北京大学统计科学中心
2020 年度总结报告

Huina Mu 北京大学统计科学中心

2020 年 12 月 21 日，北京大学统计咨询中心成立，部分嘉宾合影留念

目　录

一、综述 ····· 3
二、中心要闻 ····· 4
三、人员组成 ····· 8
四、科研项目 ····· 20
五、博士生去向 ····· 26
六、交流访问 ····· 27
七、论文发表 ····· 28
八、学术讲座 ····· 33
九、委员会 ····· 36

前言

北京大学统计科学中心（以下简称“中心”）成立于 2010 年 7 月。中心依托北京大学深厚的学术底蕴和多学科的优势，是一个跨理学部、经济管理学部、医学部的学科建设平台与交叉科研机构，2019 年 10 月完成行政团队换届。

中心的主要任务是整合全校力量加强统计学科建设，建立一支具有国际水平的统计学与数据科学的研究教学团队，促进统计学与其他学科的交叉合作并加强统计学在各学科中的应用，推动和提高各学科的实证与统计分析水平，为相关学科的发展提供数据收集和数据分析的统计方法。

中心目前设立国际顾问委员会、科学委员会和指导委员会。国际顾问委员会负责提出有关中心发展方向等方面的建议，评估中心的工作和成果；科学委员会在人才引进和学术研究等方面给予意见和建议；指导委员会在学校层面指导中心的管理运行工作。

中心的成立已在北京大学形成一个吸引海内外优秀统计人才的基地。统计科学中心将继续营造一个开放式、国际化的学术环境，广泛联系全球各地的统计学者，推动北大统计学“双一流学科”建设，希望在 2030 年北大统计学进入世界前列，亚洲领先。

著名统计学家 Peter G. Hall 教授摄

一、综述

2020 年统计科学中心在统计学研究、博士生培养、利用专业优势抗击新冠肺炎、师德师风建设以及推动统计学发展等方面开展工作。

在学科发展方面，2017 年北大统计学科入选“双一流”建设学科，并连年在教育部学科评估中获评 A+，2020 年依然保持 A+的好成绩；2020 年 10 月 15 日，由软科发布的“2020 软科中国最好学科排名”，北京大学统计学科位列第一名，这是本学科第三年连续保持第一。北京大学统计学科排名的上升，为统计学科的未来发展提供了更加广阔的舞台。与此同时，北大统计科学中心的发展也面临新的机遇与挑战。

在教研团队方面，为建设一支专业结构合理、学术素养良好、适应能力强的统计学教师队伍，中心积极引进优秀人才。2020 年，周扬加入从事博士后研究，合作导师为姚方教授。周扬博士后已获得中国博士后科学基金特别资助（站前）和中国博士后科学基金第 68 批面上资助。目前，中心现有教研人员 33 人，博士后 1 人，包括从北美全职引进的 3 位国家高层次专家和 8 名青年教师，业已形成具有国际一流水平的教研团队。

在国际合作交流方面，受全球新冠疫情的影响，2020 年中心的国内外学术交流活动受到严重影响。中心克服困难，在严格执行各级防疫政策的前提下，不断推进同国内外学者的学术交流活动，共举办学术报告 14 场，其中一半是采用线上会议形式；共接待海内外访问学者 10 余人次，参与学术报告、短期课程以及合作项目等各类学术活动。2020 年中心共邀请到来自哈佛大学 Xiao Wu 教授、南方科技大学史建清教授、中国科学技术大学王学钦教授、中科院数学与系统科学研究院张新雨研究员、武汉大学刘岩教授、香港中文大学陈昕韫教授、复旦大学林伟教授等开展了 12 场学术报告活动。这些报告拓宽了老师和学生的学术视野，活跃了统计科学中心的学术氛围，促进了统计学领域的交流。

在行政队伍建设方面，中心为加强行政工作，于 2019 年底重新组建了一支办事高效、组织决策能力强的行政团队，由姚方教授担任新一届中心主任，组建了包含周晓华和王汉生副主任、席瑞斌主任助理、行政秘书王彦懿和穆慧娜的管理团队。2020 年，中心全面加强日常运营及拓展业务等事务的管理，各项工作稳中有进。近期，中心行政团队牵头组织五年一轮的学科评估（统计学）整体工作，需要跨学部、院系协调沟通资源，中心在人员精减、资源紧张的情况下，顺利组织并完成此项复杂度极高的工作，充分展示了中心行政团队的业务能力。

在科学研究方面，2020 年中心教师累计发表学术论文 70 余篇，包含统计学及其交叉学科顶级期刊 23 篇；在研项目 53 项，包括纵向课题 38 项，横向课题 15 项。

二、中心要闻

1. 北大统计学科在 2020 软科中国最好学科获评第一

2020 年中国最好学科排名-0714 统计学：1. 北京大学；2. 南开大学；3. 中国人民大学。北京大学以 24 个中国顶尖学科（前 2%）位居榜单第一。

软科 中国最好学科排名 2020　0714 统计学 前10%

排名层次	排名区间	学校名称	
中国顶尖学科（前2%）	1~2	北京大学	南开大学
中国一流学科（前5%）	3~6	中国人民大学 东北师范大学	北京师范大学 厦门大学
中国一流学科（前10%）	7~13	复旦大学 华东师范大学 清华大学 中国科学技术大学	西南财经大学 中山大学 上海财经大学

2020 年 10 月 15 日，高等教育评价专业机构软科中国基于对学科人才培养、科研项目、成果获奖、学术论文、高端人才五个指标的综合评分，发布“2020 软科中国最好学科排名”，北大统计学科排名第一。这已是北大统计学科连续第三年保持第一名（2018-2020）。

北大统计学科拥有雄厚的科研实力和光荣的历史底蕴，自 1940 年许宝騄先生从英国获统计学博士学位回国，首次在我国大学开设数理统计课程开始，几十年筚路蓝缕，北京大学为中国统计学科人才培养、学科发展做出了巨大的贡献。2010 年在数理统计研究所的基础上成立统计科学中心，整合全校力量加强统计学科建设；2011 年设立数量经济与数理金融教育部重点实验室；2018 年成立跨学部的生物统计系。今天的北大统计站在了一个前所未有的高度，也必将不忘初心，砥砺奋进，开创更辉煌的事业。

2. 姚方教授当选数理统计学会理事会成员（IMS Council Member）

在 2020 年数理统计学会（Instituteof Mathematical Statistics）选举中，北京大学数学科学学院教授、统计科学中心主任姚方当选为新一届数理统计学会理事会成员（IMS Council Member），任期 3 年。

数理统计学会创立于 1935 年，与北美统计学会（American StatisticalAssociation）并列为为国际统计学界两大著名学会。理事会成员共 15 名，每年更换其中三分之一。今年同时当选的另外四名理事会成员为 Tony Cai、Richard Davis、 Alice Guionnet、Martin Wainwright。北大统计科学中心陈松蹊教授也曾担任此职务。

3. 周晓华教授当选数理统计学院会士（IMS Fellow）

据数理统计学院（Institute of Mathematical Statistics, IMS）官网 2020 年公布的消息，北京大学讲席教授周晓华教授当选为该学院会士(Fellow），以表彰他在诊断医学、临床试验中的因果推断、医疗费用数据分析、以及在人才培养和学术刊物服务等方面的贡献。周晓华教授现任北京大学北京大学公共卫生学院生物统计系主任、统计科学中心副主任。他在诊断医学统计学和因果推断方面做出了许多重要的贡献，是国际著名的生物统计学家。

如前文提到，数理统计学会创立于 1935 年，与北美统计学会（American StatisticalAssociation）并列为为国际统计学界两大著名学会。IMS 目前在全球约有 3500 名会员，其中 12%经过遴选当选为该机构会士。北大统计科学中心目前拥有三名 IMS 会士，另两位为北大讲席教授陈松蹊与姚方，他们三位也均已当选 ASA 会士。

4. 2020 年统计科学中心暑期课程成功举办

2020 年 8 月 10 日-14 日，北京大学统计科学中心成功举办面向本科生的夏令营暨暑期课程活动。本次活动开展了短课讲授、师生线上交流和课程报告等多种形式的活动，共选拔了来自全国各院校的 25 名优秀本科生参加，来自各类院校、企业的近 200 名学生、科研人员和业界人士同步观看了课程直播。

8 月 10 日-14 日的暑期课程阶段开设“数据科学中的随机矩阵理论”和“理论机器学习”两门课程，分别由加州大学戴维斯分校丁秀才助理教授和北京大学数学科学学院张志华教授、林伟长聘副教授讲授。在“数据科学中的随机矩阵理论”课程中，丁秀才老师从协方差矩阵估计、矩阵去噪等数据科学问题中的大维矩阵及其算法出发，详细讲解了随机矩阵理论的经典和最新结果，特别是高维情况下样本协方差矩阵特征值和特征向量的性质。在“理论机器学习”课程中，张志华老师介绍了集中不等式、贝叶斯分类器等机器学习的基本概念和理论工具，林伟老师介绍了偏差—方差权衡、Lasso 方法、图模型和过参数化模型等基本原理和方法，详细讲解了机器学习领域的经典结果和最新研究进展。

8 月 14 日下午短课结束后，参加暑期学校活动的同学与北京大学统计科学中心的老师开展线上交流。交流活动在统计中心主任姚方教授的主持下开始，统计中心艾明要、张志华、席瑞斌、林伟等老师参与了线上交流。师生之间就课程内容、学科发展、本科生科研、人才培养等话题问答互动，气氛热烈。

5. “全国统计学优秀大学生”夏令营成功举办

在 2020 年 8 月中旬的暑期课程中，中心邀请了 25 名来自全国各大高校统计学及相关专业的优秀大学生参加课程学习，并与中心教师做了深度交流，为夏令营的举办做了非常重要的铺垫工作。

8 月 24 日上午，参与暑期课程的 25 名本科学员分组进行线上课程报告展示。经评议，所有小组均考核通过，并给表现优异的学员颁发“最佳报告奖”、“优秀报告奖”。

本次夏令营和暑期课程活动帮助一大批统计学相关专业的优秀本科生梳理了数据科学和机器学习领域的基础知识，引导他们接触和了解学术前沿问题，提高他们的研究兴趣和科学品味。同学们纷纷表示，通过此次暑期活动加深了对现代统计学科发展的认识，极大激发了参与学术研究的热情。

6. 第一时间研究新冠传播规律，为疫情防控提供科学依据

中心第一时间积极组织力量，充分发挥中国学者的担当与使命感，利用学科优势对新冠病毒的传染、防控等方面展开分析，取得一系列成果，对新冠疫情的防控做出贡献。

周晓华团队快速组织概率、统计、公共卫生的科研力量，充分发挥中心的交叉科研优势，针对新冠病毒的传播特征以及发展规律、重点感染区域的感染人数、中国及海外疫情发展和控制时间等关键问题进行研究。在武汉封城次日，即根据交通大数据给出武汉病例数的估计，并随即发表在《China CDC Weekly》。团队获得多项基金资助用于新冠病毒的研究，包括国家自然科学基金新冠应急专项基金、比尔及梅琳达·盖茨基金等。团队估计了新型冠状病毒的潜伏期、武汉病例最早出现时间、发病趋势等，在《Science Advances》等期刊发表 10 余篇相关论文，为新冠病毒流行病学调查、疫情监测等做出重要贡献。陈松蹊团队基于新开发的 vSEIdR 统计模型，发表于《National Science Review》的研究对较早爆发疫情的 25 个国家防治新冠病毒的效果进行比较分析，揭示了快速应对和高强度政策对控制疫情的作用，对于防控可能发生的疫情反弹有重要借鉴意义。

7. 成立统计咨询中心，为更好服务校内外科研工作做出贡献

为更好地对北大各院系、政府部门与工业界提供数据分析和咨询服务，推动北大统计学科的整体建设及影响力，提升统计学在科研探索、政府决策、商业服务的实际应用，在数学科学学院的大力支持下，2020 年 12 月正式成立统计咨询中心。统计咨询中心将依托北大统计学科在概率统计、商务统计、精算和金融、生物统计、计算机与生命科学等领域的雄厚基础，统筹协调各项相关事宜。由中国现场统计研究会理事长房祥忠担任统计咨询中心主任。

统计咨询中心将努力探索优化服务方式，凝练并解决校内兄弟院系、校外企事业单位提出的重要统计问题。咨询中心的成立将进一步推动统计咨询模式创新，促进产学研一体化，为北大校内院系、政府部门与工业界等单位提供高质量的统计咨询服务。

8. 统计科学中心科研工作进展

统计科学中心成员近年来在统计研究和统计应用方面取得了一系列重要的、国际领先的成果和奖项。2020 年，中心教师累计发表学术论文 70 余篇，包含统计学及其交叉学科顶级期刊 23 篇。

（1）函数型数据分析方面

姚方教授带领的研究团队在复杂结构函数型数据分析领域深入研究了流形上的数据分析，发表于 2020 年《Biometrika》的成果提出了对带有误差项的函数型自变量的非参回归模型进行流形估计的思想；相关模型与方法的实际应用广泛，例如用于分光光度法研究肉块的脂肪含量问题、在胼胝体多发性硬化患者中的认知功能与大脑微结构的关系等。

（2）超高维数据方面

陈松蹊教授带领的环境统计科研团队发布了《空气质量评估报告(七)》，本报告的空气质量数据相比第六份报告新增 23 个城市，达到 68 个城市；报告分析了新冠肺炎疫情的影响，并同样剔除了气象因素对数据形成的干扰，呈现出更清晰、立体的区域空气质量演变过程。

（3）生物统计方面

周晓华教授团队估计了新型冠状病毒的潜伏期、武汉病例最早出现时间、发病趋势等，其发表于《Biometrics》的成果，基给出了潜伏期分布的估计，然后基于潜伏期分布和序列间隔，给出了再生时间分布的一个估计；相关成果对新冠病毒流行病学调查、疫情监测等抗疫工作具有重要意义。

（4）网络数据研究方面

王汉生教授带领的商务统计团队，在 2020 年对带有网络结构的数据做了持续的研究，尤其在带有网络结构超高维数据的变量筛选问题、空间自回归模型的快速估计计算方法、带有网络结构的时间序列模型以及网络结构数据本身的产生机制建模等方向上取得了较好的成果。

9. 统计科学中心教师课程教学创新成效显著

在疫情常态化的特殊国情下，中心教师创新课程教学方法，开设多门网络公开课和讨论班。开设网络公开课如《统计机器学习》、《机器学习导论》、《深度学习：算法与应用》、《生物统计》等，受众覆盖多所高校学生及企事业单位人员等，访问量达 100 万次，获得广泛好评。中心教师还在线上开设统计案例实务课程和智能大数据统计讨论班，面向社会开放，播放量累计达万余次。

10. 统计科学中心师德师风建设成果丰硕

北大统计学人坚持“以德为先”，重教研、德先行，全面落实学校师德师风评估要求。中心教师全面贯彻落实习近平总书记系列重要讲话精神以及《关于加强和改进新时代师德师风建设的意见》等文件要求，提高政治站位，坚持完善制度建设与日常教育督导相结合，打造政治素质过硬、业务能力精湛、育人水平高超的高素质教师队伍。

在2020年，中心教师争先创优，在师德师风建设中取得丰硕成果。周晓华获2020年“北京大学抗击新冠疫情标兵”。林伟获“北京大学优秀共产党员”。刘勇当选北京大学教师代表大会代表。张原获“北京大学优秀班主任”、“北京大学最受学生欢迎奖”、“北京大学优秀工会积极分子”。

三、人员组成

1. 人才引进

张成助理教授：1987 年生。2017 年于美国加利福利亚大学尔湾分校获得博士学位，现任北京大学数学科学学院助理教授（研究员学术头衔）, 2019 年 8 月入职，统计科学中心与数学科学学院合聘。

研究方向：主要涵盖统计学习、贝叶斯推断、计算生物学及其交叉学科应用。

张成于 2008 年在北京大学获得理学学士学位，并于 2011 年在北京大学获得计算数学硕士学位。此后，其就读于加利福利亚大学尔湾分校继续学习计算数学，于 2017 年获得哲学博士。2017 年 1 月至 2019 年 7 月，张成在美国福瑞德哈金森癌症研究中心担任博士后研究员，于 2019 年 8 月全职加入北京大学，任职助理教授。

研究介绍:

主要集中在大数据贝叶斯推断方法 (如 scalable MCMC, VI 等) 以及现代统计计算方法在计算生物学中的应用 (如分子进化模型等)。针对大数据提出的替换函数加速算法可以在保持 MCMC 抽样效率的同时降低运算量，同时也为结合贝叶斯推断中的两个核心算法 (MCMC 和 VI) 提供了一种可能。在进化推断这一现代计算生物学的重要领域中，提出了可用于进化树拓扑空间上的一类高效可泛化的概率图模型: subsplit Bayesian network。在此基础上，首次将 VI 这一强大工具引进到贝叶斯进化推断领域，并在一些基准数据集上取得超过此领域当前主流 MCMC 方法的表现。相关研究发表在 Statistics and Computing, Bayesian Analysis 等统计计算顶级杂志上，以及 NeurIPS, ICML, ICLR 等机器学习顶级会议中。E.g.:

Figure 4: Performance on DS1. **Left:** KL divergence for methods that use the simple split-based parameterization for the branch length distributions. **Middle:** KL divergence for methods that use PSP. **Right:** Per-tree marginal likelihood estimation (in nats): VBPI vs GSS. The number in brackets specifies the number of samples used in the training objective. MCMC results are averaged over 10 independent runs. The results for VBPI were obtained using 1000 samples and the error bar shows one standard deviation over 100 independent runs.

张原助理教授：1988 年生。2015 年于美国杜克大学获得博士学位，现任北京大学数学科学学院助理教授（研究员学术头衔），2018 年 7 月入职，统计科学中心与数学科学学院合聘。

研究方向：主要涵盖随机游动、调和测度及其相关随机增长模型、随机相互作用粒子系统 及其交叉学科应用。

张原于 2010 年在北京大学获得理学学士学位，2010-2015 年就读于美国杜克大学，并于 2015 年 5 月获得哲学博士学位。期间于 2014-2015 年获得美国 SAMSI（the Statistical and Applied Mathematical Institute）研究院 Grad. Research Fellow。在 2015-2016 年在美国加州大学洛杉矶分校（UCLA）担任附属助理教授（Assistant Adjunct Professor），2016-2018 年在美国德州农机（工）大学（Texas A&M University）任访问助理教授（Visiting Assistant Professor ）。

研究介绍：

集中于互相作用无穷粒子系统及其在生态学与社会科学的应用、基于随机游动的扩散极限聚集模型（Diffusion Limited Aggregation）以及随机编织（Random Interlacement）等方面的研究。相关研究成果在 Proceedings of the National Academy of Sciences、The Annals of Applied Probability、Electronic Journal of Probability、Electronic Communications of Probability、Latin American Journal of Probability and Mathematical Statistics 等杂志发表。

苗旺助理教授：1990 年生。2017 年在北京大学获统计学博士学位，2017 年 7 月-2018 年 7 月在哈佛大学做统计学博士后研究，2018 年 9 月入职，统计中心与光华管理学院合聘。

研究方向：因果推断和缺失数据分析，及其在经济学、社会学、流行病学、生物医药中的应用。

苗旺于 2012 年在北京大学获数学学士学位，2017 年在北京大学获统计学博士学位。于 2017 年 7 月-2018 年 7 月在哈佛大学做统计学博士后研究。苗旺曾于 2016 年荣获中国概率统计学会颁发的“钟家庆优秀论文奖“，并参与中国疾病预防控制中心主持的中国艾滋病毒感染率调查，以及国家食品安全风险评估中心主持的食源性疾病爆发监控的研究；曾受邀在多个大型国际统计会议和多所国际知名大学报告研究工作，如 ENAR2015、ICSA2016、ACIC2017、哈佛大学、沃顿商学院、香港大学。

研究介绍：

因果推断和缺失数据分析及其在经济学、社会学、流行病学、生物医药中的应用，是当前统计学和大数据研究中的重要课题。作为主要负责人已完成的研究课题包括：非随机缺失数据的识别性和在计量经济中的应用，半参数模型的估计和双稳健估计方法，因果推断的阴性对照变量方法，以及，空气污染对死亡率的影响。这些研究成果已发表在统计学及其应用领域的国际著名期刊 Journal of the American Statistical Association、Biometrika 、 Statistica Sinica、American Journal of Epidemiology 和国内核心期刊《中国科学》。

贾金柱研究员： 统计中心与公共卫生学院合聘。2009 年北京大学概率统计系博士，2009-2010 年在 UC Berkeley 做博士后，2011 年加入统计中心。2011 年 1 月至 2018 年 1 月任职于北京大学数学科学学院概率统计系和北京大学统计中心，期间访问哈佛大学一年。2018 年 2 月加入北京大学公共卫生学院。

研究方向：高维统计推断、大数据分析、统计机器学习、因果推断、生物统计等。

研究介绍：

高维统计的变量选择一直是统计学家关心的问题。尽管目前 L1 方法如 LASSO 等已经非常成熟。经过统计学家的努力，大家对 LASSO 方法的性质的研究已经非常深入。不过 LASSO 本身存在着一些问题。比如它的符号相合需要 irreprensentable condition，它的 L2 相合性需要 restricted eigenvalue condition。Lasso 本身主要研究齐方差的情形。即使在其方差的时候，Lasso 的解的好坏也取决于调整参数 lambda 的选取。而 lambda 的选取通常通过交叉验证来完成。这种方法通常具有不稳定性，且非常耗时。后来有研究表明 square root lasso 在选择调整参数时，可以事先给定而不需要不断调整。我们详细深入地研究了 square root lasso 的理论，发现该方法本质上是在加权得分函数上使用惩罚项来选择变量。根据这个原理，我们研究了 Possion 回归的变量选择，提出了一种新的最优化目标函数。 该方法与传统的带惩罚的极大似然估计相比，更加稳健，更容易设置调整参数。与 square root lasso 类似，我们的新方法也可以事先指定调整参数 lambda 的选取，而不需要交叉验证。我们的新方法得到的结果也更加精确。通过模拟 CT 图像的重建， 结果显示本文提出的方法很大程度上改进了已有方法的重建精度。下面是几个 CT 图像重建算法的对比。无论视觉效果还是重建精度 RMSE 都显示我们的新方法更有效。

左：我们的新方法得到的重建图。右：传统重建方法得到的。

参考论文：

Jinzhu Jia, Fang Xie, Lihu Xu. Sparse Poisson Regression with Penalized Weighted Score Function[J]. Electronic Journal of Statistics, 2019, 13(2), 2898-2920.

李程高级研究员：统计中心与生命科学学院合聘。2013 年加入统计中心。

研究方向：生物信息学和计算生物学，癌症基因组学和个体化医疗，高通量生物数据分析中的统计建模。目标为使用计算生物和基因组学来帮助预测、发现、验证新的癌症基因和机理， 从而为新药物和个体化医疗作出贡献。

研究介绍：

随着三维基因组技术的快速发展以及它的广泛应用前景，美国国立卫生研究院在 2014 年制定了 4D Nucleome 计划，资助多个研究团队从三维空间和时间尺度上研究细胞核内染色质的组织结构和功能，体现了该领域的重要性和前沿性。我们基于对癌症基因组中非整倍体变异频繁出现的原因和后果的研究兴趣和基础，在 2014 年开始将研究重心聚焦在癌症三维基因组学领域，通过自主建立的 Hi-C 实验和分析流程，首先研究多发性骨髓瘤细胞中非整倍体变异对三维基因组和表达谱的影响。在过去四年间，我们在三维基因组学实验技术开发、分析算法、数据库网站、与临床研究团队开展合作方面都取得了重要进展。

近期发表：

Wu P, Li T, Li R, Jia L, Zhu P, Liu Y, Chen Q, Tang D, Yu Y, Li C. (2017) 3D genome of multiple myeloma reveals spatial genome disorganization associated with copy number variations. Nat Commun. 8(1):1937.

Hydbring P, Wang Y, Fassl A, Li X, Matia V, Otto T, Choi YJ, Sweeney KE, Suski JM, Yin H, Bogorad RL, Goel S, Yuzugullu H, Kauffman KJ, Yang J, Jin C, Li Y, Floris D, Swanson R, Ng K, Sicinska E, Anders L, Zhao JJ, Polyak K, Anderson DG, Li C, Sicinski P. (2017) Cell-Cycle-Targeting MicroRNAs as Therapeutic Tools against Refractory Cancers. Cancer Cell. 31(4):576-590.e8.

Wang Y, Fan C, Zheng Y, Li C. (2017) Dynamic chromatin accessibility modeled by Markov process of randomly-moving molecules in the 3D genome. Nucleic Acids Res. 45(10): e85.

Li R, Liu Y, Li T, Li C. (2016) 3Disease Browser: A Web server for integrating 3D genome and disease-associated chromosome rearrangement data. Sci Rep. 6:34651.

林伟助理教授：统计中心与数学学院合聘。2011年南加州大学应用数学（统计方向）博士，2011-2014年在宾夕法尼亚大学生物统计与流行病学系进行博士后研究工作；2014年加入统计中心。

研究方向：主要是高维统计和大数据建模与分析，此外还包括因果分析、生存分析、成分数据分析、时空统计等，及其在基因组学、宏基因组学和环境科学中的应用。

研究介绍：

Fig. 1. Analysis of two microbiome datasets: empirical power curves of the proposed test (triangles) and the test based on log-transformed (circles) and raw (squares) compositions with $\alpha = 0{\cdot}05$ are shown in (a) for the obesity data and (c) for the Crohn's disease data; histograms of p-values from the proposed test in the back-testing are shown in (b) for the obesity data and (d) for the Crohn's disease data, for 1000 replicates.

林伟研究员及合作者在 Biometrika 上发表论文，针对高维成分数据提出一种两样本均值检验方法，为基于成分数据的科学研究和数据分析增添了新的研究手段。成分数据是指各分量为正且加和为一的数据，在宏基因组学中测量微生物组成、经济学中描述家庭支出构成、文本分析中刻画词频分布等应用中都十分常见。成分数据的特点导致其各分量之间复杂的相依关系，造成许多经典统计方法的失效甚至得到误导性的结果。该论文在两样本均值检验的一般框架下，首先定义了一种可检验的成分等价性假设，然后基于成分数据的中心化对数比变换发展了一种成分等价性检验方法，并推导了检验统计量的渐近分布及功效。论文将该方法用于肥胖症和克隆氏症中的肠道微生物组成分检验问题，取得了良好的效果。

宋晓军助理教授：统计中心与光华管理学院合聘。2014 年西班牙马德里卡洛斯三世大学经济系经济学博士，同年加入北大统计中心。

研究方向：计量经济学理论（非参数与半参数方法， 模型设定检验，自助方法，时间序列分析和实证资产定价）及应用计量经济学。

研究介绍：

宋晓军研究员及合作者撰写论文(Journal of Econometrics,已接受)，针对条件分布函数的对称性检验提出了一个新方法。由于条件对称性假设在计量经济学和统计学上的重要性，如 何有效的检验该假设至关重要。本文提出的方法不仅适用于截面数据也适用于时间序列数据， 并且能够有效的避免模型误设的问题。传统检验在验证条件对称性时，通常需要假设对称中 心是已知的常数或者函数形式，而当函数形式并不是所假设的形式时，基于该参数假设所构 建的检验显然会出现检验水平和能效方面的问题。本文利用非参数方法估计对称中心，然后 基于非参数残差构建统计量。另外由于该论文运用了条件特征函数来检验对称性，所以并不 需要假设残差是独立于解释变量的。这样构建检验统计量的话，就允许高阶相依性的存在，而这正是一般金融时间序列的特征。

该论文进一步推导了检验统计量的大样本性质，提出一个乘数自助法来实现该检验方法。大量蒙特卡洛模拟表明该方法能够很好的保持水平，并且对各种各样的非对称性备择假设都具有较好的能效。

最后，论文将该检验方法用于检验四个主要股票市场日回报率在 2008 年经济危机前后的表现， 以及检验商业周期是否对称等问题上，取得良好的实证结果。此外，本论文所发展的技术还可以检验其它非参数类型的约束， 例如在数据具有相依性情况下的条件相关性，条件独立性检验等。

涂云东副教授： 统计中心与光华管理学院合聘。2012 年加州大学经济学博士，同年加入统计中心。

研究方向：非参数建模，高维数据建模，信息计量经济学，金融计量学，经济预测，生产效率建模，宏观计量经济学等。

研究介绍：

涂云东研究员和他的合作者 2017 年在 Journal of Econometrics 上发表论文（Tu and Yi, 2017, JoE），利用模型选择和平均的方法，来解决协整误差修正模型中如何利用误差修正项 的关键问题。诺奖得主 Robert Engle 和 Clive Granger 于 1987 年在 Econometrica 上发表文章，提出了非平稳单位根过程之间存在协整，即单位根过程的线性组合成为平稳的序列，以及其对应的误差修正模型（Engle and Granger，1987）。在协整的诸多表述中，误差修正模型尤为重要，特别是对于理解经济结构和预测来讲。因为它描述了不平稳变量之间的长期均衡，即协整关系，是如何通过误差修正项来影响它们的短期波动的（差分变量）。因此，误差修正模型被用来构建对非平稳协整变量的预测。然而，研究表明，误差修正项的系数往往在统计上是不显著的，由此所导出的预测也并不是总比简单的差分模型好（Hoffman and Rasche, 1996; Jansen and Wang, 2006）。如何利用误差修正项来构建预测成为研究的关键问题。Tu and Yi（2017）首次在协整的误差修正模型中讨论使用模型选择和平均的方法来解决如何利用误 差修正项的问题，并且建立了模型选择和模型平均预测的风险渐近表达式，提出了可行的模型选择和模型平均估计量，通过数值模拟和应用实例说明模型平均方法带来的预测风险更小。

席瑞斌副教授：统计中心与数学学院合聘。2009 年华盛顿大学数学博士，2012 年加入统计中心。

研究方向：生物数据，特别是高通量生物数据的统计计算和分析，大数据及高维数据的分析以及贝叶斯统计。

研究介绍：

结构变异是指会导致基因组结构发生变化的基因组变异。结构变异在正常人基因组中广泛存在，并可能与包括癌症在内的多种疾病相关。目前探测结构变异主要基于第二代测序数据，研究者已经发展了多个基于二代测序数据探测结构变异的方法。然而，这些方法的特异性及敏感性都有待进一步提高，不同算法所预测的结构变异常常有很大区别，因此造成后续分析极大的困难。席瑞斌发展了一个探测结构变异的算法 SVmine，此算法通过整合多个算法预测结果从而提高探测结构变异的能力。大量的模拟和实际数据分析表明，无论是结构变异检测的特异性、敏感性还是断点的准确性，SVmine 均显著好于其他算法（图 1）。此文发表于 Bioinforamtics 杂志。Bioinformatics 是由牛津大学出版社出版的生物信息学杂志，主要发表生物信息及计算生物方面的学术论文，是国际公认的生物信息学权威杂志，2016 年影响因子 7.3。北京大学博雅博士后夏禹超是论文的第一作者，博士生刘芸为该文做出了重要贡献。

图 1. A. SVmine 在一个结直肠癌基因组中探测到的基因融合 VTI1A-TCF7L2，其预测的断点位置与 Sanger 测序验证的

断点位置完全相同。B. 在 8 个结直肠癌基因组中，6 个算法的体细胞结构变异预测的两两相交情况，图中的数字表示横轴和纵轴算法共同预测的结构变异在横轴算法预测结构变异中的比例。例如，一行二列的数字 27 代表 27%的 Delly 预测的结构变异同时也在 BreakDancer 预测的结构变异中。C. 经过 SVmine 整合分析后，算法两两预测交集有显著的提高。图中的数字与 B 类似。例如，经过 SVmine 整合后，现在 47% Delly 预测的结构变异同时也在 BreakDancer 预测的结构变异中。D. 不同算法预测结构变异的断点位置与 Sanger 测序验证断点位置的距离的箱式图，SVmine 所预测的断点位置在所有算法中最准确。

2. 中心部分成员

姓名	职务	研究方向	学校院系
姚方	主任	函数型数据分析、高维与复杂结构数据分析、动态数据建模、半参数与非参数方法	数学科学学院概率统计系
周晓华	副主任	随机临床试验中的设计及统计方法、因果推断分析	北京国际数学研究中心
王汉生	副主任	高维数据分析、变量选择、数据降维、值理论、以及半参数模型	光华管理学院商务统计与经济计量系
陈松蹊	科学委员会主席	超高维数据统计推断、大数据统计算法、环境统计、随机过程统计推断、计量经济	光华管理学院商务统计与经济计量系
耿　直	中心教员	因果推断、数理统计、生物医学统计、因果网络、贝叶斯网络、图模型	数学科学学院概率统计系
李　程	中心教员	生物信息学、癌症基因组学、大数据统计建模	生命科学学院生物信息专业
任艳霞	中心教员	测度值马氏过程、分枝粒子系统、概率位势理论	数学科学学院概率统计系
席瑞斌	中心教员	生物医学大数据、生物信息学、高维统计、网络模型、贝叶斯统计	数学科学学院概率统计系
杨静平	中心教员	信用风险、风险管理、债券的理论及应用、风险相关性、精算学	数学科学学院金融数学系

虞吉海	中心教员	空间计量、面板数据	光华管理学院商务统计与经济计量系
张志华	中心教员	机器学习、贝叶斯统计、自然语言处理	数学科学学院概率统计系
吴　岚	中心教员	精算学、金融统计	数学科学学院金融数学系
贾金柱	中心教员	高维统计推断、统计机器学习、因果推断	数学科学学院概率统计系
涂云东	中心教员	金融计量理论、理论计量经济学、应用计量经济学	光华管理学院商务统计与经济计量系
林　伟	中心教员	高维统计、统计机器学习	数学科学学院概率统计系
宋晓军	中心教员	计量经济学理论及应用计量经济学	光华管理学院商务统计与经济计量系
邓明华	中心教员	生物信息学、计算生物学	数学科学学院概率统计系
张俊妮	中心教员	因果推断、贝叶斯分析、蒙特卡洛方法、数据挖掘及其应用	光华管理学院商务统计与经济计量系
房祥忠	中心教员	生存分析、可靠性、纵向数据、基尼系数、空间统计等	数学科学学院概率统计系
张　原	中心教员	互相作用粒子系统、随机游动的覆盖问题以及 DLA 研究上提出新的理论	数学科学学院概率统计系
苗　旺	中心教员	大数据的因果推断方法及其在经济学、环境科学中的应用	光华管理学院商务统计与经济计量系
张　成	中心教员	统计学习、贝叶斯推断、计算生物学	数学科学学院概率统计系

陈　嵘	特聘教授	非线性和多变量时间序列分析、蒙特卡洛方法、统计计算和贝叶斯分析及其应用	Rutgers University
金加顺	特聘教授	大规模多元假设检验、 癌症分类、 变量选择、 谱聚类和主成分分析（PCA）、图论和精度矩阵、网络分析、随机矩阵理论	Carnegie Mellon University

3. 中心行政团队

穆慧娜，2015 年武汉大学经济与管理学院硕士毕业。2019 年 9 月加入中心。

王彦懿，2016 年中国人民大学环境学院硕士毕业。2018 年 6 月加入中心。

四、统计中心科研项目

统计科学中心 2020 在研国家级项目：

序号	项目（课题）名称	负责人	项目来源	项目类型	开始时间	结题时间
1	空气质量统计诊断模型	陈松蹊	国家重点研发计划	重大项目	201607	202006
2	空气质量统计评估方法（课题负责人）	陈松蹊	国家重点研发计划	重大项目	201607	202006
3	大气污染物的时空统计模型（课题负责人）	黄辉	国家重点研发计划	重大项目	201607	202006
4	基于统计与数值模式的多污染物数据场构建（课题负责人）	林伟	国家重点研发计划	重大项目	201607	202006
5	污染与宏观经济、人口健康的计量分析（课题负责人）	虞吉海	国家重点研发计划	重大项目	201607	202006
6	大气污染数据质量监控（课题负责人）	王汉生	国家重点研发计划	重大项目	201607	202006
7	部分可观测马尔科夫决策过程的随机规划算法（课题负责人）	张志华	国家重点研发计划	重大项目	201912	202212
8	粒子系统与渗透模型（课题负责人）	陈大岳	国家重点研发计划	重大项目	202012	202511
9	复杂结构的函数型数据分析	姚方	国家自然科学基金	重点项目	202001	202412

10	大数据驱动的管理决策模型与算	陈松蹊	国家自然科学基金	重点项目	201601	202012
11	面向管理决策大数据分析的理论与方	陈松蹊	国家自然科学基金	重点项目	202101	202112
12	面向儿童脑发育障碍性疾病的神经机制建模与辅助诊疗算法	陈松蹊	国家自然科学基金	数学天元基金	202101	202412
13	面向儿童脑发育障碍性疾病的神经机制建模与辅助诊疗算法	周晓华	国家自然科学基金	数学天元基金	202101	202212
14	非局部随机系统的平均化原理	刘 勇	国家自然科学基金	数学天元基金	202001	202012
15	基于新型数学和统计模型的新型冠状病毒(2019-nCoV）发生和发展规律的研究	周晓华	国家自然科学基金	应急专项	202001	202112
16	超过程的极限理论	任艳霞	国家自然科学基金	面上项目	201701	202012
17	高维复杂数据的稀疏与低秩建模及推断	林伟	国家自然科学基金	面上项目	201701	202012
18	试验设计的统计推断基础理论研究	艾明要	国家自然科学基金	面上项目	201701	202012
19	高维图模型的结构空间及学习方法	何洋波	国家自然科学基金	面上项目	201701	202012
20	金融中的动态 copula 理论及其应用研究	杨静平	国家自然科学基金	面上项目	201701	202012

21	非对称马氏过程的理论研究及应用	蒋达权	国家自然科学基金	面上项目	201801	202212
22	高维函数型数据的检验、回归与分类	姚 方	国家自然科学基金	面上项目	201801	202212
23	基于因果推断框架下的最优治疗方案选择的统计方法研究	周晓华	国家自然科学基金	面上项目	201801	202112
24	大规模优化问题的近似牛顿方法：理论与实现	张志华	国家自然科学基金	面上项目	201801	202112
25	二维高斯场及其相关问题	章复熹	国家自然科学基金	面上项目	201801	202112
26	因果推断方法研究及在食品安全与生物医学中的应用	耿直	国家自然科学基金	面上项目	201801	202112
27	成分数据分析方法及其在宏基因组数据中的应用	邓明华	国家自然科学基金	面上项目	201801	202212
28	树上随机游动中的多层分析问题	陈大岳	国家自然科学基金	面上项目	202001	202312
29	基于单细胞转录组测序的网络分析方法和理论	席瑞斌	国家自然科学基金	面上项目	202001	202312
30	与时间变换有关的资产定价和最优执行策略相关问题研究	程 雪	国家自然科学基金	面上项目	202001	202312

31	现代非平衡态热力学和统计物理中若干重要问题的随机数学理论	葛颢	国家自然科学基金	面上项目	202001	202312
32	测度值分枝马氏过程及相关 Fisher-KPP 方程性质研究	任艳霞	国家自然科学基金	面上项目	202101	202412
33	大数据分析中的试验设计和抽样技术研究	艾明要	国家自然科学基金	面上项目	202101	202412
34	基于风险相依性的风险组合的若干问题	杨静平	国家自然科学基金	面上项目	202101	202412
35	关于无对照研究和中介分析的因果建模与推断	苗 旺	国家自然科学基金	面上项目	202101	202412
36	三维自排斥游走的标度极限的探索	李欣意	国家自然科学基金	面上项目	202101	202412
37	市场冲击模型中引入执行风险后之最优执行策略及相关问题的研究	程 雪	国家自然科学基金	青年科学基金项目	201701	201912
38	带有边界条件的扩散限制凝聚模型(DLA)研究	张 原	国家自然科学基金	青年科学基金	202001	202212

统计科学中心 2020 在研其他代表性项目：

序号	项目（课题）名称	负责人	项目来源	项目类型	开始时间	结题时间
1	实时数据流的统计学习理论与方法	席瑞斌	国防科技委	重点项目	201809	202109
2	针对复杂军事系统的强化学习方法	张志华	国防科技委	重点项目	201809	202109
3	人工智能的统计理论与算法基础	林伟	北京市自然科学基金	重点项目	201910	202310
4	过参数化模型与因果学习的统计理论	林伟	北京智源人工智能研究院	横向项目	201906	202105
5	非参数贝叶斯机器学习方法与概率编程	张志华	北京智源人工智能研究院	横向项目	201906	202105
6	Large Scale Causal Structure Learning	耿直	华为技术有限公司	横向项目	201910	202210
7	深度学习的理论算法项目	张志华	华为技术有限公司	横向项目	201910	202010
8	A study on the outbreak and pattern of development of 2019 Novel Coronavirus, based on new mathematical and statistical models	周晓华	比尔及梅琳达&盖茨基金会	横向项目	202006	202205

9	基于多元信息的新型冠状病毒传播预警预测模型构建和决策平台研究	周晓华	浙大新冠应急专项	横向项目	202002	202008
10	北京大学—中债估值研究中心	杨静平	中债金融估值中心有限公司	横向项目	201912	202412
11	三方共建智慧体育大数据联合实验室合作框架协议	马尽文	国家体育总局体育科学研究所；杭州明霖信息技术有限公司	横向项目	201705	202005
12	典型统计模型及可靠性评估算法梳理及确认	房祥忠	中国航天标准化研究所	横向项目	202006	202112
13	量化投资策略和证券公司客户量化分析研究	吴岚	方正证券股份有限公司	横向项目	201801	202101
14	面向岩性岩相预测、地质体雕刻的深度学习模型研究	马尽文	中国石油集团东方地球物理勘探有限责任公司	横向项目	201905	202012
15	深度学习理论与算法研究项目	张志华	北京旷视科技有限公司	横向项目	201908	202007

五、中心毕业生就业去向

2020 年，北京大学统计、概率及金融数学专业共有 14 名博士毕业生，5 名硕士毕业生。毕业时 19 人均已落实工作，其中 4 人在中国知名高校获得教职工作。以下是中心教员指导的毕业生去向。

学位类别	姓名	专业	单位名称	指导教师
学术学位博士	陈天遥	统计学	中国农业银行总行	杨静平
	何顺	统计学	工银瑞信基金管理有限公司	邓明华
	黄家盛	概率论与数理统计	天津理工大学	姚方
	黄译旻	统计学	北京华为数字技术有限公司	艾明要
	梁德才	统计学	南开大学	姚方
	王伟楠	统计学	北京大学生物医学前沿创新中心	邓明华
	武林杰	统计学	中国医学科学院血液学研究所	席瑞斌
	肖一君	概率论与数理统计	杭州学军中学教育集团文渊中学	周晓华
	张慧铭	统计学	澳门大学	姚方
	张雨	统计学	银华基金管理股份有限公司	房祥忠
	赵洪鑫	统计学	绛霄信息科技（上海）有限公司	吴岚
	赵建杰	概率论与数理统计	北京华为数字技术有限公司	任艳霞
	赵林杰	概率论与数理统计	法国国家信息与自动化研究所（INRIA）	陈大岳
	周正雍	统计学	摩根大通投资咨询（北京）有限公司	何杨波
学术学位硕士	邓玉颖	统计学	腾讯科技（深圳）有限公司	周晓华
	董安澜	统计学	九坤投资（北京）有限公司	姚方
	李鹏程	概率论与数理统计	中信证券股份有限公司	蒋达权
	陆嘉盛	概率论与数理统计	百度在线网络技术（北京）有限公司	蒋达权
	罗艺	概率论与数理统计	中国农业银行股份有限公司	蒋达权

六、交流访问

受全球新冠疫情的影响，2020 年中心的国内外学术交流活动受到严重影响。中心克服困难，在严格执行各级防疫政策的前提下，不断推进同国内外学者的学术交流活动，2020 年统计科学中心共接待海内外访问学者 14 人次，参与学术报告、短期课程以及合作项目等各类学术活动。

统计科学中心 2020 年来访人员名单

序号	来访人员	时间
长期访问人员		
1	丁秀才，加州大学戴维斯分校	2020/8/10-2020/8/24
短期来访人员		
1	王学钦，中国科学技术大学管理学院	2020/05/29
2	张新雨，中科院数学与系统科学研究院/预测中心	2020/06/08
3	郑术蓉，东北师范大学	2020/06/17
4	刘卫东，上海交通大学数学科学学院	2020/06/26
5	刘岩，武汉大学	2020/10/9
6	陈昕韫，香港中文大学（深圳）	2020/10/9
7	林伟，复旦大学	2020/10/22
8	张海伦，香港中文大学（深圳）	2020/10/29
9	张世华，中国科学院数学与系统科学研究院	2020/11/26
10	Xiao Wu, Harvard University	2020/12/3
11	史建清，南方科技大学	2020/12/10
12	高凤楠，复旦大学	2020/12/17
13	马占宇，北京邮电大学	2020/12/24

七、论文发表

自 2020 年 1 月至 12 月，统计科学中心教员共发表论文 61 篇，按第一作者姓氏字母排序排列：

[1] Cheng, Jing and **Ai, Mingyao**. *Optimal designs for panel data linear regressions.* Statistics & Probability Letters, 163 (2020).

[2] Chen, Cheng; Gu, Ming; **Zhang, Zhihua**; Zhang, Weinan; Yu, Yong. *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*. PMLR, 108, (2020),766-775.

[3] Chen, Liang; He, Shun; Zhai, Yuyao; Deng, Minghua. *Direct interaction network inference for compositional data via codaloss*. Journal of Bioinformatics and Computational Biology, 14(7), (2020), e0207731.

[4] Chen, Chong; Wu, Changjing; Wu, Linjie; Wang, Xiaochen; **Deng, Minghua** and **Xi, Ruibin**, *ScRMD: Imputation for single cell RNA-seq data via robust matrix decomposition.* Bioinformatics, 36 (10), (2020), 3156-3161.

[5] Chen, Liang; Wang, Weinan; Zhai, Yuyao and **Deng, Minghua**. *Single-cell transcriptome data clustering via multinomial modeling and adaptive fuzzy K-means algorithm.* Frontiers in Genetics, 11, (2020), 295.

[6] Chen, Liang; Zhai, Yuyao; He, Qiuyan; Wang, Weinan and **Deng, Minghua**. *Integrating deep supervised, self-supervised and unsupervised learning for single-cell RNA-seq clustering and annotation*. Genes, 11(7), (2020), 792.

[7] Chen, Liang; Wang, Weinan; Zhai, Yuyao; **Deng, Minghu**. *Deep soft K-means clustering with self-training for single-cell RNA sequence data*. NAR Genomics and Bioinformatics, 2 (2), (2020), lqaa039.

[8] Chen, Liang.; Zhai, Yuyao; He, Qiuyan; Wang, Weinan; **Deng, Minghua.** *Integrating deep supervised, self-supervised and unsupervised learning for single-cell RNA-seq clustering and annotation*. Genes, 11(7), (2020), 792

[9] Chen, Tianyao; **Cheng, Xue** and **Yang, Jingping**. *Decomposing correlated random walks on common and counter movements.* Statistics & Probability Letters, 158, (2020), 108616.

[10] Cheng, Jing; **Ai, Mingyao**. *Optimal designs for panel data linear regressions*. Statistics & Probability Letters, 163, (2020), 108769.

[11] Chi, Weilai and **Deng, Minghua**. *Sparsity-penalized stacked denoising autoencoders for imputing single-cell RNA-seq data*. Genes, 11 (5), (2020), 532.

[12] Cong, Tianshu; Xia, Aihua and **Zhang, Fuxi**. *A large sample property in approximating the superposition of i.i.d. Finite point processes*. Stochastic Processes and Their Applications, 130 (7), (2020), 4493-4511.

[13] Cortes-Ciriano, Isidro; Lee, Jake June-Koo; **Xi, Ruibin**; et al. *Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing*. Nature Genetics, 52 (3), (2020), 331-+.

[14] Cui, Yongping; Chen, Hongyan; **Xi, Ruibin**; et al. *Whole-genome sequencing of 508 patients identifies key molecular features associated with poor prognosis in esophageal squamous cell carcinoma*. Cell Research, 30 (10), (2020), 902-913.

[15] Deng, Yuhao; You, Chong; Liu, Yukun; Qin, Jin; **Zhou, Xiaohua.** *Estimation of incubation period and generation time based on observed length-biased epidemic cohort with censoring for COVID-19 outbreak in China.* Biometrics, (2020), Doi: 10.1111/biom.13325.

[16] Fang, Jun; Jiang, Fan; Liu, Yong and **Yang, Jingping**. *Copula-based Markov process, Insurance Mathematics & Economics*, 91, (2020), 166-187.

[17] Fang, Zhuangyan; He, Yangbo. *In IDA with Background Knowledge*. Conference on Uncertainty in Artificial Intelligence, PMLR (2020), 270-279.

[18] Fu, Jingxue and **Wu, Lan**. *Consistent estimation of the number of regimes in Markov-switching autoregressive models*. Communications in Statistics-Theory and Methods, (2020), https://doi.org/10.1080/03610926.2020.1777304.

[19] Gu, Jia; Yan, Han; Huang, Yaxuan; Zhu, Yuru; Sun, Haoxuan; Qiu, Yumou; **Chen, Songxi.** *Comparing containment measures among nations by epidemiological effects of COVID-19.* National Science Review, 7 (12), (2020), 1847-1851.

[20] Huang, Yimin; Kong, Xiangshun and **Ai, Mingyao**. *Optimal designs in sparse linear models*. Metrika, 83 (2020), no. 2, 255-273.

[21] Kuang, Kun; Li, Lian; **Geng, Zhi**; Xu, Lei; Zhang, Kun; Liao, Beishui; Huang, Huaxin; Ding, Peng; **Miao, Wang** and Jiang, Zhichao. *Causal inference*. Engineering, 6 (3), (2020), 253-263.

[22] Li, Hongkai; **Geng, Zhi**; Sun, Xiaoru; Yu, Yuanyuan and Xue, Fuzhong, A novel path-specific effect statistic for identifying the differential specific paths in systems epidemiology. BMC Genetics, 21(1), (2020), 85.

[23] Li, Zhiming; Kong, Qingxun and **Ai, Mingyao**. *Construction of some s-level regular designs with general minimum lower-order confounding*. Statistics & Probability Letters, 167 (2020).

[24] Li, Hongkai; **Miao, Wang**; Cai, Zheng; Liu, Xinhui; Zhang, Tao; Xue, Fuzhong and **Geng, Zhi**. *Causal data fusion methods using summary-level statistics for a continuous outcome*. Statistics in Medicine, 39 (8), (2020), 1054-1067.

[25] Li, Xiang; Huang, Kaixuan; Yang, Wenhao; Wang, Shusen; **Zhang, Zhihua**. *On the convergence of fedavg on non-iid data*. International Conference on Learning Representations, (2020).

[26] Li, Xiang; Wang, Shusen; **Zhang, Zhihua**. *In Do Subsampled Newton Methods Work for High-Dimensional Data*? Proceedings of the AAAI Conference on Artificial Intelligence, (2020), 4723-4730.

[27] Lin, Zhenhua; **Yao, Fang**. *Functional regression on the manifold with contamination.* Biometrika, (2020), https://doi.org/10.1093/biomet/asaa041.

[28] Liu, Yue; Fang, Zhuangyan; He, Yangbo; **Geng, Zhi**. *In Collapsible IDA: Collapsing parental sets for locally estimating possible causal effects*. Conference on Uncertainty in Artificial Intelligence, PMLR (2020), 290-299.

[29] Liu, Yue; Fang, Zhuangyan; **He, Yangbo**; **Geng, Zhi** and Liu, Chunchen. *Local causal network learning for finding pairs of total and direct effects*. Journal of Machine Learning Research, 21 (148), (2020), 1-37.

[30] Luo, Xiao; Tu, Xinming; Ding, Yang; Gao, Ge and **Deng, Minghua**. *Expectation pooling: An effective and interpretable pooling method for predicting DNA-protein binding*. Bioinformatics, 36(5), (2020), 1405-1412.

[31] Mu, Y. and **Zhang, Yuan**. *On some threshold-one attractive interacting particle systems on homogeneous trees.* 2018. *Journal of Applied Probability*, 57(3), (2020), 866-898.

[32] Procaccia, Eviatar B.; Rosenthal, Ron and **Zhang, Yuan**. *Stabilization of DLA in a wedge*. Electronic Journal of Probability, 25, (2020).

[33] Procaccia, E. B.; Ye, J.; **Zhang, Yuan**. *Stationary DLA is well defined.* Journal of Statistical Physics, 181 (4), (2020), 1089-1111.

[34] Procaccia, E. B.; **Zhang, Yuan**. *On sets of zero stationary harmonic measure*. Stochastic Processes and their Applications, 131, (2021), 236-252.

[35] Qin, Jing; You, Chong; Lin, Qiushi; Hu, Taojun; Yu, Shicheng; **Zhou, Xiaohua**. *Estimation of incubation period distribution of COVID-19 using disease onset forward time: a novel cross-sectional and forward follow-up study*. Science Advances, 6 (33), (2020), eabc1202.

[36] **Ren, Yan-Xia**; Song, Renming and Sun, Zhenyao. *Spine decompositions and limit theorems for a class of critical superprocesses*. Acta Applicandae Mathematicae, 165 (1), (2020), 91-131.

[37] **Ren, Yan-Xia**; Song, Renming and Sun, Zhenyao. *Limit theorems for a class of critical superprocesses with stable branching*. Stochastic Processes and Their Applications, 130(7), (2020), 4358-4391.

[38] Shi, Xu; **Miao, Wang**; Nelson, Jennifer C.; Tchetgen Tchetgen, Eric. *Multiply robust causal inference with double-negative control adjustment for categorical unmeasured confounding*. Journal of the Royal Statistical Society Series B (Statistical Methodology), 82 (2), (2020), 521-540.

[39] Sieverling, Lina; Hong, Chen; Koser, Sandra D.; Ginsbach, Philip; Kleinheinz, Kortine; Hutter, Barbara; Braun, Delia M.; Cortes-Ciriano, Isidro; **Xi, Ruibin**; et al. *Genomic footprints of activated telomere maintenance mechanisms in cancer.* Nature Communications, 11 (1), (2020).

[40] Sun, Zhihua; Luo, Dongshan; **Zhou, Xiaohua** and Zhang, Qingzhao. *Comparative studies on the adequacy check of parametric measurement error models with auxiliary variable*. Statistical Papers, (2020), DOI: 10.1007/s00362-019-01154-3.

[41] Wan, Yating; Xu, Minya; Huang, Hui; **Chen, Song Xi.** *A spatio-temporal model for the analysis and prediction of fine particulate matter concentration in Beijing*. Enviromentrics, (2020), e2648.

[42] Wang, C.; **Zhang, Y**.; Bertozzi, A. L.; Short, M. B. A stochastic-statistical residential burglary model with independent Poisson clocks. European Journal of Applied Mathematics, (2020), 1-27.

[43] Wang, Tianbing; Wang, Yanhua; Xu, Tingmin; Li, Libing; Huo, Mingli; Li, Xian; He, Yingdong; Lin, Qiushi; Mei, Bingsong; **Zhou, Xiaohua** and Jiang, Baoguo. *Epidemiological and clinical characteristics of 3327 cases of traffic trauma deaths in Beijing from 2008 to 2017 a retrospective analysis*. Medicine, 99 (1), (2020), e18567.

[44] Wang, Xinpei; **Jia, Jinzhu** and Huang, Tao. *Shared genetic architecture and casual relationship between leptin levels and type 2 diabetes: Large-scale cross-trait meta-analysis and mendelian randomization analysis*. BMJ Open Diabetes Research & Care, 8 (1), (2020), e001140.

[45] Wu, Huangjian; Zheng, Xiaogu; Zhu, Jiang; **Lin, Wei**; Zheng, Haitao; Chen, Xueshun; Wang, Wei; Wang, Zifa and **Chen, Song Xi**. *Improving PM2.5 forecasts in China using an initial error transport model.* Environmental Science & Technology, 54 (17), (2020), 10493-10501.

[46] Wu, Linjie; Wang, Han; Xia, Yuchao and **Xi, Ruibin**. *CNV-BAC: Copy number variation detection in bacterial circular genome*. Bioinformatics, 36 (12), (2020), 3890-3891.

[47] **Wu, Lan**; Zang, Xin and Zhao, Hongxin, Analytic value function for a pairs trading strategy with a levy-driven Ornstein-Uhlenbeck process, Quantitative Finance 20 (8), (2020), 1285-1306.

[48] Xie, Guangzeng; Luo, Luo; Lian, Yijiang; **Zhang, Zhihua**. *In Lower complexity bounds for finite-sum convex-concave minimax optimization problems*. International Conference on Machine Learning, PMLR (2020); 10504-10513.

[49] Xu, Ziping; **Chen, Song Xi** and Wu, Xiaoqing. *Meteorological change and impacts on air pollution: Results from north China*. Journal of Geophysical Research-Atmospheres, 125(16), (2020), e2020JD032423.

[50] Xue, Kaijie and **Yao, Fang**. *Distribution and correlation-free two-sample test of high-dimensional means*. Annals of Statistics, 48 (3), (2020), 1304-1328.

[51] Ye, Haishan; Luo, Luo; Zhang, Zhihua. *Accelerated Proximal Subsampled Newton Method*. IEEE Transactions on Neural Networks and Learning Systems, (2020), 1-15.

[52] Ye, Haishan; Luo, Luo and **Zhang, Zhihua**. *Nesterov's acceleration for approximate Newton*. Journal of Machine Learning Research, 21 (142), (2020), 1-37.

[53] Yin, Shenyi; **Xi, Ruibin**; Wu, Aiwen; Wang, Shu; Li, Yingjie; Wang, Chaobin; Tang, Lei; Xia, Yuchao; Yang, Di; Li, Juan; et al. *Patient-derived tumor-like cell clusters for drug testing in cancer therapy*. Science Translational Medicine, 12 (549), (2020).

[54] Yin, Yunjian; Liu, Lan; **Geng, Zhi** and Luo, Peng. *Novel criteria to exclude the surrogate paradox and their optimalities*. Scandinavian Journal of Statistics, 47 (1,) (2020), 84-103.

[55] Yu, Jun; Wang, Haiying; **Ai, Mingyao** and Zhang, Huiming. *Optimal distributed subsampling for maximum quasi-likelihood estimators with massive data*. Journal of the American Statistical Association, (2020), 1-12.

[56] **Zhang, Cheng**; Dinh, Vu and Matsen, Frederick A. *Nonbifurcating phylogenetic tree inference via the adaptive LASSO*. Journal of the American Statistical Association, (2020), 1-16.

[57] **Zhang, Cheng**. *Improved Variational Bayesian Phylogenetic Inference with Normalizing Flows.* NeurIPS, (2020).

[58] **Zhang, Yunjun**; Li, Yuying; Wang, Lu; Li, Mingyuan and **Zhou, Xiaohua**. *Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a metropolis of China*. International Journal of Environmental Research and Public Health, 17 (10), (2020), 3705.

[59] Zheng, Zemin; Lv, Jinchi; **Lin, Wei**. *Nonsparse learning with latent variables*. Operations Research, (2020), https://doi.org/10.1287/opre.2020.2005.

[60] 张澍一; **陈松蹊**; 郭斌; 王恒放; **林伟**. *气象调整下的区域空气质量评估.* 中国科学: 数学, 50, (2020), 527–558..

[61] **张原**; 尤翀; 蔡振豪; 孙嘉瑞; 胡文杰; **周晓华**. *新冠肺炎 (COVID-19) 新型随机传播动力学模型及应用.* 应用数学学报, 43 (2), (2020), 440-451.

八、学术讲座

统计中心 2020 年共组织学术报告 14 场，报告人来自国内各高校和研究机构。其中来自海外的学者 1 场，国内的学者 13 场。

[1] Use of Random Integration to Test Equality of High Dimensional Covariance Matrices

报告人： 王学钦，中国科学技术大学管理学院教授

时间：2020-05-29 15:00-16:00

地点：腾讯会议

[2] Pilot CEO and innovation: A model averaging approachfor divergent-dimensional Poisson regressions

报告人： 张新雨，中科院数学与系统科学研究院/预测中心研究员

时间：2020-06-08 10:00-11:00

地点：腾讯会议

[3] 高维相关矩阵的统计推断

报告人： 郑术蓉，东北师范大学教授、博士生导师

时间：2020-06-17 10:00-11:00

地点：腾讯会议

[4] Variance Reduced Median-of-Means Estimator for Byzantine-Robust Distributed Inference

报告人： 刘卫东，上海交通大学数学科学学院副院长，特聘教授，国家级重大人才计划（基金委）获得者

时间：2020-06-26 15:00-16:00

地点：腾讯会议

[5] School Colloquium——Prediction, Computation, and Representation---The Nature of Machine Learning

报告人： 北京大学数学科学学院、统计科学中心张志华教授

时间：2020-09-18 15:00-16:00

地点：腾讯会议

[6] 面向经济与金融应用的监督学习算法比较与推测函数特征刻画

报告人：刘岩，武汉大学

时间：2020-10-09 14:00-15:00

地点：光华 2 号楼 216 教室

[7] SGD-based Online Pricing and Capacity Sizing of Queueing Systems

报告人：林华珍 (Joint work with Wei Liu, Wei Lan), 西南财经大学

时间：陈昕韫博士，香港中文大学（深圳）

地点：光华 2 号楼 216 教室

[8] Time Series Analytics: Causation Detections and Dynamics Predictions

报告人：林伟，复旦大学

时间：2020-10-22 14:00-15:00

地点：腾讯会议

[9] On-Demand Ride-Matching in a Spatial Model with Abandonment and Cancellation

报告人： 张海伦，香港中文大学（深圳）

时间：2020-10-29 14:00-15:00

地点：腾讯会议

[10] Probabilistic Matrix Decomposition for Complex Noise Modeling

报告人： 张世华，中国科学院

时间：2020-11-26 14:00-15:00

地点：理科 1 号楼 1304

[11] Air Pollution, COVID-19 Pandemic, and Human Health: Connecting the Science with Statistics and Causal Inference

报告人： Xiao Wu，Harvard University

时间：2020-12-03 10:00-11:00

地点：腾讯会议

[12] Modelling Function-valued Processes with Non-separable and/or Non-stationary Covariance Structure

报告人： 史建清，南方科技大学

时间：2020-12-10 14:00-15:00

地点：理科 1 号楼 1114

[13] Loss Two-sample testing of high-dimensional linear regression coefficients via complementary sketching

报告人： 高凤楠，复旦大学

时间：2020-12-17 14:00-15:00

地点：理科 1 号楼 1114

[14] 细粒度图像分类

报告人： 马占宇，北京邮电大学

时间：2020-12-24 14:00-15:00

地点：理科一号楼 1114

九、委员会

1. 科学委员会

陈松蹊

主席，现任 Bernoulli Society 的 Scientific Secretary、国家统计局咨询委员，讲席教授、国家高层次专家；教育部自然科学一等奖；美国科学促进会、数理统计学会、美国统计学会会士；IMS-中国候任主席。主要研究方向：超高维数据统计推断、大数据统计算法、环境统计、随机过程统计推断、计量经济。

李程

委员，美国加州大学洛杉矶分校统计系博士。2002 至 2013 年在哈佛大学生物统计系、Dana-Farber Cancer Institute 作为助理教授、副教授从事研究与教学工作。2013 年 4 月全职加入北京大学生命科学学院、生命科学联合中心、生物信息中心、统计科学中心。

任艳霞

委员，现任北京大学数学科学学院教授、概率统计系主任，《应用数学学报》编委、《应用概率统计》编委。主要研究方向：测度值马氏过程、分枝粒子系统、概率位势理论。

王汉生

委员，光华管理学院商务统计与经济计量系，嘉茂荣聘讲席教授，博导，系主任。北京大学商务智能研究中心、主任。美国统计学会会士，现为国际统计研究所、美国数理统计研究所、英国皇家统计协会以及泛华统计学会会员。

杨静平

委员，杨静平现任北京大学数学科学学院教授、金融数学系主任，中国工业与应用数学学会第七届理事会理事、中国工业与应用数学学会“金融数学与工程和精算保险专业委员会”副主任、数量经济与数理金融教育部重点实验室（北京大学）副主任、《数学进展》编委。主要研究方向：信用风险、风险管理、债券的理论及应用、风险相关性、精算学。

姚方

委员，数理统计学会 IMS 理事、《加拿大统计期刊》主编；讲席教授、国家高层次专家；数理统计学会与北美统计学会会士；至今任 9 个国际统计学期刊编委，包括 JASA、AOS 等。研究方向：函数型数据分析、高维与复杂结构数据分析、动态数据建模、半参数与非参数方法。

虞吉海

委员，现任北京大学光华管理学院商务统计与经济计量系教授。他本科和研究生均毕业于复旦大学经济系，后在俄亥俄州立大学获得经济学的硕士和博士学位。

张志华

委员，现任北京大学数学科学学院教授；Journal of Machine Learning Research 执行编委；ICML 2020 领域主席。主要研究方向：机器学习、贝叶斯统计、自然语言处理。

周晓华

委员，现任国际生物统计学会中国分会会长；中国现场统计研究会生物医学统计分会会长；讲席教授、国家高层次专家；美国科学促进会、数理统计学会、北美统计学会会士；中国数学会医学数学专业委员会主任委员。研究方向：随机临床试验中的设计及统计方法、因果推断分析。

2．指导委员会

张平文

主席，北京大学副校长、大数据分析与应用技术国家工程实验室主任，数学科学学院教授、博士生导师，中国科学院院士，北京大学数学及其应用教育部重点实验室主任。曾获国家自然科学二等奖、教育部高等学校自然科学一等奖、入选国家级重大人才计划（基金委）、冯康科学计算奖、“百千万人才工程”国家级人选、国家自然科学基金委“创新研究群体”学术带头人等多项荣誉。

陈大岳

委员，北京大学数学科学学院院长，中国数学会副监事长；中国概率统计学会副理事长；入选国家级重大人才计划（基金委）；国务院学位委员会学科评议组秘书长；曾任中国数学会概率统计学会副理事长、秘书长、《数学进展》主编。

刘俏

委员，北京大学光华管理学院院长，入选国家级重大人才计划（基金委）。兼任深圳证券交易所专家评审委员会委员和中国证监会、中国金融期货交易所、民生银行、及深交所博士后站指导导师。在公司金融，实证资产定价、实际期权、市场微观结构和中国经济研究等方面拥有众多著述。