

北京大学统计科学中心

Center for Statistical Science, Peking University

2018年度报告

2018 Annual Report



北京大学统计科学中心

CENTER FOR STATISTICAL SCIENCE AT PKU



2018 年 4 月 14 日，统计科学论坛暨北京大学“双一流”学科建设研讨会

目 录

一、综述	• • • • • 1
二、中心要闻	• • • • • 2
三、人员组成	• • • • • 18
四、科研项目	• • • • • 27
五、博士生去向	• • • • • 28
六、交流访问	• • • • • 29
七、论文发表	• • • • • 36
八、研究报告	• • • • • 40
九、学术讲座	• • • • • 42
十、委员会	• • • • • 49

北京大学统计科学中心（Center for Statistical Science, Peking University）成立于2010年7月，由数学科学学院、光华管理学院及其它从事统计研究的相关教师构成，是北京大学实行特殊机制的跨学院交叉学科研究机构。

统计科学中心依托北京大学深厚的学术底蕴和多学科的优势，整合全校统计学研究力量，致力于培养综合型的统计和数据科学人才，扩展统计学在其它学科中的应用，推动统计学研究与教育走向世界，建设国内领先、世界一流的研究机构。

统计科学中心的主要任务是建立一支具有国际水平的统计学研究教学团队，促进统计学与其他学科的交叉合作以及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，为这些学科的发展提供数据收集和数据分析的统计方法。

中心设国际顾问委员会、科学委员会、执行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；执行委员会参与中心的管理运行工作。中心在这几年中利用新体制共引进教员11人，其中3人入选中组部“国家级重大人才青年计划”，3人入选“国家级重大人才计划”长期项目，3人入选“国家级重大人才计划”短期项目。

统计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“一流学科”建设，我们希望在2030年将北京大学的统计学科建设成为世界一流学科。



Peter G. Hall 教授摄

一、综述

2018 年统计科学中心继续在统计学研究、教员引进、博士生培养以及推动统计学发展等方面开展工作。

2017 年，北大统计学科入选教育部“世界一流建设”学科系列；2018 软科发布的“中国最好学科”排名中北京大学位列统计学科第一名，并入选 2018 年新增的前 1% 学科；2018 QS 世界大学排名之统计学与运筹学专业排名中，北大统计学科位列全球第 34。统计学专业的上升，为未来学科建设的发展提供了更加广阔的舞台，北大统计科学中心的发展也面临新机遇与挑战。

今年 4 月，由中心牵头召开的 2018 统计科学论坛暨北京大学“双一流”学科建设研讨会，来自中外知名高校的统计学者与关心中国统计学专业发展的各界领导齐聚一堂，为北京大学统计学专业世界一流学科建设建言献策。北京大学公共卫生学院在 11 月成立了生物统计系。首届系主任由北京大学讲席教授、国际知名的生物统计学家周晓华担任，这标志这统计学科力量进一步增强。

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，统计中心与光华管理学院合聘了苗旺博士，与数学科学学院合聘了张原博士，他们于 2018 年 8 月入职。这是中心时隔几年后首次招聘教员，也反映了目前统计学科教员招聘的国际竞争日益加剧的态势。来自宾夕法尼亚州立大学的暴乐和牛晓月两位访问教授，在中心协助开展教学活动。同时，中心的行政团队有了更新和替换，由于原行政主管张卫香老师的健康原因，李楠和王彦懿加入中心，分别担任行政主管和行政秘书。

6 月，统计中心成功举办了第四届优秀大学生夏令营，十几所高校的 28 名三年级本科生参加；9 月，中心第四批 5 名博士研究生入学，目前中心在读研究生规模已达 15 名，扩充了北京大学在统计学科人才培养上的体量。

2018 年统计中心积极开展同国际和国内统计学界的学术交流与合作，共举办学术报告 45 场，其中来自海外的学者 33 场，国内的学者 12 场。国际著名统计学家、美国国家科学院院士、加州大学伯克利分校统计系教授郁彬教授作为北京大学“大学堂”顶尖学者来北大讲学，并发表了数据科学相关的两场系列演讲。国际著名统计学家、荷兰皇家与艺术科学院院士、莱顿大学 Aad van der Vaart 教授开设了关于高维与非参数模型的贝叶斯统计的短期课程。来自海内外知名学府的统计学与机器学习的 4 位专家 Jiashun Jin、Cheng Yong Tang、Yuan Yao、Zheng Tracy Ke 为统计专业的学生开设了暑期课程，介绍了高维统计，机器学习、深度学习等领域的一些最新进展。中心举办的国内会议——第三届北大-清华统计论坛，在课余时间开展内部学术交流活动“万花统”，进一步开拓了老师和学生的学术视野，促进了统计学领域的交流与合作。

2018 年中心教师共发表了 59 篇学术论文，其中 A 类 15 篇，有 8 篇收录在统计 4 大期刊。中心教师共发表了 34 篇科研报告。在交叉学科研究方面，中心环境统计科研团队发布了《空气质量评估报告（五）--“2+31”城市 2013-2017 年区域污染状况评估》。10 月，关于京津冀空气污染状况评估的论文在大气环境领域英文国际期刊《Atmospheric Environment》上发表。本年度，房祥忠教授当选中国现场统计研究会理事长，陈松蹊教授当选美国科学促进会（AAAS）会士。

二、中心要闻

1. 北大统计学科在软科 2018 中国最好学科获评第一

2018 中国最好学科排名

统计学：1、北京大学 2、北京师范大学 3、中国人民大学

（2018 年北京大学有 11 个前 1% 学科数，有 14 个第一名学科数）

2018中国最好学科排名 各高校第一名学科数统计		
序号	学校名称	第一名学科数
1	北京大学	14
2	清华大学	12
3	中国人民大学	6

2017 年 12 月 28 日，教育部学位与研究生教育发展中心公布全国第四轮学科评估结果显示，北京大学共有 21 个一级学科获评 A+，统计学科是其中一员。全国共有 85 所开设统计学专业的大学参与了 2017-2018 统计学专业排名，其中北京大学统计学科获评 A+。今年 11 月 1 日，软科发布的 2018 “中国最好学科排名”。北京大学统计学科位列第一名。

北京大学的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任教，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。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，砥砺前行，开创更辉煌的事业。

2. 2018 统计科学论坛暨北京大学“双一流”学科建设研讨会成功举办

4 月 14 日，2018 统计科学论坛暨北京大学“双一流”学科建设研讨会在北京大学英杰交流中心举行，本次大会由北京大学统计科学中心、数学科学学院概率统计系、光华管理学院商务统计与经济计量系、数量经济与数理金融教育部重点实验室共同主办。

上午 8 时 30 分，大会在英杰交流中心月光厅拉开序幕，姚方教授主持开幕式，北京大学副校长高松，国家统计局党组成员、副局长鲜祖德，中国概率统计学会理事长何书元，中国现场统计学会理事长房祥忠，北京大学统计科学中心联席主任陈松蹊出席了开幕式并分别发表致辞。各位领导和老师回顾了近年来北京大学统计学在学科发展上所取得的成绩，肯定了对统计学的学科建设和人才培养所作出的贡献，表达了对北京大学统计学进一步更好发展的祝愿和期待。

上午的学术报告环节从 9 时开始，北京大学耿直教授、艾明要教授分别主持了两组专题报告，姚方教授、周晓华教授给了特邀报告。来自密歇根大学、北京大学、东北师范大学、上海交通大学、中国人民大学和南开大学的 7 位学者在大会上报告了自己最新的研究进展。研究内容涵盖高维的高斯图模型、基于倾向值得分的因果推断、黎曼型函数数据的内在方法分析、随机梯度法与统计优化、高维情形下两样本分布的非参数检验、变点分析等方向。

下午 2 时，“双一流”学科建设座谈会在英杰交流中心星光厅举行。座谈会由北京大学刘力平教授主持，陈松蹊教授、耿直教授、房祥忠教授、周晓华教授、刘力平教授、姚方教授参加了这次座谈。与会嘉宾围绕北京大学统计学专业建设世界一流学科展开讨论，总结了北大统计学科目前取得的成就，充分分析了建设一流学科过程中可能存在的挑战，为北大统计学专业建设提出了宝贵意见。陈松蹊教授在座谈会中指出，北京大学为学校统计学科的发展提出了更高的目标和要求，这促使着北京大学统计科学中心更加努力地引进和培养人才，向建设世界一流学科不断迈进。

与此同时，在英杰交流中心月光厅的分组学术报告仍然继续进行，北京大学席瑞斌研究员、邓明华教授分别主持了两组报告，来自中国科学技术大学、复旦大学、北京大学、中山大学、中国科学院的 7 位学者在大会上分享自己的研究成果，内容涉及高维图模型恢复、马尔科夫随机场分析、多项分布的随机效应模型、伪回归、巴拿赫空间上的相关性、分布函数的非参数估计、基于 lasso 的网络差异分析等方面。



大会在全体与会人员的大合影中闭幕，经过一天的会议讨论，国内外的统计学者深入开展交流，沟通成果，增进友谊，至此，“双一流”学科建设研讨会取得了圆满的成功！

3. 郁彬教授入选 2018 北京大学“大学堂”顶尖学者讲学计划

郁彬，国际著名统计学家，美国艺术与科学学院院士，美国国家科学院院士。郁彬教授 1984 年毕业于北京大学数学系。在 2009 年到 2012 年间担任加州大学伯克利分校统计系系主任。郁彬 2006 年当选 Guggenheim Fellow, 2012 年作了伯努利协会的图基纪念演讲 (Tukey Memorial Lecturer)。她还是泛华统计协会 2012 年首届许宝騄奖的三位获得者之一。她也是 AAAS (American Association for the Advancement of Science)、IEEE (Institute of Electrical and Electronics Engineers)、IMS (Institute of Mathematical Statistics) 和 ASA (American Statistical Association) 的会士。也曾是 2013-2014 年度数理统计协会 (IMS) 主席，并于 2013 年度当选美国艺术与科学学院 (American Academy of Arts and Science) 院士，2014 年当选美国国家科学院院士。当选院士被认为是美国学术界最高荣誉之一。



郁彬教授在统计理论、高维数据分析、机器学习等方面成绩斐然，享有很高的国际声誉。她对交叉学科研究即广泛又深入。郁彬教授一直关心北大统计学科的发展，先后担任我校的讲座教授和国家级重大人才计划专家 (短期项目)，指导多名我院青年教师和研究生；是北大微软统计和信息技术实验室的创办者和主任之一，为我校统计学科发展、建设、人才培养做出了巨大贡献。

应北京大学“大学堂”顶尖学者讲学计划的邀请，国际著名统计学家、美国国家科学院院士、加州大学伯克利分校统计系教授郁彬教授，于 2018 年 4 月来我校讲学。讲学期间，郁彬教授发表了数据科学相关的系列演讲。

4 月 25 日晚 7 点，郁彬教授的第一场讲座在理教 103 开始，本场讲座的题目是“数据科学三原则：可预测性、可计算性和稳定性”。讲座在北京大学数学科学学院院长陈大岳教授的主持下开始，在这次讲座中，郁彬教授介绍了数据科学三原则，即可预测性、可计算性和稳定性”。郁彬教授首先高屋建瓴地解释了这三原则之间的错综复杂的联系和它们的重要性。郁彬教授在讲座中通过两个神经科学的交叉学科项目来展示了数据科学的三原则。第一个项目是使用核磁共振的脑信号来重构电影，郁彬教授团队通过加入稳定性的预测模型，获得了比以往效果更好的图像反演效果。第二个项目是使用可预测的迁移学习和稳定的深度图像来刻画脑区域的 V4 区域，V4 区域一直是脑神经科学重点研究的对象，郁彬教授通过与脑神经科学家合作，帮助发现了对 V4 区域脑神经能产生刺激的对特定图像类型。郁彬教授表示，这两个项目的成功离不开数据科学的三原则。



5月9日晚上讲座开始之前，理教103教室内组织起了小型交流会，来自校内外的师生齐聚一堂，互相交流，增进了解，会场气氛热烈，互动频繁。晚上7点，郁彬教授的第二场讲座正式开始，本场讲座的题目是“用迭代随机森林算法发现具有预测性的稳定高阶交互作用”。讲座在统计科学中心联席主任陈松蹊教授的主持下拉开序幕。郁彬教授首先在讲座中介绍了有关基因组学的背景知识，她指出，基因组学已经彻底改革了生物学，这一学科研究整个转录组、全基因组结合位点的蛋白质和其他分子的过程。然而，理解这些高级相互作用是如何驱动基因表达的，仍然是一个重大的统计挑战。郁彬教授的研究团队基于随机森林和随机交互树，以及通过大量的生物启发的仿生仿真，开发了迭代的随机森林算法（iRF）。iRF训练特征加权集成的决策树来检测稳定的、高阶交互作用，该算法的计算复杂度和随机森林相似。在讲座中，郁彬教授介绍了两个预测项目来展示iRF算法的有效性：早期的果蝇胚胎的增强子活性项目和人源细胞系原代转录单选择性剪接项目。在果蝇中，iRF稳定地发现了20对转录因子的相互作用，其中80%此前已被报道证实是存在着物理的相互作用。此外，三级交互作用，例如，在Zelda（ZLD），Giant（GT）和Twist（TWI）之间显示有高阶的交互作用，郁彬教授通过自己的研究结论与生物学家合作，将对这三者之间的交互作用做更多的后续实验。在人源细胞系项目中，iRF重新发现了H3K36me3在染色质介导剪接调控中的核心作用，并且发现了有趣的五阶和六阶交互作用，研究结果表明了多价核小体在剪接调控中的特定角色。通过从计算中识别交互作用，iRF打开了探寻分子机制之下的基因组生物学。

本次郁彬教授讲学计划的两次讲座共吸引了来自校内外三百余名师生的参加，两次活动中理教103教室座无虚席，来自统计、计算机、工程、生物信息学、神经科学等不同学科背景的同学和老师相聚一堂，深入展开交流，沟通成果，增进友谊，郁彬教授此次“大学堂”讲学计划取得了圆满成功。

4. 2018年“全国统计学优秀大学生”夏令营成功举办

2018年5月31日至6月2日，由北京大学统计科学中心主办的2018年“全国统计学优秀大学生”夏令营成功举行。夏令营共邀请28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。

5月31日中午一点，夏令营营员前往理科1号楼报到并领取材料。下午两点，夏令营笔试在理科一号楼1560教室开始，笔试内容涵盖数学、统计学、计量经济学等多学科基础知识，不同知识背景的同学各展所长，交出了满意的答卷。晚上六点，在同一教室举办了入营学生欢迎会，师生们欢聚一堂，其乐融融。

6月1日上午进行的是教员科研宣讲环节，统计科学中心联席主任陈松蹊教授首先向入营同学的到来表示欢迎，回顾了北京大学统计学科70年的发展历程，向大家展示了北大统计学专业深厚的历史底蕴、强大的师资力量和丰硕的科研成果。陈松蹊教授强调，北京大学统计科学中心始终致力于协调北大全校统计研究力量，促进统计学与其他学科的交叉与融合，建设世界一流的统计研究机构；更重要的是中心“建设一流的统计学科需要一流的学生”，希望同学们能够加入到北大统计科学中心团队中。

随后，北京大学统计科学中心的全教职员先后向大家介绍了各自的研究方向与最近的研究进展，台下同学们就自己关心的培养方式、学科前沿问题与老师们进行了深入的交流。

6月1日中午，统计科学中心的全教职员、在读博士生与参营同学们在艺园餐厅燕园厅进行了师生聚餐，席间大家互相交流学习生活，老师们向大家传授科研经验。师生频频互动，交流融洽，同学们也对未来的科研生涯有了更加明晰的了解与规划。

6月1日下午到6月2日上午进行了参营同学的自我展示环节。28位同学五人一组先后进行20分钟的个人展示，将自己本科期间参与过的学习、科研经历以及对未来的研究计划做了详尽展示，并回答了中心老师提出的问题，师生之间的印象也随之加深。6月2日下午进行了营员复试，进入复试的同学先后就自己在夏令营期间阅读的英文文献进行答辩。

为期三天的夏令营汇集了中国统计学专业优秀的青年学生，大家展示风采，互相交流，开阔眼界，也坚定了自己未来对统计学专业研究的信心。



5. 第三届北大-清华统计论坛成功举办

第三届北大-清华统计论坛于6月3日在北京大学光华管理学院1号楼举行。来自北大、清华的统计专业的师生今年再聚首，共同分享统计科学的前沿报告，通过交流与合作，促进统计学者的共同进步，推动统计学科的发展。来自海内外多所高校的150余名老师及学生参加了本次活动。

北大光华管理学院商务统计与经济计量系联合系主任、北京大学统计科学中心联席主任陈松蹊教授致开幕词，回顾了北大与清华统计学研究合作的深厚渊源，表达了对本论坛未来发展的美好愿景。随后由北大和清华两大院校的周晓华教授和杨立坚教授为大家带来精彩的学术报告。

周晓华教授在大数据分析方面有丰富经验，在医学检验、因果推断分析、缺失数据、脑科学、卫生经济、卫生服务等领域发展了许多新的统计方法。本次报告中介绍了现在很热门的精准医学的统计分析方法。精确医学是“一种新兴的医学实践，它使用个体的基因图谱来指导关于疾病的预防、诊断和治疗的决定”。受基因组学和医学成像等领域的推动，近十年来，从治疗乳腺癌到治疗重度抑郁症，精确医学取得了令人瞩目和显著的进展。精确医学的成功关键取决于精确的和可靠的统计工具的发展，用于估计从临床试验或观察研究收集的数据的最佳治疗方案。本次讨论两个主要的统计方法在精确医学的应用。一种是基于生物指标物预测和相关的协变量特异性治疗效应曲线，另一种是基于导出最佳个体化治疗规则（ITRS）。杨立坚教授主要研究时间序列，函数型及高维数据的统计推断，以及统计学对经济学、金融学、农学、食品科学、地理学和遗传学的应用。本次报告介绍了“广义加性模型（GAM）”的概念和术语。GAM从环境研究到警务前瞻系统，从信用评级到生存分析都得到了广泛的应用。GAM的统计理论，它的发展反映了过去30年统计学作为一门学科的多样性。杨教授还介绍了一些GAM理论在现实生活中的有趣应用。



北京大学统计科学中心讲席教授周晓华教授做特邀报告 清华大学统计学研究中心讲席教授杨立坚教授做特邀报告

报告结束后，北大与清华两校的同学进行了学术海报展示，其中北京大学共展出15张海报，清华大学共展出14张海报，每张海报都集结了同学们的思考与智慧，展现了同学们的专业知识的深度和广度，观点鲜明，并在现有的工作基础上提出改进和创新，具有一定现实意义或学术价值。期间参会师生自由走动，互相交流各自领域的最新进展，论坛现场互动频繁，气氛热烈。

最后，老师们从海报质量、工作难度、创新性和科学性4个维度对两校同学的海报进行点评，经过严格的评分，北京大学李伟同学、清华大学张园园与姜瑛同学恺分获“优秀海报奖”，北京大学王亚平同学荣获“优秀毕业生”奖。

本次论坛通过一个下午的交流活动使北大、清华两校师生之间加深了解、增进友谊，两校统计学领域的研究也有了更加密切的联系，会议取得了圆满成功。



6. 荷兰莱顿大学 Aad van der Vaart 院士短课成功举行

应北京大学统计科学中心邀请，荷兰莱顿大学 Aad van der Vaart 院士于 6 月 4 日-9 日访问北京大学并成功举办了主题为“高维与非参数模型的贝叶斯统计”的短期课程。Aad van der Vaart 教授是当今具有国际声望的统计学家，是荷兰皇家艺术与科学院院士。他在经验过程和半参数统计方面做出过重要的基础性贡献，所著 *Weak Convergence and Empirical Processes* (1996, 与 Jon Wellner 合著) 和 *Asymptotic Statistics* (1998) 已成为这一领域的经典文献和必读书目。近年来他的兴趣主要集中在非参数贝叶斯方法、高维统计及统计遗传学、复杂网络等应用领域。

本次课程内容主要涵盖传统贝叶斯方法、Dirichlet 过程、contraction rate、Dirichlet mixture、高斯过程先验。Aad van der Vaart 教授通过定理推导、举例应用等方式深入浅出地向大家讲解了非参数贝叶斯方法的基础理论与部分前沿知识，同时，他通过本次课程深刻揭示了频率学派观点和贝叶斯方法之间的内在联系。

本次课程共吸引了海内外近两百名师生参加，大家通过课堂学习与提问互动加深了对非参数贝叶斯方法的理解，为今后在该方面的科研工作建立了更为坚实的理论基础。



7. 暑期课程



上: Jiashun Jin ,Carnegie Mellon University

左: Cheng Yong Tang, Temple University

中: Yuan Yao, Hong Kong University of Science and Technology

右: Zheng Tracy Ke, University of Chicago

不管是传统的统计分析，还是现在火热的机器学习和深度学习，无数的数据分析师、数据工程师和数据科学家们其实都在追求一个能够完美利用数据解决现实各种问题的模型或者方法。

北大统计科学中心在 6 月 29 日-7 月 1 日举办的暑期短课，邀请来自海内外知名学府的统计学与机器学习的专家 **Jiashun Jin**、**Cheng Yong Tang**、**Yuan Yao**、**Zheng Tracy Ke** 开设短期课程，介绍高维统计，机器学习、深度学习等领域的一些最新进展。

Jiashun Jin 是卡内基梅隆大学统计学教授和机器学习列席教授，他的专长是对大数据中罕见和弱信号的统计推断。他于 2007 年获得 NSF CAREER 奖，2009 年获得 IMS Tweedie 奖，并于 2011 年当选为 IMS 研究员。他还在 2016 年至 2017 年期间在 Two-Sigma Investment 进行了两年的研究，从而获得了金融行业的宝贵经验。本次他授课的主题是：Higher Criticism for Large Scale Inference, especially for Rare and Weak effects。

Cheng Yong Tang 博士是天普大学统计科学系副教授。他的研究兴趣包括纵向数据分析，高维数据分析，非参数统计方法，经验可能性，财务数据分析，调查数据和缺失数据分析。Cheng Yong Tang 是国际统计学会的成员，也是皇家统计学会的成员。本次他授课的主题是：Parsimonious statistical modeling approaches for longitudinal studies。

Yuan Yao 是香港科技大学计算机科学与工程系的数学，化学与生物工程副教授。他目前的研究兴趣包括机器学习和高维数据分析，特别是拓扑和几何方法，应用于计算生物学，计算机视觉和信息检索等。本次他授课主题是：Differential Inclusion Method in high dimensional statistics and Deep Learning toward deeper understanding。

Zheng Tracy Ke 目前是哈佛大学统计学助理教授。她最近的工作是分析社交网络，并开发了一个名为 Mixed-SCORE 的程序。本次她授课主题是：New Tools for Analyzing Complicated and High-dimensional Data。

这次暑期短课采取线上预约的形式，来报名同学多来自北京大学及各地高校的统计与计算机大数据专业，还有社会企业相关专业人士。通过本次学习，同学们有机会与学科前沿的研究学者交流，分享学习心得，统计中心暑期短课在学界获得了广泛的支持与好评，也是中心为统计学科爱好者提供一个平等交流、学术启发的良好契机，从而促进学科建设的发展。

8. 北大环境统计研究团队关于京津冀空气污染状况评估的论文在大气环境领域英文国际期刊《Atmospheric Environment》上发表

由北京大学统计科学中心陈松蹊教授带领的环境统计团队关于京津冀地区空气质量状况评估的研究论文 Assessing air-quality in Beijing-Tianjin-Hebei region: The method and mixed tales of PM2.5 and O3 被大气环境领域英文国际期刊《Atmospheric Environment》正式发表。该论文主要是基于该团队 2017 年 8 月发布的《空气质量评估报告：京津冀 2013-2016 年区域污染状况评估》。

研究背景：随着经济发展，我国大气污染问题日益突出，大气污染形势面临严峻挑战。2013 年 9 月 10 日，国务院印发《大气污染防治行动计划》，又称《大气十条》，全面治理我国大气污染问题。该计划对全国大气污染最严重的区域——京津冀地区——提出明确目标：到 2017 年底，京津冀地区细颗粒物浓度较 2012 年下降 25%，其中，北京市细颗粒物浓度年均控制在 60 微克/立方米。由于空气污染物浓度显著的受到气象条件的影响，因而空气质量评价方法应考虑不同年份间气象条件的差异，客观、科学、公平的评估空气质量的改善状况。该文章利用气象观测的历史数据构建基准气象条件，对污染物和气象变量应用非参数回归方法，得到了一个能对污染物浓度进行气象调整的模式，并将该模型应用于 PM2.5, SO2, NO2 和 O3 四种污染物，更加科学的评估了自 2013 年春季至 2017 年夏季共 17 个季节的平均值浓度变化情况，对京津冀地区 13 个城市空气质量的改善情况做了细致分析和评估。

该论文的第一作者光华管理学院商务统计与经济计量系 2015 级硕博士生陈磊，其导师是陈松蹊教授。文章的通讯作者是郭斌副教授（西南财经大学统计研究中心、统计学院，北京大学光华管理学院商务统计与经济计量系 2015 届博士毕业生）和光华管理学院陈松蹊教授。其他作者是黄家盛（北京大学数学科学学院统计专业 2015 级直博士生）、何靖助理教授（西南财经大学统计学院，北京大学光华管理学院商务统计与经济计量系 2016 届博士毕业生）、王恒放（美国 Iowa state University 统计系 2015 级博士生）和张澍一（北京大学光华管理学院商务统计与经济计量系 2014 级直博士生）。本研究得到了北京大学统计科学中心和光华管理



学院的支持，及国家重点研发计划《大气污染成因与控制技术研究》专项项目2016YFC0207701的资助。

9. 跨学部生物统计系成立

2018年11月16日，北京大学公共卫生学院成立跨学部生物统计系。成立仪式在北京大学英杰交流中心报告厅举行。

北京大学副校长田刚、北京大学医学部副主任张宁、中国科学院院士马志明、北京大学公共卫生学院院长孟庆跃、北京大学数学科学学院院长陈大岳、北京大学统计科学中心联席主任陈松蹊发表致辞。

田刚指出，北大公共卫生学院生物统计系的成立是具有里程碑意义的一刻，标志着北京大学的学科建设和院系合作迈向了一个新的台阶。他表示，北京大学成立生物统计系，将对提高我国生物统计专业人才培养的质量、促进多学科交叉和融合等方面具有非常重要的现实意义和长远意义。



公共卫生学院生物统计系首届系主任由北京大学讲席教授、国际知名的生物统计学家周晓华担任。周晓华在全职加入北大之前曾担任美国华盛顿大学生物统计系的正教授，在因果推断、医疗大数据分析、缺失数据分析等领域作出了许多重要贡献。周晓华介绍了生物统计系的筹备情况及未来发展规划。



来自国内知名高校、科研院所、医疗机构的领导和专家，企业界精英和媒体代表等百余名嘉宾共同见证了公共卫生学院生物统计系的成立。截至成立仪式举行之时，北京大学公共卫生学院生物统计系收到来自全国高校的统计系、生物统计系及相关研究单位发来的共计 17 封贺信。



首届生物统计和数据科学论坛于当日在北京大学举行。美国国家科学院院士 Donald Rubin, 美国国家医学院院士 Xihong Li, 周晓华教授, 美国耶鲁大学生物统计系主任 Hongyu Zhao, 美国华盛顿大学生物统计系主任 Patrick Heagerty, 美国北卡大学生物统计系主任 Michael Kosorok, 美国北卡大学生物统计教授 Jason Fine, 美国哥伦比亚大学教授 Ying Kuen Cheung、Min Qian、Todd Ogden, 丹麦哥本哈根大学教授 Theis Lange, 瑞典乌普萨拉大学教授 Per Johansson 以及国内外该领域知名专家学者齐聚北京大学, 共同探讨因果推断、精准医疗等生物统计和数据科学领域的前沿问题。

10. 统计中心联席主任陈松蹊教授当选美国科学促进会会士

11 月 27 日, 世界知名学术组织—美国科学促进会 (AAAS) 选举产生了 2018 年新一届会士 (AAAS Fellow), 北京大学统计科学中心联席主任、光华管理学院陈松蹊教授成功当选。2018 年诺贝尔奖得主中有 4 位也获此殊荣。

成立于 1848 年, AAAS 是全球最大的科学学会, 并主办和出版国际著名期刊《Science》。在全球 91 个国家拥有个人会员, 下设 24 个专业分会, 涉及的学科包括数学、物理、化学、天文、地理、生物等自然科学和社会科学。其宗旨是“促进科学, 服务社会”, 即通过在全球范围内促进科学和创新的发展以造福人类。

AAAS 每年会根据学术影响及科学贡献从其会员中推选优秀的科学家, 授予“美国科学促进会会士”这一终身荣誉, 以表彰他们在促进科学方面取得的非凡成就。众多著名科学家, 如爱迪生 (发明电灯)、沃森 (发现 DNA 双螺旋结构) 及多位诺贝尔奖获得者, 都是美国科学促进会会士。在“统计学专业分会” (SECTION ON STATISTICS) 中, 陈松蹊教授是 2018 年入选的 7 位会员之一。他将参加 2019 年 2 月 16 日在美国华盛顿举行的“2019 年美国科学促进会年会论坛”, 届时将授予象征杰出成就的蓝金玫瑰奖章及证书。



SECTION ON STATISTICS

- A. John Bailer, Miami University
- **Song Xi Chen, Peking University (China)**
- Dianne M. Finkelstein, Massachusetts General Hospital/Harvard T.H. Chan School of Public Health
- Edward L. Ionides, University of Michigan
- David A. Marker, Westat
- Sharon-Lise Teresa Normand, Harvard Medical School
- Giovanni Parmigiani, Dana-Farber Cancer Institute

陈松蹊教授于 2008 年归国加入北京大学光华管理学院。目前任北京大学统计科学中心联席主任、光华管理学院商务统计与经济计量系联合系主任。

陈松蹊教授是数理统计学会(Institute of Mathematical Statistics) 资深会员 (fellow)、美国统计学会会士 (fellow)、国际统计学会 (International Statistics Institute) 当选会员 (elected member)、国际数理统计学会 (IMS) 理事会常务理事 (Council member)。同时任 The Annals of Statistics 副主编 (自 2010 年); Journal of Business and Economic Statistics 副主编 (自 2013 年); 曾任 Statistics and Its Interface 的联席主编 (2010-2013)。

陈松蹊教授迄今已在国际学术杂志发表论文 88 篇。根据 Web of Science 的评价, 陈松蹊教授的论文引用情况为: H-Index 27, I-10 Index 47, 均篇引用次数 25.9。

在进行高维数据统计理论、经济计量研究的同时, 陈松蹊教授也带领研究团队对中国的大气污染问题进行研究, 近年发布了系列空气质量报告, 产生了广泛的社会影响力。

11. “万花统”学术沙龙正式上线



北大统计科学中心“万花统”第一期学术沙龙

统计咨询与合作
打开统计走向世界的大门

2018年12月5日 (周三)
12: 00-13: 30 理科5号楼506室

Host: 牛晓月
宾夕法尼亚州立大学统计系副教授、统计咨询中心主任

作为开放窗口的统计信息咨询服务使社会由陌生到认知、由单纯索取资料到深度合作开发,
为提高统计影响力和统计公信力以及促进社会经济发展做出了自己应有的贡献,
其前景更有广阔发展的空间。

开放、自由、交流、共享

想了解统计咨询的同学们,
想深入开展项目合作的老师们,
请来北大统计科学中心“万花统”学术沙龙,
和牛老师一起探讨分享统计专业的广阔发展前景!

本次学术沙龙备有午餐, 请老师和同学们点击“阅读原文”报名参加!
也可发邮件至: statcenter@pku.edu.cn, 请备注您的姓名+所属学院+专业
12月4日周二下午5点之前截止报名, 场地有限, 参与资格以邮件通知为准。
如有疑问, 敬请咨询: 62760736, 王彦懿老师

12月5日, 由统计科学中心行政团队打造的第一期“万花统”学术沙龙正式上线, 旨在增进统计中心的学术交流氛围, 推动统计与各学科的沟通与合作, 加强师生间的交流探讨, 在常规学术报告的基础上, 统计中心拟搭建一个开放性、非正式的学术交流平台。

首期邀请到 Penn State 大学统计咨询中心的牛晓月副教授为大家抛砖引玉, 介绍国外专业院校统计咨询的起步、发展、应用。来自清华、中科院等兄弟院校的师生齐聚一堂, 一同讨论统计专业的发展前景和深度合作开发。

第二期, 北京大学心理与认知科学学院院长方方出席, 延展上期话题, 进一步讨论统计学在交

叉学科中的应用。中心希望将统计咨询发展成为资源共享，强强联手，跨学科深度合作，提升科研服务应用的专业平台。

12. 统计中心四位青年教师在北京大学第十二期“青年教师学术沙龙”汇报研究成果

12月6日，人事部组织开展了北京大学“青年教师学术沙龙”第十二期活动。本次沙龙，统计中心席瑞斌、李程、贾金柱、徐敏亚四位青年教师做了主题演讲，周晓华教授和陈松蹊教授作为特邀嘉宾分别做了开篇介绍和总结发言。

本期沙龙以“数学、生命科学、社会学、管理学等——基于统计学的学科发展与应用研究”为题，旨在加强校内以统计学为基础开展相关研究的青年教师之间的交流与合作，促进统计学基础研究和应用研究的交叉融合。来自数学科学学院、生命科学学院、公共卫生学院、光华管理学院、中国社会科学调查中心等院系的30余位青年教师参加了活动。北京大学统计科学中心联席主任、光华管理学院陈松蹊教授，北京大学生物统计系系主任、北京国际数学中心周晓华教授，北京大学科学研究部副部长蔡晖，北京大学社会科学部副部长王周谊，北京大学学科建设办公室副主任贺飞等出席活动。人事部部长刘波全程参加沙龙，本期沙龙由人事部副部长俞蓁主持。

沙龙开始后，周晓华教授首先介绍了生物统计学发展的现状，并从生物医学统计、医疗大数据等方面介绍了面向未来的生物统计学研究的重点问题。在报告分享环节，统计中心有四位老师做了主题演讲。其中，席瑞斌老师以生物大数据统计分析为题，介绍了他用统计方法和模型、生物信息学方法等研究手段在基因组学、单细胞测序数据填补、癌基因组异质性等方面的研究工作。李程老师作了“基因组学数据分析——单细胞 RNS-seq 分析”的报告，认为生物科学研究使用的最新的单细胞测序技术产生了高通量、高维度的基因组学数据，能从细胞异质性、细胞间通讯等层面揭示生命和疾病现象的规律，同时也为相关统计分析方法的发展提供了机遇，如大数据降维、批次效应矫正、缺失数据的填补算法。贾金柱老师从统计机器学习的理论和案例两个角度介绍了他的研究工作。徐敏亚老师分享了她利用统计学方法对组织管理中的问题进行实证研究的成果与心得。沙龙期间，与会教师围绕研究对象的选择、研究方法、统计学在医学研究中的应用等问题进行了热烈讨论和交流。



陈松蹊教授对沙龙进行了总结发言，他高度肯定了几位青年教师的出色工作，建议大家在已有研究工作的基础上继续发挥优势、突出特色，并从学校统计学科发展需求出发，希望通过学校支持使统计科学中心能为相关研究更好地提供统计咨询等支撑。

“青年教师学术沙龙”是青年教师培训发展的重要平台，统计中心也将继续围绕青年教师职业发展、学术前沿、科研合作、团队建设等主题定期开设学术沙龙，不断优化青年教师发展平台、为青年教师成长创造良好条件，积极助力学校高水平人才培养和“双一流”建设。

13. 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。

（一）超高维数据方面

陈松蹊教授研究小组建立了一套高维统计检验的理论和方法，解决了一系列重要的数理统计问题，取得了具有国际影响力的创新成果，并在实际中得到了重要应用。

[Li and Chen (2012)]首次在一般超高维非参数数据下，提出了两样本高维协方差矩阵 Σ_1 和 Σ_2 是否相等的检验，即检验 $\Sigma_1 = \Sigma_2$ 。该检验普遍适用于超高维数据。这一检验提出后，在生物信息等领域得到了广泛应用。它同陈-秦均值检验一起是目前对两个高维样本分布差异性的主要检验方法。该论文是 WoS 高引论文。

[Zhong, Chen and Xu(2013)] 考虑在更一般的高维数据分布，并允许不同数据维度具有弱相关性（所谓的 α -mixing（ α 混合）数据）。证明在这更一般的条件下，经过阈值处理的陈-秦检验的检验边界也达到最优检验边界。这一研究极大地提高了人们在挑战性最强的稀疏和弱信号检验问题上的认知，并给出了具有更高的检验效用的高维均值检验。

（二）函数型数据分析方面

姚方教授主要研究包括无限维和高维空间的函数型数据分析，例如函数主因子分析和各类函数回归，分类和聚类模型；复杂结构数据的模型选择和降维方法；对非独立相关型的纵向数据和生存分析的综合模型的研究。函数型数据分析的兴起主要是因为传统数理统计以向量型变量为研究和观测的对象，而实际应用中很多数据记录都是随时间和空间发生连续性变化的，例如环境研究中对空气污染的监测记录，医学研究中的脑扫描图像的变化等。现阶段的主要研究方向是把连续的函数型数据和具有高维，网络或者流形结构的模型及方法进行有机的结合，以及最新的机器学习和深度学习技术，提出创新的统计模型和解决相关的理论问题，从而可以有效的应用到大规模的数据处理中去。

由于在函数型数据分析领域所做出的奠基性和开创性的贡献，2014 年获得由加拿大统计学会和数学研究中心联合颁发的授予博士毕业 15 年内在加拿大做出突出贡献的统计学家的 CRM-SSC 奖，2017 年入选国际数理统计学会（Institute of Mathematical Statistics）的 Fellow。现任 Journal of the American Statistical Association, Statistica Sinica, Journal of Computational and Graphical Statistics, Journal of Multivariate Statistics, Canadian Journal of Statistics 副主编；曾任 Annals of Statistics, Bernoulli, Electrical Journal of Statistics, Journal of Statistical Planning and Inference 副主编。

（三）因果推断方面

耿直教授和苗旺博士等人 2018 年发表在 *Biometrika* 的论文研究了使用混杂的代理识别因果作用的方法。当存在未观测的混杂因素时,或者混杂因素的测量有误差时,因果作用一般不可识别。苗旺和耿直等人的论文研究存在混杂的代理时识别因果作用的方法,证明当存在混杂的两个独立的测量时,因果作用可识别;给出了保证识别性的一般条件。这一方法应用到经济学和流行病学等观察性研究中具有重要意义,通过这种方法,可以使用阴性对照变量(negative control)提高因果作用的识别性,改变了过去只能用阴性对照变量检验混杂的做法。他们发表在 *Annual Review of Statistics and Its Application* 和《中国科学》的论文综述了因果作用推断和因果图模型学习的问题、方法和前沿的研究。与因果推断密切相关的是缺失数据问题。随机缺失数据问题的研究有很多,但是非随机缺失数据的研究比较困难。苗旺博士和合作者 2018 年发表在 *Statistica Sinica* 的两篇论文研究了非随机缺失数据问题中的识别性方法和半参数化估计方法。

（四）网络数据研究方面

王汉生教授过去这五年(2013-2017)最主要的工作是对网络数据(Network Data)的理论研究,以及相关的产业实践。在理论研究方面,我们研究了网络拓扑结构的产生机制,以及概率建模的可能性。在网络数据方面,主要是研究了复杂相依关系同拓扑结构的参数化表达。共发表相关英文学术论文 29 篇,获得国家基金委资助 300 余万。

（五）生物信息方面

席瑞斌研究员在过去几年中发展了一系列统计、生物信息方法,很多都发表到了统计、生物信息的顶级杂志。例如在 Xia et al. (2017) 中,席瑞斌和他的合作者发展了一个整合多个算法探测结构变异的方法,显著提高了结构变异探测的敏感度、特异性及断点准确性;在 Xi et al. (2017) 中,他们发展了一个新的基于全基因组测序数据探测拷贝数变异的算法,并将此算法应用到了约 1000 个癌症病人的正常基因组上,发现了一些包括 *GOLPH3* 在内的关键的癌症易感变异位点;Yuan et al. (2017)则发展了一个用来探测网络差异的方法,此方法主要基于一个新的 D-trace 凸损失函数,证明了在超高维数下此方法的相合性。另外,席瑞斌还和他的合作者一起做了很多基因组、癌基因组的大数据分析研究,这些文章很多也都发表到了顶级杂志上。在 Chang et al. (2017)中,他们分析了~100 对食管癌基因组,发现与饮酒相关的点突变指纹,并找到了一些包括 *EGFR* 等基因在内的关键可靶向用药变异;在 Gao et al. (2017)中,席瑞斌和他的合作者分析了肝癌基因组的异质性,发现肝癌细胞有极高的异质性,构建了相应的癌细胞进化树,并发现癌细胞的异质性显著的影响了癌细胞对药物的响应;Kim et al. (2013)则对 8000 对个癌基因组进行了拷贝数变异大数据分析,发现了一些泛癌基因组中的关键拷贝数变异位点,并研究了染色体碎裂现象的多种癌症中的基本特征;最后,De Los Angeles et al. (2015)系统研究了干细胞研究中的关键检验准则,提出并展示了基因组分析在干细胞研究中的关键作用。

（六）计量经济方面

涂云东博士针对诺奖获得者 Engle and Granger 1987 年在 *Econometrica* 上提出的误差修正模型,提出了模型选择和平均的策略,改善了其在预测中的表现。该成果形成一篇文章, Tu and Yi (2017), 发表在国际一流杂志 *Journal of Econometrics*。

宋晓军博士与 bderrahim Taamouti 教授提出利用非参数方法来度量和检验时间序列中可能存在的非线性和线性格兰杰因果关系。另外,该方法由于不假设任何参数模型,所以并不会受到模型错误设定的影响。该文章于 2016 年发表在 *Journal of Business & Economic Statistics*。

（七）高维建模理论及其应用

贾金柱副研究员针对高位稀疏图模型,创新性地设计了 MCMC 算法,该方法可以有效地对高维稀疏图模型进行抽样 (Yangbo He, Jinzhu Jia and Bin Yu 2013, *Annals of Statistics*)。使用该 MCMC 算法,我们研究了如何计算因果图中等价类的数量 (Yangbo He, Jinzhu Jia and Bin Yu 2015, *Journal of Machine Learning Research*)。我们研究了非齐次模型对于变量选择方法的影响,理论上保证了变量选择方法的广泛适用性 (Jinzhu Jia, Karl Rohe and Bin Yu 2013, *Statistica Sinica*)。我们创新性地提出了预处理方法,该方法可以更好地解决变量选择问题 (Jinzhu Jia and Karl Rohe 2015, *Electronic Journal of Statistics*)。我们应用稀疏学习方法很好地分析了文本新闻数据,得到有意义的结果,为社会学家分析新闻数据提供了很大方便 (Jinzhu Jia, Luke Miratrix, Bin Yu, Brian Gawalt, Laurent El Ghaoui, Luke Barnesmoore and Sophie Clavier 2014, *The Annals of Applied Statistics*)。

三、人员组成

1、人才引进



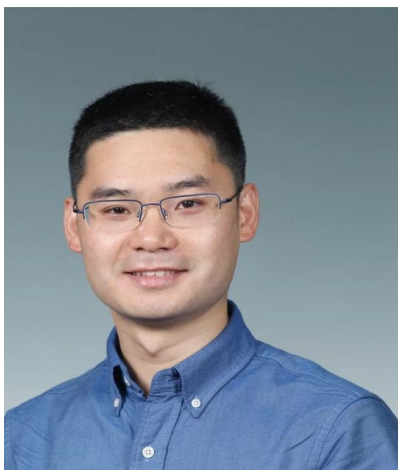
张原研究员：1988 年生。2015 年于美国杜克大学获得博士学位，现任北京大学数学科学学院助理教授（研究员学术头衔），2018 年 7 月入职，统计科学中心与数学科学学院合聘。

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。



苗旺研究员：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。我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：统计中心与医学院公卫学院合聘。2009 年北京大学概率统计系博士，2009-2010 年在 UC Berkeley 统计系进行博士后研究工作，2011 年加入统计中心。

研究方向：高维统计推断、统计机器学习和因果推断等。

研究介绍：

判别分析是一个经典的统计学和机器学习问题。在模式识别、生物信息学中有广泛的应用。LDA (Linear Discriminant Analysis) 是应用广泛的判别分析方法。LDA 的原理是将带上标签的数据点投影到维度更低的空间中，使得投影后的点会形成按类别区分，相同类别的点将会在投影后的空间中更接近。经典统计教科书告诉我们，在一些条件下 Fisher 线性判别分析(LDA)是最优的。然而当维度很高时，LDA 的表现并不好。如 Bickel and Levina (2004) 指出在高维情形 LDA 甚至不如随机猜测。

问题出现在高维协方差矩阵的估计上，因为要估计的参数非常多，高维协方差矩阵出现了很大的误差，甚至出现协方差阵不可逆的情况。为了解决这一难题，统计学家提出修正的协方差估计，并证明了许多优良的性质。也有统计学家指出可以通过规避估计协方差来解决这一难题[CL2011]。其实，Fisher 最早建立了判别分析和回归分析之间的联系 [F1936]：二分类的 LDA 等价于线性最小二乘方法——即将两个类别分别用 0 和 1 表示，然后求解线性最小二乘。有了这一联系，带有变量选择的高维线性判别就转化成了传统的 Lasso 求解问题。虽然这一转化很自然，其理论结果有待进一步探讨。原因是 Lasso 的理论结果局限于线性模型，其误差的高斯型尾概率分布起决定作用。LDA 虽然可以转化为类似线性回归的最小二乘方法，但是其本质的模型已经很不一样。Mai et al(2012) 建立了 Lasso 求解 LDA 所得模型选择相合性的结果。通过细致分析，我们发现使用 random matrix 理论，可以证明，如果起决定作用的变量个数很少时，Lasso 可以得到最优的判别结果，这一点和低维经典 LDA 结论类似。

参考论文：

Yanfang Li and Jinzhu Jia.(2017) L1 least squares for sparse high-dimensional LDA. Electronic Journal of Statistics, 2017, 11(1):2499-2518.



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。

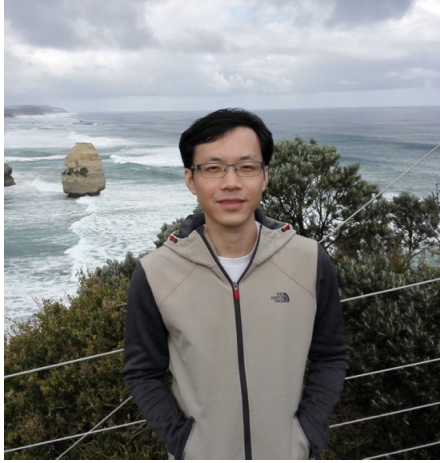
近期发表：

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.



林伟研究员：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，2015 年入选“国家级重大人才青年计划”。

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：

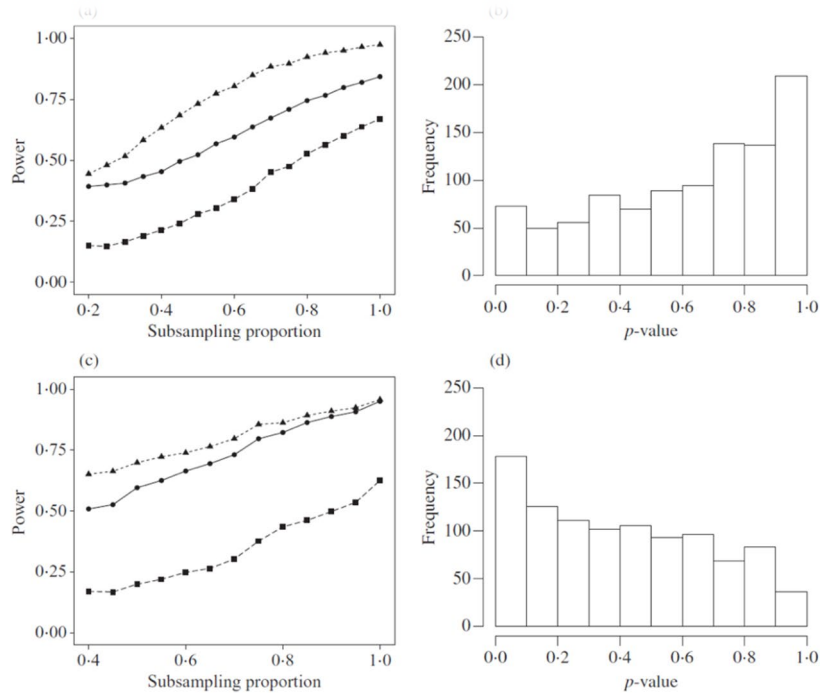


Fig. 1. Analysis of two microbiome datasets: empirical power curves of the proposed test (triangles) and the test based on log-transformed (circles) and raw (squares) compositions with $\alpha = 0.05$ are shown in (a) for the obesity data and (c) for the Crohn's disease data; histograms of p -values 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成分检验问题，取得了良好的效果。



宋晓军研究员：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。

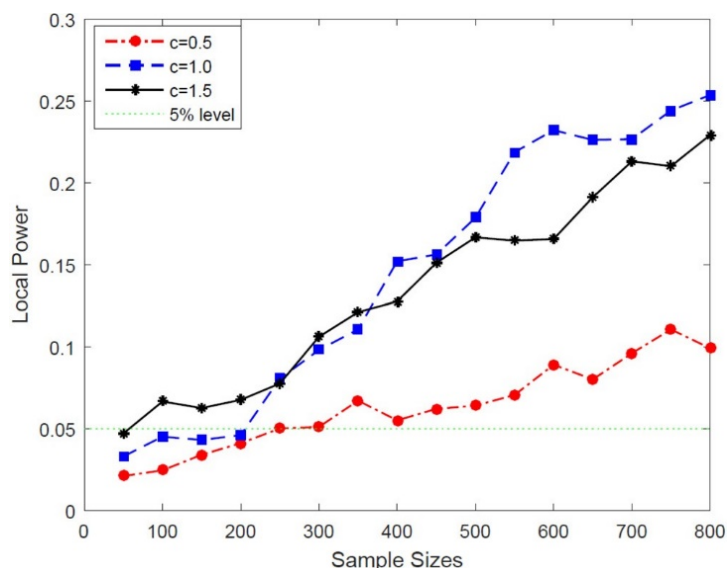
研究方向：计量经济学理论（非参数与半参数方法，模型设定检验，自助方法，时间序列分析和实证资产定价）及应用计量经济学。

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。



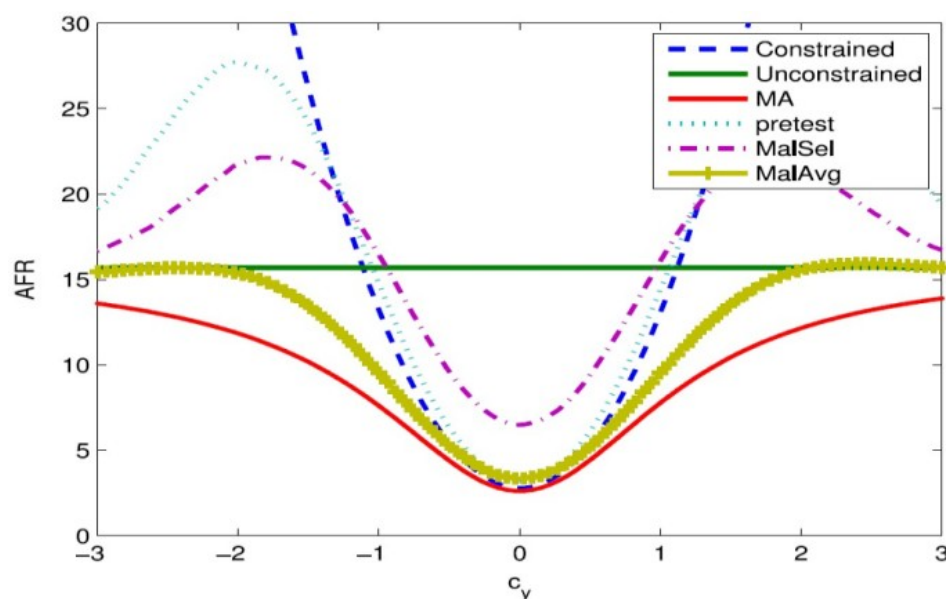


涂云东研究员：统计中心与光华管理学院合聘。2012 年加州大学经济学博士，同年加入统计中心。

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。





席瑞斌研究员：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，2013 年入选“国家级重大人才青年计划”。

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 Bioinformatics 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。

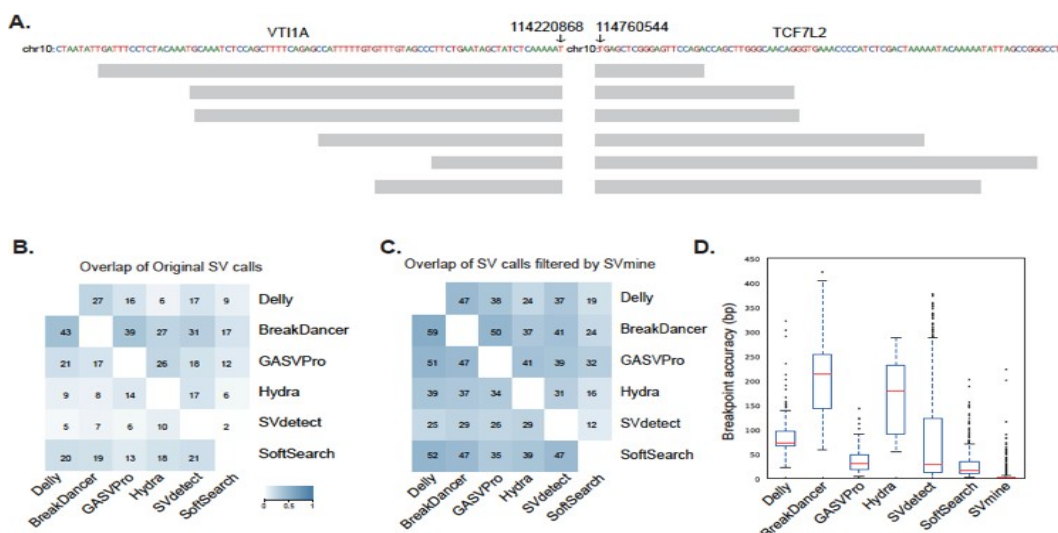


图 1. A. SVmine 在一个结直肠癌基因组中探测到的基因融合 VTI1A-TCF7L2，其预测的断点位置与 Sanger 测序验证的断点位置完全相同。B. 在 8 个结直肠癌基因组中，6 个算法的体细胞结构变异预测的两两相交情况，图中的数字表示横轴和纵轴算法共同预测的结构变异在横轴算法预测结构变异中的比例。例如，一行二列的数字 27 代表 27% 的 Delly 预测的结构变异同时也在 BreakDancer 预测的结构变异中。C. 经过 SVmine 整合分析后，算法两两预测交集有显著的提高。图中的数字与 B 类似。例如，经过 SVmine 整合后，现在 47% Delly 预测的结构变异同时也在 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	工作	研究方向	学校院系
陈松蹊	联席主任, 执行委员	金融计量学, 经济计量学, 风险度量, 随机过程, 统计推断, 高维数据分析, 抽样方法	光华管理学院商务统计与经济计量系
耿 直	联席主任, 执行委员	因果推断, 数理统计, 生物医学统计, 因果网络, 贝叶斯网络	数学科学学院概率统计系
姚 方	副主任, 执行委员	函数型数据分析, 复杂结构数据, 高维数据, 非参数回归	数学科学学院概率统计系
李 程	执行委员, 中心全职教员	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
王汉生	执行委员	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
吴 岚	执行委员	金融资产定价, 金融统计, 经济资本模型, 精算学	数学科学学院金融数学系
任艳霞	执行委员	测度值马氏过程、分支粒子系统等	数学科学学院概率统计系
周晓华	执行委员	缺失数据, 因果推断分析, 大数据分析, 半参数模型, 医学检验, 脑科学, 卫生经济, 卫生服务领域发展新的统计方法	北京国际数学研究中心
贾金柱	中心全职教员	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	中心全职教员	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
席瑞斌	中心全职教员	统计计算, 生物信息	数学科学学院概率统计系
林 伟	中心全职教员	高维统计, 大数据问题, 因果推断, 生存分析	数学科学学院概率统计系
宋晓军	中心全职教员	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
张志华	中心教员	机器学习与应用统计学	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 and 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
邓明华	统计咨询召集人	生物信息学, 计算生物学	数学科学学院概率统计系
张俊妮	统计咨询召集人	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	政府统计工作室召集人 执行委员	生存分析, 可靠性, 应用统计	数学科学学院概率统计系
虞吉海	政府统计工作室召集人	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

3. 中心行政团队



李楠，2016 年加入北大光华管理学院，曾任活动会议总监，现任中心行政主任。



王彦懿，2016 年人大环境学院硕士毕业。2018 年 6 月加入中心，任职行政秘书。

四、统计中心科研项目

统计科学中心在研项目：

序号	项目名称	负责人	项目来源	类别	开始时间	结题时间
1	生物数据统计分析的方法、理论与应用	艾明要	基金委	协作项目	2014.01	2018.12
2	试验设计的统计推断基础理论研究	艾明要	基金委	面上项目	2017.01	2020.12
3	大数据驱动的管理决策模型与算法	陈松蹊	基金委	重点项目	2016.01	2020.12
4	空气质量统计评估方法	陈松蹊	科技部	重点研发计划	2016.07	2020.06
5	空气质量分析测算技术服务	陈松蹊	中国环境监测总站	合作合作	2017.1	2018.1
6	2013-2017 年环境空气质量监测数据统计	陈松蹊	中华人民共和国环境保护部	技术合作	2017.11	2018.11
7	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	基金委	青年基金	2017.01	2019.12
8	与 Levy 过程驱动的倒向随机微分方程相关的随机控制和金融问题	程雪	基金委	协作项目	2015.01	2018.12
9	基于网络的全基因组关联分析方法	邓明华	基金委	面上项目	2015.01	2018.12
10	疾病相关生物分子网络的构建、动态演化及调控规律（中科院生物物理所）	邓明华	科技部	重大研究计划	2015.01	2019.08
11	生物数据统计分析的方法、理论与应用	耿直	基金委	协作项目	2014.01	2018.12
12	高维图模型的结构空间及学习方法习方法	何洋波	基金委	面上项目	2017.01	2020.12

13	科研启动费	苗旺	教育部	人才专项	2018.07	2020.07
14	科研启动费	张原	教育部	人才专项	2018.07	2020.07
15	文本挖掘的统计建模	贾金柱	基金委	面上项目	2016.01	2019.12
16	金融衍生品定价中的计量经济学分析	李辰旭	基金委	重点研发计划	2017.01	2020.12
17	高维复杂数据的稀疏与低秩建模及推断	林伟	基金委	面上项目	2017.01	2020.12
18	基于统计与数值模式的多污染物数据场构建	林伟	科技部	重点研发计划	2017.01	2021.12
19	超过程的极限理论	任艳霞	基金委	面上项目	2017.01	2020.12
20	高维非线性模型中的因子分析：理论和应用	涂云东	基金委	面上项目	2017.01	2020.12
21	高维复杂数据的理论与应用	王汉生	基金委	杰出青年基金	2016.01	2020.12
22	大气污染数据质量监控	王汉生	科技部	重点研发计划	2016.07	2020.06
23	北京大学-华农天农共建“北京大学农业大数据研究中心”	吴岚	企事业单位	技术服务	2016.12	2026.12
24	基于高通量测序数据研究基因组变异的统计问题	席瑞斌	基金委	面上项目	2015.01	2018.12
25	金融中的动态 copula 理论及其应用研究	杨静平	基金委	面上项目	2017.01	2020.12
26	污染与宏观经济、人口健康的计量分析	虞吉海	科技部	重点研发计划	201607	2020.06
27	实时数据流的统计学理论与方法	姚方、 席瑞斌	国防科工委	研发计划	2018.11	2020.12

五、博士生毕业去向

2018 年，北京大学统计及金融数学专业共有 20 名博士毕业生，毕业时 14 名已落实工作，其中 6 人在国内外知名高校获得教职或从事博士后工作，8 人在知名互联网企业或银行等大型金融机构工作，以下是中心教员指导的毕业生去向。

姓名	专业	工作单位/继续深造	指导教师
傅晶雪	应用数学	新加坡 dynamic technology lab	吴 岚
郭南	统计学	中债资信	杨静平
王亚平	概率论与数理统计	华东师范大学	艾明要
孙鑫伟	概率论与数理统计	北京大学信息科学技术学院	姚 远
吴昌晶	概率论与数理统计	阿里巴巴(中国)软件有限公司	邓明华
蔡政	概率论与数理统计	腾讯科技(深圳)有限公司	耿直
潘育	概率论与数理统计	申万宏源证券有限公司	陈大岳
李伟	概率论与数理统计	科磊半导体设备技术(上海)有限公司	耿直

六、交流访问

2018 年统计科学中心共接待海内外访问学者 44 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。

1. 国际著名统计学家，哈佛大学 Donald B. Rubin 教授：Models for Imputing Missing Data

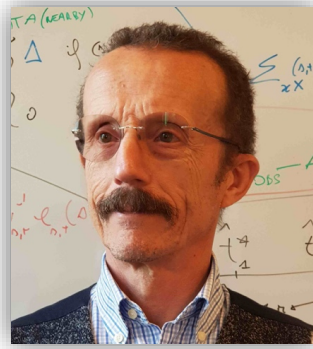
3 月 16 日，来自哈佛大学的 Donald B. Rubin 教授为统计中心的师生带了一场精彩的讲座——缺失数据的统计处理模型。有两种相对标准的方法来处理统计中的缺失数据，一种基于“选择模型”，另一种基于“模式混合”模型。前者侧重于制定完整数据模型，输入缺失数据，结合观察到的和缺失的数据符合完整数据的假设模型。相反，后者有效地拟合了每种观察和缺失数据模式的不同模型，从而直接揭示了结论对没有分布的假设的敏感性。可用于估算的实际观测数据。还有第三类模型，基于“吉布斯”因子分解；虽然这些并不意味着有效的联合分布，但由于它们易于使用，它们在教育中取得了成功。由 MCMC 计算机软件实现的多重补差，如 SAS, STATA 和 MICE。考虑结论与观察数据无法剔除的假设，无论是隐性的，如同选择模型，还是显性的，如同混合模型一样，是对具有缺失值的数据集进行令人满意的分析的关键因素。本次报告具有很强的理论性与启发性青年教师就大数据分析，数据挖掘，缺失数据等问题与 Rubin 教授进行了深入地交流。



Donald B. Rubin 教授，美国国家科学院院士，美国科学与艺术学院院士，现任哈佛大学统计系 John L. Loeb 讲席教授。他获得过统计学领域几乎所有著名奖项，包括著名的 Wilks 奖章、Parzen 奖、Snedecor 奖等，是当今世界最具影响力的统计学泰斗。他对科学的贡献已超出统计学范畴，其统计思想对生物医学、经济学、心理学、教育学、社会学及计算机科学等众多领域产生了划时代的影响。他发表了 400 余篇论文，这些论文被多次引用，仅在 2016 年一年内的引用次数就超过了 20000 次。

2. 贝加莫大学 Alessandro Fasso 教授来访期间短期课程

贝加莫大学 Alessandro Fasso 教授于 3 月 13 日-3 月 19 日，以及 3 月 24 日-3 月 31 日访问北京大学。在此期间，他开设关于 environmental time series and geostatistics, functional data analysis and splines 的 3 次短课程。



Alessandro Fassò 是贝加莫大学 2000 年的统计学教授。国际环境学会(TIES)主席(2017-2019)。GRASPA 的创始人和前任协调员(2013-2015)是意大利统计学会(SIS)环境统计的常设工作组。国际统计学会(ISI)理事会成员(2013-2017)。世界气象组织(www.wmo.int)推动的大气参考观测工作组 WG-GRUAN 成员。他主要研究环境学的统计方法和应用,空气质量,气候变量,环境模型的敏感性分析,环境时间序列,时空数据,随机监测,结构和岩土监测,工业统计,质量控制和财务时间序列分析。

特别值得一提的是, Alessandro Fassò 对空气质量监测和理解做出了贡献。考虑到空气质量监测,他开发了基于地面异构监测网络,计算机输出和卫星数据监测的颗粒物和氮氧化物的多变量时空模型,开发了统计动态绘图方法、风险评估、人体暴露分布和环境政策评估。他的研究方法允许在不同地方观察到多种污染物,并且具有不同的空间尺度,具有大量“结构缺失数据”。此外,他还讨论了异构和不平衡监测网络框架中的空气质量指标和基于模型的指标。他提出了一些臭氧时间序列的非线性模型等方法。

3. 爱荷华州立大学 Carriquiry 院士: 刑事证据评估的统计学习算法

5 月 21 日, Alicia Carriquiry, 爱荷华州立大学统计学杰出讲席教授、美国医学院院士、美国国家刑事侦探统计研究所主任。Carriquiry 教授的主要研究方向包括贝叶斯方法、刑事侦探统计、营养和饮食评估、基因组学、法医学等问题。Carriquiry 教授是国际统计研究所的当选会士、美国统计协会会士、美国数理统计研究所会士,她也曾担任国际贝叶斯分析学会(ISBA)主席。



证据评估产生于刑事和民事司法程序。近年来,大多数法医分析的科学有效性受到质疑,现在正在努力发展各种法医工具的科学和统计基础。模式证据包括指纹,鞋底印记,子弹和等其他特征,因为这类证据通常表示为图像,并不适合传统的统计建模方法。甚至诸如玻璃或涂料的化学成分之类的数值证据也难以建模,因为它通常是高维的。

在本次报告中, Alicia 关注的是证据源头问题: 两个证据项有共同的来源吗? Alicia 开发了非参数分类算法,并展示了它们在玻璃和鞋印中的应用。她认为学习算法使法医从业者能够量化两项证据之间的相似程度。此外,通过计算基于计分的似然比,从业者可以评估证据的权重,以支持起诉或辩护的证据。

这次报告打开了统计应用的思想大门，量化统计方法可以应用到社会中的各个领域，同学们在思索中感受新知识、新视野的同时，也可以将所学的知识帮助社会各个群体，解决实际的问题。

4. 滴滴首席统计学家朱宏图教授：统计运用在大数据上的机遇挑战与策略



6月14日，滴滴出行首席统计学家朱宏图来校访为师生们做了一场深入人心的报告，朱宏图教授是美国 MD 安德森癌症中心生物统计系 Bao-Shan Jing 冠名教授，北卡罗来纳大学教堂山分校生物统计学系终身教授。美国统计学会图像统计方向八位创始人之一。朱宏图教授多年来致力于医学图像统计分析、潜变量模型、统计诊断、缺失数据分析、神经信息学及数据整合等方向的学术研究，同时担任 7 种国际统计界顶级学术期刊副主编，至今已主持美国国家科学基金及国立卫生研究院的国家重要项目七项，

先后被选为美国统计学会及国际数理统计学会终身会士。目前已在统计、生物医学成像、公共卫生等领域发表期刊论文两百余篇。

随着现代科技的快速发展，生物医学领域涌现出大量跨平台数据（如影像、基因、临床信息等）。这些高维数据的整合运用带来了新的统计学挑战：大部分现有统计模型和数据平台已不再适用，急需发展更为准确全面的统计方法。面临这些挑战，朱宏图教授将从数个相关项目着手，提出海量数据整合与解析中存在的重要统计机遇和应对策略。

本次报告的实用性极高，使同学们大开眼界，了解业界的科技渗入和未来的职能发展。朱宏图教授分享了滴滴的 AI 布局，以及滴滴在智能派单、供需预测、智慧交通等领域的实践经验；大数据、人工智能技术以及云计算已经广泛应用于滴滴的各类场景，当用户打开 APP，在看不见的背后，其实已经有一系列复杂的智能算法模型在默默地提供服务。滴滴一直在持续用科技、用 AI 技术来改善城市的交通，来改善大众的出行。

5. 中科院遥感与数字地球研究所陈良富研究员：大气污染卫星监测技术的现状与发展态势

10月15日，中心特为环境统计团队和学生们准备了一场关于大气污染的学术报告，本次报告邀请的是陈良富，遥感科学国家重点实验室副主任，中国科学院遥感与数字地球研究所研究员，博士生导师，中国科学院大学岗位教授，享受国家政府津贴专家。“十三五”国家重点研发计划“大气污染成因与控制技术研究”重点专项总体专家组成员。

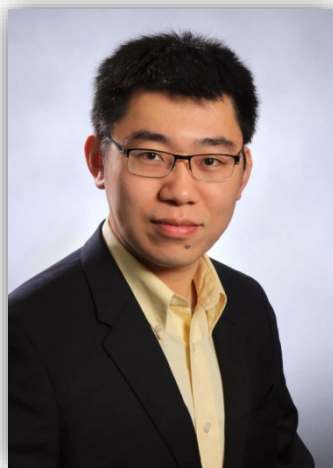


陈良富研究员自 2005 年以来主要从大气环境遥感研究，构建了针对我国重污染环境的大气环境空气质量卫星遥感基础理论与方法体系，形成气溶胶、霾、颗粒物浓度、沙尘、秸秆焚烧监测，以及二氧化氮、二氧化硫、臭氧、甲醛和乙二醛等污染气体，以及二氧化碳、甲烷、一氧化碳、氧化亚氮等温室气体卫星遥感反演算法与模型。研发了一套环境空气质量卫星遥感业务化工程化技术，开发了第一套国家级和省级环境空气质量多源遥感监测业务系统。

在座的师生，尤其是从事环境污染研究的统计专业的老师和同学通过此次讲座，从宏观监测的角度审视了自己的研究，同时既丰富了专业知识，又启发了研究报告的撰写。

6. Penn State University 暴乐副教授：艾滋病高危人群规模估计

暴乐博士是宾夕法尼亚州立大学统计学副教授和高级数据同化和可预测性技术中心副主任。他还担任艾滋病规划署参考小组的关键技术顾问， 他的研究重点是：1.使用统计模型来解决全球健康问题，例如评估艾滋病毒流行，健康指标，年龄和导致特定儿童死亡率；2. 开发大数据的快速算法；3.开发离散数据分析方法。



全球健康方向是一个医疗、药理、传染病学和统计学的交叉学科。报告以艾滋病高危人群规模估计为例，介绍了全球健康研究中主要的统计方法。在过去的 15 年中，全球 HIV 的预防措施有效地降低了传播率，治疗措施有效地降低了死亡率。然而，HIV 防治仍有许多挑战：传染率高的地方往往是贫穷的地区，这里数据很难获取；缺少数据质量诊断；缺少学术上的交流；一些特定人群如同性恋，吸毒者，性工作者很难通过传统的抽样调查方法获得大

样本数据。针对以上问题,作者应用贝叶斯模型同时考虑多种数据来源,如 Mapping method, Multiplier Method, Capture recapture method 等,并在数据严重缺失的情况下引入先验信息帮助模型推断。分层模型被用来联系各地区的模型参数,使得数据缺失的地区可以借用数据相对丰富的地区的信息。由于经费原因,很多国家只有个别年份有详细的调查。因此随机游走模型被用来推断每年感染人数的波动情况。报告以巴基斯坦数据为例,对不同时间,不同地点,不同来源的数据进行整合,同时估计出全国和各地区的高危人群规模以及置信区间。根据模型提供 R 包和 shiny 界面方便非统计专业的人进行分析。

7. UC Berkeley 丁鹏教授: Randomization Inference for Peer Effects

12 月 13 日,北京大学统计科学中心邀请来自 UC Berkeley 统计学助理教授丁鹏,带来 "Randomization Inference for Peer Effects" 研究报告,对室友效应做了有趣的分析和讲解。这篇论文即将在统计学顶级期刊 "Journal of the American Statistical Association" 发表。



北大一种典型的学生宿舍是本科生的四人间:从入学一开始的四个互相陌生的学生,在一起学习和生活四年。那么,四人之间的相互促进和影响就非常重要,也是大家关心的问题。

北大大部分学生是通过高考录取的,另外有小部分通过保送录取的(2018 年北大保送生有 190 人左右)。保送生是通过奥赛和科技创新比赛的成绩录取,他(她)们主要集中在理工科院系。这两种学生常常混合住在四人的宿舍里面。我们很自然地问:高考生和保送生的成绩是否有差别?高考生和保送生混合在一个宿舍是否会对相互的成绩产生影响?

针对北大本科生分配宿舍过程中把高考生和保送生混合在一个宿舍的"同宿舍"效应,通过对北大信息科学学院和物理学院 2013 年的毕业男生(女生样本不足)第一年 GPA 的分析,发现了如下一些有趣的结果。总体来讲,保送生第一年的平均 GPA 较高。进一步分析如下图所示,在信科学院,当保送生的三个室友全为保送生时,相比其他三种分配,第一年的平均成绩有显著提高。

在物理学院,当高考生的两个室友为保送生时比只有一个室友是保送生的分配下的平均成绩有显著提高。总的来说,保送生的室友也是保送生时,他的成绩会有显著提高。但是对高考生而言,室友是高考生还是保送生,并没有非常明确的结果。

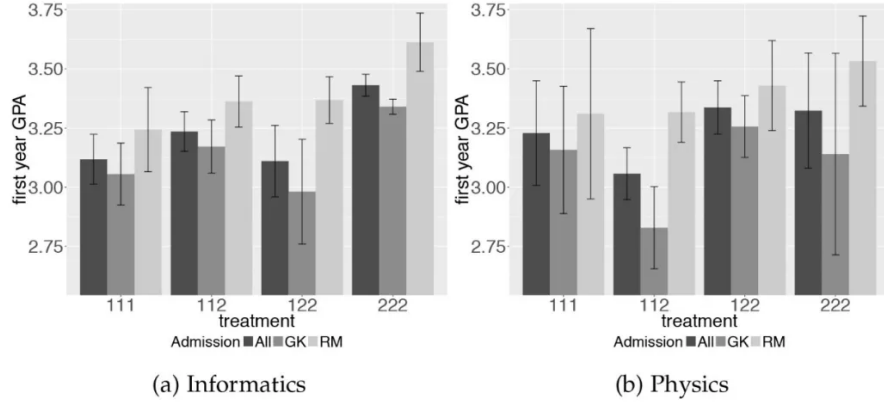


Table 2: Estimated peer effects with $\mathcal{R} = \{r_1, r_2, r_3, r_4\} = \{111, 112, 122, 222\}$. $\hat{\tau}$, $\hat{\tau}_{[1]}$ and $\hat{\tau}_{[2]}$ are the estimated average peer effects (subgroup average peer effects) for students from Gaokao and recommendation, respectively. The numbers in the parentheses are the estimated standard errors. The bold numbers correspond to those peer effects significantly different from zero at level 0.05.

Department	Estimator	(r_1, r_2)	(r_1, r_3)	(r_1, r_4)	(r_2, r_3)	(r_2, r_4)	(r_3, r_4)
Informatics	$\hat{\tau}$	-0.117 (0.086)	0.008 (0.109)	-0.313 (0.071)	0.125 (0.101)	-0.196 (0.059)	-0.321 (0.089)
	$\hat{\tau}_{[1]}$	-0.117 (0.112)	0.074 (0.152)	-0.285 (0.087)	0.191 (0.145)	-0.168 (0.074)	-0.359 (0.127)
	$\hat{\tau}_{[2]}$	-0.119 (0.126)	-0.125 (0.120)	-0.369 (0.123)	-0.006 (0.092)	-0.250 (0.096)	-0.244 (0.088)
Physics	$\hat{\tau}$	0.172 (0.142)	-0.108 (0.140)	-0.095 (0.177)	-0.280 (0.103)	-0.267 (0.150)	0.013 (0.148)
	$\hat{\tau}_{[1]}$	0.329 (0.185)	-0.099 (0.167)	0.017 (0.268)	-0.427 (0.145)	-0.311 (0.255)	0.116 (0.243)
	$\hat{\tau}_{[2]}$	-0.007 (0.219)	-0.119 (0.231)	-0.222 (0.225)	-0.112 (0.145)	-0.215 (0.135)	-0.103 (0.153)

在上图表中, Informatics 指信息科学学院, Physics 指物理学院。表中加粗的数字表示对平均 GPA 有显著不同作用的分配方案, 如在信息科学学院 (r_1, r_2) 这组数据中, 如果一个学生是高考生, 那么从一个只有 1 个保送生室友的宿舍搬入有 3 名保送生室友的宿舍, 其学习成绩将会提高 0.285, 如果是保送生, 其学习成绩将会提高 0.369, 平均来说会提高 0.313。

该论文对因果分析和政策评价中干涉作用的推断方法做出了重要贡献。当存在干涉作用时, 根据实际背景提出了新的假定来简化和处理问题: 每个个体只受室友的属性 (高考生或者保送生) 的影响; 在此基础上, 当高考生和保送生随机混合到一个宿舍时, 可以通过逆概率加权得到干涉作用的估计以及对应的方差, 并可以进一步给出置信区间和反过来找到同宿舍效应的度量。这项研究工作对北大宿舍分配方式或有启发作用。

(注: 本次报告由于女生样本不足, 所采用的数据均为男生数据, 且只针对北大物理学院和信息科技学院的保送生和高考生展开调查研究。)

8. 统计中心 2018 年来访人员名单

No	来访人员	时间
长期访问人员		
1	暴乐, 宾夕法尼亚大学	2018/7/1-2019/7/30
2	牛晓月, 宾夕法尼亚大学	2018/7/1-2019/7/30

短期来访人员		
1	Young Min Kim, Kyungpook National University, Korea	2018/3/1
2	Prof. Donald B. Rubin , Harvard University	2018/3/16
3	Xiangyu Cui, Shanghai University of Finance and Economics	2018/3/22
4	Prof. Wang Xin , City University of Hong Kong	2018/3/29
5	Alessandro Fassò, University of Bergamo	2018/3/29
6	Gang Han, Texas A&M University, School of Public Health	2018/4/12
7	Yuhong Yang, School of Statistics, University of Minnesota	2018/4/19
8	Bin Yu, UC Berkeley	2018/4/25-5/9
9	Chi-Lun Cheng, Institute of Statistical Science, Academia Sinica, Taiwan	2018/4/26
10	金焰, 纯信资产管理(深圳)有限公司	2018/5/10
11	李国星讲师, 北京大学公共卫生学院	2018/5/10
12	Bai Jiang, 普林斯顿大学	2018/5/14
13	Xiaohu Wang, Chinese University of Hong Kong	2018/5/17
14	Alicia Carriquiry, Iowa State University	2018/5/21
15	Junhui Wang, City University of Hong Kong	2018/5/24
16	吴月华, 加拿大约克大学	2018/5/24
17	Kaizheng Wang, Princeton University	2018/5/31
18	Kashinath Chatterjee, Visva-Bharati University, INDIA	2018/6/5
19	Lan Liu, University of Minnesota	2018/6/7
20	Dachuan Chen, University of Illinois at Chicago	2018/6/11
21	朱宏图教授, 滴滴首席统计学家	2018/6/14
22	Hongyuan Cao, University of Missouri-Columbia	2018/6/21
23	Jun Yan (康涅狄格大学)	2018/6/25
24	Jingshu Wang, University of Pennsylvania.	2018/6/28
25	Jin Jiashun, Carnegie Mellon University	2018/6/29-7/1
26	Cheng Yong Tang, Temple University	2018/6/29-7/1
27	Yuan Yao, Hong Kong University of Science and Technology	2018/6/29-7/1
28	Zheng Tracy Ke, University of Chicago	2018/6/29-7/1
29	Sanat K. Sarkar, Temple University	2018/7/2
30	Danyu Lin, University of North Carolina at Chapel Hill	2018/7/2
31	Ting Wang, Washington University Medical School	2018/7/3
32	王曦, 上海立信会计金融学院	2018/9/20
33	陈良富, 中国科学院遥感与数字地球研究所	2018/10/15
34	蔡知令, 加州大学戴维斯分校	2018/10/18
35	毛晓军, 复旦大学	2018/11/27
36	牛晓月, 宾夕法尼亚大学	2018/12/5
37	王静姝, 宾夕法尼亚大学沃顿商学院	2018/12/8
38	严加安, 中科院院士	2018/12/12
39	Peng Ding, UC Berkeley	2018/12/13
40	Weichen Wang, Two Sigma Investments	2018/12/14
41	李勇, 中国人民大学汉青研究院金融学院	2018/12/20
42	Renyu Zhang, New York University Shanghai	2018/12/27

七、论文发表

自 2018 年 1 月至 12 月，统计科学中心教员共发表论文 59 篇，按第一作者姓氏字母排序排列：

- [1] Arnbjerg, CJ ; Vestad, B ; Hov, JR ; Pedersen, KK; Jespersen, S ; Johannesen, HH ; Holm, K; Halvorsen, B; Fallentin, E ; Hansen, AE ; **Lange, T** ; Kjaer, A ; Troseid, M ; Fischer, BM; Nielsen, SD . Effect of Lactobacillus rhamnosus GG Supplementation on Intestinal Inflammation Assessed by PET/MRI Scans and Gut Microbiota Composition in HIV-Infected Individuals, JAIDS-JOURNAL OF ACQUIRED IMMUNE DEFICIENCY SYNDROMES, 78 期: 4 页: 450-457.
- [2] Cai, W., Guan, G., Pan, R., Zhu, X., and **Wang, H.** (2018), Network linear discriminant analysis, Computational Statistics and Data Analysis, 117, 32-44.
- [3] Cao, Y., **Lin, W.** and Li, H. (2018). Two-sample tests of high-dimensional means for compositional data. Biometrika, 105, 115-132.
- [4] Cao, Y., **Lin, W.** and Li, H. (2018+). Large covariance estimation for compositional data via composition-adjusted thresholding. Journal of the American Statistical Association, to appear.
- [5] Zhen-Qing Chen, **Ren, Y.-X.** and Renming Song: Llog L criterion for a class of multitype superdiffusions with non-local branching mechanisms, Science China Mathematics, accepted.
- [6] Zhen-Qing Chen, **Ren, Y.-X.** and Ting Yang: Skeleton decomposition and law of large numbers for supercritical superprocesses, Acta Applicandae Mathematicae, accepted.
- [7] **Chen, L.**, Guo, B., Huang, J., He, J., Wang, H., Shuyi Zhang, and S.X. Chen (2018). Assessing air-quality in Beijing-Tianjing-Hebei region: the method and mixed tales of PM2.5 and O3. Atmospheric Environment, 193, 290-301.
- [8] **S.X. Chen**, J. Li and P.-S. Zhong (2018) Two-Sample and ANOVA Tests for High Dimensional Means, The Annals of Statistics, to appear.
- [9] **Zhi Geng**, Yue Liu, Chunchen Liu, and Wang Miao (2018). Evaluation of Causal Effects and Local Structure Learning of Causal Networks, Annual Review of Statistics and Its Application, accepted.
- [10] J. He and **S. X. Chen** (2018) High-Dimensional Two-Sample Covariance Matrix Testing via Super-diagonals, Statistica Sinica, 28, 2671-2696.
- [11] Huang, M., Wang, S., **Wang, H.**, and Jin, T. (2018), Maximum smoothed likelihood estimation for a class of semiparametric pareto mixture densities, Statistics & Its Interface, 11(1):31-40.
- [12] Huang, D., Chang, X., and **Wang, H.** (2018), Spatial autoregression with repeated measurements for social networks, Communications in Statistics - Theory and Methods, 47(15), 3715-3727.

- [13] Huang, D., Zhou, J., and **Wang, H.** (2018), RFMS method for credit scoring based on bank card transaction data, *Statistica Sinica*, 28, 2903-2919.
- [14] Huang, M., Wang, S., **Wang, H.**, and Jin, T. (2018), Maximum smoothed likelihood estimation for a class of semiparametric Pareto mixture densities, *Statistics and Its Interface*, 11(1), 31-40.
- [15] Huang, D., Guan, G., Zhou, J., and **Wang, H.** (2018), Network-based naive Bayes model for social network, *Science China Mathematics*, 61 (4):627-640.
- [16] Kong, Xiangshun; **Ai, Mingyao**; Tsui, Kwok Leung (2018). Design for sequential follow-up experiments in computer emulations. *Technometrics* 60(1), 61-69.
- [17] Kong, Xiangshun; **Ai, Mingyao**; Tsui, Kwok Leung (2018). Flexible sliced designs for computer experiments. *Annals of the Institute of Statistical Mathematic* 70(3), 631-646.
- [18] A.E. Kyprianou, Sandra Pala and **Ren, Y.-X.**(2018).Almost sure growth of supercritical multi-type continuous state branching process , *ALEA, Lat. Am. J. Probab. Math. Stat.* 15, 409-428.
- [19] W., Fang, Z., **Wang, H.**, and Tsai, C. L. (2018), Covariance matrix estimation via network structure, *Journal of Business and Economics Statistics*, 36(2), 359--369.
- [20] Lan, W., Ma, Y., Zhao, J., **Wang, H.** and Tsai, C. L. (2018), Sequential model averaging for high dimensional linear regression models, *Statistica Sinica*, 28, 449-469.
- [21] Lyu G, Zhang C, Ling T, Liu R, Zong L, Guan Y, Huang X, Sun L, Zhang L, **Li C**, Nie Y, Tao W. Genome and epigenome analysis of monozygotic twins discordant for congenital heart disease. *BMC Genomics*. 2018; 19(1):428.
- [22] Li R, Liu Y, Hou Y, Gan J, Wu P, **Li C**. 3D genome and its disorganization in diseases. *Cell Biol Toxicol*. 2018; doi: 10.1007/s10565-018-9430-4.
- [23] Li T, Jia L, Cao Y, Chen Q, **Li C**. OCEAN-C: mapping hubs of open chromatin interactions across the genome reveals gene regulatory networks. *Genome Biol*. 2018; 19(1):54.
- [24] Li, W., Jiang, Z. C., **Geng, Z.** and Zhou, X. H. (2018) Identification of causal effects with latent confounding and classical additive errors in treatment. *Biometrical Journal*, 1-18, DOI: 10.1002/bimj.201700048
- [25] Lin, Z. and **Yao, F.** (2018) Intrinsic Riemannian functional data analysis. To appear *Annals of Statistics*.
- [26] Luo, P., Cai, Z. and **Geng, Z.** (2018) Criteria for multiple surrogates. To appear in *Statistica Sinica*
- [27] Martin Schultz, Line Jee Hartmann Rasmussen, Malene H. Andersen, Jakob S. Stefansson, Alexander C. Falkentoft, Morten Alstrup, 1 Andreas Sandø, 1 Sarah L. K. Holle, Jeppe Meyer, Peter B. S. Törnkvist, Thomas Høi-Hansen, Erik Kjølner, 1 Birgitte Nybo Jensen, Morten Lind, Lisbet Ravn, Thomas Kallemose, **Theis Lange**, Lars Køber, Lars Simon Rasmussen, Jesper Eugen-Olsen, and Kasper Karmark Iversen. Use of the prognostic biomarker suPAR in the emergency department improves risk stratification but has no effect on mortality: a cluster-

randomized clinical trial (TRIAGE III), SCANDINAVIAN JOURNAL OF TRAUMA RESUSCITATION & EMERGENCY MEDICINE(2018) , 69.

[28] Mortensen, J; Dich, N; **Lange, T**; Ramlau-Hansen, CH; Head, J; Kivimaki, M; Leineweber, C; Rod, NH(2018). Weekly hours of informal caregiving and paid work, and the risk of cardiovascular disease, EUROPEAN JOURNAL OF PUBLIC HEALTH, 28 期: 4 页: 743-747.

[29] Peter Møller Hansen, Kristoffer Lindskov Hansen, Mads Møller Pedersen, **Theis Lange**, Lars Lönn, Jørgen Arendt Jensen, and Michael Bachmann Nielsen. Atherosclerotic Lesions in the Superficial Femoral Artery (SFA) Characterized with Velocity Ratios using Vector Velocity Ultrasound, Ultrasound Int Open, p; 4(3): E79–E84.

[30] Petersen GL, Mortensen EL, Rod NH, **Lange T**, Flensburg-Madsen T, Hansen ÅM, Lund R. Occupational Social Class and Personality Traits in Relation to Leisure-Time Physical Activity Level: Cross-Sectional Results From the Copenhagen Aging and Midlife Biobank. J Aging Health.Sep;30(8):1263-1283.

[31] Zhenhua Lin, Hans-Georg Müller, and **Fang Yao**. Mixture inner product spaces and their application to functional data analysis. The Annals of Statistics, Volume 46, Number 1 (2018), 370-400.

[32] Koudstaal, M.*, and **Yao, F**. From mutiple Gaussian Sequences to functional data and beyond: a Stein estimation approach (supplementary material). Journal of the Royal Statistical Society, Series B, volume 46, number 1(48).

[33] **Wang Miao, Zhi Geng**, ERIC TCHETGEN (2017). Identifying Causal Effects With Proxy Variables of an Unmeasured Confounder, Biometrika , 105(4):987-993.

[34] **Wang Miao**, ERIC TCHETGEN (2018). Identification and inference with nonignorable missing covariate data, Statistica Sinica 28, 2049-2067.

[35] **Minya Xu**, With Xin Qin, Peter Hom. Am I a peasant or a worker? An identity strain perspective on why and when developing-world migrants quit. Human Relations (accepted).

[36] **Minya Xu**, With Dong Ju, Xin Qin and Paul Spector. A multilevel study of abusive supervision, norms and personal control on counterproductive work behavior: A theory of planned behavior approach. Journal of Leadership and Organizational Studies (accepted).

[37] Mao, X., **Chen, SX** and Wong, R. (2018) Matrix Completion with Covariate Information, Journal of the American Statistical Association, to appear.

[38] Pan, R., Guan, R., Zhu, X., and **Wang, H**. (2018), A latent moving average model for network regression, Statistics and Its Interface, Volume 11, 641–648.

[39] Qiu, Y., **Chen, S.X.** and Nettleton, D. (2018) Detecting Rare and Faint Signals via Thresholding Maximum Likelihood Estimators, Annals of Statistics, 46, 895-923.

- [40] **Ren, Y.-X.**, Renming Song and Rui Zhang(2018).William decomposition for superprocesses, Electron. J. Probab. 23, No. 23, 1-33.
- [41] **Ren, Y.-X.**, Renming Song and Zhenyao Sun (2018). A 2-spine decomposition of the critical Galton-Watson tree and a probabilistic proof of Yaglom's theorem, Electron. Commun. Probab. Volume 23, paper no. 42, 12 pp
- [42] **Ren, Y.-X.**, Renming Song and Rui Zhang: Super critical super processes: proper normalization and non-degenerate strong limit Science China Mathematics, accepted.
- [43] BaoLuo Sun, Lan Liu, **Wang Miao**, Kathleen Wirth, James Robins and Eric J. Tchetgen (2018). Semiparametric estimation with data missing not at random using an instrumental variable, Statistica Sinica ,28, 1965-1983
- [44] Traen, Bente; Stulhofer, Aleksandar; Janssen, Erick; Carvalheira, Ana Alexandra; Hald, Gert Martin; **Lange, Theis**; Graham, Cynthia(2018). Sexual Activity and Sexual Satisfaction Among Older Adults in Four European Countries, Archives of sexual behavior, 10.1007/s10508-018-1256-x.
- [45] C. Ho, W. Wang and **J. Yu**, 2018. International Knowledge Spillover through Trade: A Time-Varying Spatial Panel Data Approach. Economics Letters 162, 30-33.
- [46] Wang H, Wang F, Wu S, Liu Z, Li T, Mao L, Zhang J, **Li C**, Liu C, Yang Y. Traditional herbal medicine-derived sulforaphene promotes mitophagic cell death in lymphoma cells through CRM1-mediated p62/SQSTM1 accumulation and AMPK activation. Chem Biol Interact. 2018; 281:11-23.
- [47] **Wang, Yaping**; Yang, Jianfeng; Xu, Hongquan (2018). On the connection between maximin distance designs and orthogonal designs. Biometrika 105(2), 471-477.
- [48] Xu, K., Sun, J., Liu, J., and **Wang, H.** (2018), An empirical investigation of taxi driver response behavior to ride-hailing requests: A spatio-temporal perspective, Plos One, 13(6): e0198605.
- [49] Yin, Y. J., Liu, L. and **Geng, Z.** (2018) Assessing the treatment effect heterogeneity with a latent variable.Statistica Sinica 28, 115-135.
- [50] Yin, Y. J., Zhou, X. H., **Geng, Z.** and Lu, F. (2018) Assessing the heterogeneity of treatment effects by identifying the harm rate. Statistica Sinica 28, 137-156.
- [51] Yu, Jun; Kong, Xiangshun; **Ai, Mingyao**; Tsui, Kwok Leung (2018). Optimal designs for dose-response models with linear effects of covariates. Computational Statistics and Data Analysis 127, 217-228.
- [52] Yu, Jun; **Ai, Mingyao**; Wang, Yaping (2018). Optimal designs for linear models with Fredholm-type errors. Journal of Statistical Planning and Inference 194, 65-74.
- [53] Zhang, X. and **J. Yu**, 2018. Spatial Weights Matrix Selection and Model Averaging for Spatial Autoregressive Models. Journal of Econometrics 203, 1-18.

- [54] Zhao T, Fu Y, Zhu J, Liu Y, Zhang Q, Yi Z, Chen S, Jiao Z, Xu X, Xu J, Duo S, Bai Y, Tang C, **Li C**, Deng H. Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. *Cell Stem Cell*. 2018; 23(1):31-45.e7.
- [55] Zhang C, Fan C, Gan J, Zhu P, Kong L, **Li C**. iSeq: Web-Based RNA-seq Data Analysis and Visualization. *Methods Mol Biol*. 2018; 1754:167-181.
- [56] Zhou, J., Huang, D., and **Wang, H.** (2018), A note on estimating network dependence in a discrete choice model, *Statistics and its Interface*, Volume 11, No. 3, 433-439.
- [57] 李晨煦, **李辰旭** (2018) 随机杠杆效应对隐含波动率曲面期限结构的影响分析. *金融学季刊*, 第 12 卷第 1 期.
- [58] 苗旺, 刘春辰, **耿直** (2018). 因果推断的统计方法, 《中国科学: 数学》, 第 48 卷: 1 ~ 33.
- [59] **任艳霞**, 宋仁明, 张蕊: 上临界超过程的一类强极限的性质, 中国科学, 录用。

八、研究报告

- [1] Lei Chen, Bin Guo, Jiasheng Huang, Jing He, Hengfang Wang, Shuyi Zhang, **Song Xi Chen**
Assessing Air-Quality in Beijing-Tianjin-Hebei Region: the Method and Mixed Effects of PM_{2.5} and O₃
- [2] **Zhi Geng**, Yue Liu, Chunchen Liu, and **Wang Miao**
Evaluation of Causal Effects and Local Structure Learning of Causal Networks
- [3] **Wang Miao** and Eric Tchetgen Tchetgen
Invited Commentary: Bias Attenuation and Identification of Causal Effects With Multiple Negative Controls
- [4] **Wang Miao** and Eric Tchetgen Tchetgen
Identification and Inference With Nonignorable Missing Covariate Data
- [5] BaoLuo Sun, Lan Liu, **Wang Miao**, Kathleen Wirth, James Robins and Eric J. Tchetgen
Tchetgen
SEMI-PARAMETRIC ESTIMATION WITH DATA MISSING NOT AT RANDOM USING AN INSTRUMENTAL VARIABLE
- [6] Zengmiao Wang, Huaying Fang, Nelson Leung-Sang Tang, and **Minghua Deng**
VCNet: vector-based gene co-expression network construction and its application to RNA-seq data
- [7] HUAYING FANG, CHENGCHENG HUANG, HONGYU ZHAO, and **MINGHUA DENG**
gCoda: Conditional Dependence Network Inference for Compositional Data
- [8] Changjing Wu, Hongyu Zhao, Huaying Fang and **MINGHUA**
Graphical model selection with latent variables

- [9] Jiang Chang, Wenle Tan, Zhiqiang Ling, **Ruibin Xi**, Mingming Shao, Mengjie Chen, Yingying Luo, Yanjie Zhao, Yun Liu, Xiancong Huang, Yuchao Xia, Jinglin Hu, Joel S Parker, David Marron, Qionghua Cui, Linna Peng, Jiahui Chu, Hongmin Li, Zhongli Du, Yaling Han, Wen Tan, Zhihua Liu, Qimin Zhan, Yun Li.
Comprehensive analysis of esophageal squamous-cell carcinoma identifies alcohol drinking-related mutation signature and novel genomic alterations
- [10] Yuchao Xia, Yun Liu, Minghua Deng and **Ruibin Xi**
Pysim-sv: a package for simulating structural variation data with GC-biases
- [11] Haiyang Liu, Jack Ting-Ju Chiang, Ryan Fehr, **Minya Xu** and Siting Wang
How do leaders react when treated unfairly? Leader narcissism and self-interested behavior in response to unfair treatment.
- [12] X. Dai, G.H. Müller and **F. Yao**
Optimal Bayes classifiers for functional data and density ratios
- [13] Z. Lin, G.H. Müller, and **F. Yao**
Mixture inner product spaces and their application to functional data analysis
- [14] Yuanpei Cao, **Wei Lin** and Hongzhe Li
Two-sample tests of high-dimensional means for compositional data
- [15] **Yanxia Ren**, Renming Song and Rui Zhang
Supercritical Superprocesses: Proper Normalization and Non-degenerate Strong Limit
- [16] Zhen-Qing Chen, **Yanxia Ren** and Renming Song
Llog L criterion for a class of multitype superdiffusions with non-local branching mechanisms
- [17] Zhen-Qing Chen, **Yanxia Ren** and Ting Yang
Skeleton decomposition and law of large numbers for supercritical superprocesses
- [18] Xiaojun Mao, **Song Xi Chen** and Raymond K. W. Wong Matrix Completion with Covariate Information
- [19] M. Koudstaal, and **F. Yao**
From Multiple Gaussian Sequences to Functional Data and Beyond: A Stein Estimation Approach
- [20] M. Koudstaal, and **F. Yao**
Supplementary Material for “From Multiple Gaussian Sequences to Functional Data and Beyond: A Stein Estimation Approach”
- [21] Shuyi Zhang, Bin Guo, **Anlan Dong**, Jing He, Ziping Xu, **Song Xi Chen**
Cautionary Tales on Air Quality Improvement in Beijing
- [22] Chong Chen, Ruibin Xi and Nan Lin
Community Detection by L0-penalized Graph Laplacian
- [23] Wei Lan, Yingying Ma, Junlong Zhao, **Hansheng Wang** and Chih-Ling Tsai Sequential Model Averaging for High Dimensional Linear Regression Models

- [24] Lujun Li, Hui Shao, Ruodu Wang and **Jingping Yang**
Worst-case Range Value-at-Risk with Partial Information
- [25] D. Huang, X. Chang and **Hansheng Wang**
Spatial Autoregression with Repeated Measurements for Social Networks
- [26] Yacine Ait-Sahalia, **Chenxu Li**
Closed-Form Implied Volatility Surfaces for Stochastic Volatility Models
- [27] **Lei Chen**, Bin Guo, Jiasheng Huang, Shuyi Zhang, Hengfang Wang, Jing He, **Songxi Chen**
空气质量评估报告(五)
- [28] **Huiming Zhang**, Bo Li and G. Jay Kerns
A characterization of signed discrete infinitely divisible distributions
- [29] Kangjie Zhou and **Jinzhu Jia**
Variance Reduction for Causal Inference
- [30] Kangjie Zhou, **Jinzhu Jia**
Propensity Score Adapted Covariate Selection for Causal Inference
- [31] **Yundong Tu**, Qiwei Yao, Rongmao Zhang
Error-Correction Factor Models for High-dimensional Cointegrated Time Series
- [32] **Yundong Tu** and Li shuo
A Joint Test for Parametric Specification and Independence in Nonlinear Regression Models
- [33] **Yundong Tu** and Wng siwei
函数型核加权最小二乘估计法及其在经济学的应用
- [34] Bingyan Wang and **Jinzhu Jia**
Identifiability of Generalized Linear Model with Nonignorable Missing Data

九、学术讲座

北京大学统计科学中心 2018 年共组织学术报告 45 场，报告人来自美国、英国、丹麦、加拿大、香港以及国内各高校、研究机构和知名企业。其中来自海外的学者 33 场，国内的学者 12 场。

- [1] Nonparametric Spectral Density Estimation under Long-Range Dependence Young Min Kim, Kyungpook National University, Korea
2018-03-01 14:00-15:00
217, Guanghua Building 2
- [2] Models for Imputing Missing Data, Including Methods for Assessing Sensitivity of Conclusions to Them
Prof. Donald B. Rubin , Harvard University

2018-03-16 14:00-15:30

Lecture Hall, Jiayibing Building, Jingchunyuan 82, BICMR

- [3] Forecasting security's volatility using low-frequency and high-frequency historical data and option-implied information

Xiangyu Cui, Shanghai University of Finance and Economics

2018-03-22 14:00-15:00

217, Guanghua Building 2

- [4] A multidimensional network approach reveals microRNAs as determinants of the mesenchymal colorectal cancer subtype

Prof. Wang Xin, City University of Hong Kong

2018-03-29 14:00-15:00

1114M, Science Building 1

- [5] Case studies on functional control charts and atmospheric profiles uncertainty Alessandro Fassò, University of Bergamo

2018-03-29 15:15-16:15

217, Guanghua Building 2

- [6] Bayesian Hybrid Inference and Its Application in HIV Viral Load Modeling and Computer Experiments

Gang Han, Associate Professor of Biostatistics Texas A&M University, School of Public Health

2018-04-12 15:00-16:00

1303, Science Building 1

- [7] Treatment Allocations Based on Multi-Armed Bandit Strategies

Yuhong Yang, School of Statistics, University of Minnesota

2018-04-19 14:00-15:00

217, Guanghua Building 2

- [8] 数据科学三原则：可预测性、可计算性、稳定性。

郁彬，著名统计学家、美国国家科学院院士，加州大学伯克利分校

2018-04-25 19:00-21:00

理教 103

- [9] Polynomial regression with heteroscedastic measurement errors in both axes: Estimation and hypothesis testing

Chi-Lun Cheng, Institute of Statistical Science, Academia Sinica, Taipei, Taiwan
2018-04-26 14:00-15:00
1303, Science Building 1

[10] 用迭代随机森林算法发现具有预测性的稳定高阶交互作用
郁彬, 著名统计学家、美国国家科学院院士, 加州大学伯克利分校
2018-05-09 19:00-21:00
理教103

[11] Quantitative Investment in China
金焰, 纯信资产管理(深圳)有限公司
2018-05-10 14:00-15:00
202, Guanghua Building 1

[12] 大气污染&气候变化与人群健康的关系: 基于案例的介绍
李国星, 北京大学公共卫生学院
2018-05-10 15:00-16:00
1303, 理科一号楼

[13] Convergence of Contrastive Divergence in Exponential Family
Bai Jiang, Princeton University
2018-05-14 14:00-15:00
1114, 理科一号楼

[14] Estimation for Persistence Matrix in Multivariate Diffusion Processes
Xiaohu Wang, Chinese University of Hong Kong
2018-05-17 14:00-15:00
217, Guanghua Building 2

[15] Learning algorithms to evaluate forensic evidence
Alicia Carriquiry, Iowa State University
2018-05-21 16:00-17:00
光华管理学院 1 号楼 201 室

[16] Smooth collaborative recommender systems
Junhui Wang, City University of Hong Kong
2018-05-24 14:00-15:00
理科一号楼 1303

- [17] Some spatio-temporal modeling problems and recent advances
吴月华, 加拿大约克大学
2018-05-24 15:00-16:00
1303, 理科一号楼
- [18] Entrywise eigenvector analysis of random matrices with applications
Kaizheng Wang, Princeton University
2018-05-31 14:00-15:00
1303, 理科一号楼
- [19] Bayesian Statistics for High-Dimensional and Nonparametric Models
Aad van der Vaart, 荷兰皇家与艺术科学院院士、莱顿大学
2018-06-04-06-09
202, 光华 1 号楼
- [20] General constructions of optimal supersaturated designs
Kashinath Chatterjee, Visva-Bharati University, INDIA
2018-06-05 14:00-15:00
理科一号楼 1303
- [21] Efficient Inference for task fMRI studies
Lan Liu, assistant professor in the department of statistics at University of Minnesota
2018-06-07 14:00-15:00
1303, 理科一号楼
- [22] The Five Trolls under the Bridge: Principal Component Analysis with Asynchronous and Noisy High Frequency Data
Dachuan Chen, University of Illinois at Chicago
2018-06-11 14:00 - 15:00
217, Guanghua Building 2
- [23] 统计运用在大数据上的机遇挑战与策略
朱宏图, 滴滴首席统计学家
2018-06-14 14:00-15:00
1114, 理科一号楼
- [24] Statistical Methods for Integrative Analysis of Multi-Omics Data
Hongyuan Cao, University of Missouri-Columbia
2018-06-21 14:00-15:00

理科一号楼 1303

- [25] Combined Score Equations for Spatial Extremes Modeling with Application to Detection and Attribution of Climate Change
Jun Yan, 康涅狄格大学
2018-06-25 14:00-15:00
217, Guanghua Building 2
- [26] A comprehensive Framework for Genome-wide Mendelian Randomization under Pervasive Pleiotropy
Jingshu Wang, the Wharton School, University of Pennsylvania
2018-06-28 14:00-15:00
理科一号楼 1303
- [27] Higher Criticism for Large Scale Inference, especially for Rare and Weak effects
Jiashun Jin, Carnegie Mellon University
2018-06-29 -07-01
202, 光华 1 号楼
- [28] Parsimonious statistical modeling approaches for longitudinal studies
Cheng Yong Tang, Temple University
2018-06-29 -07-01
202, 光华 1 号楼
- [29] Differential Inclusion Method in high dimensional statistics and Deep Learning toward deeper understanding
Yuan Yao, Hong Kong University of Science and Technology
2018-06-29 -07-01
202, 光华 1 号楼
- [30] New Tools for Analyzing Complicated and High-dimensional Data。
Zheng Tracy Ke, University of Chicago
2018-06-29 -07-01
202, 光华 1 号楼
- [31] Local False Discovery Rate Based Methods for Multiple Testing of One-Way Classified Hypotheses
Sanat K. Sarkar, Department of Statistical Science, Temple University
2018-07-02 14:00-15:30
理科一号楼 1114

- [32] Research and Training in Biostatistics
Danyu Lin, Department of Biostatistics, University of North Carolina at Chapel Hill
2018-07-02 10:00-12:00
理科一号楼 1303
- [33] Transposable elements in normal and cancer epigenome
Ting Wang, Washington University Medical School
2018-07-03 14:00-15:00
理科一号楼 1304
- [34] Nonparametric Estimation of Generalized Transformation Panel Data Models with Fixed Effects
王曦, 上海立信会计金融学院
2018-09-20 14:00--15:00
光华管理学院 2 号楼 216
- [35] 大气污染卫星监测技术的现状与发展态势
陈良富, 中国科学院遥感与数字地球研究所
2018-10-15 14:00-15:00
光华 2 号楼 217 教室
- [36] Business Report Writing
蔡知令, 加州大学戴维斯分校
2018-10-18 14:00-15:00
光华 2 号楼 K02
- [37] Leaving No One Behind: Size Estimation of People at High Risk for HIV Infections
暴乐, 宾夕法尼亚大学
2018-11-01 14:00-15:00
光华 2 号楼 216 教室
- [38] Matrix Completion under Low-Rank Missing Mechanism
毛晓军, 复旦大学
2018-11-27 14:00-15:00
光华 2 号楼 216 教室
- [39] Scalable bioinformatics methods for single-cell RNA-seq analysis
Joshua Ho, University of Hong Kong

- 2018-11-29 14:00-15:00
理科一号楼 1303
- [40] Circuit Breakers and Contagion
曾旭东, 上海财经大学
2018-12-06 14:00-15:00
光华 2 号楼 K02 教室
- [41] 我心目中的科学与艺术
严加安, 中国科学院院士
2018-12-12 19:00-20:00
光华 1 号楼 216 教室
- [42] Randomization inference for peer effects, with application to the roommate assignment in Peking University
Peng Ding, UC Berkeley
2018-12-13 14:00-15:00
光华 2 号楼 K02 教室
- [43] Robust Large Covariance Estimation for Heavy-tailed Factor Models
Weichen Wang, Princeton University
2018-12-14 14:00-15:00
光华 2 号楼 217 教室
- [44] Integrated Deviance Information Criterion for Latent Variable Models
李勇, 中国人民大学汉青研究院
2018-12-20 14:00-15:00
光华 2 号楼 K02 教室
- [45] Recommender Systems on Platforms (Joint work with Xuan Wang and Dennis J. Zhang)
Renyu Zhang, New York University Shanghai
2018-12-27 14:00-15:00
光华 2 号楼 K02 教室

十、委员会

1. 国际顾问委员会



王永雄

主席，斯坦福大学统计系及健康研究和政策系教授，美国科学院院士，曾获得过统计学界最高奖“考普斯会长奖”。



范剑青

委员，普林斯顿大学运筹与金融工程系系主任，统计研究委员会主任，中央研究院院士，曾获得过统计学界最高奖“考普斯会长奖”、洪堡研究奖、世界华人数学家大会晨兴数学金奖等奖项。



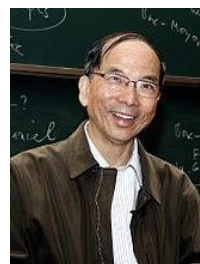
马志明

委员，中国科学院数学与系统科学研究院应用数学研究所研究员，国家“973”计划重点基础研究发展规划“核心数学的前沿问题”项目首席科学家，曾任中国数学会理事长，中国科学院院士，第三世界科学院院士。



Peter J. Bickel

委员，加州大学伯克利分校统计系教授，美国艺术与科学院院士，美国科学院院士。



吴建福

委员，乔治亚理工学院工程统计讲座教授，数理统计学会、美国统计学会、运筹学和管理学研究协会等学会会士，美国工程院院士，曾获得过统计学界最高奖“考普斯会长奖”。



黎子良

委员，现为美国斯坦福大学教授。1983 年获国际统计学界的“考普斯会长奖”。

2. 科学委员会



陈嵘

委员，罗格斯大学，北京大学光华管理学院特聘教授，美国统计学会荣誉会员，美国统计学会学报、商务经济统计学报编委，兼泛华统计学会董事。



陈松蹊

委员，北京大学统计科学中心联席主任，北京大学光华管理学院商务统计与经济计量系联席系主任，国家首批“国家级重大人才计划”入选者，现为北京大学讲席教授。



刁锦寰

委员，国际著名统计学家，现为美国芝加哥大学商学院经济计量学与统计学荣休讲席教授。他是泛华统计协会的创始人暨首任会长，是美国统计学会高级会员、台湾中央研究院院士，北京大学名誉教授。



耿直

委员，北京大学统计科学中心联席主任，北京大学数学科学学院概率统计系教授，获评国家级重大人才计划（基金委）。现任中国现场统计研究会第九界理事会理事长（2013-2017），IMS-China 主席，中国统计学会副会长，国家社会科学基金学科评审组专家等职。



胡永华

委员，北京大学公共卫生学院常务副院长、流行病与卫生统计学系主任，教育部流行病学重点实验室主任，北京大学循证医学中心常务副主任，北京大学公共卫生与预防医学教育委员会主任委员，被聘为香港中文大学荣誉教授，亚太地区公共卫生科学理事会中国区域主任。



陈大岳

委员，北京大学数学科学学院院长，中国数学会秘书长，中国概率统计学会副理事长。曾获国家级重大人才计划（基金委）等多项资助，曾任中国数学会副理事长和《数学进展》主编等职。



姚琦伟

委员，伦敦政治经济学院统计系主任，北京大学光华管理学院特聘教授，他还是皇家统计学会荣誉会员、国际统计研究所推选成员、美国统计协会、数理统计学会、泛华统计学会会员。

3. 指导委员会



高松

主席，北京大学党委常委、副校长，中国科学院化学学部院士，英国皇家化学会会士 (FRSC, 2007-), 2013 年当选为发展中国家科学院院士，2013 年荣获“何梁何利基金科学与技术进步奖”。



陈大岳

委员，北京大学数学科学学院院长，中国数学会秘书长，中国概率统计学会副理事长。曾获国家级重大人才计划（基金委）等多项资助，曾任中国数学会副理事长和《数学进展》主编等职。



刘俏

委员，北京大学光华管理学院院长，国家级重大人才计划 2 项（教育部、基金委）人选。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。



张平文

委员，现任北京大学副教务长、学科建设办公室主任，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、国家级重大人才计划 2 项（教育部、基金委）人选、冯康科学计算奖、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。



编辑：李楠 王彦懿

Tel: 0086-010-62760736

Mail: stat-center@pku.edu.cn

Http: www.stat-center.pku.edu.cn

Add: 北京大学理科五号楼

Postcode: 100871