



2021 年高维数据统计分析及其应用 学术研讨会

(2021 年 5 月 14 日-15 日, 南京)

会 议 手 册

主办单位:

中国工业与应用数学学会

承办单位:

南京信息工程大学

中国工业与应用数学学会体育数学专业委员会

江苏省工业与应用数学学会



目 录

会议通知.....	1
会议日程安排.....	2
报告摘要与专家介绍.....	3
承办单位简介.....	10
交通指南.....	13

会议通知

2021 年高维数据统计分析及其应用学术研讨会

为促进高维数据分析等方向的交叉研究与发展，增强和国内外同行专家的学术交流与合作，提高相关领域青年教师和学生的学术水平和学术视野，由中国工业与应用数学学会（CSIAM）主办，CSIAM 体育数学专业委员会、南京信息工程大学、江苏省工业与应用数学学会承办的“2021 年高维数据统计分析及其应用学术研讨会”将于 2021 年 5 月 14 日-15 日在南京召开。会议采用线上线下混合模式，线上直播通过腾讯会议平台举行。会议将邀请相关领域的专家和学者与会，交流最新研究成果与动态。

有关会议事宜如下：

一、会议时间与地点、在线参会方式

会议时间：2021 年 5 月 14-15 日；

会议地点：南京信息工程大学藕舫楼 818 会议室；

腾讯会议 ID：175 633 013；

链接：<https://meeting.tencent.com/s/zNkTqmxxt6wIb>.



二、会议学术（组织）委员会

曹春正 南京信息工程大学

来 鹏 南京信息工程大学

史建清 南方科技大学

三、会议内容及日程安排

1. 5 月 14 日（周五）下午报到注册；
2. 5 月 15 日（周六）上午开幕式、学术报告；下午学术报告、闭幕式。

四、会议注册、住宿

1. 注册费：800 元/人，CSIAM 正式会员 600 元/人，会议报告人、学生、在线参会人免注册费；
2. 会议注册、住宿地点：南气宾馆（会务组统一安排食宿，费用自理）。

五、联系人

曹春正（13951756040，caochunzheng@163.com）

来 鹏（13913873207）

中国工业与应用数学学会
南京信息工程大学
中国工业与应用数学学会体育数学专业委员会
江苏省工业与应用数学学会

会议日程安排

2021 年 5 月 15 日，星期六上午（藕舫楼 818 会议室）				
08:45-09:00	开幕式		曹春正 南京信息工程大学数学与统计学院副院长	
会议报告				
时间	报告人	单位	题目	主持人
09:00-09:45	郑 明	复旦大学	Regression Analysis of Doubly Truncated Data	曹春正
09:50-10:00	茶歇			
10:00-10:45	周迎春	华东师范大学	基于图像数据的因果中介效应分析	
10:50-11:35	王 涛	上海交通大学	A Zero-Inflated Logistic Normal Multinomial Model for Extracting Microbial Compositions	
12:00-13:30	午餐（南气宾馆）			
2021 年 5 月 15 日，星期六下午（藕舫楼 818 会议室）				
会议报告				
时间	报告人	单位	题目	主持人
14:00-14:45	林金官	南京审计大学	某地高校人才引进待遇的统计建模与分析	来 鹏
14:50-15:35	张新雨	中国科学院	Model Averaging Estimation for High-Dimensional Covariance Matrices with a Network Structure	
15:40-15:50	茶歇			
15:50-16:35	史建清	南方科技大学	可穿戴设备的统计学习方法和精准医疗	
16:40-17:25	王启华	中国科学院	A Beyond Multiple Robust Approach for Missing Response Problem	
17:30-17:40	闭幕式		曹春正	
18:00-19:30	晚餐（离会返程）			

报告摘要与专家介绍

郑 明

Regression Analysis of Doubly Truncated Data

Abstract: Doubly truncated data are found in astronomy, econometrics, and survival analysis literature. They arise when each observation is confined to an interval, that is, only those which fall within their respective intervals are observed along with the intervals. Unlike the one-sided truncation that can be handled by counting process-based approach, doubly truncated data are much more difficult to handle. In their analysis of an astronomical dataset, Efron and Petrosian proposed some nonparametric methods for doubly truncated data. Motivated by their approach, as well as by the work of Bhattacharya et al. for right truncated data, we propose a general method for estimating the regression parameter when the dependent variable is subject to the double truncation. It extends the Mann–Whitney-type rank estimator and can be computed easily by existing software packages. Weighted rank estimation is also considered for improving estimation efficiency. We show that the resulting estimators are consistent and asymptotically normal. Resampling schemes are proposed with large sample justification for approximating the limiting distributions. The quasar data in Efron and Petrosian and an AIDS incubation data are analyzed by the new method. Simulation results show that the proposed method works well.

报告人简介: 郑明, 民建上海市委委员, 民建市科教委员会主任, 教授、博士生导师, 复旦大学管理学院副院长; 香港城市大学、香港中文大学、美国 MIT、美国哈佛大学、哥伦比亚大学等访问学者; 曾获国家统计局颁发的全国优秀统计教师奖, 复旦大学研究生心目中的好导师奖, 复旦大学优秀研究生导师奖, 上海市科教系统“三八红旗手”, 上海市科教系统“比翼双飞奖”等。兼职: 教育部高等学校统计学教学指导委员会, 全国应用统计专业学位研究生教育指导委员会委员, 全国统计教材编审委员会委员、常务组成员, 中国统计教育学会高等教育分会副会长, 上海市科协委员会委员, 上海市质量技术应用统计学会副理事长, 上海市统计专业高级专业技术职务任职资格审定委员会委员、专家库成员, 上海市统计学会副会长, 上海数学会常务理事等学术职务。

周迎春

基于图像数据的因果中介效应分析

摘要：在生物学、心理学、医学等很多领域，因果中介分析研究“某个处理”和“潜在结果”之间是否存在中介效应，即处理是否通过中介导致了结果。已有研究的中介类型通常为标量、向量或曲线。然而，在心理学和神经科学领域，功能性磁共振图像（fMRI）被广泛使用，它能很好地反映出脑部的应激变化，而脑部的变化常常可作为从某种刺激/处理到最终行为结果的中介，因此把图像型变量引入到因果中介分析对于此类问题的研究尤为重要。本报告提出了一种以三维图像为中介的因果中介分析方法，主要分为两部分：第一部分构造了以三维图像为中介的线性图像型结构方程模型，并为之设计了合理的参数可识别性假设；第二部分建立了有效的参数估计方法。之后，以大量的模拟研究来验证所提方法的有效性，并将此方法应用到以三维磁共振脑图像为中介的观测性数据当中，得到了具有启发意义的结果。

报告人简介：周迎春，华东师范大学统计学教授、博士生导师。复旦大学学士，美国波士顿大学（Boston University）统计学博士。主要研究函数型数据分析、因果推断以及生物统计中的各类问题。主持国家自然科学基金项目两项、参与国家自然科学基金重点项目两项，获上海市“浦江人才计划”，承担了多项医院与制药公司的企事业横向项目。在 *Annals of Applied Statistics*, *Statistics in Medicine* 等国际优秀期刊发表论文 30 篇。担任中国工程概率统计学会理事、中国现场统计研究会高维数据统计分会理事等。2018 年获第三届全国应用统计专业学位研究生教育教学成果奖一等奖，2019 年获宝钢教育基金优秀教师奖。

王 涛

A Zero-Inflated Logistic Normal Multinomial Model for Extracting Microbial Compositions

Abstract : High throughput sequencing data collected to study the microbiome provide information in the form of relative abundances and should be treated as compositions. Although many approaches including scaling and rarefaction have been proposed for converting raw count data into microbial compositions, most of these methods simply return zero values for zero counts. However, zeros can distort downstream analyses, and they can also pose problems for composition-aware methods. This problem is exacerbated with microbiome abundance data because they are sparse with excessive zeros. In addition to data sparsity, microbial composition estimation depends on other data characteristics such as high dimensionality, over-dispersion, and complex co-occurrence relationships. To address these challenges, we introduce a zero-inflated probabilistic PCA model that accounts for the compositional nature of microbiome data, and propose an empirical Bayes approach to estimating microbial compositions. An efficient algorithm, based on variational approximation, is developed for carrying out maximum likelihood estimation. Simulations and an application to a data set from the Human Microbiome Project are presented to compare the performance of the proposed method with that of the existing methods.

报告人简介: 王涛, 上海交通大学, 生物信息和生物统计系长聘教轨副教授, 统计系兼聘教授, 博士生导师。2013 年在香港浸会大学取得博士学位, 2014-2015 年在美国耶鲁大学从事博士后研究工作。现任上海交大-耶鲁大学生物统计与数据科学联合中心研究员和美国耶鲁大学生物统计系客座助理教授。主要研究方向包括高维数据统计学、生物统计学、以及统计机器学习, 主持国家自然科学基金面上项目等, 在统计学顶尖期刊 JASA、JRSSB、Biometrika 等发表学术论文多篇。

林金官

某地高校人才引进待遇的统计建模与分析

摘要：本报告以某地高校近五年的人才引进待遇数据为研究对象，展示了统计建模与分析的过程。报告分别用单指标模型、带跳单指标模型、自适应带跳单指标模型等拟合引进待遇数据，并做了理论分析，从回归模型的角度对高校教师的薪酬水平进行分析发现：首先，高校教师的薪酬水平受诸多因素的影响，其中学术排名对薪酬水平的作用较为明显；其次，薪酬水平的变化曲线在某些特殊点位置并非是连续的，而是呈现了跳跃现象，这种不连续现象的刻画在对模型假设时非常重要。因此，带跳单指标模型不仅很好地刻画了各因素对薪酬水平的影响，同时可以检测薪酬水平发生跳跃的特殊位置，从而更好地解释实际数据。

报告人简介：林金官，男，南京审计大学统计学教授、博士生导师，统计科学与大数据研究院院长。现主要从事非线性统计、计量经济、金融统计与风险度量、统计诊断、面板数据分析和统计应用等方面的研究工作。2000 年以来，在国内外核心期刊上发表论文一百余篇，其中 SCI、SSCI 收录论文八十余篇。荣获第十一届全国统计科研成果二等奖，第十二届江苏省统计科研成果二等奖，等等。主要兼职：教育部统计学类教学指导委员会委员；中国工程概率统计学会副理事长；江苏省概率统计学会理事长；江苏省现场统计研究会副理事长；中文核心期刊《应用概率统计》《数理统计与管理》杂志编委等。

张新雨

**Model Averaging Estimation for High-Dimensional Covariance Matrices
with a Network Structure**

Abstract: In this paper, we develop a model averaging method to estimate a high-dimensional covariance matrix, where the candidate models are constructed by different orders of polynomial functions. We propose a Mallows-type model averaging criterion and select the weights by minimizing this criterion, which is an unbiased estimator of the expected in-sample squared error plus a constant. Then, we prove the asymptotic optimality of the resulting model average covariance estimators. Finally, we conduct numerical simulations and a case study on Chinese airport network structure data to demonstrate the usefulness of the proposed approaches.

报告人简介: 张新雨，中科院数学与系统科学研究院/预测中心研究员，主要研究方向为模型平均、模型选择、组合预测等。2010 年在中科院系统所获博士学位，曾是德州农工大学 (TAMU) 博士后和宾夕法尼亚州立大学 (PSU) 的 Research Fellow。担任期刊《Journal of Systems Science and Complexity》领域主编、期刊《Statistical Analysis and Data Mining》Associate Editor、期刊《系统科学与数学》和《应用概率统计》编委，是中国统筹法优选法与经济数学研究会数据科学分会副理事长和国际统计学会当选会员。先后主持国家自然科学基金委优秀和杰出青年研究基金项目，曾获得中国管理学青年奖和中科院优秀博士学位论文等奖励。

史建清

可穿戴设备的统计学习方法和精准医疗

摘要：可穿戴设备已经被越来越多地用于日常生活和健康医疗中，但对其海量数据的统计学习方法还处在初期阶段。可穿戴设备收集日常生活中的数据（即 free-living 数据）。相对于实验数据，free-living 数据有以下特点（以 AX6 数据为例子）：（1）Free-living 数据的随意性：受试者收集的数据是多段长短不同的，非标准的多维时间序列和函数型数据，我们要对此类非标准数据建立标准，和建立个体化的精准预测模型；（2）巨大的差异性（非齐性）：除了个体间的差异外，还有环境文化差异，以及同一个体日常生活的非规律性等；（3）数据结构异常复杂：例如原始数据可以是对应各类动作步态的过程变量（加速度或角速度），通常还包含较大的噪音信号，使得数据在不同时间和频率上都有非常不同的均值和方差结构；（4）海量数据：受试者佩戴可穿戴设备七天产生的原始数据总量超过 5G，如何去除海量数据中的无用信息，保留建模所需的有效信息是待解决的问题。本报告将会利用一些实际例子讨论如何解决这些问题的思路。

报告人简介：史建清，南方科技大学统计与数据科学系教授，生物医学统计中心主任，博士生导师。曾任英国纽卡斯尔大学（Newcastle University）统计学教授，英国国家艾伦图灵研究院图灵研究员。主要研究方向包括函数型数据分析，生物医学统计，缺失数据分析，Meta-Analysis 等。在国际学术刊物上发表高水平学术论文多篇，包括统计顶级期刊 JRSSB, JASA, Biometrika 和 Biostatistics。曾任英国皇家统计协会《应用统计》副主编，Guest AE for JRSS Discussion Paper，英国纽卡斯尔大学云计算和大数据研究培训中心副主任。曾获邀任剑桥大学世界最顶级数学学院之一牛顿学院访问研究员，获美国统计协会非参数统计分会年度最佳论文奖，2012 年获英国 Wellcome Trust Health Innovation Challenge Fund，共计 210 万英镑。2011 年在著名统计学出版社 Chapman & Hall 出版专著：Gaussian Process Regression Analysis for Functional Data。

王启华

A Beyond Multiple Robust Approach for Missing Response Problem

Abstract: Imputation and the inverse probability weighting are two commonly used approaches in missing data analysis. Parametric versions of them are not robust due to model misspecification of some unknown functions. Nonparametric ones are robust but are impractical when the number of covariates is large due to the problem of “curse of dimension”. A beyond multiple robust method is proposed in this paper. This method balances the parametric and nonparametric methods by using some model information contained in the outcome regression function and the selection probability function, and hence alleviates the model misspecification problem and “curse of dimension” problem simultaneously. To illustrate the proposed method, we focus on the estimating problem of response mean in the presence of missing responses. A beyond multiple robust estimator of the response mean is defined, which is proved to be consistent and asymptotically normal as long as one of the true models for the outcome regression or selection probability functions can be some function of its assumed models, without the requirement that one of the true models is correctly specified. Also, it is shown that the asymptotic variance of the proposed estimator is equal to the semiparametric efficiency bound established by Hahn (1998, *Econometrica*, pp315-331) when both the selection probability function and the outcome regression function are the functions of their assumed models, respectively. The finite sample properties of the proposed estimator are evaluated by simulation studies and the proposed method is illustrated by a real data analysis.

报告人简介: 王启华, 中国科学院数学与系统科学研究院研究员, 博士生导师, 国家杰出青年基金获得者, 教育部长江学者奖励计划特聘教授, 中科院“百人计划”入选者, 首届全国百篇优秀博士论文奖获得者。曾在北京大学与香港大学任教, 先后访问加拿大 Carleton 大学、美国 California 大学戴维斯分校、美国 California 大学洛杉矶分校、美国 Yale 大学、美国华盛顿大学、美国西北大学、德国 Humboldt 大学、澳大利亚国立大学及澳大利亚悉尼大学等。主要从事生存分析、缺失数据分析、高维数据统计分析及非-半参数统计推断等方面的研究。出版专著三部, 在 *The Annals of Statistics*, *JASA* 及 *Biometrika* 等国际重要刊物发表论文百余篇。是高维统计分会理事长, 分别是生存分析及生物统计分会副理事长, 先后是 IMS-China 和 IBS-China 委员会委员, 是一些国际与国内学术期刊的编委。

承办单位简介

南京信息工程大学数学与统计学院

南京信息工程大学享有“中国气象人才摇篮”之美誉，其前身南京气象学院始建于1960年。1978年被国务院确定为全国重点大学，2017年入选国家“双一流”建设高校和江苏高水平大学建设高校，学校由江苏省人民政府和教育部、中国气象局、国家海洋局共建。大气科学学科入选国家“世界一流学科”建设学科，在教育部近两次一级学科评估中先后获评全国第一和 A+等级，地球科学、工程学、计算机科学、环境科学与生态学、化学、农业科学等6个学科跻身 ESI 学科排名全球前1%。拥有地球科学、工程学、计算机科学、环境科学与生态学等4个 ESI 全球排名前1%的学科，拥有大气科学、环境科学与工程、信息与通信工程、管理科学与工程、数学、科学技术史6个一级学科博士点，22个一级学科硕士点、14个专业学位硕士点，75个本科专业。

数学与统计学院是与学校同时期起源且实力雄厚的二级学院。学院源于上世纪60年代的数学教研室（组），历经基础部、数学系、数理学院等发展历程，于2011年11月更名为数学与统计学院。目前设有3个专业系、1个研究所及公共数学教学部，拥有1个省级科研平台“江苏省统计科学研究基地”、1个省级实验示范中心“数学教育实验中心”、1个共建国家应用数学中心、1所空间天气研究院、1所数据科学研究院、4个教育部企业协同育人平台、5个江苏省企业研究生工作站以及多个校企研究院。在校学生千余人，其中博士、硕士研究生192人。

数学与统计学院拥有较为完整的学科专业和良好的教学条件。在学科专业方面，学院拥有“数学”一级学科博士点、博士后流动站；“空间天气学”二级学科博士点；“应用统计”专业学位硕士点；“信息与计算科学”国家一流专业建设点；“应用统计学”国家一流专业建设点；“数学”中国气象局重点学科等。学院设有信息与计算科学、应用统计学、数学与应用数学、信息与计算科学（嵌入式培养）四个本科专业。信息与计算科学、应用统计学、数学与应用数学专业全部为江苏省重点专业。学院具备优良的教学条件。学院实验中心为江苏省实验示范中心，拥有一流的软硬件发展环境，实验室面积1000余平方米，包括数学建模、数学教育、空间天气学等国家与地方共建的实验室。拥有每秒浮点计算能力达到126万亿次的高性能计算系统，数十台工作站及服务器。数学与统计学院教师先后获得江苏省教育教学成果特等奖、江苏省教学成果奖一等奖、江苏省研究生培养模式改革成果二等奖等省部级教学奖励30余项。

学院还拥有高水平的师资队伍和较强的科研实力。学院拥有全时教职工120余人，其

中教授27人、海外非全时教授12人、副教授43人，博导 14人、硕导 40人。45岁以下专业教师100%拥有博士学位，85%以上的专业教师具有海外工作、学习经历。教师队伍中拥有教育部新世纪优秀人才1人、国务院特殊津贴专家3人，海外杰青3人，另有霍英东教育基金项目获得者、江苏省双创人才、江苏省“333人才工程”、“六大人才高峰”、“青蓝工程”中青年学术带头人、“青蓝工程”优秀青年骨干教师等60多人。自2018年以来，学院教师主持国家自然科学基金、国际合作等项目60余项，主持省部级项目26项，承担企事业单位委托项目28项，主持与主要参加“973”、“863”和国家公益性行业专项等课题多项，科技经费超过2千万；在国内外重要学术刊物上发表论文400余篇，其中 SCI 收录近300篇；出版专著、教材20余部，参编研究生教材5部。学院教师获世界气象组织颁发的“Norbert Gerbier-Mumm”奖、国家计委、国家科委、财政部颁发的科技攻关重大科技成果奖、国家统计局全国科技进步（课题）奖等多项荣誉。

近年来，学院始终秉持“笃行以生为本，厚植大学精神”科学理念，坚持“数学与应用数学高层次人才、气象特色应用数学与统计人才以及信息大数据与科学计算高技术人才”培养导向，立足专业学习、社会实践以及学科竞赛培养契合社会需求的复合型学生人才。学生在国际和全国学科竞赛中多次获得重要奖项。自2011年以来共获得全国大学生数学建模竞赛特等奖1项（2011年获，全国唯一本科类“高教社”杯），一、二等奖79项；获得美国大学生数学建模竞赛特等奖3项，特等提名奖3项，一、二等奖149项；获得全国研究生数学建模竞赛一、二等奖49项。毕业生总就业率超过98%，其中高质量就业率超过95%，毕业生主要在政府机关、国有企业、科研机构、教育、IT 行业、气象、通信、金融等部门从事科研、教学、技术开发及生产应用等工作。同时一大批毕业生考入英国帝国理工、北大、清华、浙大、复旦、同济、中科院等国内外著名高校及研究所，考研出国率达30%以上。学院现与美国佛蒙特大学、英国雷丁大学、纽卡斯尔大学、伯明翰大学、西苏格兰大学等联合培养本科生和研究生，学生综合发展成效显著。

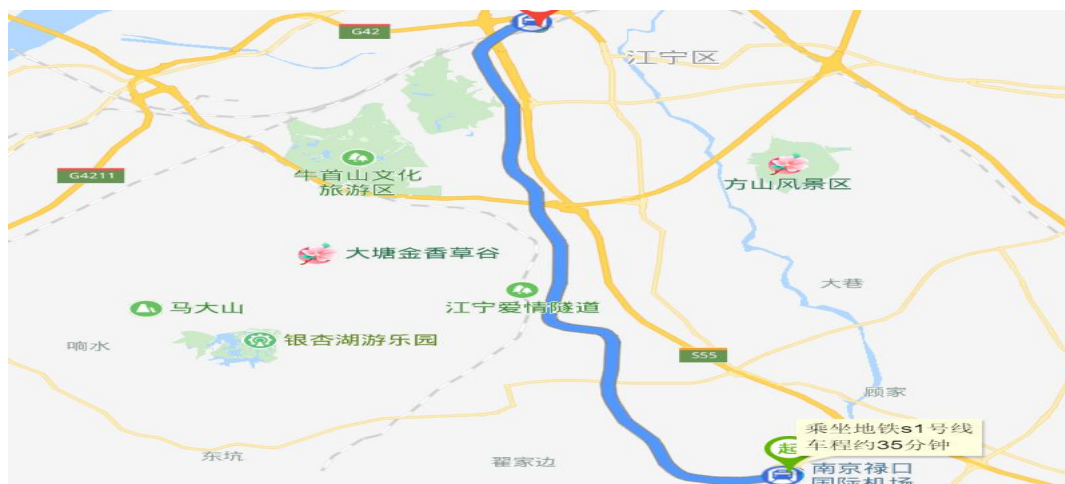


南信大数统院真诚期待海内外高层次人才的加盟，携手共创美好未来!

交通指南

(期待您的随时莅临指导)







祝您生活愉快!

Wish you a happy life.