



贝叶斯统计与人工智能 前沿论坛

会议手册

主办：中国现场统计研究会贝叶斯统计分会

承办：复旦大学管理学院

上海·2026年8月3日-5日



目 录

论坛介绍	1
论坛日程	2
分组报告详细安排	4
特邀主旨报告嘉宾简介	16
分组报告摘要	18
承办单位简介	65
交通参考	71



中国现场统计研究会贝叶斯统计分会

贝叶斯统计与人工智能前沿论坛

中国现场统计研究会贝叶斯统计分会拟于 2026 年 8 月 3 日-5 日在上海举办“贝叶斯统计与人工智能前沿论坛”。论坛旨在为全国贝叶斯统计与人工智能相关领域的学者、科研人员、业界专家与研究生提供高水平学术交流平台，共同探讨贝叶斯统计理论与方法的前沿进展，及其与人工智能技术的交叉成果。本次论坛邀请到南方科技大学孙建国教授、西安交通大学孙剑教授做论坛主旨报告，并组织分会学术报告等多种形式的学术活动。

一、主办单位

主办单位：中国现场统计研究会贝叶斯统计分会

承办单位：复旦大学管理学院

二、论坛内容

（一）特邀论坛主旨报告

（二）分组学术报告

三、论坛时间地点

报到时间：2026 年 8 月 3 日，全天注册报到

报到地点：复旦大学管理学院政立路院区 A 楼（上海市杨浦区政立路 558 号）

论坛时间：2026 年 8 月 4 日至 8 月 5 日

论坛地点：复旦大学管理学院政立路院区（上海市杨浦区政立路 558 号）

四、论坛日程

8 月 3 日注册报到		
地点：复旦大学管理学院政立院区 A 楼 1 楼		
时间	日程安排	
10:00-20:00	注册报到	
16:00-20:00	晚餐	
8 月 4 日上午：主题论坛		
地点：复旦大学管理学院政立院区 A 楼 B1 报告厅		
08:00-08:30	注册报到	
开幕式		
08:30-09:00	主持人介绍参会嘉宾	主持人：郁 文 教授 复旦大学
	复旦大学管理学院院长张付强教授致辞	
	中国现场统计研究会副理事长艾明要教授致辞	
	云南大学副校长唐年胜教授致辞	
	论坛合影	
09:00-10:00	主旨报告 1：Transfer Learning Estimation for Regression Analysis of Failure Time Data 演讲人：孙建国 教授 南方科技大学	主持人：朱仲义 教授 复旦大学
10:00-10:20	茶歇	
10:20-11:20	主旨报告 2：生成式人工智能：数学基础与交叉应用 演讲人：孙 剑 教授 西安交通大学	主持人：夏 寅 教授 复旦大学
11:20-13:30	午餐	A 楼 B1 大食代
8 月 4 日下午：分组报告		
时间	议程安排	地点
13:30-15:10	分组报告 1 上半场 主持人：郦泽雨 东南大学	A203 教室
	分组报告 2 上半场 主持人：魏颖颖 香港中文大学	A206 教室
	分组报告 3 上半场 主持人：朱文圣 东北师范大学	A208 教室
	分组报告 4 上半场 主持人：朱雪宁 复旦大学	A212 教室
	分组报告 5 上半场 主持人：冯项楠 复旦大学	A213 教室

	分组报告 6 上半场 主持人：吴 尚 复旦大学	A305 教室
15:10-15:30	茶歇	
15:30-17:10	分组报告 1 下半场 主持人：王纯杰 长春工业大学	A203 教室
	分组报告 2 下半场 主持人：刘 旭 上海财经大学	A206 教室
	分组报告 3 下半场 主持人：林晓蕾 复旦大学	A208 教室
	分组报告 4 下半场 主持人：刘 彬 复旦大学	A212 教室
	分组报告 5 下半场 主持人：戴国榕 复旦大学	A213 教室
	分组报告 6 下半场 主持人：章 玮 复旦大学	A305 教室
18:00-20:00	晚宴	C 楼 7 楼
8 月 5 日上午：分组报告		
08:30-10:10	分组报告 7 上半场 主持人：马 莉 厦门大学	A203 教室
	分组报告 8 上半场 主持人：包亚杰 复旦大学	A206 教室
	分组报告 9 上半场 主持人：陈志勇 福建师范大学	A208 教室
	分组报告 10 上半场 主持人：李 婷 上海财经大学	A212 教室
	分组报告 11 上半场 主持人：周建军 云南大学	A213 教室
	分组报告 12 主持人：时 琛 复旦大学	A305 教室
10:10-10:30	茶歇	
10:30-12:10	分组报告 7 下半场 主持人：马宇恒 华东师范大学	A203 教室
	分组报告 8 下半场 主持人：夏 寅 复旦大学	A206 教室
	分组报告 9 下半场 主持人：方冠华 复旦大学	A208 教室
	分组报告 10 下半场 主持人：邵凌轩 复旦大学	A212 教室
	分组报告 11 下半场 主持人：沈 娟 复旦大学	A213 教室
12:10-13:30	午餐	A 楼 B1 大食代

分组报告详细安排：

8月4日分组报告1（A203教室）

时间	报告人	单位	报告题目
上半场 主持人：郦泽雨 东南大学			
13:30-13:55	张立文	上海财经大学	BizFinBench.v2: A Unified Dual-Mode Bilingual Benchmark for Expert-Level Financial Capability Alignment.
13:55-14:20	周 灿	南京审计大学	Learning Semi-Parametric Tree Models from Mixed Data
14:20-14:45	毛俊凡	东南大学	Generating Tensor Data by Diffusion Models with Tucker U-net
14:45-15:10	郦泽雨	东南大学	Distributional Counterfactual Inference by Conditional Tensor Diffusion Model
15:10-15:30	茶歇		
下半场 主持人：王纯杰 长春工业大学			
15:30-15:55	陈雪蓉	西南财经大学	A General Distributed Privacy-Preserving Statistical Inference Framework for Massive Data
15:55-16:20	杨 凯	长春工业大学	Bayesian Forecasting for a Logistic Mixture Double Autoregressive Model
16:20-16:45	杜明月	吉林大学	Regression Analysis of Interval-Censored Failure Time Data with Change Points and a Cured Subgroup
16:45-17:10	刘 蓉	长春工业大学	Statistical Inference for Interval-Censored Data for Linear Transformation Models within Ordinary Differential Equation Framework

8月4日分组报告2（A206 教室）

时间	报告人	单位	报告题目
上半场 主持人：魏颖颖 香港中文大学			
13:30-13:55	吴梦云	上海财经大学	Sparse Bayesian Deep Functional Learning with Structured Region Selection
13:55-14:20	唐 荣	香港科技大学	Subsampled Pseudo-posteriors for Scalable Bayesian Moment-condition Inference
14:20-14:45	葛淑菲	上海科技大学	A Bayesian Approach to Edge-Based Brain Connectivity Analysis
14:45-15:10	魏颖颖	香港中文大学	Decontamination of Transcript Leakage in Spatial Transcriptomic Data
15:10-15:30	茶歇		
下半场 主持人：刘 旭 上海财经大学			
15:30-15:55	杨 莹	复旦大学	Online Sparse Regression with Expanding Observables
15:55-16:20	周叶青	同济大学	Power Enhancement for Test of Multivariate Independence
16:20-16:45	曹智苗	山东大学	Decoding AI Precision Medicine Paradox: A Fairness-Aware Optimal Transport
16:45-17:10	严晓东	西安交通大学	AI Safe: Statistical Methods for Security Protection from Small to Large Models

8月4日分组报告3 (A208 教室)

时间	报告人	单位	报告题目
上半场 主持人：朱文圣 东北师范大学			
13:30-13:55	谌自奇	华东师范大学	Conditionally Whitenened Generative Models for Probabilistic Time Series Forecasting
13:55-14:20	刘鹤飞	云南财经大学	Bayesian Fréchet Regression
14:20-14:45	陆 婧	东北师范大学	A Robust Bootstrap-CUSUM Method for Detecting Test Speededness and Pinpointing Changepoints in Computerized Testing
14:45-15:10	崔婷婷	浙江财经大学	A Bayesian Variable Selection Approach for Dynamic Treatment Regimes
15:10-15:30	茶歇		
下半场 主持人：林晓蕾 复旦大学			
15:30-15:55	牟荣吉	上海交通大学	A Bayesian Phase I/II PK-integrated Regimen Optimization Design for Survival Efficacy and Late-onset Toxicity
15:55-16:20	陈中媛	南京大学	Data-Driven Statistical Methods for Cancer Treatment Recommendations and Causal Anti-Cancer Drug Sensitive Genes Selection
16:20-16:45	童培峰	北京大学	Residual Covariance Unlocks Robust Mapping of Dynamic Causal Brain Network
16:45-17:10	施润叶	复旦大学	A Bayesian Clustering Framework for Decoding Heterogeneous Genetic Architectures

8月4日分组报告4（A212教室）

时间	报告人	单位	报告题目
上半场 主持人：朱雪宁 复旦大学			
13:30-13:55	毛晓军	上海交通大学	Fast and Flexible Covariance Estimation for Diffusion Posterior Sampling in Imaging Inverse Problems
13:55-14:20	郁淼淼	华东师范大学	Statistical Inference for A General Federated Learning Framework under Local Differential Privacy
14:20-14:45	张心雨	华东师范大学	Two-way Matrix Autoregressive Model with Thresholds
14:45-15:10	滕佳烨	上海财经大学	AutoSTAT: 全自动数据分析助手
15:10-15:30	茶歇		
下半场 主持人：刘 彬 复旦大学			
15:30-15:55	虞龙	上海财经大学	High-dimensional Factor and Clustering Analysis via Projected and Truncated Sample Covariance Matrix
15:55-16:20	付盛	南开大学	Multinomial Logistic Regression: Old and New
16:20-16:45	任明旸	上海交通大学	Model-agnostic Information Transfer for Classification with Label Noise
16:45-17:10	骆钰澐	上海财经大学	Rate-Optimal Online Learning for Dynamic Assortment Selection with Positioning

8月4日分组报告5（A213教室）

时间	报告人	单位	报告题目
上半场 主持人：冯项楠 复旦大学			
13:30-13:55	潘文亮	中国科学院 数学与系统科学 研究院	Ball Impurity: Measuring Heterogeneity in General Metric Spaces
13:55-14:20	李 睿	上海对外经贸 大学	Quantile Regression and Homogeneity Identification of a Semiparametric Panel Data Model
14:20-14:45	王超杰	江苏大学	STransfer: A Transfer Learning-Enhanced GCN for Spatial Transcriptomics Data
14:45-15:10	涂霁原	上海财经大学	Robust Variational Bayes by Min-Max Median Aggregation
15:10-15:30	茶歇		
下半场 主持人：戴国榕 复旦大学			
15:30-15:55	刘 林	上海交通大学	A Common Honest Mistake in Semiparametric Theory
15:55-16:20	马慧娟	华东师范大学	Efficient Estimation of Quantile Treatment Effects by Incorporating External Controls
16:20-16:45	郑潇逸	中国科技大学	Fairness Evaluation for Risk Prediction Models under Heterogeneous Risk Distributions
16:45-17:10	周慧娟	上海财经大学	BMDD: A Probabilistic Framework for Accurate Imputation of Zero-inflated Microbiome Sequencing Data

8月4日分组报告6（A305教室）

时间	报告人	单位	报告题目
上半场 主持人：吴 尚 复旦大学			
13:30-13:55	郭 旭	北京师范大学	MATCH: Multiplier-Assisted Tests for Conditional Hypotheses in Non-Euclidean Data
13:55-14:20	李树威	广州大学	A Factor-Augmented Regression Approach to Learning Bivariate Interval-Censored Data
14:20-14:45	谭承利	西安交通大学	Emergent Short-range Memory in Stochastic Gradient Noise and Its Implications on Generalization
14:45-15:10	刘弘之	上海对外经贸大学	Asymptotic Properties of Adaptive Quantum Phase Estimation
15:10-15:30	茶歇		
下半场 主持人：章 玮 复旦大学			
15:30-15:55	贺 莘	上海财经大学	Transfer Learning for Kernel-based Regression
15:55-16:20	张宇谦	中国人民大学	Quadruply Robust Methods for Causal Mediation Analysis
16:20-16:45	陈自元	北京大学	Deconfounding via Profiled Transfer Learning
16:45-17:10	贾骏雄	西安交通大学	Nonparametric Prior Learning in Differential Equation Modeling

8月5日分组报告7（A203教室）

时间	报告人	单位	报告题目
上半场 主持人：马 莉 厦门大学			
08:30-08:55	段志斌	西安交通大学	结构化贝叶斯建模：从可信表征学习到大模型稳健对齐
08:55-09:20	张龄月	东北财经大学	ATM: An Aggregation Test of Moments Approach for Assessing High-Dimensional Normality
09:20-09:45	夏新涛	浙江大学	Anytime-Valid Inference for Online Ranking of Large Language Models
09:45-10:10	何凤翔	爱丁堡大学	Bayesian Inference Forgetting
10:10-10:30	茶歇		
下半场 主持人：马宇恒 华东师范大学			
10:30-10:55	龚 欢	国防科技大学	Robust and Efficient Inference for GARCH Models with Heavy-Tailed and Asymmetric Innovations
10:55-11:20	马宇恒	华东师范大学	Locally Private Estimation with Public Features
11:20-11:45	江 声	港中文深圳	Errors-in-variables Gaussian Processes for Mixed-input Regression
11:45-12:10	汪时嘉	上海科技大学	A Multifidelity Approximate Bayesian Computation with Pre-Filtering

8月5日分组报告8 (A206 教室)

时间	报告人	单位	报告题目
上半场 主持人：包亚杰 复旦大学			
08:30-08:55	吴未迟	清华大学	Bayesian Online Collective Anomaly and Change Point Detection in Fine-grained Time Series
08:55-09:20	田 焯	浙江大学	Conformalized Multiple Testing under Unknown Null Distribution with Symmetric Errors: An Empirical Bayes Perspective
09:20-09:45	张智涵	华东师范大学	Nonparametric Statistical Inference for Permutation-Based Models: Optimality and Adaptivity
09:45-10:10	包亚杰	复旦大学	Tilted Conformal Prediction for Coverage Parity Control
10:10-10:30	茶歇		
下半场 主持人：夏 寅 复旦大学			
10:30-10:55	冯 龙	南开大学	High Dimensional Alpha Test for Linear Factor Pricing Model with L_q -norm
10:55-11:20	李 赛	清华大学	Efficient Machine Unlearning with Minimax Optimality
11:20-11:45	李易诚	华东师范大学	Minimax Lower Bounds in Differential Privacy
11:45-12:10	李竞阳	复旦大学	Online Tensor Learning: Computational and Statistical Trade-offs, Adaptivity and Optimal Regret

8月5日分组报告9（A208教室）

时间	报告人	单位	报告题目
上半场 主持人： 陈志勇 福建师范大学			
08:30-08:55	饶昊宸	福建师范大学	Renewable High-dimensional Expected Shortfall Regression
08:55-09:20	李铂桂	上海海关学院	具有可分时空过滤器的固定效应半参数单指标面板模型的估计和检验
09:20-09:45	李 芬	福建农林大学	Penalized Quadratic Inference Functions Estimation for Fixed Effects Partially Linear Single Index Spatial Error Model
09:45-10:10	陈志勇	福建师范大学	Latent Single-Index Models with Factor Structure for Multivariate Ordinal Data
10:10-10:30	茶歇		
下半场 主持人：方冠华 复旦大学			
10:30-10:55	张思亮	华东师范大学	A Cumulative Ordered Spike-and-Slab Prior for Adaptive Dimension Selection in Joint Latent Space Models
10:55-11:20	郭金鑫	中央民族大学	基于竞争风险模型的中间跳过项目缺失与作答反应联合建模
11:20-11:45	许 鑫	中央民族大学	协变量和层级学习轨迹的高阶隐马尔可夫模型的多值扩展
11:45-12:10	欧阳靖	香港大学	Statistical Analysis of Large-scale Item Response Data under Measurement Non-invariance

8月5日分组报告10（A212教室）

时间	报告人	单位	报告题目
上半场 主持人：李 婷 上海财经大学			
08:30-08:55	柏 杨	上海财经大学	Functional Graph Inference via Residual-Based Distance Covariance
08:55-09:20	李 婷	上海财经大学	Efficient Online LLM Watermark Detection via Rao–Blackwellized E-Processes
09:20-09:45	裘予琦	华东师范大学	Budgeted Dynamic Monitoring for Timely Actionable Detection Under Action-Induced Observation
09:45-10:10	张颖杰	华东师范大学	Nodewise Loreg: Nodewise L_0 Regression for High-dimensional Sparse Precision Matrix Estimation
10:10-10:30	茶歇		
下半场 主持人：邵凌轩 复旦大学			
10:30-10:55	伍书缘	上海财经大学	Bias-Correction for Privacy-Protected Spatial Autoregressive Models with Application to Restaurant Network Analysis
10:55-11:20	田琳琳	东华大学	Optimal Investment Problem in A Renewal Risk Model with Generalized Erlang Distributed Interarrival Times
11:20-11:45	赵花丽	华中科技大学	Generalizing Transfer Learning: A Flexible Doubly Robust Estimation Approach for Missing Data
11:45-12:10	张国宇	北京大学	A Nonparametric Potential Outcomes Framework for Bidirectional Causal Inference

8月5日分组报告11（A213教室）

时间	报告人	单位	报告题目
上半场 主持人：周建军 云南大学			
08:30-08:55	徐礼柏	苏州大学	Score-based Generative Models with Gamma Processes
08:55-09:20	易凤婷	云南大学	Variational Bayesian Estimation for Joint Models of Longitudinal and Interval-Censored Fairure Time Data with Random Change-Points
09:20-09:45	赵 波	长春工业大学	Regularization Copula Link-based Semiparametric Regression Models with Bivariate Interval-censored Data
09:45-10:10	袁 泉	云南大学	Strongly Consistent Community Detection in Popularity Adjusted Block Models
10:10-10:30	茶歇		
下半场 主持人：沈 娟 复旦大学			
10:30-10:55	朱蔚萱	厦门大学	DAG Trend Filtering for Genomic Denoising via Higher-Order Bayesian Networks and DAG Shrinkage Processes
10:55-11:20	张艺馨	中国科学技术大学	Merits and Challenges of Plasma Proteomics on Association Replicability
11:20-11:45	陈奕帆	香港浸会大学	A Recipe for Causal Graph Regression: Confounding Effects Revisited
11:45-12:10	张如谦	清华大学	Bayesian Environment Invariant Regression

8 月 5 分组报告 12 (A305 教室)

时间	报告人	单位	报告题目
主持人：时 琛 复旦大学			
08:30-08:50	王川杰	上海财经大学	Satellite-Informed Extreme Climate Risks and Volatility in Crop Futures
08:50-09:10	熊子扬	复旦大学	Scalable Bayesian Calibration of High-Resolution Multi-Output Building Energy Models
09:10-09:30	毛若轩	复旦大学	Fast Bayesian Inference on Spatial Blind Source Separation with the Relaxed Threshold Gaussian Process
09:30-09:50	时 琛	复旦大学	Valuing Sustainability in China
09:50-10:10	罗嘉雯	华南理工大学	Forecasting Downside Betas with Machine Learning Models
10:10-10:30	茶歇		

五、特邀主旨报告嘉宾简介



特邀主旨报告 1

报告人：孙建国 教授 南方科技大学

报告题目：Transfer Learning Estimation for Regression Analysis of Failure Time Data

个人简介：孙建国教授现任南方科技大学讲席教授，长期致力于生物统计、基因数据分析、复杂数据建模方法及其在临床医学与社会科学领域的交叉应用研究。他在 Journal of the American Statistical Association (JASA), Journal of the Royal Statistical Society: Series B (JRSSB), Biometrika, Annals of Statistics, Biometrics, Statistica Sinica 及 Statistics in Medicine 等国际顶级统计期刊发表论文 300 余篇，并出版统计学著作 4 部。他是美国统计学会 (ASA) Fellow、国际数理统计学会 (IMS) Fellow, 及国际统计学会 (ISI) Elected Member。曾担任美国统计学会生存分析分会主席 (ASA Lifetime Data Science Section)、国际泛华统计协会 (ICSA) 主席等多个重要学术组织职务。目前或曾经担任包括 JASA, Statistics in Medicine, Lifetime Data Analysis, The International Journal of Biostatistics, Computational Statistics & Data Analysis 等国际顶级统计期刊副主编，现在担任 Statistics in Biosciences 主编。

报告摘要：Transfer learning has recently attracted a great deal of attention as a powerful tool to borrow information from existing or source studies to enhance the analysis of a current or target study and correspondingly, many methods have been developed under various situations. In this talk, we will discuss transfer learning estimation for regression analysis of failure time data and present some new estimation methods.



特邀主旨报告 2

报告人：孙剑 教授 西安交通大学

报告题目：生成式人工智能：数学基础与交叉应用

个人简介：孙剑，西安交通大学数学与统计学院副院长、教授，西安数学与数学技术研究院常务副院长。担任教育部科技委委员，人工智能领域顶级国际期刊 IEEE TPAMI 和 IJCV 编委、ICCV/ECCV/ICLR/CVPR 等领域主席，曾在微软亚洲研究院、法国巴黎高师、法国国家信息与自动化研究院等从事博士后或访问学者工作。长期从事人工智能(尤其是图像和医学影像分析)中的数学模型与算法研究，主要包括成像数学反问题与医学辅助诊断、人工智能生成与泛化性问题的基础模型与算法研究等，提出模型驱动深度学习理论及其代表算法 ADMM-Net。成果发表于 IEEE Trans. PAMI, NSR, SIAM J. Imaging Sci., NeurIPS, CVPR 等交叉领域顶级期刊和会议；获得国家杰出青年科学基金，陕西省自然科学一等奖(第一完成人)、国家自然科学二等奖(第二完成人)，承担国家重点研发计划项目、基金委企业联合基金重点项目等。

报告摘要：生成式人工智能是当前通用人工智能发展的重要方向，主要通过设计人工智能算法实现对多模态、高维复杂样本分布的学习与新样本的生成，是当前人工智能应用于自动问答、跨模态生成、AI for science 等方法基础。生成式人工智能的底层基础是数学与统计学，本报告将介绍生成式人工智能的发展背景、数理原理及面临的挑战，重点阐述基于最优传输的可控与条件生成方法，并进一步介绍多模态与动态生成的基本方法及其在医学影像生成、多模态图文对齐、分子结构生成等任务中的应用。最后，报告总结并展望生成式 AI 未来发展前景，着重关注数学方法驱动的生成式人工智能在科学研究与交叉应用中的潜力。

六、分组报告摘要：

报告人：张立文 上海财经大学

报告题目：BizFinBench. v2: A Unified Dual-Mode Bilingual Benchmark for Expert-Level Financial Capability Alignment

报告摘要：Large language models have become an increasingly important technological foundation for intelligent financial services. However, existing financial benchmarks largely rely on generic or simulated data and focus primarily on static offline tasks, making it difficult to accurately assess model capabilities in real-world business environments. To address this limitation, we introduce BizFinBench.v2, a large-scale financial benchmark built from authentic user queries in the Chinese and U.S. stock markets and integrating both offline and online evaluation tasks. The benchmark covers four core business scenarios, eight offline tasks, and two online tasks, comprising 29,578 expert-annotated question-answer pairs. Experimental results show that ChatGPT-5 achieves an average accuracy of 61.5% on the offline tasks, still substantially below the performance of financial experts, while DeepSeek-R1 delivers the strongest results on the online tasks. Further error analysis reveals key limitations of current models in financial knowledge understanding, complex reasoning, and dynamic decision-making. BizFinBench.v2 provides a systematic, business-oriented framework for evaluating financial large language models and offers practical guidance for their deployment and continued improvement in real-world financial applications.

报告人：周灿 南京审计大学

报告题目：Learning Semi-Parametric Tree Models from Mixed Data

报告摘要：Causal discovery and representation involving latent variables and structures have attracted growing interest in the era of artificial intelligence, particularly for their critical role in understanding real-world data. While many existing methods focus exclusively on either purely continuous or purely discrete data, this paper addresses the challenge of learning latent structures from mixed data. We propose a novel semi-parametric tree model capable of handling mixed data and develop an algorithm for learning the structure of this model using additive information distances. We demonstrate that this algorithm efficiently and accurately recovers the true structure, given the information distances. Additionally, the sample-based version of the structural learning algorithm achieves probabilistic approximate correctness, with a finite sample bound established for exact structural recovery. Both simulated and real data are used to assess the performance of our proposed algorithm, with experimental results showing that our algorithm can effectively discover latent hierarchical structures behind mixed data.

报告人：毛俊凡 东南大学

报告题目：Generating Tensor Data by Diffusion Models with Tucker U-net

报告摘要：Large dimensional factor models are widely used in asset pricing, macro economic analysis, international trading, social network analysis, machine learning, and so on. In this talk, we first show how to generate tensors with low-rank factors via the diffusion generative model with our carefully tailored Tucker Unet. We theoretically present the accuracy of the deep generative artificially networks, including the approximation error, estimation error and the TV error of the large-dimensional distribution function of the generated tensors. Empirical studies in picture generation and Newyork taxi data show the good performance of our Tucker Unet. Next, we give a new statistical testing for the existence of the tensor factor structure in large dimension. We show that our test is powerful which is theoretically underpinned compared with the most recent works.

报告人：郦泽雨 东南大学

报告题目：Distributional Counterfactual Inference by Conditional Tensor Diffusion Model

报告摘要：Counterfactual analysis with panel and multi-way data often requires recovering an unobserved block of potential outcomes from its factual complement. Existing factor-based completion methods principally produce point predictions or common components, whereas distributional counterfactual analysis requires the conditional law of the missing block. We develop Counterfactual Tucker Diffusion (CFT-Diff), a conditional tensor-diffusion method for this task. A prespecified mask partitions a low-Tucker-rank tensor into treated counterfactual entries and factual controls; the forward process corrupts only the treated entries and the reverse process conditions on the fixed controls. CFT-Diff fits its score over a structured masked Tucker class: masked Tucker encoders, a low-dimensional core regression, a Tucker decoder, and a skip connection. The resulting reverse diffusion generates repeated counterfactual completions and hence distributional summaries of the missing block. Our theory provides distributional guarantees for treated potential outcomes, and the convergence rates rely on the Tucker core dimension, the largest mode dimension, and the effective treated dimension rather than the ambient dimension alone. Thus, CFT-Diff provides a statistically and computationally parsimonious approach to distributional counterfactual imputation.

报告人：吴梦云 上海财经大学

报告题目：Sparse Bayesian Deep Functional Learning with Structured Region Selection

报告摘要：In modern applications such as ECG monitoring, neuroimaging, wearable sensing, and industrial equipment diagnostics, complex and continuously structured data are ubiquitous, presenting both challenges and opportunities for functional data analysis. However, existing methods face a critical trade-off: conventional functional models are limited by linearity, whereas deep learning approaches lack interpretable region selection for sparse effects. To bridge these gaps, we propose a sparse Bayesian functional deep neural network (sBayFDNN). It learns adaptive functional embeddings through a deep Bayesian architecture to capture complex nonlinear relationships, while a structured prior enables interpretable, region-wise selection of influential domains with quantified uncertainty. Theoretically, we establish rigorous approximation error bounds, posterior consistency, and region selection consistency. These results provide the first theoretical guarantees for a Bayesian deep functional model, ensuring its reliability and statistical rigor. Empirically, comprehensive simulations and real-world studies confirm the effectiveness and superiority of sBayFDNN. Crucially, sBayFDNN excels in recognizing intricate dependencies for accurate predictions and more precisely identifies functionally meaningful regions, capabilities fundamentally beyond existing approaches.

报告人：唐荣 香港科技大学

报告题目：Subsampled Pseudo-posteriors for Scalable Bayesian Moment-condition Inference

报告摘要：Bayesian pseudo-posteriors based on moment conditions, such as Bayesian empirical likelihood (EL) and Bayesian exponentially tilted empirical likelihood (ETEL), provide a robust route to Bayesian inference when the model is specified only through moment restrictions, but their computation is often prohibitive. In this work, we address two barriers to Bayesian pseudo-inference under moment restrictions. First, scalable sampling is challenging for tall data because the pseudo-likelihoods are not log-additive across observations, so standard likelihood-subsampling methods do not apply. Second, in simulation-based and latent-variable problems, the moment function itself may be defined through an intractable expectation, making full-data pseudo-posterior inference infeasible. We develop a new family of subsampled pseudo-posteriors that replaces the full-data pseudo-posterior with an aggregate of mini-batch pseudo-posteriors. We show that naive mini-batching induces a shifted-mixture distortion, yielding aggregates that are correctly centered but overly diffuse. To remove this distortion, we introduce moment-function-level control variates that align mini-batch moment equations around a reference estimator and recover the full-data posterior shape. For Bayesian ETEL, we establish finite-sample total variation bounds to the full-data pseudo-posterior, allowing discontinuous and intractable moment functions. Numerical experiments show that the method applies broadly across moment-condition-based Bayesian pseudo-inference, closely approximates full-data pseudo-posteriors, and substantially reduces computational cost.

报告人：葛淑菲 上海科技大学

报告题目：A Bayesian Approach to Edge-Based Brain Connectivity Analysis

报告摘要：Brain connectivity analysis using MRI faces the curse of dimensionality in estimating high-dimensional covariance matrices, as existing methods often overlook latent topological structures among brain edges. We propose a Bayesian hierarchical model using a finite-dimensional Dirichlet distribution to model network topology with constant parameter dimensionality, improving stability. The covariance matrix is reformulated for positive definiteness, and a Metropolis-Hastings algorithm jointly infers topology and correlation. Optimizing the likelihood function reduces computational complexity. Simulations confirm effective recovery of topology and parameters. Compared to Graphical Lasso and Dirichlet process models, our method offers flexible tuning with better accuracy and convergence stability. Sensitivity analysis shows minimal impact from λ and sampling order. Applied to Alzheimer's Disease Neuroimaging Initiative data, the model identified structural subnetworks validated by anatomical metrics and literature, with stronger intragroup than intergroup connections. This framework provides an efficient, reliable tool for large-scale brain connectivity studies.

报告人：魏颖颖 香港中文大学

报告题目：Decontamination of Transcript Leakage in Spatial Transcriptomic Data

报告摘要：Spatial transcriptomics offers unprecedented opportunities to investigate the tissue organization, cell-cell interactions and the tumour microenvironment. In particular, the latest imaging-based spatial transcriptomic platforms have achieved cellular resolution. However, recent benchmarking studies of imaging-based spatial transcriptomic platforms have reported that certain genes appear to be expressed in cell types where such expression is biologically unexpected. Our data analysis suggests that this phenomenon is likely attributable to transcript leakage from neighboring cells rather than imperfect cell segmentation. As a result, the observed transcript counts for a given cell represent the sum of its endogenous expression and transcripts leaked from neighboring cells. Unfortunately, if the transcript leakage problem is not properly adjusted, it can lead to serious consequences for the quantification of expression levels, cell type annotation, differential gene expression detection and identification of spatially variable genes. However, despite active research on the deconvolution of bulk genomic data and spot-based spatial transcriptomic data, statistical methods for correcting transcript leakage are lacking. Here, we develop a Bayesian hierarchical model, DeLeakage, to Decontaminate transcript Leakage in imaging-based spatial transcriptomic data. DeLeakage does not require any reference data as input and allows transcript leakage levels to vary across genes. We prove the model identifiability of DeLeakage, which shows that contamination due to transcript leakage can be distinguished from the true underlying gene expression. Moreover, we develop an efficient Markov chain Monte Carlo algorithm with GPU parallelization for inference. Application to simulated and real data show that DeLeakage successfully adjusts transcript leakage and outperforms existing methods.

报告人：湛自奇 华东师范大学

报告题目：Conditionally Whitenened Generative Models for Probabilistic Time Series Forecasting

报告摘要：Probabilistic forecasting of multivariate time series is challenging due to non-stationarity, inter-variable dependencies, and distribution shifts. While recent diffusion and flow matching models have shown promise, they often ignore informative priors such as conditional means and covariances. In this work, we propose Conditionally Whitenened Generative Models (CW-Gen), a framework that incorporates prior information through conditional whitening. Theoretically, we establish sufficient conditions under which replacing the traditional terminal distribution of diffusion models, namely the standard multivariate normal, with a multivariate normal distribution parameterized by estimators of the conditional mean and covariance improves sample quality. Guided by this analysis, we design a novel Joint Mean-Covariance Estimator (JMCE) that simultaneously learns the conditional mean and sliding-window covariance. Building on JMCE, we introduce Conditionally Whitenened Diffusion Models (CW-Diff) and extend them to Conditionally Whitenened Flow Matching (CW-Flow). Experiments on five real-world datasets with six state-of-the-art generative models demonstrate that CW-Gen consistently enhances predictive performance, capturing non-stationary dynamics and inter-variable correlations more effectively than prior-free approaches. Empirical results further demonstrate that CW-Gen can effectively mitigate the effects of distribution shift.

报告人：刘鹤飞 云南财经大学

报告题目：Bayesian Fréchet Regression

报告摘要：Fréchet regression extends traditional linear regression to accommodate response variables in general metric spaces, aligning with current trends in data science and attracting growing attention across disciplines. To date, nearly all statistical inference for Fréchet regression has been developed within the frequentist framework, with no prior work exploring a Bayesian formulation. This article introduces a Bayesian inference approach for Fréchet regression. Unlike parametric models, Fréchet regression is defined solely through a conditional expectation rather than an explicit parametric form, making it infeasible to assign priors to model parameters. To address this challenge, we place priors directly on the predicted values of the response variable. Furthermore, since Fréchet regression does not assume an error distribution, traditional likelihood-based methods cannot be applied. We therefore construct a pseudo-likelihood function to capture sample information. In addition, we propose a modified Bayesian Information Criterion (BIC) tailored for variable selection in Fréchet regression. Theoretical results establish both the consistency of the estimator and the validity of variable selection under the modified BIC. Simulation studies and real-data applications further demonstrate the reliability and practical utility of the proposed Bayesian method.

报告人：陆婧 东北师范大学

报告题目：A Robust Bootstrap-CUSUM Method for Detecting Test Speededness and Pinpointing Change-points in Computerized Testing

报告摘要：Test speededness, caused by time constraints, can impact examinees' performance, leading to decreased response accuracy, particularly toward the end of the test. Most existing methods for detecting test speededness rely on specific distributional assumptions for response times (RTs), such as the lognormal distribution, which may lead to incorrect statistical inference if the true data distribution deviates from these assumptions. This paper proposes a novel Bootstrap-CUSUM method for detecting test speededness, which is robust to non-normality in log-RTs. By constructing a CUSUM person-fit statistic for log-RTs and using the multiplier bootstrap to estimate its empirical distribution, our method facilitates individual level detection and change-point estimation. We prove the theoretical consistency of the method under both null and alternative hypotheses. Simulation studies show that the Bootstrap-CUSUM method outperforms the likelihood ratio test, Wald test, and score test in terms of correct classification rate, true detection rate, and false positive rate, demonstrating superior robustness and adaptability across different data distributions. The real data analysis further demonstrates the practical utility of the proposed method for detecting test speededness.

报告人：崔婷婷 浙江财经大学

报告题目：A Bayesian Variable Selection Approach for Dynamic Treatment Regimes

报告摘要：In recent years, the increasing availability of high-dimensional data has offered new opportunities, but also posed significant challenges for estimating dynamic treatment regimes. Existing high-dimensional methods are largely frequentist and rely on penalization, while Bayesian approaches with rigorous selection guarantees remain limited, especially when the number of covariates exceeds the sample size. To address this gap, we propose a high-dimensional Bayesian Q-learning (HBQL) approach based on stage-wise Bayesian hierarchical models with shrinking-and-diffusing priors. Inference at stage 1 is conducted using an augmented dataset, in which pseudo-outcomes are multiply imputed from the stage-2 posterior predictive distribution via a high-dimensional backward-induction Gibbs (HBIG) sampler. Furthermore, we show that strong selection consistency holds for the stage-1 procedure. Extensive simulation studies demonstrate that the proposed approach maintains competitive statistical performance in high-dimensional multi-stage settings. The application of HBQL to the real-world China Rural Hypertension Control Project (CRHCP) dataset further illustrates its practical utility in precision medicine.

报告人：毛晓军 上海交通大学

报告题目：Fast and Flexible Covariance Estimation for Diffusion Posterior Sampling in Imaging Inverse Problems

报告摘要：Imaging inverse problems such as inpainting and deblurring seek to recover images from noisy measurements produced by degradation operators. Diffusion posterior sampling provides a flexible zero-shot solution using a pretrained unconditional diffusion prior, with likelihood-score approximation as its key bottleneck: first-order methods use heuristic covariance choices, whereas second-order corrections require extra covariance training or costly computation. We propose a fast, plug-and-play covariance estimation module for diffusion posterior sampling. It uses intermediate trajectory variables to estimate the covariance term associated with the likelihood-score approximation, and integrates with existing first-order samplers without separately trained covariance estimators or costly repeated gradient computations. Theoretically, we establish a structured local likelihood-score error bound for the proposed module and a posterior sampling error bound that links this local approximation error to the distributional error. Experiments use FFHQ and ImageNet datasets and include linear and nonlinear tasks, Gaussian and Poisson noise, pixel- and latent-diffusion solvers, and varying first-order hyperparameters, the proposed module consistently improves the corresponding first-order samplers in perceptual and distortion-based metrics, with limited additional computation.

报告人：郁淼淼 华东师范大学

报告题目：Statistical Inference for A General Federated Learning Framework under Local Differential Privacy

报告摘要：To address privacy leakage in intelligent decision-making process, this paper proposes a general federated learning framework based on local differential privacy (LDP) with four main contributions. First, we develop a bounded transformation that maps observations into a compact interval, thereby eliminating the restrictive bounded-support assumption commonly imposed in existing differential privacy methods. Second, after the privacy mechanism is applied, the privatized data are truncated to ensure that they lie within the domain of the inverse transformation and improves the robustness of the reconstructed private data. Third, we propose a robust noise-averaging privatization mechanism that averages multiple independent noise variables. This approach reduces the distortion caused by extreme noise realizations and achieves a convergence rate comparable to that of the oracle estimator. Finally, we develop a communication-efficient one-shot federated learning procedure in which clients transmit only local estimators. We further establish the asymptotic normality of the aggregated estimator, enabling valid statistical inference. By jointly addressing privatization instability and communication efficiency, the proposed framework provides a principled trade-off between privacy protection and statistical efficiency.

报告人：张心雨 华东师范大学

报告题目：Two-way Matrix Autoregressive Model with Thresholds

报告摘要： Recently, matrix-valued time series data have attracted significant attention in the literature with the recognition of threshold nonlinearity representing a significant advance. However, given the fact that a matrix is a two-array structure, it is unfortunate, perhaps even surprising, for the threshold literature to focus on using the same threshold variable for the rows and the columns. In fact, evidence in economic, financial, environmental and other data shows the necessity and advantages of allowing the possibilities of different threshold variables for the rows and columns, with possibly different threshold parameters, hence the need for a Two-way Matrix AutoRegressive model with Thresholds (2-MART). The resulting model poses new theoretical and computational challenges, since the two threshold variables interact with the bilinear row-column dynamics in both estimation and asymptotic analysis. In this paper, we develop a comprehensive methodology for the 2-MART model, by overcoming various challenges. Compared with existing models in the literature, the new model can achieve greater dimension reduction, much better model fitting, more accurate predictions, and more plausible interpretations.

报告人：滕佳烨 上海财经大学

报告题目：AutoSTAT: 全自动数据分析助手

报告摘要： 数据分析是一项具有重要价值的工作，其开展往往需要投入大量的人力与物力成本。随着大模型技术的飞速发展，尽管大模型能够辅助人类完成数据分析任务，但生成高质量的数据分析报告仍高度依赖使用者的专业经验与操作技巧。基于此背景，本次汇报将重点介绍一款全自动数据分析助手 ——AutoSTAT。该工具通过结构化处理流程，将人类先验知识融入智能体的构建过程，有效提升了智能体的自主规划能力。有关本项目的详细信息，包括部署方案、项目试用等内容，均可通过官网 autostat.cc 进行查询。

报告人：潘文亮 中国科学院数学与系统科学研究院

报告题目：Ball Impurity: Measuring Heterogeneity in General Metric Spaces

报告摘要：Data in various domains, such as neuroimaging and network data analysis, often come in complex forms without possessing a Hilbert structure. The complexity necessitates innovative approaches for effective analysis. We propose a novel measure of heterogeneity, ball impurity, which is designed to work with complex non-Euclidean objects. Our approach extends the notion of impurity to general metric spaces, providing a versatile tool for feature selection and tree models. The ball impurity measure exhibits desirable properties, such as the triangular inequality, and is computationally tractable, enhancing its practicality and usefulness. Extensive experiments on synthetic data and real data from the UK Biobank validate the efficacy of our approach in capturing data heterogeneity. Remarkably, our results compare favorably with state-of-the-art methods in metric spaces, highlighting the potential of ball impurity as a valuable tool for addressing complex data analysis tasks.

报告人：李睿 上海对外经贸大学

报告题目：Quantile Regression and Homogeneity Identification of a Semiparametric Panel Data Model

报告摘要：In this article, we study the quantile regression and homogeneity detection of a varying index coefficient panel data model with fixed individual effect that shows nonlinear time trend. Based on spline approximation, we get the estimates of trend function, link function and index parameters, and establish the convergence rate and asymptotic normality accordingly. Noting that the subjects within a group may share the same trend functions, which motivates us to further identify the possible homogeneity in this trend by integrating regularization with binary segmentation algorithm. For determining the thresholding parameter in homogeneity identification, we propose a generalized Bayesian information criterion following line in Chen(2019). Further, we propose a penalized method to identify the constant and linear structure of nonparametric functions in our model. Finally, we provide more efficient estimates via grouped observations and improve the large sample properties. Simulations and a hybrid real data analysis are conducted to illustrate the finite sample performance of the proposed approach.

报告人：王超杰 江苏大学

报告题目：STransfer: A Transfer Learning-Enhanced GCN for Spatial Transcriptomics Data

报告摘要：Capturing spatial structure is fundamental to the analysis of spatial transcriptomics data. However, most existing methods focus on clustering within individual tissue slices and often ignore the high inter-slice similarity inherent in multi-slice datasets. To address this limitation, we propose STransfer, a novel transfer learning framework that combines graph convolutional networks (GCNs) with positive pointwise mutual information (PPMI) to model both local and global spatial dependencies. An attention-based module is introduced to fuse features from multiple graphs into unified node representations, facilitating the learning of low-dimensional embeddings that jointly encode gene expression and spatial context. By transferring knowledge from labeled slices to adjacent unlabeled ones, STransfer significantly enhances clustering accuracy while reducing manual annotation costs.

报告人：涂霖原 上海财经大学

报告题目：Robust Variational Bayes by Min-Max Median Aggregation

报告摘要：We propose a robust and scalable variational Bayes (VB) framework to handle data contamination and outliers. Our approach partitions the dataset into m disjoint subsets and formulates a joint optimization problem based on robust aggregation. A key insight is that the full posterior distribution can be characterized as the minimizer of the mean Kullback-Leibler (KL) divergence from the m -powered local posteriors. To achieve robustness, we replace this mean KL divergence with a min-max median formulation. This formulation ensures consistency between the KL minimizer and the Evidence Lower Bound (ELBO) maximizer, while facilitating sharper statistical rates for the variational posterior mean. We also observe that the behavior of the m -powered marginal log-likelihood depends heavily on whether local latent variables are present. To guarantee the consistency of the aggregated variational posterior, we treat these two scenarios separately; specifically, when local latent variables are present, we introduce an aggregate-and-rescale strategy. Theoretically, we provide a finite-sample analysis of the proposed posterior, including a refined Bernstein-von Mises (BvM) theorem that accommodates a diverging number of subsets m . Our findings indicate that this two-stage approach yields a smaller approximation error than directly aggregating the m -powered local posteriors. Furthermore, we establish a nearly optimal statistical rate for the variational posterior mean, advancing existing theories for min-max median estimators. Extensive simulation studies demonstrate the efficacy of our method.

报告人：郭旭 北京师范大学

报告题目：MATCH: Multiplier-Assisted Tests for Conditional Hypotheses in Non-Euclidean Data

报告摘要：We propose a new procedure MATCH (Multiplier-Assisted Tests for Conditional Hypotheses) to test whether the non-Euclidean data match the target model, which is a general framework for significance and specification testing in Fréchet regression. MATCH covers global significance, partial significance, and the adequacy of global Fréchet regression, providing a unified way to compare unrestricted conditional Fréchet means with restricted alternatives. One of the key challenges is that the ordinary held-out loss difference is first-order degenerate under the null: the oracle losses coincide, and plug-in statistics is dominated by nuisance estimation error. MATCH uses sample splitting and independent random multipliers on held-out losses to create a nondegenerate Gaussian leading term without residuals or tangent-space coordinates. To improve data use and stability, we further develop cross-fitted tests and repeated cross-fitting with p-value merging. We establish asymptotic null validity, consistency under fixed alternatives, and local power guarantees. Simulations for distributional, symmetric positive-definite (SPD) matrix-valued, and spherical responses support the theoretical findings, and applications to county-level household income distributions and North Atlantic tropical-cyclone locations demonstrate the practical use of the proposed tests.

报告人：李树威 广州大学

报告题目：A Factor-Augmented Regression Approach to Learning Bivariate Interval-Censored Data

报告摘要：Comorbidity, such as the coexistence of hypertension and diabetes, poses a major challenge in modern medicine, as it reflects the systemic and holistic nature of human health. Understanding the latent risk factors underlying comorbidity is essential for the prevention and treatment of diseases from a more comprehensive and integrated perspective. In real-life disease surveillance, comorbidity occurrence is often monitored through periodic examinations, leading to the frequent emergence of bivariate interval-censored data. In this article, we propose a generalized exploratory factor analysis model within a likelihood-based framework and integrate it with a class of time-to-event frailty models. This joint modeling framework enables the extraction of latent factors from various types of covariates while simultaneously assessing their effects on bivariate interval-censored failure events. We focus on nonparametric maximum likelihood estimation and develop an expectation-maximization algorithm to achieve reliable and efficient implementation. Sufficient conditions for model identifiability are established, and the asymptotic properties of the proposed estimators are derived. The proposed method demonstrates satisfactory performance in simulation studies and is further applied to a real-world data analysis of Type 2 diabetes, showing its desirable abilities in addressing the multicollinearity issue and uncovering the latent covariate structure.



报告人：谭承利 西安交通大学

报告题目：Emergent Short-range Memory in Stochastic Gradient Noise and Its Implications on Generalization

报告摘要：Investigating stochastic gradient descent (SGD) from the perspective of stochastic differential equations (SDEs) is quite popular in the deep learning community. In this talk, I will present an analytical result on modeling SGD with SDEs driven by fractional Brownian motion, which reveals the escaping efficiency when trapped in local minima. From the optimization point of view, I will further show how we can relate the smoothness of the optimization pathway to the generalization ability.

报告人：刘弘之 上海对外经贸大学

报告题目：Asymptotic Properties of Adaptive Quantum Phase Estimation

报告摘要：Quantum phase estimation algorithms are core sub-algorithms in many important quantum algorithms, having application value in the field of quantum computing. Bayesian quantum phase estimation algorithms are noise-robust quantum phase estimation algorithms suitable for medium-sized quantum bits. This report will explore the asymptotic properties of adaptive Bayesian quantum phase estimation algorithms, and efficiently improve existing algorithms based on these asymptotic properties.

报告人：陈雪蓉 西南财经大学

报告题目：A General Distributed Privacy-Preserving Statistical Inference Framework for Massive Data

报告摘要：In the era of massive data, privacy-preserving distributed inference plays a key role in balancing data utility, computational efficiency, and privacy protection in statistical analysis. However, current differentially private distributed inference and federated learning methods either focus solely on parameter estimation, assume i.i.d. data, or are restricted to specific models (e.g., linear models), failing to provide comprehensive statistical inference for general models with heterogeneous data. Moreover, sensitivity calculations in existing differential privacy methods often rely on data truncation or boundedness assumptions, tailored to each model on a case-by-case basis. To address these issues, we propose a unified distributed privacy-preserving statistical inference framework suitable for both homogeneous and heterogeneous data. The framework leverages an encryption technique that integrates concepts from symmetric encryption and differential privacy. It is computationally efficient, achieves high accuracy, and eliminates the need for case-by-case sensitivity calculations. The framework is applicable to widely used models such as linear regression, logistic regression, the Cox model, and quantile regression, among others. We also provide guidance on selecting encryption functions that can simplify both sensitivity calculations and theoretical derivations. Theoretical properties, including differential privacy guarantees, convergence rates, and the asymptotic normality of the proposed private estimators, are established. The effectiveness of our method is demonstrated through both numerical simulations and real data analyses.

报告人：杨凯 长春工业大学

报告题目：Bayesian Forecasting for A Logistic Mixture Double Autoregressive Model

报告摘要：To capture the dynamic relationship between financial time series and covariates, we consider a logistic mixture double autoregressive model with explanatory variables. The model is composed of two double autoregressive models whose mixing ratio is time-varying and is driven by a logistic regression structure. By introducing a series of Bernoulli distributed latent variables, a complete data likelihood is obtained, which makes the Bayesian inference feasible. Based on this likelihood, a new Markov chain Monte Carlo algorithm is developed to address the parameter estimation problem. The heteroscedasticity test problem for the underlying process is also addressed by means of Bayes factor. The performances of the proposed methods are evaluated via simulations. Finally, the proposed model is applied to the Shanghai Stock Exchange Index data set.

报告人：杜明月 吉林大学

报告题目：Regression Analysis of Interval-Censored Failure Time Data with Change Points and a Cured Subgroup

报告摘要：There exists a substantial body of literature that discusses regression analysis of interval-censored failure time data and also many methods have been proposed for handling the presence of a cured subgroup. However, only limited research exists on the problems incorporating change points, with or without a cured subgroup, which can occur in various contexts such as clinical trials where disease risks may shift dramatically when certain biological indicators exceed specific thresholds. To fill this gap, we consider a class of partly linear transformation models within the mixture cure model framework and propose a sieve maximum likelihood estimation approach using Bernstein polynomials and piecewise linear functions for inference. Additionally, we provide a data-driven adaptive procedure to identify the number and locations of change points and establish the asymptotic properties of the proposed method. Extensive simulation studies demonstrate the effectiveness and practical utility of the proposed methods, which are applied to the real data from a breast cancer study that motivated this work.

报告人：刘蓉 长春工业大学

报告题目：Statistical Inference for Interval-Censored Data for Linear Transformation Models within Ordinary Differential Equation Framework

报告摘要：Interval-censored data commonly arise in clinical follow-up and epidemiological studies. For linear transformation models, this paper proposes a estimation method that combines Bernstein polynomial approximation with ordinary differential equations to flexibly characterize the unknown baseline hazard function. The consistency and convergence rate of the estimators are established, and the estimator of the regression parameters is shown to be asymptotically normal and semiparametrically efficient. Simulation results demonstrate that the proposed method yields small estimation bias, accurate standard error estimates, and coverage probabilities close to the nominal level. Finally, the method is applied to hypertension onset data from the China Health and Retirement Longitudinal Study, illustrating its practical applicability.

报告人：杨莹 复旦大学

报告题目：Online Sparse Regression with Expanding Observables

报告摘要：Online high-dimensional regression has gained increasing attention in recent years, yet existing methods typically assume that all candidate features, including important ones, are observed from the outset of data collection. This assumption is often violated in real-world scenarios, where new variables become available gradually as data accumulate. To address this gap, we introduce a novel framework, Recurrent Adaptive Variable Selection (RAVAS), for online regression with expanding observability. RAVAS employs a recurrent procedure that dynamically updates feature selection as both the sample size and the observable feature set grow. The algorithm is designed to be computationally efficient and memory-light, relying only on low-dimensional sufficient statistics that are updated online. A key advantage of the method lies in its ability to detect and incorporate important variables that emerge later, thereby mitigating the effect of early-stage missingness. We establish theoretical guarantees on model selection, estimation error, and feature coverage, and develop an adaptive online tuning strategy. Extensive simulations and real-world experiments verify the effectiveness of RAVAS for high-dimensional streaming data.

报告人：周叶青 同济大学

报告题目：Power Enhancement for Test of Multivariate Independence

报告摘要：Testing for independence between two random vectors is a fundamental problem in statistics. It is observed from empirical studies that many existing omnibus consistent tests may not perform well for some strongly nonmonotonic and nonlinear dependence. To get insights into this issue, this paper reveals that a class of existing multivariate independence tests may lose their power due to cancellation of positive and negative terms in dependence metrics. The cancellation leads to the sum of these terms very close to zero. Motivated by this finding, we propose a class of consistent metrics indexed by a positive integer γ to characterize independence. We further prove that the metrics with even or infinity γ can effectively avoid the cancellation, and have better powers under the alternatives in which two mean differences offset each other. In practice, it is desirable to target at a wide range of dependence scenarios. Thus, we further advocate to combine the p-values of test statistics with different γ 's. The advantages of the newly proposed tests are illustrated by numerical studies.

报告人：曹智苗 山东大学

报告题目：Decoding AI Precision Medicine Paradox: A Fairness-Aware Optimal Transport

报告摘要： Precision oncology faces a critical paradox where the pursuit of maximal individual-level precision may inadvertently widen healthcare disparities at the group level. This tension necessitates ethically grounded frameworks that reconcile individualized precision with therapeutic equity. In real world oncology, the inherent discreteness of clinical follow-up results in unobserved event times where the exact timing of disease progression remains unknown. This temporal uncertainty compromises the estimation of individualized treatment rules and often leads to suboptimal therapeutic guidance. To bridge this gap, we propose a robust framework that identifies optimal treatment regimes by maximizing the cumulative expected log survival probability. This optimization objective utilizes the logarithmic survival function as the primary metric to ensure the pursuit of maximal individual survival benefits despite the presence of censored clinical outcomes. To maintain therapeutic equity, we incorporate optimal transport to enforce demographic parity by aligning treatment distributions across patient subgroups. This mechanism mitigates treatment disparities tied to sensitive attributes such as gender or ethnicity without sacrificing individual survival benefits. The resulting methodology provides clinicians with a mathematically rigorous tool for equitable precision medicine. By integrating the proposed maximization of individual survival benefits with the mandate for clinical fairness, this work offers a scalable solution for delivering reliable individualized care in practical oncology settings.

报告人：严晓东 西安交通大学

报告题目：AI Safe: Statistical Methods for Security Protection from Small to Large Models

报告摘要： This report explores AI security protection statistical methods spanning small and large models. Firstly, it reviews the evolution of classic security issues of small models, covering adversarial examples, backdoor poisoning and algorithmic bias. Secondly, it establishes a statistical framework for large model security, and discusses the correlation between large model security risks and statistical principles from the data, alignment and downstream task layers. Finally, this paper proposes a security optimization method based on preference-enhanced distribution alignment for large model fine-tuning. Relevant experimental results verify the effectiveness of the proposed approach in enhancing the security performance of large AI models.

报告人：牟荣吉 上海交通大学

报告题目：A Bayesian Phase I/II PK-integrated Regimen Optimization Design for Survival Efficacy and Late-onset Toxicity

报告摘要：Optimizing dose-schedule regimens is critical in modern oncology trials, yet existing Phase I/II designs often neglect pharmacokinetic (PK) information, rely on short-term surrogates, and assume rapid outcomes—a flawed premise for immuno- and targeted therapies. To address these limitations, we propose B-PROSE, a novel Bayesian Phase I/II PK-guided design optimizing for survival efficacy and late-onset toxicity. Following the exposure-response pathway, B-PROSE employs a Bayesian hierarchical PK model to summarize cumulative exposure (AUC) from longitudinal concentrations. Conditional models then link AUC to toxicity and time-to-event efficacy via proportional hazards, uniquely allowing efficacy to depend on both exposure and toxicity. To address delayed outcomes, Bayesian data augmentation leverages early PK and partial follow-up to impute pending toxicities, enabling continuous posterior updating. A utility function balancing efficacy and toxicity guides adaptive allocation while discarding suboptimal regimens. Simulations demonstrate that B-PROSE achieves operating characteristics comparable to the ideal design with complete observations, enhancing trial efficiency and decision-making in dose-schedule optimization.

报告人：陈中媛 南京大学

报告题目：Data-Driven Statistical Methods for Cancer Treatment Recommendations and Causal Anti-Cancer Drug Sensitive Genes Selection

报告摘要：Despite the availability of large amounts of genomic/clinical data, medical treatment recommendations and causal anti-cancer drug sensitive genes selection have yet to successfully take good advantage of them. In this talk, I will first introduce a data-guided statistical machine learning approach for cancer treatment recommendations (also called individualized treatment rules, ITRs) with feature scores by applying a dimension reduction method (Sliced Inverse Regression, SIR). Second, I will show a Bayesian learning method to iteratively impute the hidden confounding factors when estimating causal effects for high-dimensional regression. Such approach is utilized in the selection of causal anti-cancer drug sensitive genes.

报告人：童培峰 北京大学

报告题目：Residual Covariance Unlocks Robust Mapping of Dynamic Causal Brain Network

报告摘要：Mapping brain's causal networks is fundamental to neuroscience. Existing methods for estimating the effective connectivity are largely based on long data segments, which is incapable of capturing finer temporal network dynamics of transient neural events. We introduce a semi-parametric framework for constructing dynamic brain causal networks with short window data that integrates a residual-rotated Lasso method for sparse vector autoregressive modeling, a scale-invariant generalized variance decomposition for network estimation, and a semi-parametric prewhitening step to mitigate potential model misspecification. The approach reduces the data length required for robust causal network inference from hundreds of seconds to less than two seconds. When applied to stereo-electroencephalography data from three patients with temporal lobe epilepsy, it provides the first quantitative characterization of a seizure's entire frequency-specific network evolution, matching the clinical annotations. Our non-interventional network analysis reveals a high network in-degree in high-frequency bands at the epileptogenic zone, a finding consistent with the interventional cortico-cortical evoked potential studies.

报告人：施润叶 复旦大学

报告题目：A Bayesian Clustering Framework for Decoding Heterogeneous Genetic Architectures

报告摘要：The heterogeneous nature of genetic effects is continuously shaped by environmental factors through the modulation of multi-omics regulatory nodes. Understanding such heterogeneity is critical for informing precision health and personalized health management. However, existing methods often rely on pre-specified modifying environmental variables, limiting their ability to capture complex heterogeneous genetic architectures under undefined environmental contexts. Here, we propose Multi-Omics Clustering for Heterogeneous Association (MOCHA), a novel Bayesian analytical paradigm that aim to identify latent population subgroups with distinct genetic effects driven by multi-omics profiles, without requiring a priori marker specification. Extensive simulations showed that MOCHA was able to accurately identify true population clustering structure and uncover markers with cluster-specific genetic effects, and provided reliable parameter estimates. Finally, applying MOCHA to multi-omics data from the IMAGEN study, we identified two distinct neurodevelopmental clusters associated with adolescent inhibitory control, where post-hoc characterization of these adolescent clusters provided novel insights into the biological mechanisms of brain plasticity along with the joint influences of multidimensional environmental factors.

报告人：虞龙 上海财经大学

报告题目：High-dimensional Factor and Clustering Analysis via Projected and Truncated Sample Covariance Matrix

报告摘要：This paper addresses spike identification in high-dimensional mixture factor models, where the spiked eigenvalues of the covariance matrix may arise from clustered mean vectors and/or latent common factors. The magnitudes of the eigenvalues from these two sources can be comparable, posing a significant challenge to model identification and estimation. We introduce a novel {it projected and truncated sample covariance matrix} which, with suitable thresholding, consistently recovers the factor loading space regardless of the relative strength of the cluster-driven eigenvalues. Importantly, this consistency holds even when the subspaces induced by the factor structure and the clustered means are partially confounded. Given this estimated factor space, we propose a projected spectral clustering procedure that removes the factor component and reveals the underlying cluster structure. Furthermore, we develop an iterative refinement algorithm that builds upon these steps to further enhance estimation accuracy. All proposed methods are supported by rigorous theoretical guarantees and are validated through extensive simulations and real data analyses.

报告人：付盛 南开大学

报告题目：Multinomial Logistic Regression: Old and New

报告摘要：Multinomial logistic regression (MLR) is a widely used method for modeling categorical outcomes, but it often faces challenges such as parameter nonidentifiability, particularly when applied to high-dimensional group-structured covariates. To address these issues, we introduce the group simplex-based MLR (GSMLR), which offers efficient parameterization and computational advantages. GSMLR eliminates the need for subjective reference category selection and avoids the cumbersome constraints typical of traditional MLR. We establish a definitive oracle error bound for GSMLR estimators providing strong theoretical guarantees, and propose an efficient algorithm that combines groupwise coordinate descent with refined safe screening rules. Extensive numerical experiments show that GSMLR significantly outperforms existing methods in both prediction accuracy and computational efficiency, making it a robust tool for high-dimensional data analysis and decision support in complex domains.

报告人: 任明旻 上海交通大学

报告题目: Model-agnostic Information Transfer for Classification with Label Noise

报告摘要: Label noise presents a fundamental challenge in modern machine learning, especially when large-scale datasets are generated via automated processes. An increasingly common and important data paradigm, particularly in domains like medical imaging, involves learning from a large dataset with coarse, noisy labels supplemented by a small, expert-verified, clean dataset. This setting constitutes a typical information transfer and fusion problem. However, the significant distribution shift between the noisy and clean data violates the core overall parametric similarity assumptions of existing statistical transfer learning methods, while their reliance on parametric models is ill-suited for complex data like images. To address these limitations, this paper develops a generic model-agnostic nonparametric framework for classification with label noise, which applies to a broad class of classifiers. Our approach leverages the small clean dataset to “purify” the large noisy one and carefully manages the remaining ambiguous samples. This framework is underpinned by a rigorous statistical theory. Its empirical performance is demonstrated through simulations and a real-world application to medical image analysis for pneumonia diagnosis.

报告人: 骆钊澐 上海财经大学

报告题目: Rate-Optimal Online Learning for Dynamic Assortment Selection with Positioning

报告摘要: In online retailing, the seller aims to offer assortment of items with maximized revenue. We introduce a new online learning problem called Dynamic Assortment Selection with Positioning (DAP) that additionally learns the optimal positioning within the assortment. Specifically, the customers make purchases based on the item attractiveness as the product of the position effect and unknown preference parameter through a multinomial logit choice model. We first demonstrate that any assortment-only algorithm that neglects position effects results in linear regrets. To address this gap, we propose the Truncated Linear Regression Upper Confidence Bound (TLR-UCB) policy. TLR-UCB utilizes a novel geometrical linear-bandit-type feedback structure for UCB construction under random and adaptive position effects. In addition, TLR-UCB conducts well-designed truncations before applying linear regression to handle conditional geometric responses. In theory, we establish a regret upper bound of $\tilde{O}(T^{1/2})$ for TLR-UCB, matching our derived $\Omega(T^{1/2})$ lower bound. Moreover, we develop an Explore-In-TLR-UCB (EI-TLR) policy to tackle unknown position effects. It first conducts a joint learning procedure to estimate unknown preferences and position effects, and then implements a generalized TLR-UCB procedure driven by estimated position effects. Extensive experiments demonstrate the superior performance of TLR-UCB and EI-TLR over other benchmark policies.

报告人：刘林 上海交通大学

报告题目：A Common Honest Mistake in Semiparametric Theory

报告摘要：We discuss a common lack of rigor in semiparametric theory and some of its applications. We will show its applications in instrumental variable analysis and causal graphs.

报告人：马慧娟 华东师范大学

报告题目：Efficient Estimation of Quantile Treatment Effects by Incorporating External Controls

报告摘要：Quantile treatment effect (QTE) is increasingly recognized as an essential tool in understanding the heterogeneous impact of interventions across different segments of the outcome distribution and providing a richer perspective beyond average treatment effect. Despite its importance, obtaining precise QTE estimates can be challenging in randomized clinical trials with limited sample sizes. External controls have grown in popularity in recent years as a tool to empower the clinical trials. In this paper, we propose a novel approach that combines clinical trial data with external controls to enhance the efficiency of QTE estimation. By leveraging external controls, we first derive the efficient influence function of the QTE estimand and then propose a sieve-based weighted M-estimator that achieves the semiparametric efficiency bound. The proposed estimator is shown to be more efficient than the one solely using trial data. We also extend the proposed framework to other QTE estimands defined in different populations and provide an alternative approach for more efficient estimation of average treatment effects. Extensive simulation studies and application to the job training data demonstrate the practical utility of the proposed methods.

报告人：郑潇逸 中国科技大学

报告题目：Fairness Evaluation for Risk Prediction Models under Heterogeneous Risk Distributions

报告摘要：As predictive algorithms increasingly mediate access to essential opportunities, ensuring algorithmic fairness has become a societal imperative. To date, a growing body of literature has focused on evaluating the fairness of predictive models and promoting fairness during model development. Group fairness metrics, which assess whether model performance is comparable across groups defined by sensitive attributes, have been widely applied to evaluate clinical prediction models. However, these metrics suffer from an intrinsic limitation: differences in predictive performance across subgroups may arise not only from model-induced biases but also from differences in underlying risk distributions. This issue, long recognized in the machine learning literature and often referred to as "infra-marginality", is particularly salient in clinical risk prediction, where both outcome-predictor associations and predictor distributions often vary across population subgroups. In previous work, we provided the first statistical formulation of this issue in the context of the "equality of opportunity" criterion and proposed a framework based on adjusting risk distributions to address it. In this work, we extend this approach to a broad class of group fairness metrics and establish their large-sample properties. The resulting adjusted fairness metrics vanish when, conditional on the same underlying risk, predictions are assigned in the same probabilistic manner across subgroups. We demonstrate the proposed method through simulation studies and an application to a real-world mortality risk prediction model.

报告人：周慧娟 上海财经大学

报告题目：BMDD: A Probabilistic Framework for Accurate Imputation of Zero-inflated Microbiome Sequencing Data

报告摘要：Microbiome sequencing data are inherently sparse and compositional, with excessive zeros arising from biological absence or insufficient sampling. These zeros pose significant challenges for downstream analyses, particularly those that require log-transformation. We introduce BMDD (BiModal Dirichlet Distribution), a novel probabilistic modeling framework for accurate imputation of microbiome sequencing data. Unlike existing imputation approaches that assume unimodal abundance, BMDD captures the bimodal abundance distribution of the taxa via a mixture of Dirichlet priors. It uses variational inference and a scalable expectation-maximization algorithm for efficient imputation. Through simulations and real microbiome datasets, we demonstrate that BMDD outperforms competing methods in reconstructing true abundances and improves the performance of differential abundance analysis. Through multiple posterior samples, BMDD enables robust inference by accounting for uncertainty in zero imputation. Our method offers a principled and computationally efficient solution for analyzing high-dimensional, zero-inflated microbiome sequencing data and is broadly applicable in microbial biomarker discovery and host-microbiome interaction studies. BMDD is available at: <https://github.com/zhohuj1994/BMDD>.

报告人：贺莘 上海财经大学

报告题目：Transfer Learning for Kernel-based Regression

报告摘要： In recent years, transfer learning has garnered significant attention. Its ability to leverage knowledge from related studies to improve generalization performance in a target study has made it highly appealing. This paper focuses on investigating the transfer learning problem within the context of nonparametric regression over a reproducing kernel Hilbert space. The aim is to bridge the gap between practical effectiveness and theoretical guarantees. We specifically consider two scenarios: one where the transferable sources are known and another where they are unknown. For the known transferable source case, we propose a two-step kernel-based estimator by solely using kernel ridge regression. For the unknown case, we develop a novel method based on an efficient aggregation algorithm, which can automatically detect and alleviate the effects of negative sources. This paper provides the statistical properties of the desired estimators and establishes the minimax rate. Through extensive numerical experiments on synthetic data and real examples, we validate our theoretical findings and demonstrate the effectiveness of our proposed method.

报告人：张宇谦 中国人民大学

报告题目：Quadruply Robust Methods for Causal Mediation Analysis

报告摘要： Estimating natural effects is a core task in causal mediation analysis. Existing triply robust (TR) frameworks (Tchetgen Tchetgen & Shpitser 2012) and their extensions have been developed to estimate the natural effects. In this work, we introduce a new quadruply robust (QR) framework that enlarges the model class for unbiased identification. We study two modeling strategies. The first is a nonparametric modeling approach, under which we propose a general QR estimator that supports the use of machine learning methods for nuisance estimation. We also study high-dimensional settings, where the dimensions of covariates and mediators may both be large. In these settings, we adopt a parametric modeling strategy and develop a model quadruply robust (MQR) estimator to limit the impact of model misspecification. Simulation studies and a real data application demonstrate the finite-sample performance of the proposed methods.

报告人: 陈自元 北京大学

报告题目: Deconfounding via Profiled Transfer Learning

报告摘要: Unmeasured confounders are a major source of bias in regression-based effect estimation and causal inference. In this paper, we propose a new profiled transfer learning framework, ProTrans, to address confounding effects in the target dataset, when additional source datasets with similar confounding structures are available. We introduce the concept of profiled residuals to characterize the shared confounding patterns between source and target datasets. By incorporating these profiled residuals into the target debiasing step, we effectively mitigate the latent confounding effects. We also propose a source selection strategy to enhance the robustness of ProTrans to noninformative sources. As a byproduct, ProTrans can also be used to estimate treatment effects in the presence of potential confounders, without the use of auxiliary features such as instrumental or proxy variables, which are often challenging to select in practice. Theoretically, we prove that the resulting estimated model shift from the sources to the target is confounding-free without imposing specific assumptions on the true confounding structure, and that the target parameter estimation achieves the minimax optimal rate under mild conditions. Simulated and real-world experiments validate the effectiveness of ProTrans and support the theoretical findings.

报告人: 贾骏雄 西安交通大学

报告题目: Nonparametric Prior Learning in Differential Equation Modeling

报告摘要: The statistical inverse problems of partial differential equations (PDEs) can be seen as the PDE-constrained regression problem. From this perspective, we propose general generalization bounds for learning infinite-dimensionally defined prior measures in the style of the probability approximately correct Bayesian learning theory. The theoretical framework is rigorously defined on infinite-dimensional separable function space, which makes the theories intimately connected to the usual infinite-dimensional Bayesian inverse approach. Inspired by the concept of differential privacy, a generalized condition has been proposed, which allows the learned prior measures to depend on the measured data. After illustrating the general theories, the specific settings of linear and nonlinear problems have been given and can be easily casted into our general theories to obtain concrete generalization bounds. Based on the obtained generalization bounds, infinite-dimensionally well-defined practical algorithms are formulated.

报告人：段志斌 西安交通大学

报告题目：结构化贝叶斯建模：从可信表征学习到大模型稳健对齐

报告摘要：深度学习虽然取得了显著进展，但特征冗余、决策机理不透明，以及噪声和分布偏移条件下不确定性难以刻画等问题仍较为突出。贝叶斯建模将模型参数和潜变量视为随机变量，借助后验分布描述不确定性及其传播过程，为处理模型与低质数据之间的失配提供了有效途径。在此基础上，稀疏、非负和层次等结构先验的引入，可以进一步提高模型的可解释性、鲁棒性和数据利用效率。本报告以结构化贝叶斯建模为主线，讨论其在可信特征学习与大模型稳健对齐中的应用。首先，通过将结构先验融入深度特征学习和决策过程，实现特征选择、语义解耦与不确定性量化，并利用高效变分推理提高模型的可扩展性。随后，针对对齐数据规模有限、噪声较多等问题，在参数调节与奖励优化过程中引入结构化先验，以改善大模型微调的泛化能力和奖励模型的稳健性，并进一步分析相关方法的泛化机理。

报告人：张龄月 东北财经大学

报告题目：ATM: An Aggregation Test of Moments Approach for Assessing High-Dimensional Normality

报告摘要：The Gaussian assumption is the most common and widely used distribution in statistical methodology. Various affine invariant tests for normality rely on the inverse of the covariance matrix and do not apply to high-dimensional data. The Gaussian moments are equal to the specific values, which implies an exclusive quantitative relationship. This paper introduces two novel indices, the co-third moment and the co-fourth moment, to characterize the shape of the distribution relative to the high-dimensional Gaussian family. Using these indices, two new tests with asymptotic properties under mild regularity conditions defining the subfamily for which tests are theoretically valid are proposed, which substantially avoid using the inverse covariance matrix and are easy to implement. By aggregating the strengths of the two tests and using the power enhancement technique, this study develops a more sensitive test of high-dimensional normality. Numerical studies and real data analyses provide evidence that the proposed methods achieve well-controlled size and competitive power, and illustrate their ability to detect departures from Gaussianity across different dimensional settings. The practical steps for empirically checking the regularity conditions are also discussed.

报告人：夏新涛 浙江大学

报告题目：Anytime-Valid Inference for Online Ranking of Large Language Models

报告摘要：Online evaluation of large language models increasingly relies on sequentially collected pairwise preferences, enabling human-aligned assessment and continuous data collection until closely performing models can be reliably distinguished. However, adaptive sampling and continuous monitoring invalidate classical fixed-sample inference, rendering existing ranking procedures largely heuristic. We propose SERPANT (Sequential E-value Ranking and Pruning via Adaptive Null Testing), a principled framework for online LLM ranking with anytime-valid guarantees. SERPANT formulates model comparison as a collection of pairwise hypothesis tests and constructs e-processes to ensure family-wise error rate control at any monitoring time. Anytime validity provides a theoretical justification for early stopping, enabling substantial cost savings from expensive human annotation. To improve efficiency, we introduce a novel tournament-based sampling strategy that adaptively selects comparisons based on past outcomes. The proposed framework is also extended to top-k model identification.

报告人：何凤翔 爱丁堡大学

报告题目：Bayesian Inference Forgetting

报告摘要：The right to be forgotten has been legislated in many countries but the enforcement in machine learning would cause unbearable costs: companies may need to delete whole models learned from massive resources due to single individual requests. We propose a {it Bayesian inference forgetting} (BIF) framework, the first that realizes the right to be forgotten in Bayesian inference. In the BIF framework, we develop forgetting algorithms for variational inference and Markov chain Monte Carlo. Theoretical analysis demonstrates that (1) our algorithms can provably remove the influence of single datums on the learned models; and (2) our algorithms have guaranteed generalizability. Experiments of Gaussian mixture models on the synthetic dataset and Bayesian neural networks on the real-world data verify the feasibility of our methods.

报告人：吴未迟 清华大学

报告题目：Bayesian Online Collective Anomaly and Change Point Detection in Fine-grained Time Series

报告摘要：Fine-grained time series data are crucial for accurate and timely online change detection. While both collective anomalies and change points can coexist in such data, their joint online detection has received limited attention. In this research, we develop a Bayesian model that captures time series with collective anomalies and change points, and introduce a recursive online inference algorithm to jointly detect the most recent collective anomaly and change point. Based on this, we further propose a second algorithm enhanced with a novel collective anomaly removal that effectively reduces the time and space complexities to linear with comparable performance. We establish theoretical guarantees in a tractable setting, including high-probability bounds on false alarms and detection delay. Extensive simulation experiments show that both of our algorithms achieve significantly higher accuracy and lower delay for both types of distributional shifts compared with multiple baseline models, including those focusing on change points or anomalies only. The usefulness and effectiveness of our methods are demonstrated by applications to stock market attention data and electricity consumption data.

报告人：田烱 浙江大学

报告题目：Conformalized Multiple Testing under Unknown Null Distribution with Symmetric Errors: an Empirical Bayes Perspective

报告摘要：This article addresses an aspect of a fundamental concern regarding the selection of null distributions in large-scale multiple testing. In modern data-intensive applications involving thousands or even millions of hypotheses, the theoretical null distribution of the test statistics often deviates from the true underlying null distribution, severely compromising the false discovery rate (FDR) analysis. We propose a conformalized empirical Bayes method using self-calibrated empirical null samples (SENS) for both one-sample and two-sample multiple testing problems under centered symmetric null errors. The new framework sidesteps the use of potentially erroneous theoretical null distributions and leverages conformal tools to achieve finite-sample FDR control, requiring only symmetry of the null error distribution rather than the stronger parametric assumptions common in existing empirical Bayes methods. Moreover, SENS incorporates structural insights from empirical Bayes into inference, exhibiting competitive empirical performance relative to existing approaches. We complement the finite-sample validity theory with a supporting power analysis under a specific Gaussian working model, and investigate the robustness of SENS to asymmetric errors and data-driven re-centering. Numerical experiments on both simulated and real data illustrate the strengths and limitations of the proposed method.

报告人: 张智涵 华东师范大学

报告题目: Nonparametric Statistical Inference for Permutation-Based Models: Optimality and Adaptivity

报告摘要: Structured binary matrices with an unknown ordering of candidates arise in crowdsourcing, educational assessment, and related problems, but existing work on permutation-based models has focused mainly on estimation rather than uncertainty quantification. We study nonparametric statistical inference for the identity of the best candidate in a noisy binary permutation-based model. Our two goals are to construct a confidence set that contains the latent best candidate with prescribed coverage and to test whether a prespecified candidate is truly the best. For confidence-set construction, we derive information-theoretic lower bounds over the general permutation class and propose a score-based procedure whose expected size is controlled at the optimal resolution order. We further show that shape constraints, specifically unimodal and more general s-modal structure, can dramatically sharpen inference. Under these assumptions, we develop an interval-scan procedure, establish its near-optimality up to logarithmic factors and prove its adaptivity to the unknown modality levels. For hypothesis testing, we obtain near-optimal separation guarantees under structured alternatives. Simulations show that the proposed procedures maintain validity and achieve sharper confidence sets and higher testing power than heuristic or unstructured baselines, and analyses of the EdNet and Eedi educational data sets illustrate the practical gains of shape-aware inference.

报告人: 包亚杰 复旦大学

报告题目: Tilted Conformal Prediction for Coverage Parity Control

报告摘要: We study fair predictive inference when the sensitive attribute is observed during training but unavailable at test time. Coverage parity is measured by the difference in conditional coverage rates between the two sensitive groups. Although conformal prediction provides finite-sample marginal coverage guarantees, it may exhibit substantial coverage disparities across groups. Motivated by a population oracle characterization, we propose a tilted conformal prediction method that adjusts the conformity score according to the estimated local group imbalance. The resulting prediction set preserves finite-sample marginal coverage and provides asymptotic control of the population coverage disparity without requiring access to the sensitive attribute at prediction time. Experiments demonstrate that the proposed approach improves coverage parity while maintaining competitive prediction efficiency.

报告人：饶昊宸 福建师范大学

报告题目：Renewable High-dimensional Expected Shortfall Regression

报告摘要：Expected Shortfall (ES) has become a core coherent risk measure in finance and statistics, and high-dimensional ES regression is crucial for characterizing heterogeneous tail risk with massive covariates. Existing offline methods for high-dimensional ES regression rely on access to full data, which fails under streaming data scenarios with sequential batch arrival and limited storage. To address this issue, this paper proposes a renewable estimation and inference framework for high-dimensional ES regression tailored to streaming data. By optimizing a surrogate loss function determined only by current data and historical information, the proposed procedure updates the estimator of ES regression coefficients without storing full raw data. Based on the online estimator, we design an online debiased estimator and further construct valid Wald-type confidence intervals using consistent variance estimation. Theoretically, we establish non-asymptotic error bounds for the online high-dimensional ES estimator and verify the asymptotic normality of the online debiased estimator. Extensive simulations show that the proposed method achieves estimation accuracy and inference performance comparable to the offline benchmark. Moreover, an application on the car insurance claim dataset demonstrates strong practical value in insurance risk management.

报告人：李铂桂 上海海关学院

报告题目：具有可分时空过滤器的固定效应半参数单指标面板模型的估计和检验

报告摘要：针对所提出的具有可分时空过滤器的固定效应半参数单指标面板模型，本文构建模型未知参数和单指标函数的截面拟极大似然估计和检验模型非线性的 F 检验统计量，并从理论和模拟两方面分别考察估计量和检验统计量的大样本性质和有限样本表现。研究结果表明所构建的截面拟极大似然估计具有相合性和渐近正态性， F 检验统计量渐近服从 χ^2 分布；估计量和检验统计量具有良好的有限样本表现；忽略误差潜在的时空相关性会导致估计结果偏误较大，且有效性下降。本文还将所提出的模型估计技术和检验方法用于实证分析我国重工业碳排放强度的影响因素。

报告人：李芬 福建农林大学

报告题目：Penalized Quadratic Inference Functions Estimation for Fixed Effects Partially Linear Single Index Spatial Error Model

报告摘要：This paper focuses on studying fixed effects partially linear single index spatial error model with correlated structure within individuals. By combining spline basis function approximation, variable transformation, penalty function and quadratic inference function, penalized quadratic inference functions estimators for unknowns are constructed. By appropriately selecting tuning parameters, we derive that the parametric estimators satisfy consistency and asymptotic normality, and the nonparametric estimator has the optimal convergence rate. Numerical simulation implies the estimates have excellent small sample performance. The proposed model is applied to analyze the driving forces of China's provincial digital economy development.

报告人：陈志勇 福建师范大学

报告题目：Latent Single-Index Models with Factor Structure for Multivariate Ordinal Data

报告摘要：We propose a latent semi-parametric model for multivariate ordinal data, in which the single-index models with factor structure are used to assess the effects of the latent covariates on the latent responses and to explore the covariance structure of the latent responses. Since the indices may vary from minus infinity to plus infinity, the traditional spline can not be applied directly to approximate the unknown link function. We employ a modified version to tackle this pervasive problem by transforming the indices into the unit interval via a continuously cumulative distribution function and constructing a novel combination of the spline. A fully Bayesian modelling is performed by facilitating efficient MCMC approach with free-knot splines to analyze the proposed model. We also consider a proposed distribution augmentation algorithm to sample from the conditional posterior of the correlation matrix. To accelerate the convergence, we make use of the partial-collapse and parameter expansion as well as reparameterization techniques in our algorithm. The performance of the proposed model and estimation method are checked through a simulation study and applied to a real dataset.

报告人：柏杨 上海财经大学

报告题目：Functional Graph Inference via Residual-Based Distance Covariance

报告摘要：Functional graphical models provide a powerful framework for uncovering network structures among random functions, with broad applications in neuroimaging, biomedical research, and environmental monitoring. We propose an edge-wise inference framework for recovering functional graphical models through conditional independence testing based on residual functional distance covariance (fDCov). For each candidate edge, the two target functional variables are regressed on the remaining nodes to obtain estimated residual processes, whose dependence is then assessed using a fDCov statistic. The resulting edge-wise p-values are combined through multiple-testing procedures to estimate the graph while controlling false discoveries. We establish the asymptotic theory of the proposed statistic when residual processes are estimated from discretely observed functional data, proving its asymptotic equivalence to an oracle statistic and deriving its null limiting distribution. To avoid direct evaluation of this complicated limit, we develop Satterthwaite approximation, permutation, and bootstrap calibration procedures, and establish the consistency of the bootstrap test. Extensive simulations demonstrate accurate size control and substantial power gains over competing methods in nonlinear and non-Gaussian settings. An application to EEG data further illustrates the practical utility of the proposed framework for functional network inference.

报告人：李婷 上海财经大学

报告题目：Efficient Online LLM Watermark Detection via Rao-Blackwellized E-Processes

报告摘要：As large language models (LLMs) are increasingly deployed, reliable and efficient mechanisms for distinguishing AI-generated text from human-written content have become essential. Statistical watermarking has emerged as a promising solution, yet most existing methods are typically fixed-horizon procedures, precluding valid early stopping in streaming generation. In this paper, we develop an efficient online watermark detection framework with anytime-valid inference based on Rao-Blackwellized e-processes, enabling recursive token-level evidence updates without storing the full history. In particular, we instantiate the framework for the Gumbel-max watermark and reduce the original token-level dependence testing problem to a pivot-induced sequential testing problem with an explicit null distribution. Theoretically, we prove anytime-valid type I error control under arbitrary optional stopping and establish positive asymptotic log-growth under watermarking, implying consistency of the proposed stopping rules. Simulations and experiments on real LLM-generated text demonstrate efficient online detection with rigorous anytime-valid guarantees.

报告人: 裘予琦 华东师范大学

报告题目: Budgeted Dynamic Monitoring for Timely Actionable Detection Under Action-Induced Observation

报告摘要: Clinical monitoring policies determine both which patients are measured and which deteriorations can be detected. This action-induced observation creates a policy-evaluation problem in which the measurement process is simultaneously a source of information, burden, and statistical coarsening. We develop a framework for budgeted dynamic monitoring based on threshold rules applied to a time-varying severity score. For each threshold, we define three target curves: a natural-process final-outcome value, a monitoring burden curve, and a timely detected-event yield. The detected-event target treats observation as part of the event being valued, while final-outcome ascertainment is handled as terminal coarsening. We estimate the curves using augmented inverse-probability scores with stacked-sandwich standard errors under low-dimensional working models, and we select thresholds subject to prespecified burden and detected-yield constraints. Selection is restricted to support-admissible thresholds and uses a one-sided feasibility margin to separate evaluable feasible rules from boundary-uncertain rules. Simulations show that the curve estimators are calibrated on the support-admissible grid, that the observed-process detected-yield estimator remains stable under informative visits, and that support and margin screening reduce false-feasible recommendations relative to naive plug-in selection. We apply the framework to laboratory monitoring for acute kidney injury during the first 48 hours of intensive-care stays in MIMIC-IV. The analysis of 56,686 ICU stays shows how mortality, specimen burden, and timely detected-AKI yield vary across monitoring thresholds, identifies budget-detection regimes with stable recommendations, and uses missed-event and action-mechanism sensitivity analyses to display the main observational vulnerabilities.

报告人: 张颖杰 华东师范大学

报告题目: Nodewise Loreg: Nodewise L_0 Regression for High-dimensional Sparse Precision Matrix Estimation

报告摘要: We propose Nodewise Loreg, a nodewise L_0 regression method for estimating high-dimensional sparse precision matrices. We establish its asymptotic properties, including convergence rates, support recovery, and asymptotic normality under high-dimensional sub-Gaussian settings. Notably, the Nodewise Loreg estimator is asymptotically unbiased and normally distributed, eliminating the need for debiasing required by Nodewise Lasso. For comparison, we also develop a debiased version of Nodewise Loreg, similar to the debiased Nodewise Lasso estimator. The asymptotic variances of the unbiased Nodewise Loreg estimator are upper bounded by those of both debiased Nodewise Loreg and debiased Nodewise Lasso estimators for Gaussian data, potentially offering more powerful statistical inference. Extensive simulations show that the unbiased Nodewise Loreg estimator generally outperforms the two debiased estimators in asymptotic normal behavior. Moreover, Nodewise Loreg surpasses Nodewise Lasso, CLIME, and GLasso in most simulations in terms of matrix norm losses, support recovery, and timing performance. Application to a breast cancer gene expression dataset further demonstrates Nodewise Loreg's superiority over the three L_1 -norm based methods.

报告人：徐礼柏 苏州大学

报告题目：Score-based Generative Models with Gamma Processes

报告摘要：Diffusion models are predominantly built on Gaussian perturbations, whose symmetry and light-tailed behavior provide substantial analytical and algorithmic convenience. However, this design choice can be restrictive when the target perturbation structure is asymmetric, skewed, or exhibits localized abrupt variations. We present Gamma-Noise Diffusion Models (GNDM), a non-Gaussian diffusion framework that replaces Gaussian perturbations with gamma-driven increments while preserving the familiar score-based training and reverse-time sampling interface. Our approach constructs a forward corruption process with composable gamma perturbations, derives an approximate reverse-time stochastic dynamics using a local expansion of the reverse conditional, and further develops a probability-flow ordinary differential equation (ODE) that enables deterministic sampling. We also introduce a weighted score-matching objective adapted to the conditional gamma perturbation family. Experiments on MNIST, CIFAR-10, and CelebA demonstrate that the proposed framework is practically trainable and capable of high-quality generation. On unconditional CIFAR-10 generation, our implementation achieves an FID of 2.58 and an IS of 10.02. These results suggest that diffusion modeling can be extended beyond Gaussian assumptions without discarding the core score-based generative modeling pipeline.

报告人：易凤婷 云南大学

报告题目：Variational Bayesian Estimation for Joint Models of Longitudinal and Interval-Censored Failure Time Data with Random Change-Points

报告摘要：In clinical follow-up studies, survival outcomes are often interval-censored, and covariate effects may exhibit abrupt changes across latent thresholds, reflecting heterogeneous disease progression. These features pose substantial challenges for statistical inference due to the intractable likelihood induced by interval censoring and the nonregularity introduced by random change-points. We develop a variational Bayesian approach for estimating random change-points in covariate effects within a joint modeling framework for longitudinal and survival data. The proposed method accommodates interval-censored event times and captures subject-specific change-points in the survival component while accounting for dependence through shared random effects. To address the resulting non-conjugacy, we consider both a non-conjugate variational Laplace approximation and a reparameterization-based gradient estimator for efficient posterior approximation. Simulation studies demonstrate that the proposed methods yield accurate estimation under a range of sample sizes and censoring levels, with the gradient-based approach offering notable computational advantages. An application to Alzheimer's disease data illustrates the ability to identify individualized change-points in the effect of APOE4 on disease progression.

报告人：赵波 长春工业大学

报告题目：Regularization Copula Link-based Semiparametric Regression Models with Bivariate Interval-censored Data

报告摘要：Bivariate interval-censored data with two correlated event times and sparse covariates being imprecisely observed within predefined intervals are common in biomedical research. We propose Copula link-based semiparametric regression models and formulate the association of the bivariate event through the Archimedean Copula function. The association parameter does not need to be known. The marginal distributions are modeled through semiparametric probit models, with the accelerated failure time model, proportional hazards, or odds model being a special case. We propose a penalized sieve estimation method that involves a negative sieve log-likelihood function plus a smoothly clipped absolute deviation penalty. The objective function is approximated as a quadratic regularization function, so the minimization is easy to implement. We show that the proposed variable selection method has oracle properties and the estimator is semiparametrically efficient. Extensive simulations are conducted to evaluate the performance of the proposed method in finite samples. The proposed methodology is applied to the Atherosclerosis Risk in Communities Study.

报告人：袁泉 云南大学

报告题目：Strongly Consistent Community Detection in Popularity Adjusted Block Models

报告摘要：The Popularity Adjusted Block Model (PABM) provides a flexible framework for community detection in network data by allowing community-specific node popularity. However, this flexibility increases model complexity and raises several unresolved challenges, particularly in adapting spectral clustering techniques effectively and achieving strong consistency in label recovery. To address these challenges, we first propose the Thresholded Cosine Spectral Clustering (TCSC) algorithm and establish its weak consistency under PABM. We then introduce a one-step Refined TCSC algorithm and prove that it achieves strong consistency under PABM, correctly recovering all community labels with high probability. We further show that a two-step Refined TCSC significantly accelerates the convergence of clustering error, especially in small-sample settings. Additionally, we propose a data-driven method for selecting the number of communities, which outperforms existing approaches under PABM. The effectiveness and robustness of our methods are validated through extensive simulations and real-world applications.

报告人：王川杰 上海财经大学

报告题目：Satellite-Informed Extreme Climate Risks and Volatility in Crop Futures

报告摘要：Extreme climate events increasingly threaten global food security and financial stability, yet existing risk measures are often aggregate or lagged. We develop the first real-time, high-frequency, crop-specific index of extreme climate risk (ECR) exposure using meteorological and satellite data. Applying this framework to major crop futures in China, we show that daily ECR exposure predicts next-day volatility, with effects strongest after the market opens, fading rapidly thereafter, and differing across crops with distinct climate sensitivities. Compared with a news-based climate risk index, our ECR measure provides more timely and informative signals. These findings offer a new approach to quantifying physical climate risk and have implications for both financial stability and food system resilience.

报告人：熊子扬 复旦大学

报告题目：Scalable Bayesian Calibration of High-Resolution Multi-Output Building Energy Models

报告摘要：Bayesian calibration of high-resolution building energy models is challenging because a weather-coupled Gaussian process (GP) must represent complex temporal responses while remaining computationally tractable. In this setting, emulator error can materially affect parameter inference. We propose a sequential workflow that treats selected time points and output types as tasks and uses building parameters as the only continuous inputs to a low-rank multi-task GP. GP hyperparameters are estimated before MCMC, and final predictions are obtained by rerunning EnergyPlus at calibrated parameters, with optional residual correction. In controlled studies, a selected configuration reduced daily-energy CVRMSE from 3.59% to 1.90% in approximately 12% of the runtime required by the selected comparison configuration. A controlled misspecification study achieved a mean corrected test CVRMSE of 5.25%, while six-output calibration reduced the reported parameter CVRMSEs by about 57% relative to the average single-output results. The workflow improves scalability, but does not provide the joint uncertainty propagation of a full KOH analysis.

报告人：毛若轩 复旦大学

报告题目：Fast Bayesian inference on spatial blind source separation with the Relaxed Threshold Gaussian Process

报告摘要：Imaging techniques, such as functional Magnetic Resonance Imaging (fMRI), have been widely applied in the field of neuroimaging. While the analysis of fMRI data is based on the measurement of the Blood Oxygen Level Dependent (BOLD) signal, the BOLD signal is often affected by various sources of noise, such as noise during acquisition (respiration). So Blind Source Separation (BSS) methods are commonly used to help extract the BOLD signal from potential source signals. Popular statistical approaches of BSS, like Independent Component Analysis (ICA), typically assume spatial independence of the data. However, the independence assumption is often violated in brain imaging due to the pre-processing step of the data which is spatial smoothing, computing a weighted average of surrounding voxels. Recent studies have proposed methods that no longer require assumptions of spatial or temporal independence. For instance, the Bayesian Spatial Blind Source Separation via the Thresholded Gaussian Process (BSS-BSP) uses a novel thresholded Gaussian process as a prior in a Bayesian framework, reducing reliance on strict assumptions. However, this method suffers from high computational complexity because the full conditional distributions of its parameters are not available in closed-form. So our study proposes a novel Bayesian framework with a new prior named the Relaxed Threshold Gaussian Process (RTGP) which is more effective when analyzing fMRI data. As it ensures the full conditional distribution of each parameter is tractable, allowing more computationally efficient MCMC methods.

报告人：时琛 复旦大学

报告题目：Valuing Sustainability in China

报告摘要：Understanding how financial markets respond to climate transition risk is critical to asset pricing and portfolio choice. This paper provides a systematic analysis of China's sustainable fund market and develops a daily news-based measure of climate transition risk using a large language model (LLM)-assisted approach. Using this index, we document a pronounced dual impact of transition risk in China. Broad equity market indices respond negatively to climate policy shocks, whereas aggregate climate fund portfolios exhibit positive average exposure. However, we show that passive broad-based allocation across climate funds is ineffective for hedging, as the aggregate response lacks statistical robustness. To address this limitation, we construct Climate-Efficient Portfolios (CEP) by maximizing the Sharpe ratio subject to transition-risk exposure. In addition, exploiting the high correlation structure among Chinese climate funds, we propose a parsimonious Max5 strategy that achieves out-of-sample performance comparable to more complex optimization-based portfolios. These insights contribute to a deeper understanding of the evolving role of climate funds in China's sustainable development.

报告人：罗嘉雯 华南理工大学

报告题目：Forecasting Downside Betas with Machine Learning Models

报告摘要：Against rising global geopolitical uncertainties and heightened financial market volatility, the dynamic prediction of systemic downside risks is particularly critical for China's retail-dominated stock market, which is highly susceptible to investor sentiment fluctuations. Using 5-minute high-frequency data of CSI 300 constituent stocks spanning 2010 to 2024, this paper constructs daily downside Beta risk measures and develops forecasting models based on a multi-dimensional predictor set. The comprehensive factor system covers lagged risk terms, corporate accounting fundamentals, market technical indicators, macroeconomic variables, and industry dummy variables. To address the time-varying characteristics, multicollinearity issues, and complex nonlinear interactions inherent in intraday risk dynamics, we adopt Random Forest (RF) and Gradient Boosted Regression Trees (GBRT) to integrate multi-source information and screen effective predictive factors. Empirical results demonstrate that incorporating multi-dimensional information substantially improves the out-of-sample forecasting performance of downside Beta measures, and tree-based machine learning models consistently outperform conventional linear forecasting approaches. Specifically, RF achieves outstanding performance in numerical prediction accuracy and directional forecasting capability, while GBRT shows unique advantages in constructing market-neutral portfolios to realize effective risk control. In addition, only the RF model produces prominent long-short return spreads. Compared with traditional linear models, machine learning frameworks respond more rapidly to extreme downside market shocks and maintain more stable and robust predictive performance amid market turbulence.

报告人：龚欢 国防科技大学

报告题目：Robust and Efficient Inference for GARCH Models with Heavy-Tailed and Asymmetric Innovations

报告摘要：Financial returns often exhibit heavy tails and asymmetry, which complicate inference in GARCH models, especially when strict stationarity may fail. We develop a generalized non-Gaussian quasi-maximum likelihood estimation framework that combines robust moment normalization with a fitted non-Gaussian quasi-likelihood. The moment normalization weakens moment requirements and improves robustness to heavy-tailed innovations, while the fitted quasi-likelihood adapts to tail thickness and asymmetry through scale and shape parameters estimated from preliminary standardized residuals. Under fractional moment conditions, we show that the estimator is consistent and asymptotically normal for the full parameter under strict stationarity and for the coefficient parameter in the boundary and explosive cases. We further develop a strict stationarity test and examine two sources of asymptotic efficiency. For model diagnostic checking, we develop a score-based portmanteau test aligned with the fitted non-Gaussian quasi-likelihood and show that it has a common null distribution across the stationary, boundary, and explosive cases. Monte Carlo simulation studies and exchange rate applications illustrate the finite-sample performance and practical implications of the proposed methods.

报告人：马宇恒 华东师范大学

报告题目：Locally Private Estimation with Public Features

报告摘要：We initiate the study of locally differentially private (LDP) learning with public features. We define semi-feature LDP, where some features are publicly available while the remaining ones, along with the label, require protection under local differential privacy. Under semifeature LDP, we consider three fundamental estimation problems: non-parametric density estimation, classification, and regression. Given the smoothness assumption, we show that the minimax convergence rate is significantly improved compared to classical LDP. Then, we propose HistOfTree, an estimator that fully leverages the information contained in both public and private features. Theoretically, HistOfTree reaches the mini-max optimal convergence rate. Empirically, HistOfTree achieves superior performance on both synthetic and real data. We also explore scenarios where users have the flexibility to select features for protection manually. In such cases, we propose an estimator and a data-driven parameter tuning strategy, leading to analogous theoretical and empirical results.

报告人：江声 港中文深圳

报告题目：Errors-in-variables Gaussian Processes for Mixed-input Regression

报告摘要：Gaussian process regression with mixed inputs typically treats ordinal covariates as arguments of the response surface, through level covariance structures or deterministic embeddings. We study a different setting in which ordinal variables are coarsened or noisy measurements of latent continuous covariates that govern the response, while direct continuous measurements are available only for a calibration subset. We propose an errors-in-variables Gaussian process (EIV-GP) that places a GP prior on the response surface $f(x, u)$, where x is observed and u is latent, and links u to the observed ordinal vector through an explicit measurement model. Thus, an ordinal vector induces a conditional distribution over latent inputs. Prediction integrates over this distribution, propagating coarsening uncertainty and allowing non-Gaussian and heteroskedastic predictive laws even with homoskedastic Gaussian observation noise. We show that marginalizing the latent-input GP over the conditional latent distribution yields a valid mixed-input GP kernel, with deterministic embedding kernels recovered as point-mass special cases. Calibration measurements anchor the latent scale; without calibration, the latent coordinate is a working scale whose physical interpretation is not identified. For inference, we integrate out the finite GP vector and develop posterior samplers that jointly update kernel parameters, measurement-model parameters, and response-informed latent covariates. Simulations show improved continuous ranked probability score, interval scores, and coverage relative to deterministic embedding GPs, with further gains from calibration measurements.

报告人：汪时嘉 上海科技大学

报告题目：A Multifidelity Approximate Bayesian Computation with Pre-filtering

报告摘要：Approximate Bayesian Computation (ABC) methods often require extensive simulations, resulting in high computational costs. This paper focuses on multifidelity simulation models and proposes a pre-filtering hierarchical importance sampling algorithm. Under mild assumptions, we theoretically prove that the proposed algorithm satisfies posterior concentration properties. Furthermore, we characterize the error upper bound and establish the relationship between algorithmic efficiency and the pre-filtering criteria. Additionally, we provide a practical strategy to assess the suitability of multifidelity models for the proposed method. Finally, we develop a multifidelity ABC sequential Monte Carlo with adaptive pre-filtering strategy. Numerical experiments are used to demonstrate the effectiveness of the proposed approach.

报告人：冯龙 南开大学

报告题目：High Dimensional Alpha Test for Linear Factor Pricing Model with L_q -norm

报告摘要：We consider testing zero pricing errors in high-dimensional linear factor pricing models. Existing methods are mainly based on either an L_2 statistic, which is effective under dense alternatives, or an L_∞ statistic, which is powerful under very sparse alternatives. To bridge these two regimes, we develop a class of L_q -based tests for finite q , including the practically useful L_4 and L_6 cases. We show that larger q , leads to greater sensitivity to sparse alternatives. We further establish the asymptotic independence between the L_∞ statistic and the L_q statistic for any finite q , which motivates a Cauchy combination test that adapts to a broad range of sparsity levels. Simulation studies and a real-data analysis show that the proposed methods are more robust to the unknown sparsity of the alternative and can outperform existing procedures in finite samples.

报告人：李赛 清华大学

报告题目：Efficient Machine Unlearning with Minimax Optimality

报告摘要：There is a growing demand for efficient data removal to comply with regulations like the GDPR and to purge the statistical influence of biased or corrupted data. This has motivated the field of machine unlearning, which aims to eliminate the influence of specific data subsets without the cost of full retraining. In this work, we propose a statistical method for machine unlearning with generic convex and smooth loss functions. We establish the minimax optimality of the proposed algorithm in linear models when only the pre-trained estimate, forget samples, and a small subsample of the remaining data are available. Our results reveal that the estimation error decomposes into an oracle term and an unlearning cost determined by the forget proportion and the forget model bias. We further establish the results for logistic regression and in moderate dimensions by considering Ridge-type regularizations. We conduct numerical experiments and real-data applications on Yelp review data and UK Biobank data to validate the accuracy and computational efficiency of our proposals.

报告人: 李易诚 华东师范大学

报告题目: Minimax Lower Bounds in Differential Privacy

报告摘要: Minimax lower bounds quantify the unavoidable loss of statistical accuracy imposed by differential privacy. This talk presents an information-theoretic approach based on van Trees inequalities and the contraction of Fisher information under privacy constraints. In the central zCDP model, a private van Trees inequality converts the privacy budget into an upper bound on the information carried by a private estimator. Combined with priors tailored to bandable covariance classes, it yields matching minimax lower bounds under operator and Frobenius losses and identifies the dimension-dependent cost of privacy. We then extend this perspective to federated learning under clientwise sample-level zCDP, allowing arbitrary public-transcript interaction and repeated reuse of local samples. Scalar and matrix federated van Trees inequalities capture the total and directional information contributed by homogeneous and heterogeneous clients. Applications to mean estimation, linear regression, nonparametric regression, and functional mean estimation establish, together with existing upper bounds, minimax rates over fully interactive public-transcript protocols. These results illustrate how privacy contracts statistical information and why unrestricted interaction does not improve the minimax rates in these problems.

报告人: 李竞阳 复旦大学

报告题目: Online Tensor Learning: Computational and Statistical Trade-offs, Adaptivity and Optimal Regret

报告摘要: High-dimensional tensor data arise naturally in a wide range of applications, including spatiotemporal forecasting, image processing, scientific computing, and multimodal data analysis. Low-rank tensor models provide an effective way to reduce the computational and storage costs associated with such data, but conventional offline algorithms typically require repeated access to the entire dataset and are therefore unsuitable for streaming-data and real-time prediction settings. In this talk, I will introduce an online Riemannian gradient descent algorithm, referred to as oRGrad, for learning low-rank tensors from sequentially arriving observations. By exploiting the geometric structure of the low-rank tensor manifold, the method performs computationally efficient updates using only newly observed data while preserving the underlying low-rank structure. I will discuss its stability and convergence properties, the computational and statistical trade-offs induced by step-size selection, its adaptivity to different learning regimes, the preservation of incoherence in online tensor completion, and optimal regret guarantees for online prediction. Finally, I will briefly illustrate the method through an application to forecasting the solar F10.7 index using tensor-valued observations of ionospheric total electron content.

报告人：张思亮 华东师范大学

报告题目：A Cumulative Ordered Spike-and-Slab Prior for Adaptive Dimension Selection in Joint Latent Space Models

报告摘要：Network models are increasingly vital in psychometrics for analyzing relational data, which are often accompanied by high-dimensional node attributes. Joint latent space models (JLSM) provide an elegant framework for integrating these data sources by assuming a shared underlying latent representation; however, a persistent methodological challenge is determining the dimension of the latent space, as existing methods typically require pre-specification or rely on computationally intensive post-hoc procedures. The key innovation of this work is a cumulative ordered spike-and-slab (COSS) prior, which we incorporate within a Bayesian joint latent space modeling framework. This prior enables the latent dimension to be inferred automatically and simultaneously with all model parameters. We develop an efficient Markov Chain Monte Carlo (MCMC) algorithm for posterior computation. Theoretically, we establish that the posterior distribution concentrates on the true latent dimension and that parameter estimates achieve Hellinger consistency at a near-optimal rate that adapts to the unknown dimensionality. Through extensive simulations and three real-data applications, we demonstrate the method's superior performance in both dimension recovery and parameter estimation. Our work offers a principled, computationally efficient, and theoretically grounded solution for adaptive dimension selection in psychometric network models.

报告人：郭金鑫 中央民族大学

报告题目：基于竞争风险模型的中间跳过项目缺失与作答反应联合建模

报告摘要：项目反应理论模型是教育评估和心理测量中的一个重要工具，其研究对象主要是二值或离散型反应变量。但是，在实际应用中，反应缺失是一种非常普遍的现象，特别是在大规模测验评估中。例如，在一些情况下，被试虽然已经看到项目并且有时间思考，但仍可能会选择中间跳过一些项目，这样就会产生中间跳过的项目（omitted item，简称OI）缺失。大量的研究表明，OI引起的缺失，往往与被试的能力有关，因而是不可忽略的缺失。对于这样的缺失，如果处理不当，可能会导致不准确的统计推断，从而影响测试的有效性。对于存在OI缺失的数据，本研究基于竞争风险模型，提出了一个RA、RT和中间跳过的项目所用时间（omission time，简称OT）的联合模型。在这个方法中，RT和OT是相互竞争的。也就是说，这两个时间中较短的那个会被观测到，较长的时间则会删失。这种竞争关系导致了被试作答或中间跳过项目。模拟研究表明，在本研究提出的这个方法下，不仅可以得到准确的参数估计，而且稳健性可以得到验证。为了对比分析提出的方法和其他备选方法，本研究进一步分析了2015年PISA科学测试的数据。

报告人：许鑫 中央民族大学

报告题目：协变量和层级学习轨迹的高阶隐马尔可夫模型的多值扩展

报告摘要：在教育测评实践中，测试题目常常采用多级评分方式，以更细致地诊断学生的表现。然而，现有纵向认知诊断模型主要聚焦于二分反应数据，通常需要将多级评分数据二值化处理，因而难以充分利用多级评分所包含的诊断信息。为解决这一问题，本文提出一种适用于多级反应数据的纵向认知诊断模型。

报告人：欧阳靖 香港大学

报告题目：Statistical Analysis of Large-scale Item Response Data under Measurement Non-invariance

报告摘要：International Large-Scale Assessments collect valuable data on educational quality and performance across countries, enabling education systems to share effective techniques and policies. A key analytical tool is the Item Response Theory (IRT) model, which measures individuals' latent traits such as skills and abilities. However, a major challenge arises from Differential Item Functioning (DIF), where different groups (e.g., genders and countries) may have different probabilities of correctly answering the items after controlling for individual latent abilities. Ignoring or improperly accounting for DIF when calibrating the IRT model can lead to severe biases in the estimated performance distributions, which may further distort the ranking of the groups. Unfortunately, existing methods cannot guarantee the statistically consistent recovery of the group ranking without unrealistic assumptions for ILSA, such as the existence and knowledge of reference groups and anchor items. To fill this gap, we propose new approaches to DIF analysis across multiple groups. The approaches are computationally efficient and statistically consistent, without making strong assumptions about reference groups and anchor items. The proposed methods are applied to PISA 2022 data from the mathematics, science, and reading domains, providing insights into their DIF structures and performance rankings of countries.

报告人：伍书缘 上海财经大学

报告题目：Bias-Correction for Privacy-Protected Spatial Autoregressive Models with Application to Restaurant Network Analysis

报告摘要：Spatial autoregressive (SAR) models and their extensions are important tools for studying network effects. However, with an increasing emphasis on data privacy, data providers often implement protection measures that render standard SAR models inapplicable. In this study, we introduce a privacy-protected SAR model that incorporates noise into both the response and covariates to meet privacy requirements. With noise present in both components, the traditional quasi-maximum likelihood estimator becomes difficult to compute because the likelihood function cannot be directly formulated. To bypass this hurdle, we begin with a pseudo-likelihood approach, initially omitting the noise in the covariates. A Newton-Raphson algorithm is then applied to compute the estimator; however, the estimator is biased. To address this, we propose a bias-corrected Newton-Raphson-type algorithm that simultaneously accounts for noise in both the response and covariates. We further show, under appropriate regularity conditions, that the resulting estimator is consistent and asymptotically normal. To further enhance computational efficiency, we also develop a bias-corrected least squares estimator. Several extensions are discussed, and the finite-sample performance of the proposed methods is evaluated through extensive simulations. We apply the proposed methodology to restaurant transaction data from a third-party payment platform. Our method identifies a statistically significant competitive network effect among restaurants and further reveals meaningful restaurant-customer interaction patterns.

报告人：田琳琳 东华大学

报告题目：Optimal Investment Problem in A Renewal Risk Model with Generalized Erlang Distributed Interarrival Times

报告摘要：This paper explores the optimal investment problem of a renewal risk model with generalized Erlang distributed interarrival times. The phases of the Erlang interarrival time are assumed to be observable. The price of the risky asset is driven by the constant elasticity of variance model (CEV) and the insurer aims to maximize the exponential utility of the terminal wealth by asset allocation. By solving the corresponding Hamilton-Jacobi-Bellman (HJB) equation, we establish the concavity of the value function and derive an explicit expression for the optimal investment policy when the interest rate is zero. When the interest rate is nonzero, we obtain an explicit form of the optimal investment strategy, along with a semi-explicit expression of the value function, whose concavity is also rigorously proven.

报告人：赵花丽 华中科技大学

报告题目：Generalizing Transfer Learning: A Flexible Doubly Robust Estimation Approach for Missing Data

报告摘要：Large-scale population-level datasets, such as the UK Biobank and the All of Us Research Program, often lack covariates needed for a specific analysis, such as genetic or lifestyle measures, while related studies measure them. This creates a cross-population missing data problem in which covariates are completely unobserved in the target population, rather than partially missing within one dataset. We propose an augmented transfer regression learning method for this setting. The key identifying condition is a sub-population shift assumption: the joint distribution of the outcome and observed covariates may differ across source and target populations, but the conditional distribution of the missing covariates given observed variables is invariant. We combine importance-weighted estimating equations with imputation terms for first- and second-order moments of the missing covariates. The resulting estimator is doubly robust, remaining consistent if either the density ratio model or both imputation models are correctly specified. It is root-n consistent and asymptotically normal, and attains the semiparametric efficiency bound when both nuisance models are correctly specified.

报告人：张国宇 北京大学

报告题目：A Nonparametric Potential Outcomes Framework for Bidirectional Causal Inference

报告摘要：Identification of causal effects is a central objective across many scientific disciplines. While much of the existing causal inference literature focuses on unidirectional causal relationships, bidirectional causal relationships arise naturally in many empirical settings and introduce additional challenges for both identification and estimation. Although there is a broader literature on simultaneous equations and reduced-form identification, work that directly studies bidirectional causal effects has largely proceeded under parametric structural assumptions, which can be restrictive and difficult to generalize. In this paper, we develop a fully nonparametric potential outcomes framework for bidirectional causal inference. The framework accommodates both discrete and continuous variables and yields nonparametric identification results under weak conditions. We then propose corresponding nonparametric estimators based on these identification results. Simulation studies show that the proposed estimators perform well in finite samples.

报告人：朱蔚萱 厦门大学

报告题目：DAG Trend Filtering for Genomic Denoising via Higher-Order Bayesian Networks and DAG Shrinkage Processes

报告摘要：Graph-based denoising is a critical preprocessing step for analyzing noisy data, particularly in genomic applications where gene regulatory networks exhibit inherent directional dependencies. This paper introduces a directed acyclic graph trend filtering framework that leverages novel higher-order Bayesian networks and graphical shrinkage processes to enhance local adaptivity in signal smoothing along the directed edges of a graph. Unlike traditional graph trend filtering, which is based on undirected graphs, the proposed method explicitly respects the directional structure of graphs, improving interpretability and accuracy in capturing dependencies. We employ a Hamiltonian Monte Carlo algorithm for efficient posterior inference. Through simulations and genomic applications, the proposed method outperforms a state-of-the-art graph trend filtering algorithm in terms of mean squared error reduction and signal-to-noise ratio improvement, demonstrating its utility in recovering true signals while accounting for meaningful structural information.

报告人：张艺馨 中国科学技术大学

报告题目：Merits and Challenges of Plasma Proteomics on Association Replicability

报告摘要：Large-scale plasma proteomics enables investigation into the molecular basis of brain-related traits, including cognition, mental health and brain structure. However, the replicability of protein–phenotype associations remain insufficiently characterized. Using proteomic, imaging, and phenotypic data from 52,632 UK Biobank participants, we evaluated the internal replicability of protein–phenotype associations by a directional consistency criterion with random subsampling. Overall, 77.4% of cognition and mental health traits showed high proteome-wide replicability. Among brain structural measures, total cortical surface area and volume showed higher replicability than thickness across both hemispheres and over 20 cortical regions. Matched-cohort multi-omics analyses enabled replicability comparisons across molecular layers. Proteomics, the layer more proximal to phenotypic variation, had higher replicability than genomics across all traits and metabolomics for over half. We identified missingness, dilution level, proportion of samples below the LOD, and sample size as influential factors affecting replicability, and developed an exploratory framework to estimate replicability for future panels. At the protein level, we identified 11 highly replicable pleiotropic proteins across cognition and mental health traits. Furthermore, assay variability was associated with protein-level replicability. Collectively, these findings demonstrate the potential of plasma proteomics for mapping highly replicable biomedical signatures while highlighting challenges to sustaining replicability in future proteomic studies.

报告人：陈奕帆 香港浸会大学

报告题目：A Recipe for Causal Graph Regression: Confounding Effects Revisited

报告摘要：AI systems learning from complex networks (graphs) often struggle when faced with unfamiliar data. While "causal learning" helps these systems focus on true causes, making them more robust, it has primarily been applied to categorizing items rather than predicting specific numbers—a tougher challenge for AI on graphs. Our research introduces a new method for applying causal learning to predict numbers on graphs. This approach addresses misleading information, known as "confounders," by adapting established techniques. Furthermore, we utilize "contrastive learning." This enables us to successfully extend causal methods, which were previously specific to classification (sorting items), to the task of regression (predicting numerical values). The result is AI that can make more reliable numerical predictions on graphs, even as data changes. This is crucial for developing trustworthy AI in fields like materials science or economics, where data is inherently complex and ever-evolving.

报告人：张如谦 清华大学

报告题目：Bayesian Environment Invariant Regression

报告摘要：The availability of data from multiple heterogeneous environments has motivated methods that remain reliable under distributional shifts. When the joint distribution of response and predictors varies across environments, the response may still depend on a subset of predictors through an invariant mechanism. Existing methods typically assess candidate invariant sets through pooled stability criteria, treating environmental variation as nuisance. In this paper, we propose a Bayesian framework that explicitly separates a shared response mechanism from environment-specific or response-dependent associations, exploiting heterogeneity as evidence for structure learning. A competitive spike-and-slab prior is designed to force each predictor to compete between invariant and non-invariant spurious effects. Under a tractable working model, we establish invariant model selection consistency and posterior contraction for invariant coefficients. We further study the presence of irrelevant predictors, characterize posterior concentration on an equivalent invariant class, and introduce a post-selection refinement that consistently recovers the minimal invariant model. Simulations and a real application illustrate the robustness and finite-sample efficiency of the proposed method.

七、承办单位简介

复旦大学管理学院

开智求真，拓新领变。自 1917 年复旦大学首创商科伊始，开拓与创新的精神便深植于复旦大学管理学院的基因之中。1929 年成立商学院，开创中国现代商学教育先河。1952 年，全国高校院系调整，工商管理学科教育中断。改革开放之后，复旦大学率先恢复管理教育，1977 年开始招收管理学科专业学生，1979 年成立管理科学系，1985 年恢复组建管理学院。2025 年，学院迎来恢复建院四十周年。百年征途，始终扎根于中国实践，持续对标国际前沿，复旦大学管理学院一直屹立于中国管理教育发展的潮头。

学科体系

经过近 40 年的发展壮大，学院目前设有 8 个系，拥有一级学科博士点 5 个，二级学科博士点 12 个（含自设博士点），一级学科硕士点 5 个，二级学科硕士点 18 个（含自设硕士点）以及工商管理硕士（MBA）/高级管理人员工商管理硕士（EMBA）、会计硕士（MPAcc）、国际商务硕士、金融硕士及应用统计硕士 5 个专业学位硕士点，本科专业 7 个，博士后科研流动站 4 个。2018 年，在教育部学位与研究生教育发展中心公布的全国首次专业学位水平评估结果中，复旦大学工商管理类别获评“A+”。

师资力量

截止 2026 年 6 月底，全院共有教师 164 人，其中教授 62 人、副教授 71 人。98.8% 的教师具有博士学位，其中取得境外博士学位的有 80 人，占总人数的 48.8%。

科研实力

学院涌现出一批卓越的学术带头人，形成了十余个高质量的学术研究团队，积累了一批优秀的学术专著、论文和专利。不仅在全球公认的 UTD 和 FT 国际高等级学术期刊上发表的排名达到世界一流中游水平，而且在国家自然科学基金项目评审中展现出卓越的科研竞争力。在 UTD 发布的“全球商学院科研百强榜（2021-2025）”中，学院位列全球第 65 位，连续 11 年全球百强榜排名持续上升。

人才培养

学院严谨的治学态度和良好的学习氛围，每年都吸引了众多考生前来深造，培养了一批批管理精英。截至 2026 年 7 月，学院有本科生、博士生、专业学位硕士项目（包括 MBA、EMBA 等）各类在读学生达 6000 人。目前学院校友已近 68000 人。高层管理教育中心每年培养近 4500 名学员，迄今共计组织近 72000 名企业管理者参与培训，累计服务企业超 850 家。

国际交流

学院恢复建院以来，广泛推行“走出去、引进来”的国际化战略，积极拓展国际交流，与美国、意大利、挪威、英国、法国、新加坡、澳大利亚、韩国、香

港等 32 个国家和地区的 116 所著名高等院校和研究机构建立了多种形式的交流、合作关系。国际化战略已经覆盖到了学生、师资、科研、课程设置、企业联系、国际合作研究等各个层面。

权威认证

为了拓展更为丰富的全球化资源，加强与各国优秀商学院间的交流和科研合作，学院还积极加入诸多国际顶尖的合作组织。2007 年成为 PIM (Partnership in International Management) 成员。2009 年成为 GMAC (美国商学院协会研究生入学管理委员会) 会员。2011 年参与发起成立 CoBS (商学院商业与社会全球联盟)。2012 年作为创始会员加入 GNAM (全球高端管理联盟)。2009 年学院首次取得 EQUIS (欧洲质量发展认证体系) 国际认证并于 2012 年、2017 年、2023 年再认证中连续三次获得五年期认证。2010 年学院通过 AACSB (国际精英商学院协会) 认证，于 2015 和 2020 年的再认证中连续两次获得五年期认证，并于 2025 年再认证中获得六年期的认证。

国际声誉

在英国《金融时报》(FT) 2025 年全球 EMBA 项目排名中，学院创造历史新纪录，实现全项目、多维度的综合排名全球领先。复旦大学—华盛顿大学 EMBA 项目荣登全球第 1，复旦大学 EMBA 项目位列全球第 8，复旦大学—香港大学 IMBA 项目全球第 19，复旦大学—BI (挪威) 国际合作 MBA 项目全球第 21，4 个参评项目全部进入全球 25 强，学院由此成为全球唯一拥有 4 个全球 25 强项目的商学院。与此同时，在 FT 2025 年全球 MBA 项目排名中，复旦大学 MBA 项目位列全球第 30 位。

企业合作

学院历来将与企业的合作贯穿于人才培养和学院发展的每一个环节，从而更好地回馈社会、服务社会。企业通过校园宣讲会、招聘会、实务课程、实习基地、iLab、D-Lab 等形式参与学生职业兴趣小组、搭建职业信息平台等，展示企业品牌与文化、选拔人才。学院发起了管理大师论坛、问学讲堂、博思荟、聚劲论坛等各种论坛讲座，使之成为院企合作的纽带。学院还设立了专职团队，全方位对校企合作关系进行高效、有序的管理。众多海内外著名企业与学院开展了企业参访、学生实践、论坛合作、人才定向培养、案例研究、产学研合作、资金与物资捐助等多种形式的合作，不断加强业界和学界的互动与交流，支持学院发展。

科创战略

2020 年，学院在中国管理教育界率先启动管理赋能科创的“科创战略”，提出“无科创 无未来”的创新性口号。五年来，学院教师深入科创企业调研，开展原创研究，编写中国首部科创企业管理教材，开发科创企业管理系列案例，构建科创数据平台，开办“复旦科创企业家营”、“复旦 MBA 科创青干营”，升级并主办“长三角聚劲科创大赛”，发起成立科创企业发展生态联盟，搭建科创投融资助力平台。学院在科创管理理论研究、科创管理人才培养、赋能科创生态圈建设等方面所取得的成绩，赢得了业界好评，也引领着中国管理教育面向未来的创新与发展。

政立院区

2024 年正式启用的政立院区，以“思想新策源、文化新地标、社区新生态、建筑新典范、教育新高地”为核心理念，成功实践了大学校区、科技园区与公共社区的“三区联动”。它不仅是智慧化的教育建筑与开放创新空间，更是管理教育的高地与全球合作的重要舞台，为学院建设世界一流商学院奠定了坚实的空间基础。站在恢复建院四十周年的新起点，复旦大学管理学院正以更宏阔的视野与更坚定的信念，扎根于中国波澜壮阔的管理实践，贡献于全球商业文明的发展，致力于培养能够引领中国产业走向世界、推动全球共同增长的领导者，稳步迈向引领科创时代的世界一流商学院。

统计与数据科学系

历史简介：

复旦大学统计学专业成立于 1938 年，开创了我国大学中设立统计学系之先例。上世纪 80 年代初，在国家统计局支持下，复旦大学获批建立数理统计专业，成为全国最早恢复的数理统计专业之一。1984 年，成立统计运筹系，4 年后归属管理学院，1999 年起统计运筹系更名为统计学系。近年来，随着人工智能与数据科学的发展，为顺应学科发展趋势，体现学科建设方向，统计学系于 2021 年 8 月更名为统计与数据科学系。

复旦大学统计学科具有悠久的历史 and 优良的学术传统。在 1938 年统计学系成立之时，就云集了当时一流的统计学者二十余人。新中国成立以后，一大批杰出统计学者，如我国首位有重大国际影响的统计学家吴定良教授、对我国统计科学和军事医学做出过卓越贡献的薛仲三教授、著名概率统计学者吴立鹏教授、汪嘉冈教授、李贤平教授、最先把生存分析引入国内的郑祖康教授等，都在复旦统计专业执教授业、辛勤耕耘，奠定了复旦统计学科的扎实基础和优良传统。复旦数学与统计学科也培养出了多位国际顶尖的统计学者，如范剑青教授、应志良教授、孟晓犁教授、何旭铭教授等。

师资力量：

统计与数据科学系目前有全职教师 22 名。其中教授 6 位、副教授 9 位、助理教授 6 位、青年副研究员 1 位。教师中 95% 拥有博士学位，40% 拥有海外博士学位，45% 具有博士生导师资格。

科研成果：

近五年来（2021-2025），统计与数据科学系在国内外学术期刊上发表论文 158 篇（国际期刊 137 篇，包括在 JASA、JRSSB、AOS、Biometrika 上发表 22 篇，国际会议 11 篇，包括机器学习三大顶会 7 篇，国内期刊 10 篇）；承担国家自然科学基金 25 项（包括优秀青年科学基金项目 3 项、重点项目 1 项，2025 年获优秀青年科学基金项目立项 1 项），国家“千人计划”青年项目 1 项，省部级项目 6 项，横向项目 9 项；出版教材 1 部。获得教育部自然科学奖、教育部高校科研优秀成果奖、国家统计局统计科学技术进步奖以及国家精品课程等。

人才培养：

本科项目：上世纪 80 年代初复旦大学获批建立数理统计专业，于 1983 年招收第一批本科生，成为国内最早恢复招收本科生的三所高校之一。2021 年管理学院统计学专业获批成为国家级一流本科专业建设点。2025 年统计与数据科学系参与教育部统计学“101 计划”本科教材建设工作。统计专业本科生多次获得国家奖学金、上海市优秀毕业生称号，近三年来毕业生有超过 80% 国内外知名大学继续深造。

硕博项目：复旦大学统计专业是在全国最早设立概率论与数理统计硕士和博士点的高校之一。1985 年概率统计专业设立博士点，2000 年统计学系增设统计学硕士点，2011 年获全国首批统计学一级学科博士学位授予权。统计专业博士毕业生去到多所国内顶尖大学任教或赴业界知名企业工作。

数据科学与商务分析项目：2020 年复旦大学管理学院启动应用统计（数据科学与商务分析）专业硕士项目，以着眼未来、放眼国际、立足上海为培养目标，对接国际顶尖商学院商务分析项目培养体系，依托管理学院优势，培养能够应用定量方法分析复杂商务数据，设计人工智能算法，解决商务领域中的决策与管理问题，同时能够为政府和企业等提供数据分析与人工智能开发及应用的高端人才。项目毕业生去到科技互联网、金融、咨询等诸多行业的知名企业就业。

学科影响力：

复旦大学统计学科在“软科”世界一流大学学术专业排名中，2023 年位于 35 位、2024 年位于 48 位，2025 年位于 51-75 位，在国内高校中排名分别为第 1、第 3 和第 4 位。在 QS 世界大学统计与运筹专业排名中，2025 年位于 33 位，在国内高校中排名第 4。

教师名录（按姓氏音序）：

包亚杰 助理教授，博士毕业于上海交通大学，统计学专业；研究方向：共形预测、选择性推断、联邦学习、深度学习。

戴国榕 讲师，博士毕业于美国德克萨斯农工大学，统计学专业；研究方向：缺失数据、分位数回归、半监督推断、变量选择。

方冠华 青年副研究员，博士毕业于美国哥伦比亚大学，统计学专业；研究方向：潜变量模型，厚尾分布理论，强化学习，贝叶斯变分推断。

冯项楠 副教授，博士毕业于香港中文大学，统计学专业；研究方向：潜变量模型、函数型和图像型数据分析、大数据和复杂数据的贝叶斯统计分析、网络众筹和移动营销等问题的建模分析。

胡瑾瑾 讲师，博士毕业于复旦大学，管理科学与工程专业；研究方向：保险精算、金融统计、生物统计。

黄 达 副教授，博士毕业于北京大学，统计学专业；研究方向：时间序列分析、计算统计、多元统计、统计学习、商务统计应用。

蒋斐宇 副教授、博士生导师，博士毕业于清华大学，统计学专业；研究方向：非线性时间序列分析、金融时间序列、断点分析。

黎德元 教授、博士生导师，博士毕业于荷兰鹿特丹大学，统计学专业；研究方向：极值统计、风险管理、分布式统计推断、分位数回归。

李竞阳 助理教授，博士毕业于香港科技大学，数学专业；研究方向：高维统计推断、矩阵/张量学习、信号处理、非凸优化、在线学习、联邦学习。

刘 彬 副教授、博士生导师，博士毕业于复旦大学，概率论与数理统计专业；研究方向：高维统计推断、变点分析、数据趋动检验、高斯逼近及自助法、稳健方法。

邵凌轩 副教授，博士毕业于北京大学，统计学专业；研究方向：函数型数据分析，非欧数据分析，稳健分析。

沈 娟 副教授，博士生导师；博士毕业于美国密歇根大学，统计学专业；研究方向：混合模型、亚组分析、贝叶斯变量选择。

吴 尚 讲师，博士毕业于美国威斯康星大学，统计学专业；研究方向：机器学习、随机控制、金融计量经济与统计、量子计算。

夏 寅 教授、博士生导师，博士毕业于美国宾夕法尼亚大学，统计学专业；研究方向：高维数据的统计推断、大范围假设检验、多元统计分析、非参数方程估计、具有形态限制的统计推断。

肖志国 教授、博士生导师，博士毕业于美国威斯康星大学，统计学专业；研究方向：变量误差模型、因果推断、全球生产网络、中国经济。

徐勤丰 副教授，博士毕业于华东师范大学，概率论与数理统计专业；研究方向：贝叶斯统计、非参数与半参数统计分析、试验设计、医学统计。

郁 文 教授、博士生导师，博士毕业于复旦大学，概率论与数理统计专业；研究方向：半参数统计模型，生存分析，经验似然，两阶段抽样设计，约束统计推断。

章 玮 助理教授，博士毕业于北京大学，统计学专业；研究方向：高维统计推断，机理驱动模型，函数型数据分析，因果推断和政策评估，分布式学习。

郑 明 教授、博士生导师，博士毕业于复旦大学，概率统计专业；研究方向：生存分析、经济统计中的计量方法、抽样调查。

周 健 副教授，毕业于复旦大学，应用数学专业；研究方向：概率论与数理统计。

朱雪宁 副教授、博士生导师，博士毕业于北京大学，统计学专业；研究方向：网络数据建模、高维数据建模。

朱仲义 教授、博士生导师，博士毕业于东南大学，系统工程专业；研究方向：非参数和半参数模型、回归诊断、纵向（面板）数据模型、保险精算信度理论。



附件：交通参考



73

