Gene-Environment Interactions in Disease Modeling

报告人简介: 李晶茂, 厦门大学经济学院统计学与数据科学系 2020 级博士研究生。主要研究领域是高维数据分析、生物统计、机器学习等。研究论文发表在 *Biometrics*, *Annals of the Institute of Statistical Mathematics* 等期刊上。

报告摘要: In biomedical studies, gene-environment (G-E) interactions have been demonstrated to have important implications for disease prognosis beyond the main G and main E effects. Many approaches have been developed for G-E interaction analysis, yielding important findings. However, hierarchical multi-label classification, which provides insightful information on disease outcomes, remains unexplored in G-E analysis literature. Moreover, unlabeled data and the potential selection bias are commonly observed in practical settings but omitted by many existing methods of hierarchical multi-label classification. In this study, we introduce a novel approach for the two-layer hierarchical response with G-E interactions. To address the presence of unlabeled samples and the selection bias problem, we make specific data assumptions and propose a two-step penalized estimation for model estimation. Simulation shows that it has superior accuracy in classification and feature selection performance. The analysis of The Cancer Genome Atlas (TCGA) data on lung cancer demonstrates the practical utility of the proposed approach.

分会场六（上半场）：矩阵数据统计推断

1. 何勇

报告题目: Matrix Kendall's tau in High-dimensions: with Applications to Matrix Factor Model and 2-Dimensional (sparse) Principal Component Analysis

报告人简介: 何勇, 山东大学金融研究院, 教授, 博士生导师, 山东大学齐鲁青年学者, 山东省高等学校“金融

科技数学理论”青年创新团队负责人；山东大学学士（2012），复旦大学博士（2017），师从张新生教授；从事金融计量统计、数理统计以及机器学习等方面的研究，在国际统计学、计量经济学权威期刊 *Journal of Econometrics*, *Journal of Business and Economic Statistics*, *Biometrics* (封面文章), *Biostatistics* 等发表研究论文 30 余篇；现主持国家自然科学基金面上项目以及全国统计科学研究重点项目等。获第一届统计科学技术进步奖一等奖（第二位），担任美国数学评论评论员、中国现场统计研究会生存分析分会副理事长、中国现场统计研究会机器学习分会常务理事及 *JASA*, *JRSSB*, *AOS*, *JOE*, *JBES*, *Biometrics* 等国际学术期刊匿名审稿人。

报告摘要: In this talk, I will introduce a new type of Kendall's tau for robust statistics, names as matrix-type Kendall's tau, which generalize the spatial Kendall's tau (Marden, 1999) in the literature to deal with random matrix elliptical observations. I will elaborate on its use in robust estimation for both factor model and principal eigenvectors (under both sparse and non-sparse settings) in High-dimensions.

2. 虞龙

报告题目: Extreme eigenvalues of sample covariance matrices under generalized elliptical models with applications

报告人简介: 虞龙，现任上海财经大学统计与管理学院助理教授，博士毕业于复旦大学管理学院，曾在新加坡国立大学完成博士后研究工作。研究领域为多元统计分析、高维数理统计、计量经济学、随机矩阵理论等，学术成果发表在 *Biometrika*, *Journal of Econometrics*, *Journal of Business and Economic Statistics* 等。主持 1 项国家自然科学基金青年基金项目 and 1 项上海市浦江人才计划项目。

报告摘要: We consider the extreme eigenvalues of the sample covariance matrix $Q=YY'$ under the generalized elliptical model that $Y=TXD$. Here TT' is a bounded p by p positive definite deterministic matrix representing the population covariance structure, X is a p by n random matrix containing either independent columns sampled from the unit sphere or i.i.d. centered entries, and D is a diagonal random matrix containing i.i.d. entries which are independent of X . Assuming that p and n are comparably large, we prove that the extreme edge eigenvalues of Q can have several types of distributions depending on TT' and D asymptotically, including Gumbel, Fréchet, Weibull, Tracy-Widom, Gaussian and their mixtures. Based on our theoretical results, we consider two important applications. First, we propose some statistics and procedure to detect and estimate the possible spikes for elliptically distributed data. Second, in the context of a factor model, by using the multiplier bootstrap procedure via selecting the weights in D , we propose a new algorithm to infer and estimate the number of factors in the factor model.

3. 吴彬

报告题目: Statistical Inferences for Staleness Factor Models and Robust Volatility Estimation with Large Panel High-Frequency Data

报告人简介: 吴彬，中国科学技术大学管理学院博士后。于 2023 年 6 月取得中国科学技术大学博士学位。主要研究兴趣包括高频数据统计、随机模型和计算金融、风险管理等，其研究成果发表在 *European Journal of Operational Research*, *Probability in the Engineering and Informational Sciences* 等国际期刊上。主持一项校青年创新基金。

报告摘要: Price staleness, defined as zero return for a lack of price adjustment, is the focus of this paper. We propose stale-based approximate factor models estimated from large panel high-frequency data. Asymptotic distributions of the estimated factors and loadings are established when the dimension q and the sample size n simultaneously tend to infinity. Our emphasis is on analyzing the statistical properties of systematic staleness, defined as simultaneous staleness across multiple assets. Based on price staleness, we demonstrate that spot and integrated volatility estimates of prices are biased and provide correction estimators. Numerical experiments are conducted to showcase the accuracy of our estimation method and the good performance of the estimator.

4. 蔡雄

报告题目: Matrix-factor-augmented regression

报告人简介: 蔡雄，南京审计大学讲师、校聘副教授，硕士生导师。博士毕业于北京工业大学。研究领域为非参数/半参数统计、函数型数据分析、金融计量统计。在国际统计学期刊 *Statistica Sinica*, *The Canadian Journal of Statistics*, *ASTIN Bulletin* 等发表研究论文 10 余篇；现主持国家自然科学基金青年项目 1 项。

报告摘要: As matrix-variate observations are increasingly available, to incorporate the interplay between the multi-cross-sections, we introduce a matrix-factor-augmented regression model (M-FARM) that proposes to predict ahead of time with factors of matrix predictors augmented in the regression. We show that the estimation error in the factor matrices, estimated by the projection procedure in the first step, enters into the estimation error of the regression parameters and the prediction error of the response variable with an asymptotically negligible rate. The central limit theorems of the estimates of the regression parameters are established under some mild conditions. Forecasting intervals with a theoretical guarantee are given. Monte-Carlo simulations justify the theory. We find empirically that the augmented matrix factors do help in forecasting macroeconomic variables relative to the benchmark matrix autoregressive model and vector-factor-augmented regression model (V-FARM).

分会场六（下半场）：高维时间序列

1. 马辰辰

报告题目：A Tale of Two Types of Structural Instabilities in High Dimensional Factor Models

报告人简介：马辰辰，北京大学统计科学中心统计学专业 2019 级博士研究生，导师为涂云东教授。她的研究领域为计量经济学，研究兴趣包括高维时间序列、因子模型、结构变点和门限效应的估计以及 Lasso 估计方法等。她的文章曾发表在 Journal of Econometrics 以及《经济管理学报》上。

报告摘要：With the increasing availability of large data sets in economics and finance, the large factor model has become one of the most important tools to achieve dimension reduction in the statistical and econometric analysis. To capture the instability caused by economic condition shifts or policy reforms, factor models with structural breaks in the factor loadings are accordingly developed. On the other hand, recurring regime shifts that relate to higher frequency recurring fluctuation arise in situation where “history repeats”, and is conveniently described by threshold factor models, which allow recurring regime shifts in the factor loadings according to the magnitude of a (continuous) threshold variable. In practice, it is often difficult to decide whether structural break or threshold effect, or both types of instabilities one should employ to portray the observed data. This talk shall discuss how to model each type of instability in factor analysis separately first, and then provide a solution to distinguish the two categories in a model that simultaneously allow both types of structural instabilities. The proposed models are estimated by machine learning techniques such as group Lasso, backward elimination algorithms and information criterion-based model selection methods. The associated asymptotic properties are established and are corroborated by finite sample simulation results and empirical examples.

2. 吴潇然

报告题目：Sparse factor model for high dimensional time series

报告人简介：吴潇然，浙江大学概率论与数理统计专业在读博士生，师从张荣茂教授。本科毕业于浙江大学数学科学学院。主要研究领域为高维时间序列与计量经济学。

报告摘要：Factor models have been extensively employed in high-dimensional time series. However, little is known for the case with sparse loading matrix. This paper introduces a sparse factor model with an easy-to-implement estimation method, aiming to enhance interpretability and relax the constraints on the dimension p of the time series. In particular, it is shown that under weak conditions, the loading space could be consistently estimated with a convergence rate related to the sparseness for each column in the loading matrix and the eigenvalues used to recover the latent factor and loading matrix. In addition, a randomized sequential test is introduced to determine the number of sparse factors. Simulations and real data analysis on sea surface air pressure and stock portfolios are also provided to illustrate the performance of the proposed method.

3. 罗祥夫

报告题目：Adaptive shrinkage estimation for high dimensional change point detection

报告人简介：罗祥夫，厦门大学经济学院统计系数理统计专业 2020 级博士研究生，研究兴趣为变点估计、非参数回归、高维变量选择等。

报告摘要：This paper proposes an adaptive sparse group lasso estimator to detect the change points in a high-dimensional data sequence. It is worth noting that, under the change point setting the construction of consistent adaptive weight is highly non-trivial since many classical assumptions (such as the RE condition and the irrepresentable condition) do not hold, in this paper, we propose a novel procedure (which is named KMax procedure) to construct the adaptive weight and prove its consistency. With the aid of the KMax procedure, The proposed adaptive sparse group lasso estimator could adaptively adjust to different change point patterns (sparse or dense) under a unified framework, and simultaneously estimate the change structure (the exact coordinate where the change occurs in the vector sequence) and the mean vector in each homogeneous regions. Theoretical properties of the proposed estimators are established, The Monte Carlo simulations and a real data application demonstrate that our procedure works well in finite samples.

4. 高尚

报告题目：Conditional Factor Models with High-dimensional Instrument Characteristics: Estimation and Variable Selection

报告人简介：高尚，复旦大学经济学院博士候选人。目前主要研究方向为高维因子模型、时间序列等，已有论文发表于 Econometric Theory 期刊上。曾获国家奖学金。

报告摘要：In this paper, we consider variable (characteristic) selection in a conditional factor model with large-dimensional instrument characteristics. We propose a group LASSO estimation approach to facilitate this selection process and develop an efficient algorithm for estimator computation. We also implement a cross-validation method for optimal regularization

parameter selection. We establish asymptotic properties, such as the consistency and sparsity, of the estimators. The Monte Carlo simulation study demonstrates that the proposed estimators and variable selection procedures are valid in finite samples. Finally, we demonstrate the advantages of the proposed method in predicting asset returns.

分会场七（上半场）：高维统计推断

1. 许凯

报告题目： Testing the Effects of High-Dimensional Covariates via Aggregating Cumulative Covariances

报告人简介： 许凯，安徽师范大学数学与统计学院教师，主要从事复杂非线性相依关系度量及应用的研究，在 Science China Mathematics, AOS, JASA, BKA 等学术期刊上发表研究论文，主持国家自然科学基金项目。

报告摘要： In this article, we test for the effects of high-dimensional covariates on the response. In many applications, different components of covariates usually exhibit various levels of variation, which is ubiquitous in high dimensional data. To simultaneously accommodate such heteroscedasticity and high dimensionality, we propose a novel test based on an aggregation of the marginal cumulative covariances, requiring no prior information on the specific form of regression models. Our proposed test statistic is scale-invariance, tuning free and convenient to implement. The asymptotic normality of the proposed statistic is established under the null hypothesis. We further study the asymptotic relative efficiency of our proposed test with respect to the state-of-art universal tests in two different settings: one is designed for high-dimensional linear model and the other is introduced in a completely model-free setting. A remarkable finding reveals that, thanks to the scale-invariance property, even under the high dimensional linear models, our proposed test is asymptotically much more powerful than existing competitors for the covariates with heterogeneous variances while maintaining high efficiency for the homoscedastic ones.

2. 杨松山

报告题目： New Tests for High-Dimensional Two-sample Mean Problems with Consideration of Correlation Structure

报告人简介： 杨松山，中国人民大学统计与大数据研究院助理教授、博士生导师。研究兴趣包括高维数据分析，模型算法优化，机器学习以及统计模型在金融学、生理学和心理学中的应用。在 JASA、JOE、JCGS 等国际统计学期刊发表十余篇文章。

报告摘要： We propose a test statistic for two sample mean testing problems for high dimensional data by assuming the linear structure on high dimensional precision matrices. A new precision matrix estimation method considering its linear structure is first proposed, and the regularization method is implemented to select the true basis matrices that can further reduce the approximation error. Then the test statistic is constructed by imposing the estimation of the precision matrix. The proposed test is valid for both the low dimensional setting and high dimensional setting even if the dimension of the data is greater than the sample size. The limiting null distributions of the proposed test statistic under both null distribution and alternative distribution are derived. Extensive simulations are conducted for estimating the precision matrix and testing difference of the high dimensional mean vector. Simulation results show that the proposed estimation method enjoy low estimation error for the precision matrix and the regularization method is able to efficiently select the important basis matrix. The testing method performs well compared with the existing methods especially when the elements of the vector have unequal variances. A real data example is then provided to demonstrate the potential of the proposed method in real world applications.

3. 周叶青

报告题目： Rank-based indices for testing independence between two high-dimensional vectors.

报告人简介： 周叶青，同济大学数学科学学院特聘研究员。研究方向为高维数据降维、独立/条件独立检验。研究成果发表在 Annals of Statistics、Journal of the American Statistical Association、Journal of Econometrics、Journal of Business & Economic Statistics 等期刊上。

报告摘要： To test independence between two high-dimensional random vectors, we propose three tests based on the rank-based indices derived from Hoeffding's D, Blum-Kiefer-Rosenblatt's R and Bergsma-Dassios-Yanagimoto's τ^* . Under the null hypothesis of independence, we show that the distributions of the proposed test statistics converge to normal ones if the dimensions diverge arbitrarily with the sample size. We further derive an explicit rate of convergence. Thanks to the monotone transformation-invariant property, these distribution-free tests can be readily used to generally distributed random vectors including heavily tailed ones. We further study the local power of the proposed tests and compare their relative efficiencies with two classic distance covariance/correlation based tests in high dimensional settings. We establish explicit relationships between D, R, τ^* and Pearson's correlation for bivariate normal random variables. The relationships serve as a basis for power comparison. Our theoretical results show that under a Gaussian equicorrelation alternative, (i) the proposed tests are superior to the two classic distance covariance/correlation based tests if the components of random vectors have very different scales; (ii) the asymptotic efficiency of the proposed tests based on D, τ^* and R are sorted in a descending order.

4. 杨伟超

报告题目: High-dimensional testing based on quadratic norm statistics

报告人简介: 杨伟超, 北京师范大学统计学院应用统计专业直博四年级学生, 主要研究方向是高维假设检验, 中介效应检验等, 目前文章被 *Statistica Neerlandica*, *Communications in Mathematics and Statistics*, 南开管理评论等期刊接收。

报告摘要: Testing hypothesis for high-dimensional regression coefficients is of fundamental importance in the statistical theory and applications. In this session, we introduce some developments for testing the significance of high-dimensional coefficients in high-dimensional regression models based on quadratic norm statistics. First, we investigate 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. Second, we investigate the testing of regression coefficients within high-dimensional generalized linear models featuring general covariance structures. Third, we consider tests for indirect effect in the linear mediation models when both mediators and exposures are high-dimensional. Simulations evaluate the finite-sample performances of our introduced procedures, and real data analyses illustrate their application.

分会场七（下半场）：无模型假设的统计推断

1. 梁德才

报告题目: Tests of Partial/Weak Separability for Complex Functional Data

报告人简介: 梁德才, 南开大学统计科学学院助理教授, 北京大学统计学博士, 主要研究领域为函数型数据分析、时空统计等。在统计学期刊 *Journal of the American Statistical Association*, *Statistica Sinica* 等发表多篇论文。入选中国科协青年人才托举工程, 主持 1 项国家自然科学基金青年项目。

报告摘要: For complex functional data, it is quite challenging to model the cross-covariance structure which consists of dual aspects of multivariate and functional features. To simplify the cross-covariance analysis, the assumption of partial/weak separability is widely used to decompose the data into an additive form of functional components and multivariate or spatially correlated random coefficients. In this work, we propose hypothesis testing procedures to examine the validity of partial/weak separability. We study the asymptotic properties of the different norms of the test statistic, resulting in a chi-square type mixture test and a high-dimensional test which are suitable to finite- or high-dimensional multivariate functional data with diverse correlation structures. We assess the empirical performance of the proposed tests through simulation studies for several types of complex functional data, followed by corresponding real examples: multichannel tonnage data, electroencephalography data and China PM2.5 data.

2. 任好洁

报告题目: Online multiple changepoints detection with false discovery rate control

报告人简介: 任好洁是上海交通大学数学科学学院长聘教轨副教授, 18 年博士毕业于南开大学, 随后在宾州州立大学从事博士后研究。她的研究方向包括统计异常探查、在线学习与监控、高维数据推断等。在 *JASA*, *Biometrika*, *NIPS* 等杂志上发表学术论文 10 余篇。入选中国科协青年人才托举工程, 主持 1 项国家自然科学基金青年项目。

报告摘要: Technological advances have led to the emergence of an increasing number of applications requiring analysis of datastreams, that are characterized by an indefinitely long and time-evolving sequence. In such applications, the status of a stream can alternate, possibly many times, between a regular status and an irregular status. Consequently, it is necessary to develop statistical methodologies that constantly detect multiple changepoints in an online manner. While we may employ conventional methods of sequential change detection to trigger signals after the change occurs, no online procedure is available to quantify the uncertainty of the detected changes. In this work, we fill this gap by framing the online multiple changepoints detection into an online multiple testing problem and proposing a new framework to test the null hypothesis that there is no change between neighboring signaled points. To obtain valid p-values for online multiple testing, we propose a data-fission-based procedure that is a simple yet effective way of dealing with the post-detection uncertainty quantification. It is shown that popular online false discovery rate (FDR) control methods with those p-values can achieve finite-sample FDR control. The advantage of the proposed method is demonstrated via simulation studies and a data example.

3. 王光辉

报告题目: Reliever: Relieving the Burden of Costly Model Fits for Changepoint Detection

报告人简介: 王光辉, 华东师范大学统计学院副教授。2018 年博士毕业于南开大学, 2018-2021 年曾工作于南开大学统计与数据科学学院。研究兴趣包括变点检测和高维数据推断等。在 *AOS*, *JRSSB*, *Statistica Sinica*, *Technometrics* 和 *Journal of Quality Technology* 等国际统计学期刊上发表论文十余篇。

报告摘要: We propose a general methodology Reliever for fast and reliable changepoint detection when the model fitting is costly. Instead of fitting a sequence of models for each potential search interval, Reliever employs a substantially reduced number of proxy/relief models that are trained on a predetermined set of intervals. This approach can be seamlessly integrated with state-of-the-art changepoint search algorithms. In the context of high dimensional regression models with changepoints, we establish that the Reliever, when combined with an optimal search scheme, achieves estimators for both the changepoints and corresponding regression coefficients that attain optimal rates of convergence, up to a logarithmic factor. Through extensive numerical studies, we showcase the ability of Reliever to rapidly and accurately detect changes across a diverse range of parametric and nonparametric changepoint models.

4. 严晓东

报告题目: 策略统计学习

报告人简介: 严晓东, 山东省高等学校优秀青年创新团队负责人, 山东大学未来学者, 副研究员, 博士生导师, 香港理工大学研究员 (Research Fellow), 加拿大阿尔伯塔大学博士后研究员, 云南大学与香港理工大学联合培养博士, 全国工业统计学教学研究会理事, 中关村软联智能算法委员会秘书长, 中国现场统计研究会多元分析应用专业委员会常务理事。在统计学顶级期刊 JRSSB, AOS, JASA 计量经济顶刊 JOE 以及人工智能顶级会议 AAAI 等发表论文 30 余篇, 荣获山东省大数据研究会“优秀青年”称号和“云南省 2020 年优秀博士论文”奖, 以主持人获得了国自科面上和青年基金, 科技部重点研发 (项目骨干) 和国家统计局等项目资助。

报告摘要: 非线性期望是中国本土开辟的原创性研究方向, 对各个领域的科学研究越来越重要, 尤其是大数据和人工智能的兴起, 为非线性期望创新理论与应用研究提供了更强劲的动力。最近, 团队基于强化学习最简单模型-多臂老虎机模型, 开创了“策略极限理论”, 是非线性概率理论与强化学习的重大突破交叉研究成果, 变革了传统统计方法研究范式, 后续开展的相关研究在大数据采样、实验设计、迁移学习、在线学习和元学习等可解释和可信赖的统计理论与方法研究上均取得重大突破。

分会场八 (上半场): 案例教学专场

1. 朱雪宁

报告题目: 从《红楼梦》文本看红楼梦作者之谜

报告人简介: 朱雪宁, 复旦大学大数据学院副教授, 博士生导师。

报告摘要: 作为四大名著之一, 红楼梦具有极高的文学价值。红楼梦全书共 120 回, 前 80 回比较公认的作者是曹雪芹, 而后 40 回至今作者归属仍是谜团。本案例通过分析前后语言特征的变化, 从统计角度看后 40 回的作者归属。此次案例研讨主要讲解该案例包括的文本分析、统计分析、数据挖掘等相关知识点。

2. 王菲菲

报告题目: 案例教学模式探索: 以时间序列分析为例

报告人简介: 王菲菲, 中国人民大学统计学院副教授, 博士生导师。研究上关注文本分析、网络数据分析等, 研究论文发表于 Journal of Econometrics, Journal of Business and Econometric Statistics, Journal of Machine Learning Research, 中国科学 (数学) 等国内外高水平期刊上。主持国家自然科学基金青年和面上项目、全国统计科学研究重大项目等。曾获中国人民大学优秀科研成果奖、课外教学优秀奖等。

报告摘要: 案例教学作为一种重要的教学方法, 能够帮助学生更好地理解理论知识、培养解决问题的能力, 并将理论与实践相结合。本次报告将探讨案例教学模式在教学实践中的应用方式, 并以《时间序列分析》课程为例介绍基于案例教学的课程建设经验。我们将分享融合案例的几种方式, 以展示案例教学在教学过程中的重要作用。

3. 常莹

报告题目: 数据科学人才实践能力的培养: 本科阶段实践类教学活动的设计与实施

报告人简介: 常莹, 狗熊会在线实习项目负责人、西安欧亚学院特聘教师。专注互联网数据的分析与价值实现, 曾服务于阿里巴巴集团、美丽联合集团、宝宝树等企业。合著有《数据领导力: 人工智能时代数据化转型的关键路径》《Excel 商务数据分析与应用》。

报告摘要: 数据科学相关专业的学生在毕业后从事的工作既要求他们具备达标的专业技能, 也要求他们具备理解和解决领域实际问题的能力, 两者共同构成了学生职场入门和长期发展的关键保障。其中, 解决实际问题能力的养成需要学生在校期间就能看见、理解、模仿大量实际问题的定义和解决过程, 即, 以培养将解决问题的思路和方法系统地融入本科阶段四年的教学。本次报告展示西安欧亚学院大数据专业在本科阶段培养学生实践

能力的整体安排，包括各学年实践能力培养的侧重点、各阶段相关教学活动的协调与整合、具体教学活动的设计。

4. 黄丹阳

报告题目：研究生商务大数据案例分析课程教学

报告人简介：黄丹阳，中国人民大学统计学院教授，博士生导师。主持国家自然科学基金面上项目，北京市社会科学基金重点项目等多项省部级及以上课题，入选北京市科协青年人才托举工程，曾获北京市优秀人才培养资助。长期从事复杂网络建模、超高维数据分析、分布式计算等方向的理论研究，以及统计理论研究在中小微企业信用风险评估，企业数字化发展中的应用研究。获北京市青年教师教学基本功大赛二等奖，中国人民大学教学标兵等多项教学奖励，多次获得中国人民大学优秀科研成果奖等科研奖项。在 JRSSB, JOE, JBES 等国内外权威期刊发表论文 30 余篇，著有教材一部，独立作者专著一部。

报告摘要：随着数字经济的蓬勃发展，培养兼具商业分析思维和信息化实践能力的复合人才已成为现代教育领域的一项重要课题。针对这样的需求，商务大数据案例分析课程在传统案例分析的基础上融入更多一手数据分析实践经验，通过课堂讲授和小组作业双线推进的方式，让学生体验从数据获取、清洗到分析建模、得出决策支持的全流程。通过教学积累和迭代，课程累积了丰富的案例，总结出了多样化的教学方法，并在产学研结合等方向探索商业数据分析课程的改进方向。

分会场八（下半场）：青年学者成长论坛

1. 黄丹阳

简介：黄丹阳，中国人民大学统计学院教授，博士生导师。2015 年毕业于北京大学光华管理学院。研究兴趣包括复杂网络建模、超高维数据分析、分布式计算等，注重统计理论研究在小微企业数字化发展中的实际应用。在 JRSSB, JoE, JBES 等国内外权威期刊发表论文 30 余篇，著有教材一部，独立作者专著一部。

报告摘要：

2. 杨松山

简介：杨松山，中国人民大学统计与大数据研究院助理教授、博士生导师。2018 年毕业于美国宾夕法尼亚州立大学，获统计学博士学位。2018 年至 2021 年在美国从事对冲基金量化研究工作。研究兴趣包括高维数据分析，模型算法优化，机器学习以及统计模型在金融学、生理学和心理学中的应用。在 JASA, JOE, JCGS 等国际统计学期刊发表论文十余篇。

3. 邱怡轩

简介：邱怡轩，上海财经大学统计与管理学院副教授，博士毕业于普渡大学统计系，毕业后曾于卡内基梅隆大学担任博士后研究员。主要研究方向包括深度学习、生成式模型和大规模统计计算与优化等，科研成果发表在统计学国际权威期刊及机器学习顶级会议上，是众多开源软件（如 Spectra、LBFSGS++、ReHLIne、showtext、prettydoc 等）的开发者与维护者。

4. 邓婉璐

简介：邓婉璐，清华大学统计学研究中心副教授，本科、博士毕业于北京大学数学学院，曾于宾西法尼亚大学学习。主要研究方向为因果推断、贝叶斯网络等。主要讲授《初等概率论》、《因果推断导论》等课程，曾获清华大学清韵烛光奖、多次年度教学优秀奖，负责的《初等概率论》课程获清华大学标杆课、毕业生心目中的好课程、思政示范课等荣誉。

六、相关机构简介

全国工业统计学教学研究会青年统计学家协会

全国工业统计学教学研究会青年统计学家协会成立于 2019 年 4 月，是在全国工业统计学教学研究会领导下的、旨在促进我国统计学发展和青年统计学者的学术交流的社会团体。协会自批准成立以来，已经吸纳国内 200 余家高校成为理事单位，覆盖国内几乎所有的具有统计学（或者相关学科）的高校，是聚集青年数据科学人才的优质学术平台。

江苏师范大学

江苏师范大学是江苏省人民政府和教育部共建高校，是江苏高水平大学建设高校。学校现有 59 个本科招生专业，32 个一级学科硕士点，20 个硕士专业学位类别，1 个服务国家特殊需求博士人才培养项目，1 个博士后科研流动站，覆盖 11 个学科门类。学校现有 2 个学科入榜 2022 “软科世界一流学科排名”，10 个学科入榜 2023 “软科中国最好学科排名”，其中，中国语言文学进入 2023 “软科中国最好学科排名” 前 12%。化学、工程学、材料科学、数学、社会科学总论、计算机科学、植物与动物科学 7 个学科进入 ESI 排名全球前 1%；8 个学科入选省优势学科四期建设工程，10 个学科入选“十四五”省重点学科。近五年来，获批国家级科研项目 397 项，其中国家社会科学基金项目 158 项（重大项目 11 项、领军人才项目 2 项、重点项目 10 项），国家自然科学基金项目 231 项（重大项目 1 项、重点项目 3 项、国家杰青 2 项，国家优青 5 项），科技部重点研发计划项目 2 项。

狗熊会

狗熊会致力于成为数据产业的高端智库，以“聚数据英才，助产业振兴”为使命。狗熊会关注数据科学基础教育，通过生产优质的数据科学教育内容，提供卓越的研究、实践、就业机会，帮助相关专业的老师、同学、以及从业者们，充分享受数据分析的快乐，促进个人职业的终身幸福与成长。