



北京大学统计科学中心

Center for Statistical Science, Peking University

2016 年度报告

祝贺北大统计科学中心的成立，
希望“中心”成为吸引海外高层次人才，
培养本土拔尖人才的基地。
李源澄
6.18

目 录

一、综述	2
二、人员组成	3
三、中心要闻	12
四、交流访问	21
五、发表论文	26
六、研究报告	31
七、学术讲座	33
八、委员会	40

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、光华管理学院以及医学部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。

中心的成立得到了各级政府部门和北京大学相关部门的支持和指导，他们在中心向着建设具有国际先进水平的队伍、培养复合型统计人才、大力推进跨学科研究、促进高校与政府及企业等机构的积极互动的目标不断迈进的过程中也一直给予关注与支持。

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，整合全校的统计研究力量，致力于培养综合型的统计人才，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、世界一流的统计研究机构。

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；执行委员会负责中心的日常管理工作。中心在这几年中利用新体制共引进海外高层次人才7人，其中3人入选中组部“青年千人计划”。

统计科学中心的主要任务是建立一支具有国际先进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验设计和统计分析水平，为这些学科的发展提供数据收集和数据分析的统计方法。

中心的成立将在北京大学建立一个吸引海内外优秀统计人才的基地。中心非常重视开放式、国际化的学术环境建设，广泛联系全球各地的统计学者，吸引优秀人才到北京大学工作。目前，统计科学中心正在积极开展研究生招生工作，几年之后，中心将为我国的统计学、自然科学、人文社会科学等各个领域提供杰出的科学研究人才，为各行业提供优秀的管理和技术人才，进一步提高北京大学在国内统计学研究领域的领先优势。



Peter G. hall 教授摄

一、综述

2016 年统计科学中心继续在统计学研究、博士生培养以及推动北大及中国统计学发展等方面开展工作。

统计科学中心积极开展同国际和国内统计学界的学术交流与合作。2 月至 3 月统计科学中心邀请哥本哈根大学生物统计部副教授 Theis Lange 博士在北京大学开设了题为“Causal inference based on counterfactuals”的短期课程。5 月，为促进海峡两岸学者的交流合作，第三次“两岸经济对话”在台湾“中央研究院”经济所拉开帷幕；“两岸经济对话”系列活动由芝加哥大学商学院教授、北京大学名誉教授刁锦寰先生发起，由陈松蹊教授和周雨田教授共同组织，集中研讨交流两岸经济学研究的最新问题，针对两岸经济建模和预测等问题深入探讨。5 月 17-18 日“大数据时代下的高维统计建模与分析研讨会”在北京大学光华管理学院成功举行，本次会议由北京大学统计科学中心和北京大学数量经济与数理金融教育部重点实验室共同主办，北京大学光华管理学院承办。7 月统计科学中心成功举办了第三届优秀大学生夏令营，中国人民大学、南开大学、北京师范大学、上海财经大学、中山大学等十几所高校的 31 名三年级本科生参加了本次夏令营。2016 年 9 月，中心第二批四名博士研究生入学。12 月，数据科学与概率统计学术研讨会成功举办。12 月 16 日为促进北大和清华两校教师之间以及学生之间的交流和合作，促进统计学科的发展，迎接新时代统计学面临的机遇和挑战，第一届北大-清华统计论坛在北京大学国际数学中心甲乙丙楼多功能会议厅召开。2016 年 12 月 12 日至 2017 年 1 月 5 日，华盛顿大学的助理教授韩放老师在北京大学开设了机器（统计）学习理论课程。统计科学中心资助和举办的这些活动极大地扩大了中心在国内外的影响力。

2016 年中心教员共发表 53 篇学术论文，其中有 5 篇发表在统计学四大期刊。继 2015 年 3 月发布《空气质量评估报告：北京城区 2010-2014 年 PM2.5 污染状况研究》后，由陈松蹊教授率领的统计中心团队在三月份正式发布了《空气质量评估报告（二）：中国五城市空气污染状况之统计学分析》。报告一经发出，便引起了国内外媒体的关注，美国《纽约时报》网站和国内《参考消息》先后对其进行了报道。由统计科学中心教师为主的环境统计团队成功申请到国家重点研究专项—空气质量统计诊断模型，该大气污染专项历时四年(2016.6.1—2020.5.31)。中心在这一年中进一步同国内外的学者进行了有效的学术交流，共举办学术报告 56 场，其中来自海外的学者 31 场，国内的学者 25 场。这些学术报告拓宽了老师和学生的学术视野，促进了统计学领域的交流，对中心的学术发展起到了很大的推动作用。我们衷心感谢所有对北京大学统计科学中心发展给予关心和指导的各级领导、各界朋友、海内外同行。



二、人员组成

1. 执行委员会

陈松蹊

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席系主任，首批“千人计划”入选者，北京大学讲席教授。美国统计学会会士，数理统计协会资深会员和常务理事，国际统计学会当选会员，中国统计学会常务理事。研究方向：数理统计，经济计量学，高维数据分析，环境统计。

耿直

统计科学中心联席主任，数学科学学院概率统计系教授，国家杰出青年科学基金获得者。曾任 IMS-China 主席，中国现场统计研究会第九届理事会理事长（2013-2017），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。研究方向：因果推断，数理统计，生物医学统计。

陈大岳

统计科学中心副主任，曾获国家自然科学基金委“杰出青年”等多项资助，北京大学数学科学学院常务副院长、中国数学会副理事长和《数学进展》主编等职。研究方向：马氏过程，无穷粒子系统，渗透模型，随机游动。

房祥忠

北京大学数学科学学院概率统计系系主任。研究方向为生存分析和可靠性。1996 年获中国航天总公司（部级）科技进步二等奖。2002 年获北京市科技进步二等奖。2002 年获中国国防科学技术工业委员会国防科学技术奖三等奖。

李程

生命科学学院研究员。研究方向：生物信息和计算生物学在癌症基因组学中的应用。研究组开发的 dChip、ComBat 软件和相关数据分析方法被广泛应用于基因表达和 SNP 生物芯片的数据分析和可视化。

王汉生

光华管理学院商务统计与经济计量系系主任。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。研究方向：高维数据分析，变量选择，搜索引擎营销，社会关系网络。

吴岚

数学科学学院金融数学系系主任。研究方向：金融资产定价，金融统计，经济资本模型，精算学。中国精算师协会会员，教育委员会委员，中国精算师考试负责人。

2. 全职教员



贾金柱研究员: 2009 年在北京大学概率统计系获博士学位, 2009-2010 年在 UC Berkeley 统计系进行博士后研究工作, 2011 年加入北大统计中心和数学学院。贾金柱博士的研究方向为高维统计推断、统计机器学习和因果推断等。他在变量选择方法的理论研究、高维数据统计学习的应用和因果推断等领域发表论文多篇。

高维统计中的预处理方法:

Lasso 这一方法的大样本性质(相合性)需要条件。它的符号相合(模型选择相合性)需要 irrepresentable condition, 它的 L2 相合性需要限制特征值条件 (restricted eigenvalue condition)。Jinzhu Jia and Karl Rohe (2015)的工作 Preconditioning the Lasso for sign consistency. Electronic Journal of Statistics, pp 1150-1172. 展示和提出如何规避稀疏线性回归分析需要的条件。给定 X, Y 满足如下线性回归方程:

$$Y = X\beta + \epsilon$$

该研究提出当设计矩阵 X 不满足 irrepresentable condition 时, 通过选择一个合适的矩阵 F , 对方程进行如下的预处理: 回归方程两边都乘以 F :

$$FY = FX\beta + F\epsilon$$

去掉 Lasso 所需要的 irrepresentable condition 和 restricted eigenvalue condition. 该研究提出的 F 如下构成: 记 X 的 SVD 分解为

$$X = UDV^T \text{ 则 } F \text{ 定义为 } F = UD^{-1}U^T.$$

由于 F 的定义取决于 X 的 SVD 分解, 我们称这种预处理方法为基于 SVD 分解的预处理方法。我们的预处理使得变换后的设计矩阵(FX) 的特征值只有 0 或者 1。当变量的个数(p)小于观测数据(n)的个数时, 变换后的矩阵是正交矩阵, 这个时候 irrepresentable condition 和 restricted eigenvalue condition 都成立。当变量的个数大于观测数据个数的时候, 变换后的设计矩阵行正交, 此时我们也证明了 irrepresentable condition 和 restricted eigenvalue condition 都以很大的概率成立。我们证明了在高维回归中, 预处理可是规避 irrepresentable condition, 并且明显改进 Lasso 的性质。随机模拟也进一步证明了我们的理论。我们的早先 arxiv 版本是 Preconditioning to comply with the irrepresentable condition. google scholar 显示, 该结果目前已经被引用 11 次。



涂云东研究员: 2012 年在加州大学获经济学博士学位，同年加入北大统计中心和光华管理学院。涂云东博士的研究兴趣涵盖非参数建模，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。他的研究致力于如何利用已有的信息或数据进行更加有效的推断和经济预测，让数据自身来揭示背后的经济规律，减少经济研究中作出的不合理假设。

股权溢价的非线性预测模型和经济约束：

涂云东研究员和他的合作者 2015 年在 *Journal of Business and Economic Statistics* 上发表论文 (Lee, Tu and Ullah, 2015 JBES) 研究了金融学中资产溢价的非线性建模和预测。资产溢价的预测是实证金融研究中的热点问题，长期以来，大量的研究文献通过实证发现，资产溢价不可以被金融学中的其它变量所预测。该研究发现资产溢价和金融变量间存在非线性关系，而这一非线性关系在文献中并没有被发现。进一步，该研究结合实证金融中资产溢价非负性的约束条件，建立了局部平均的非线性预测模型以及带约束的局部平均非线性预测模型，有效的改善了文献中一直以来的最佳预测---历史平均预测。同时，该研究还利用自助平均 (bootstrap aggregating or bagging) 的方法对约束预测进行模型平均，有效的降低了约束预测模型的预测均方差。对所提出的新的预测方法的理论性质，该研究进行了渐近理论推导，推导了新的估计量的渐近偏差、方差和均方误差对局部参数的依赖关系，发现在大多数局部参数的取值空间上，新的估计量在均方误差上都有更好的表现。实证分析表明，在样本外的预测精度上，新的预测估计量相比历史平均预测有较大改善。

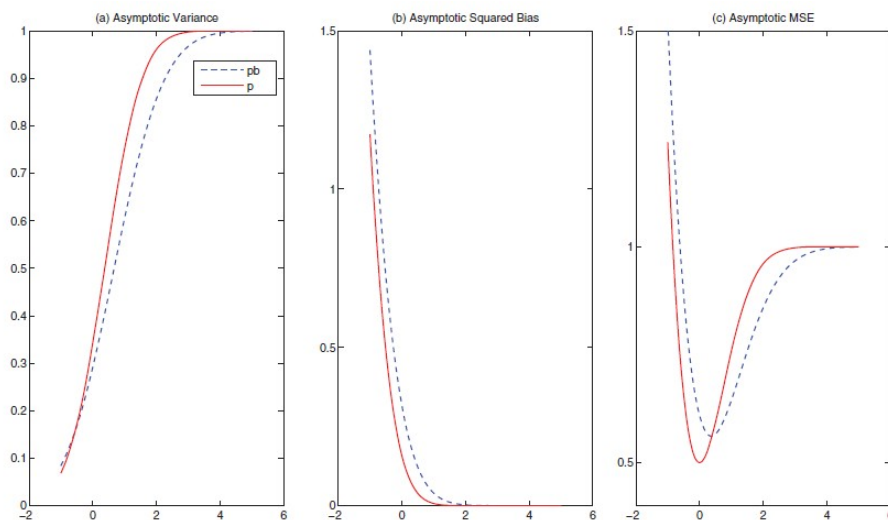
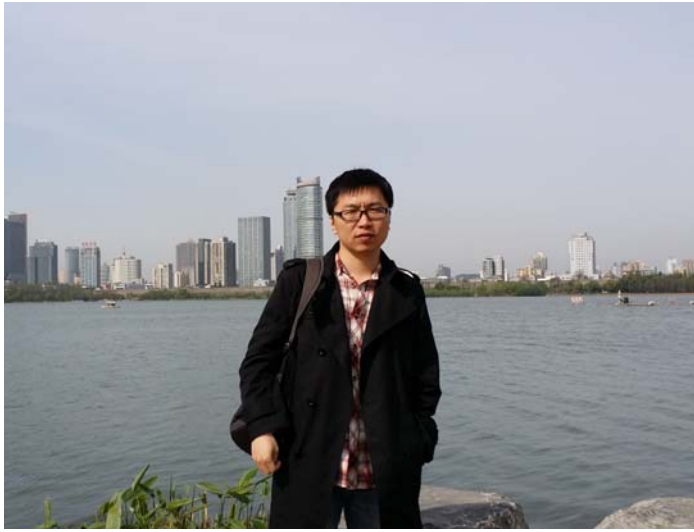


Figure 1. Asymptotic properties of the local historical average with positivity constraint (LHA_p , solid line) and the local historical average with positivity constraint and bagging (LHA_{pb} , dashed line).



席瑞斌研究员：2009 年在华盛顿大学获数学博士学位，2012 年加入北大统计中心和数学科学学院，2013 年入选“青年千人计划”。席瑞斌博士的研究方向是生物数据，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。

多功能干细胞技术：

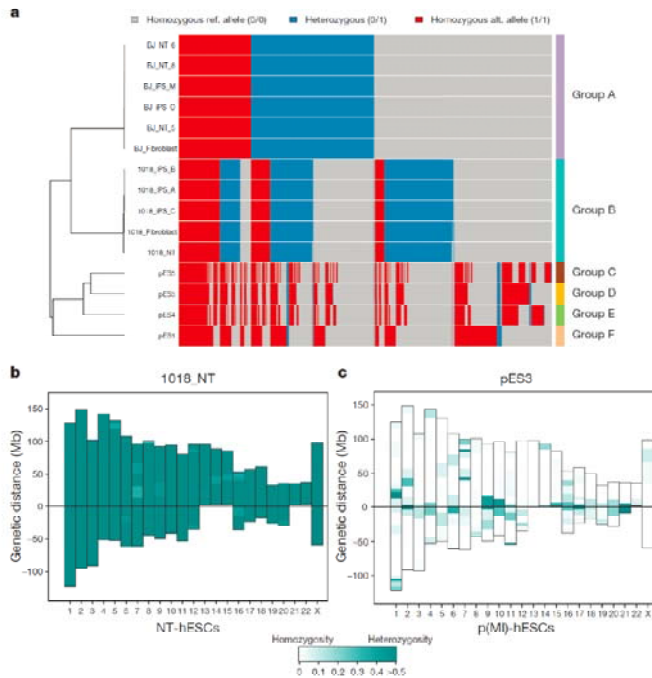


Figure 1 a. 通过分析诱导多功能细胞基因组及其来源体细胞基因组，可以排除其他细胞的污染；b,c 两种不同的 iPS 诱导技术所得到的多功能细胞具有不同的基因组特征。

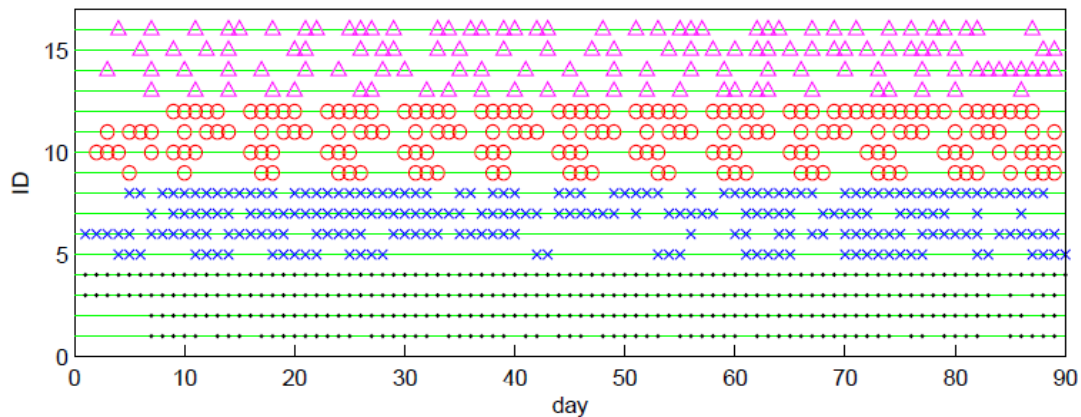
席瑞斌研究员和他的合作者在 Nature 上发文 (De Los Angeles et al. 2015 Nature) 研究了多种多功能干细胞技术。多功能干细胞技术是干细胞领域的一项重大突破，在生命科学的基础研究中占有重要的地位，有广泛的医学应用前景。然而，由于 2014 年日本理化研究所的小保方晴子涉嫌造假的事件后，如何确定通过多功能干细胞技术得到的细胞为多功能细胞引起了相关研究人员的重视。本文回顾了

多种多功能干细胞的验证方法，并提出了可以通过分析高通量测序数据的方法来确定所得到的细胞无其他细胞污染 (上图 a)。同时，由于不同技术得到的多功能细胞的基因组会有不同的特点，我们也可以通过分析高通量测序数据来确定所得到细胞确实具有有所该特点 (上图 b,c)。在图 a 中，我们可以清晰的看到由诱导多功能技术 (induced pluripotent stem cell 或 iPS 细胞) 及核转移技术 (nuclear transfer cell) 所生成的多功能细胞与其来源细胞有几乎一样的基因型 (如 BJ_NT_6 与其来源细胞 BJ_Fibroblast, 1018_iPS_A 与其来源细胞 1018_Fibroblas)，而由 parthenogenetic 技术产生的细胞 (如 pES1)，则由于 parthenogenetic 技术的特点，几次实验所产生多功能干细胞的基因型有较大不同。在图 c 中显示了由 parthenogenetic 技术产生的多功能干细胞的只在靠近 centromere 附件有同源 SNP，其余则大部分为同源 SNP，此特点与 parthenogenetic 技术点一直。相对的，由核转移技术生成的多功能干细胞则无此特点 (图 b)。



黄辉研究员：2010 年在马里兰大学获统计学博士学位，2012-2013 年在迈阿密大学管理科学系进行博士后研究工作，2014 年加入北大统计中心和数学学院，2015 年入选“青年千人计划”。黄辉博士的研究方向是函数型数据分析，纵向数据分析，空间统计学，时间序列与空间点模式分析。

多元随机场的联合统计建模：



具有时间或空间相依结构的数据在实际应用中很常见，例如纵向数据，时空数据，图谱数据等。把数据在时间或空间上的变化轨迹看成随机曲线（随机场），就可以利用函数型数据分析技术提取数据特征。上图展现的是可卡因成瘾者的毒品使用轨迹，每个时间点上的观测数据只有 0-1 数值，但其轨迹特征包含了很多重要信息，例如周期模式和高使用量模式等。同时，每个个体可能有多个相依的非高斯轨迹变量，其交叉相依结构也包含了很多数据信息。我们近期在 JASA 上的一篇文章尝试对这种多元、非高斯的随机过程进行联合统计建模和聚类分析。其基本思想是综合利用广义线形混合效应模型，函数主成分分解和样条函数逼近等方法，来达到数据降维和模型参数估计的目的。我们的方法，将观测到的非高斯随机过程与潜藏高斯过程联系起来，把现有的函数数据分析方法扩展到任意指数族分布上。这种针对非高斯多元随机曲线的一般性联合建模方法（joint modeling）具有很强的可扩展性，也可以广泛应用在图谱和大气污染等数据分析中。另一方面，对于类似行为模式这种观测较为规则的轨迹数据，还可以从点过程的角度出发将其看成复发事件（recurrent event），这样可以降低随机效应模型参数估计过程中多重积分所带来的高计算成本。针对上图中的可卡因使用轨迹数据，我们近期在 Statistics in Medicine 上的一篇文章提出了一种利用估计方程进行联合统计建模的方法，有效的提高了运算效率并减小了模型参数估计的偏差。



李程高级研究员：2013 年加入北大统计中心和生命科学学院。李程博士的研究方向是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。

三维基因组的实验和数据分析：

李程研究组目前做实验产生 Hi-C、RNA-seq、基因组测序数据来研究染色体三维构象(3D 基因组)，并结合其他基因组数据进行整合分析和数据建模，发现癌症和干细胞领域新的生物规律或药物靶点。

在许多肿瘤和非肿瘤的人类疾病中，染色体重排 (Chromosomal rearrangement, CR) 是重要的致病原因之一。CR 事件通过破坏基因和蛋白质结构导致相关疾病的发生。此外，它们还可以通过改变染色体三维结构和基因表达而导致疾病。在这项研究中，我们使用 Hi-C 数据和 ChIP-seq 数据筛选潜在的由于三维结构改变而致病的 CR。我们设计的算法可以发现已有实验验证的和疾病相关的 CRs (如多指疾病)，这些 CR 通过破坏染色体的三维结构导致基因的异常表达。我们还发现智力缺陷可能是一个三维染色体结构改变致病的候选疾病。我们开发了一个 Web 服务器 (3Disease Browser, <http://3dgb.cbi.pku.edu.cn/disease/>)，集成和可视化疾病相关的染色体重排事件和染色体的三维结构。我们从两个染色体拷贝数变异的骨髓瘤细胞系出发，分别探索拷贝数变异 (copy number variation) 对细胞三维结构以及表达量变化的影响。发现 CNV 变化可引起染色体三维结构的差异，并且 CNV 变化的边界大多位于 TAD 的边界。CNV 也会导致基因表达量的差异，同时 A、B compartment 的变化可以引起基因表达量的变化。这些发现对理解癌症的发病机理和揭示新药物靶点有重要意义

我们在生物信息平台建设方面的进展如下：

- 1、建立和完善了 RNA-seq 和甲基化测序的分析流程。完成了 20 余个生科院和生命科学联合中心科研课题中的生物信息分析工作。
- 2、在生科院/生命中心开展生物信息和测序数据分析的技术讲座，以帮助湿实验室师生了解和使用生物信息和统计数据分析方法，以及二代测序和相关分析软件。四次讲座及暑期培训班参与的师生超过 350 人次。
- 3、利用平台论坛网站 (www.iseq.me) 普及和讨论基因组学、生物信息学、统计学和科研方法的知识和技巧，目前已发布 250 个话题。



林伟研究员：2011 年在南加州大学获得应用数学（统计方向）博士学位，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 年加入北大统计中心和数学学院，2015 年入选“青年千人计划”。他的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。

高维工具变量回归的正则化方法：

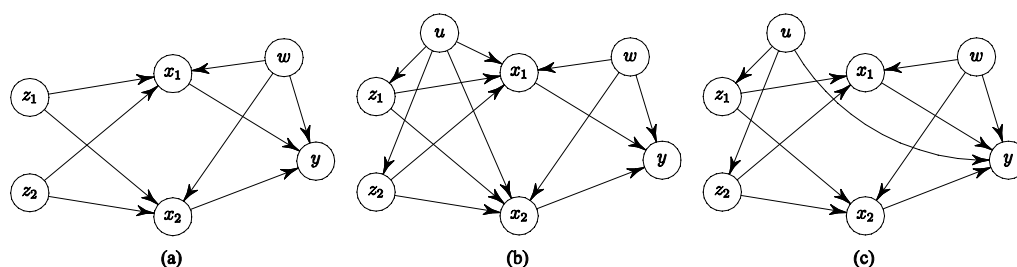


Figure 1. Causal diagrams showing the relationships between two genotypes z_1 and z_2 , two gene expression levels x_1 and x_2 , a clinical phenotype y , an unobserved phenotype w that confounds the associations between gene expression levels and the clinical phenotype, and an unobserved variable u representing possibly present population substructure. The population substructure (a) is not present, (b) affects genotypes and gene expression levels, or (c) affects genotypes and the clinical phenotype.

林伟研究员与合作者在 *Journal of the American Statistical Association* 上发表论文，提出了高维工具变量回归的一类两阶段正则化方法，为高维数据的因果推断提供了新的有力工具。这项研究工作的驱动问题来源于遗传基因组学。近年来，基因表达研究的深入进展导致了众多基因标记的发现，然而受制于混杂因素的影响，利用这些发现破解复杂遗传性状的遗传机制的研究却仍存在重重阻力。诸多因素，包括未能观测的变量、实验条件、环境扰动等，都可能对基因表达施加重要的影响，从而导致虚假的关联或对效应方向及大小的错误估计。图 1 是混杂变量与其他变量关系的一个图示。这一复杂性在高维基因组数据分析中更加突出和棘手。此外，遗传基因组数据集的广泛获取也对基因表达和遗传变异的联合分析提出了紧迫要求，以期对发现癌症、心血管疾病、肥胖等复杂疾病的致病基因提供关键指导。

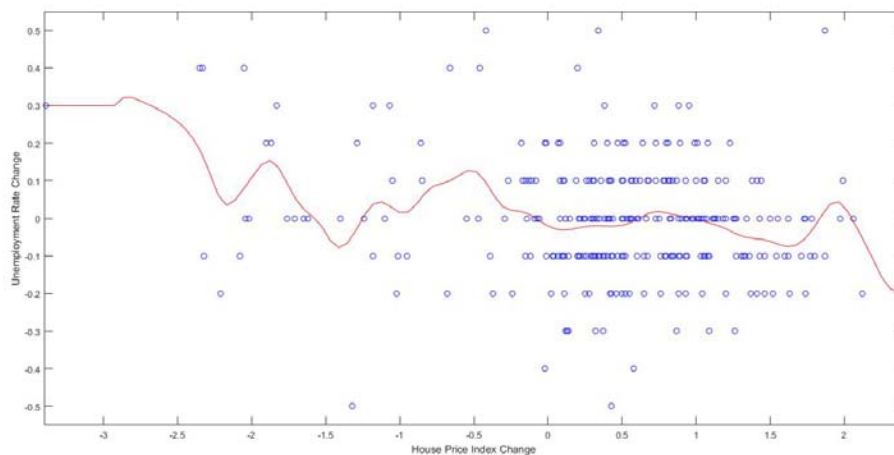
工具变量方法最早产生于计量经济学研究中并得到广泛应用，此前在生物统计中的应用较为有限，主要集中在流行病学研究中的孟德尔随机化问题。林伟研究组的论文指出，工具变量模型和方法也可以用于遗传基因组数据分析，为遗传变异、基因表达及临床性状的联合分析和混杂控制提供了一个理想框架。为了解决这一应用中协变量和工具变量的双高维度为变量选择和参数估计问题带来的方法论挑战，该论文发展了一类两阶段正则化方法，在选择与估计最优工具变量的同时，可以达到识别和估计重要协变量效应的目的。该方法可以看作经典的两阶段最小二乘法的一个重要高维推广，旨在利用稀疏性来改进两阶段的估计。该论文严格建立了这一新方法的理论基础，并通过数值实验验证了该方法的有效性。最后，该方法被用于一个小鼠肥胖数据集上，找到了若干可能导致肥胖的候选基因和遗传变异。



宋晓军研究员：2014年在西班牙马德里卡洛斯三世大学经济系获经济学博士，同年加入北大统计中心和光华管理学院。宋晓军博士的研究方向是计量经济学理论（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。

非参数模型残差的序列相关性检验：

众所周知，回归模型误差项的表现对回归函数拟合的效果是有很大影响的，如果随机误差项是独立同分布的，那么回归函数的估计和推断就可以不受误差自相关(*autocorrelation*)或者异方差(*heteroskedasticity*)的影响。然而，当随机误差项存在自相关或者异方差的情况下，回归函数的估计是一个较复杂的工作而且估计的有效性要大打折扣。如何检验误差存不存在自相关性或者异方差，或者检验误差是不是独立同分布(*independent and identically distributed, i.i.d.*)，对回归函数的估计效果以及接下来的统计推断都是至关重要的。我提出如何在非参数模型中检验误差项的独立同分布性质，即“A Nonparametric Test for Serial Independence of Errors in Nonparametric Models”。首先，传统的独立同分布检验通常是建立在参数回归模型残差(*parametric residuals*)的基础上，而非参数回归模型中关于此类的分析还很少见。其次，运用非参数模型可以避免模型误设的问题。最后，由于非参数残差(*nonparametric residuals*)的特殊性，在其基础上构建独立同分布检验所需要的分析工具也和传统参数模型不同。同时关于非参数回归模型中误差项的许多其它检验也是有待进一步研究的问题。检验非参数回归模型误差的独立同分布性质应用广泛。例如，在研究房价指数变动和失业率变动的问题上，很多人建议利用核回归方法来研究两个变量之间的动态关系，如下图所示。然而通过本文提出的检验发现，残差具有很强的相关性。这说明房价指数变动并不能很好的解释失业率变动，很容易推测出房价指数变动的滞后项也应该具有很强的解释作用。所以进一步的分析可以利用一个误差服从 p 阶自回归的非参数模型来解释两者的动态关系。



3. 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
陈大岳	副主任, 执行委员	马氏过程, 无穷粒子系统, 渗透模型, 随机游动	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
黄辉	中心全职教员	函数型数据分析, 空间统计学	数学科学学院概率统计系
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
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	执行委员, 政府统计工作室召集人	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

三、中心要闻

1. 陈松蹊教授团队研究报告深度解析五大城市 PM2.5 数据

作为中国政治经济发展中非常重要的五个大型城市，北京、上海、广州、成都和沈阳分别位于京津冀、长三角、珠三角、川渝和辽宁经济区，这五大区域的经济总量占据全国总量的 50% 以上，能源消耗量巨大，也是大气污染和雾霾天气灾害较为严重的区域。雾霾的主要成分 PM2.5 对人类健康有着极大的危害，也不同程度地影响着农业、生态、气候和居民的生活质量，因此研究中国这五个主要城市 PM2.5 的污染状况及其影响因素，将为中国大气污染的预防和治理提供重要的实证依据。

继 2015 年 3 月发布《空气质量评估报告：北京城区 2010-2014 年 PM2.5 污染状况研究》后，由陈松蹊教授率领的北大光华和北大统计科学中心团队，日前正式发布了《空气质量评估报告（二）：中国五城市空气污染状况之统计学分析》，报告一经发出，便引起了国内外媒体的关注，美国《纽约时报》（The New York Times）网站和国内《参考消息》先后对其进行了报道。

研究团队运用统计学方法交叉验证了美国使/领馆和邻近的环保部站点数据的可靠性，并结合天气条件、冬季供暖影响等因素对所获取的数据进行多方面、多层次的对比和分析，极大地保证了数据含金量，为公众准确认识和评估 PM2.5 严重性、找出空气污染原因和防治方法提供一个宝贵的、切实可行的方向标。

1) 五城市 PM2.5 污染状况对比及分析

空气质量统计评估发现，尽管五城市空气质量近年来有一定改善，最明显的改善是在 2015 年，但 PM2.5 水平还远远高于世界卫生组织所建议的上限，所以中国的大气污染防治需要更大的力度、更科学的方法。报告结合五个城市所在区域的能源消耗，研究了空气质量与能源消耗的关系，分析表明上海所在的长江三角洲地区能源消耗最多，广州在某些能源消耗指标上也排在第二、三位，四川消耗了最少的能源，但成都却是空气质量最差的城市之一。这说明，要保证良好的空气质量，不同城市和地区，有不同的能源消耗和工业活动承载的上限，而这个上限很大程度上受该地区的地理环境和气象条件影响。

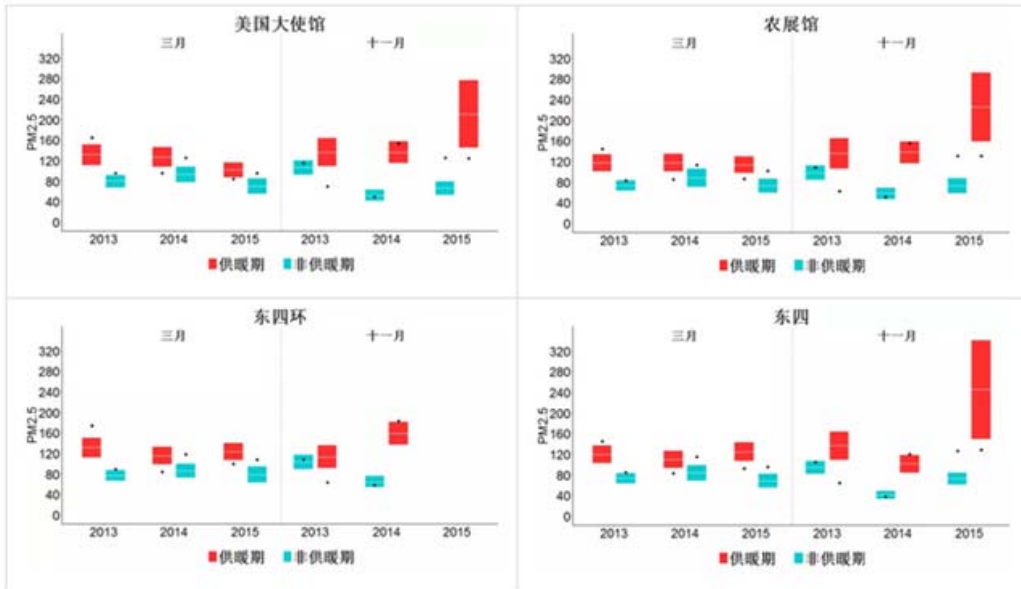
表 1：不同度量指标下五城市 PM_{2.5} 污染状况的排名

指标排名	污染状态占比	严重污染占比	优良空气占比	污染平均持续时间 (小时)
1 (最差)	成都 (0.88)	北京 (0.22)	成都 (0.12)	成都 (92.4)
2	沈阳 (0.78)	沈阳 (0.11)	沈阳 (0.22)	北京 (67.9)
3	北京 (0.76)	成都 (0.09)	北京 (0.24)	广州 (62.5)
4	广州 (0.68)	上海 (0.04)	广州 (0.32)	沈阳 (52.4)
5 (最好)	上海 (0.63)	广州 (0.02)	上海 (0.37)	上海 (52.3)

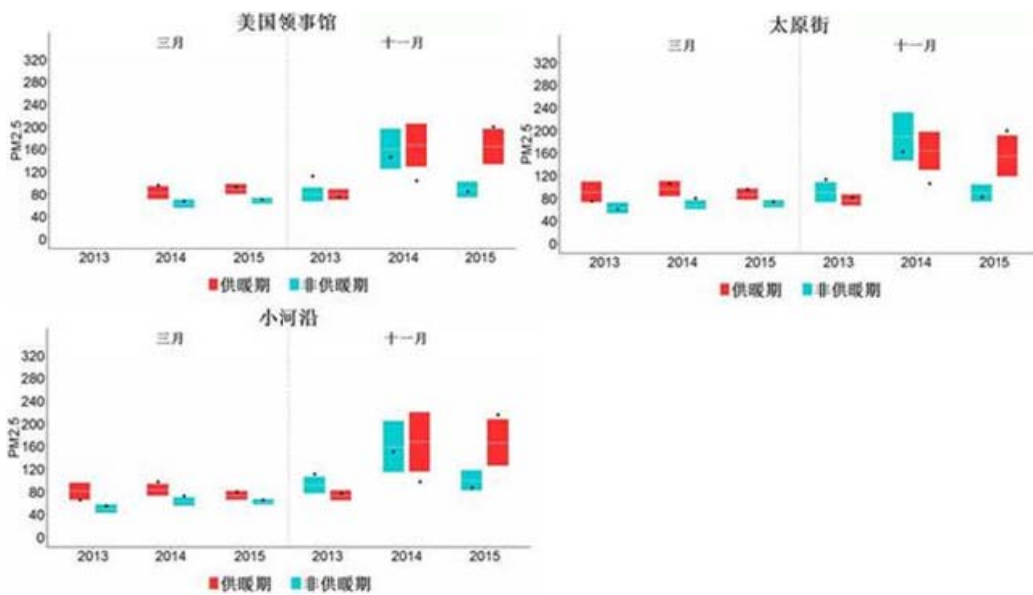
注：“优良空气”为 PM_{2.5} 小于或等于 35 微克/立方米，“污染状态”为 PM_{2.5} 大于 35 微克/立方米，“严重污染”为 PM_{2.5} 大于 150 微克/立方米。

2) 冬季供暖对 PM2.5 的影响

研究团队还利用了北京和沈阳两个城市的数据，量化和评估城市冬季供暖对 PM2.5 的影响。通过统计分析发现，北京和沈阳具有非常显著的冬季供暖效应。冬季的大气污染防治是中国北方地区面临的巨大挑战，解决这一问题的关键是大大提高天然气等低排放能源的使用，替代、减少煤炭的消耗，包括使用经济手段极大的抑制煤炭的消耗。



北京冬季供暖效应



沈阳冬季供暖效应

3) 官方数据可靠性深度验证

为了交叉验证数据的可靠性, 报告分别使用美国使/领馆和邻近环保部站点两个独立数据源的 PM2.5 数据, 分析和比较了描述大气污染不同状态的统计量的一致性和相关系数。同时, 在剔除气象因素对数据造成的影响后, 使用两个数据源的数据分别计算出各站点的 PM2.5 的月均值和百分位数浓度。

通过多方面、多层次的相互对比和相关性分析, 研究团队发现, 使用这两个数据源所得到的空气质量评价, 有着比较高的一致性。这可以在一定程度上, 为公众对中国主要城市空气污染数据质量的评估提供一个参考依据。需要强调的是, 报告的分析结果, 并不能说明在其它地区的数据也是可靠的。其它城市的数据可靠与否, 同样需要对其数据进行系统分析后才能得出结论。

2. 2016 年北京大学许宝騄讲座

4 月 7 日, 浙江大学数学学院林正炎教授在理科一号楼 1114 教室举办了主题为“弱收敛”的学术讲座。



下午 3 时, 理科一号楼 1114 报告厅中, 北大师生济济一堂。讲座在陈大岳教授的主持下拉开帷幕。林教授的讲座从随机变量序列的弱收敛性讲到随机过程(元)序列的弱收敛。对于前者, 林教授从随机变量序列的各种收敛性引入, 主要介绍依分布收敛(弱收敛)中的中心极限定理成立的充要条件及证明概要。关于充分性的证明, 林教授先提了一下特征函数法, 接下来简要介绍 Lindeberg 替换法和 Stein 方法。通过必要性证明的简述, 试图介绍概率极限理论的几种常用手段。至于随机过程序列的弱收敛理论, 包括经典和现代两部分。前者林教授从概率测度弱收敛引入, 介绍随机过程的依分布收敛(弱收敛)及 Donsker 不变原理, 对于现代理论林教授只简要介绍了鞅方法。

林正炎教授, 国际数理统计研究院(IMS) Fellow。长期从事概率统计的教学和研究, 在概率极限理论和统计大样本理论的主要方向都有系统的工作。已由 Springer 出版社、科学出版社等出版专著九部。在国内外著名期刊发表论文 230 余篇, 研究成果多次被国内外同行引用。曾获国家自然科学基金三等奖(1997)、教育部科技进步二等奖三项, 浙江省政府、教委、教委科技进步奖等多项奖励。是国家级有突出贡献中青年专家和浙江省特级专家。被授予全国“五一”劳动奖章。已培养博士 32 名、硕士近百名, 是国家百篇优秀博士论文的导师。学生中有些已在国际学界有较大的影响。曾被评为首届国家级教学名师、首届浙江省功勋教师, 获浙江省优秀教学成果一等奖、教育部优秀教材一、二等奖。宝钢特等教学奖。曾任中国概率统计学会副理事长, 现任浙江省数学会理事长、《高校应用数学学报》合作主编。

3. 美国使馆环境科技卫生处官员访问陈松蹊教授



4月19日下午美国使馆环境科技卫生处负责环境和空气污染问题的官员 Stefan Whitney 先生及其助手刘远女士前来北京大学统计科学中心访问陈松蹊教授。

Stefan Whitney 先生和刘远女士与陈松蹊教授的研究团队共同交流了近日发布的《空气质量评估报告》-中国五城市空气污染状况之统计学分析,探讨了中美如何在空气污染这一领域开展进一步的合作等问题。

4. 2016年“两岸经济对话”在台湾“中央研究院”顺利举行



日前,“两岸经济对话”在台湾“中央研究院”经济所拉开帷幕。此次对话是北京大学光华管理学院和台湾“中央研究院”共同举办的第三次两岸经济学学术研讨交流会。

北京大学光华管理学院商务统计与经济计量系联合系主任、统计科学中心联席主任陈松蹊教授,北大光华应用经济系主任周黎安教授,以及商务统计与经济计量系虞吉海教授、涂云东助理教授和宋晓军助理教授应邀参加了本次学术研讨活动,并报告了各自的最新学术成果。台湾“中央研究院经济所”所长简锦汉教授,周雨田教授,萧代基教授,蔡文祯教授,陈宜廷教授,许育进副教授等参与了研讨并作学术报告。研讨会在欢快而严谨的氛围中进行,与会人对研讨内容积极发言,会后纷纷表示收获颇丰。

随后,陈松蹊教授一行还应邀前往台湾“中华经济研究院”进行学术访问和参观,该学院院长吴中书教授亲切会见,并介绍了台湾在宏观经济建模和预测方面的方法和宝贵经验。吴院长深切关注大陆宏观经济建模和预测研究进展,强调定量分析宏观数据的重要性,并希望台湾经验能够对大陆经济的建模提供参考和借鉴。

“两岸经济对话”系列活动由芝加哥大学商学院教授、北京大学名誉教授刁锦寰先生发起,由陈松蹊教授和周雨田教授共同组织,集中研讨交流两岸经济学研究的最新问题,针对两岸经济建模和预测等问题深入探讨。前两次对话分别于2014年2月和2015年5月在台湾“中央研究院”和北京大学举行。

5. 大数据时代下的高维统计建模与分析研讨会

2016年5月17日至18日,“大数据时代下的高维统计建模与分析研讨会”在北京大学光华管理学院2号楼阿里巴巴报告厅成功举办。本次会议由北京大学统计科学中心和北京大学数量经济与数据金融教育部重点实验室联合主办、北京大学光华管理学院承办。会议旨在凝聚国内外高维统计分析领域的专家学者,提供交流合作的研究平台,促进统计学特别是大数据分析和高维统计研究的发展。

5月17日上午的开幕式由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主持。陈松蹊教授强调了大数据时代的到来给统计学提出的新的挑战与机遇,以及统计学与数学、信息科学和各个应用领域联系交叉的重要意义。随后,中国科学院院士马志明教授与北京大学光华管理学院院长蔡洪滨教授分别发表了开幕致辞。马志明院士肯定了北京大学统计科学中心在学术科研和社会贡献方面取得的巨大成果,并希望作为由北京大学光华管理学院和数学学院跨学院交叉建立的一个高端研究中心,继续脚踏实地、顶天立地,在学术研究和统计应用两个方面越做越好。蔡洪滨对参会者来到光华切磋学术、交流思想表示了热烈的欢迎,对大会组织者的辛勤付出表达了感谢。他对商务统计系在学术研究、人才培养等方面取得的成就以及对国家发展和光华管理学院建设方面作出的贡献表示了肯定。蔡洪滨表示此次会议的主题与学院一直以来所坚持的学术与现实相结合的导向相契合,并预祝会议圆满成功。



来自高维统计领域的国际顶尖专家,中国科学院院士、西安交通大学徐宗本教授, COPSS 奖获得者、普林斯顿大学范剑青教授, COPSS 奖获得者、宾夕法尼亚大学蔡天文教授出席会议并发表主旨演讲。此外,来自北京大学、耶鲁大学、明尼苏达大学、墨尔本大学、西安交通大学、南开大学、中国人民大学等国内外多所高校的多位相关领域的著名专家、优秀青年学者也进行了学术报告,报告内容涉及高维统计中的参数估计、假设检验、计算方法、统计建模等方向,广泛介绍了高维统计领域的最新前沿成果。近 300 名来自国内高校的青年统计学者和研究生参加了此次会议。

在 5 月 17 日的报告中,共有 12 位学者向与会者展示了他们在高维统计方向的最新研究成果。徐宗本院士介绍了一种基于 ADMM 和深度学习的大数据算法新框架及其应用;范剑青教授分享了高维统计中控制错误发现的新方法;宾夕法尼亚州立大学的李润泽教授展示了

他在高维协方差矩阵线性结构方面的最新检验理论；西南财经大学的常晋源教授详细讲解了高维白噪声序列的检验方法；北京大学的林伟教授分享了他在高维成分数据的两样本均值检验方面的最新成果；蔡天文教授介绍了一种两样本超高维稀疏均值检验的问题；南开大学的王兆军教授介绍了高维复杂数据的统计推断与在线监控；北京大学的王汉生教授展示了他在网络数据向量自回归方法上的创新；耶鲁大学的周慧斌教授介绍了网络数据分析中的一些最优性结果；西安交通大学的付利亚教授分享了她在纵向数据的稳健回归方面取得的成果；中国科学院和上海财经大学的周勇教授展示了长度偏差右删失数据的半参数统计分析；北京大学的涂云东教授分享了他在误差修正因子模型方面取得的最新研究成果。

5月18日，专家们进一步与大家分享高维统计的学术前沿。耶鲁大学和中山大学的张和平教授比较了含讨厌参数的带有惩罚项的回归法、条件距离相关方法和他提出的 pLASSO 方法在信息不对称情形下假设检验的表现；北京大学的杨静平教授介绍了复合伯恩斯坦 copula 的相关理论；中国人民大学的朱利平教授提出了一种改进的向前回归方法；明尼苏达大学的邹晖教授讲解了在变换正态模型中的多任务分位数回归方法；北京大学的席瑞斌教授介绍了他使用大数据分析策略在癌症基因组的拷贝数变异检测中取得的最新研究成果；西安交通大学的孟德宇教授着重谈了自定进度学习算法；东北师范大学的胡江博士生详细展示了他和白志东教授合作完成的大维架构下 AIC、BIC 准则的强相合性的研究；北京师范大学的梁宝生博士生分享了他和童行伟教授提出的对删失数据的非参数风险函数的有效估计；北京大学的陈松蹊教授讲解了在高维协方差矩阵估计的窗宽选择方面的最新研究结果。

最后，陈松蹊教授和蔡天文教授对大会作了总结发言。他们对各位报告人的精彩演讲表示感谢，指出高维统计分析和大数据分析在统计理论研究和实际应用方面具有举足轻重的地位，并对后续相关会议的举办和课程的开设表示期待。在为期两天的会议中，与会者对大数据背景下的高维统计理论与应用进行了深入探讨，充分展示了高维统计在理论、算法、应用等方面的最新国际学术前沿成果，使得大数据背景下的统计学与数学、信息科学、经济管理等应用学科的联系更加紧密。此次会议将对推动统计学大数据方向的发展、跨学科的交流、人才的培养起到积极的作用。

6. 博士在读研究生梁萱获得北京大学 2015-2016 学年度才斋奖学金

近期，北京大学光华管理学院商务统计与经济计量系、北京大学统计科学中心联合培养的博士研究生梁萱获得了“北京大学 2015-2016 学年度才斋奖学金”。她申请的课题名称为《五城市及华北平原空气污染状况的科学评估与建模》，项目编号为 CZ201507。

北京大学“才斋奖学金”是由青鸟集团公司捐赠的用于资助北京大学优秀博士生从事高水平、有难度的创新性研究而设立的专项奖学金。本年度在全校共资助了 14 个项目。

梁萱同学师从北京大学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、国家“千人计划”专家陈松蹊讲席教授。在读期间，她和陈松蹊教授率领的环境统计研究团队在《英国皇家学会会刊》(Proceedings of the Royal Society A) 发表论文一篇，并共同撰写了两篇有关“空气质量评估”的中文研究报告。

7. 统计科学中心成功举办 2016 年暑期夏令营

2016年7月8日至10日,北京大学统计科学中心举办了第三届优秀大学生夏令营活动,来自中国人民大学、南开大学、北京师范大学、上海财经大学、中山大学等十几所高校的31名三年级本科生参加了本次夏令营。

此次夏令营的活动包括科研文献阅读理解、笔试与面试考核、统计中心老师介绍各课题组的科研情况、师生聚餐、夏令营学员的英文自我介绍等环节。统计中心联席主任陈松蹊老师在开幕式上给参加夏令营的同学介绍了统计科学中心的基本情况,并做了关于空气质量统计建模评估的特邀科研报告,引起了学员们的浓厚兴趣。



陈松蹊老师,贾金柱老师和李程老师分别对光华管理学院,数学科学学院以及生命科学学院做了简单介绍;统计科学中心的其他七位全职教员也分别介绍了自己的科研方向及教学成果,并与参加夏令营的同学进行了深入的对话和交流,让学员们对统计科学中心和国际前沿的统计学研究方向有了更加全面的了解。在这次夏令营中,统计科学中心的老师选拔了表现突出的四名学员,并推荐保送北京大学直博研究生项目。

8. 第一届北大-清华统计论坛

2016年12月16日,第一届北大-清华统计论坛在北京大学国际数学中心甲乙丙楼多功能会议室成功举办。本届论坛由北京大学统计科学中心、清华大学统计学研究中心、北京大学数学科学学院、北京国际数学中心联合举办。论坛上,北京大学统计科学中心科学委员会主席、美国加州伯克利大学统计系郁彬教授和清华大学统计学研究中心主任、美国哈佛大学统计系刘军教授做大会报告,百余名来自北大清华及其他院校的师生参加论坛。

下午 2 时许，论坛在北京大学统计科学中心贾金柱老师的主持下开始，加州伯克利大学郁彬教授首先做题为“人工神经网络遭遇真实神经网络：V4 神经区域模式选择”的大会报告。郁彬教授是美国国家科学院和美国艺术与科学学院两院院士、北京大学统计科学中心科学委员会主席、北京大学微软统计与信息技术教育部-微软重点实验室的创办者及联席主任。她与基因组学、神经科学、医学领域科学家合作进行跨学科研究，开发了统计和机器学习方法、算法和理论，并与领域知识以及量化批判思维结合以解决这些领域中的数据问题。郁彬教授在报告中指出，人类与其他灵长类动物的视觉成像都由大脑中一系列的神经元组成的区域控制，其中一个重要的区域便是 V4 区域，由于 V4 神经元有高度非线性响应的特征，一直以来，科研工作者都很难准确描述视觉信息在 V4 区域的表达。郁彬教授的科研团队利用卷积神经网络模型 (CNN) 将经过变换的自然图像拟合 V4 区域 71 条独立神经元的 spike rates。为了得到每个神经元的模型，郁彬教授首先将模拟图像传递给卷积神经网络内层，将内层的活动作为高维回归的预测向量，将 V4 神经元的应答率作为响应向量。因此，每个神经元的最终模型都包含一个由 CNN 提供的多层非线性变换、一个由回归模型提供的权重线性层。研究发现，使用 CNN 模型前两层获得的模型比加博尔小波模型得到的结果预测效果更好。为了刻画每个 V4 神经元的空间和模式的选择性，郁教授团队精确优化了使得预测 spike rates 最大化的输入图像，并检验的模型的稳定性。研究结果显示，V4 神经元主要识别曲线、斑点、棋盘格等形状的图片。



下午 3 时 30 分，哈佛大学统计系刘军教授做题为“一种诊断相关性的广义 R 平方”的大会报告。刘军教授是清华大学“千人计划”讲座教授，清华大学统计学研究中心主任，是世界生物统计和生物信息学领域的著名专家。刘军教授实验室是转录因子-DNA 序列结合位点的预测课题的先驱，在贝叶斯方法、蒙特卡罗方法、生物信息学、遗传学等领域做出了一

系列奠基性工作。对统计理论、复杂系统优化、基因组学、信号处理等领域产生了非常深远的影响。

刘教授的报告从统计学最基础的诊断两随机变量相关性问题的展开，他指出，机关皮尔森相关系数在推断线性相关性上是有效的，但在面临非线性问题时便束手无策，刘军教授提出一个新的评价标准——G 平方——来检验两个随机变量的独立性并刻画两者关系。G 平方几乎与皮尔森相关系数的平方（即传统的 R 平方）一致，其意义也与 R 平方刻画两随机变量之间的关系一样具有直观性，但是在处理非线性与异方差问题时具有较高的效率。刘军教授详细介绍了 R 平方的两种估计量，并证明了他们的一致性。模拟结果显示，G 平方估计量是众多刻画变量相关性的统计量中最有效的一种。

在互动环节，多名到场师生与郁彬教授、刘军教授进行交流，就报告中的细节问题展开提问，会场气氛热烈。通过本次论坛，北京大学、清华大学等高校从事统计学研究的师生之间增进了解，深入沟通，在统计学科事业上取得了互相增益的效果。

9. 中心教员博士毕业生去向

姓名	专业	工作单位	指导教师
姜博川	概率论与数理统计	解放军装备发展部陆军装备体系研究重点实验室	艾明要
蒋智超	概率论与数理统计	普林斯顿大学博士后	耿直
董晶	统计学	中国工商银行总行读博士后	张俊妮
李硕	统计学	天津财经大学助理教授	陈松蹊、涂云东
何婧	统计学	西南财经大学助理教授	陈松蹊
邹韬	统计学	澳洲国立大学助理教授	陈松蹊

四、交流访问

1. Short course: Causal inference based on counterfactuals

哥本哈根大学生物统计部副教授 Theis Lange 博士于 2 月 22 日至 3 月 14 日在北京大学开设题为“Causal inference based on counterfactuals”的短期课程，该课程的学习目标主要为以下几个部分：1) 认识因果问题，深刻理解社会科学和健康科学中因果结论背后的假设；2) 利用有向非循环图（DAGs）方法讨论因果假设，根据提出的概念模型识别混淆变量，并使用虚拟变量标准化 DAGs；3) 理解逆概率加权和 G 估计方法，并讨论每种方法应用于实际数据集中的优势和局限性；4) 基于嵌套虚拟变量，学习并理解中介变量分析。



本次短期课程吸引了来自校内外百余人前来学习，课堂气氛活跃，前来参加学习的学生受益匪浅。

2. “高维统计与网络分析”夏季短期课程顺利完成



2016 年 6 月 23-26 日，由北京大学统计科学中心组织的“高维统计与网络分析”夏季短期课程在北京大学光华 1 号楼成功举办。本次短期课程旨在展示统计与数据科学的最新前沿进展，促进国内外学者的交流合作，培养具有国际视野、创新精神的高素质人才。本次课程共开设“高维统计”与“网络分析”两个主题，分别由 Rutgers University 的张存惠教授和 University of Michigan 的朱冀教授主讲。

Michigan 的朱冀教授主讲。

23、25 日，由 University of Michigan 朱冀教授主讲的“统计网络分析”课程开讲。朱冀教授 1996 年在北京大学获物理学硕士学位，2003 年在斯坦福大学获统计学博士学位，现任美国密歇根大学统计系教授，是机器学习与高维数据分析领域的知名学者，他曾在国际主流学术刊物上共发表 70 多篇学术论文，担任包括国际统计学顶尖刊物《Journal of the American Statistical Association》、《Biometrika》在内的多个期刊副主编。朱冀教授在课程中指出，网络数据是当前大数据时代下复杂结构数据中重要的一类数据，它广泛存在于生物学、物理学、经济学、计算机科学、社会科学等各个领域。朱教授通过两天的课程，向大家介绍了针对网络数据进行统计分析的主要前沿方法及应用。结合随机图模型、谱方法、随机矩阵理论等统计学方法，我们可以更高效准确地分析与解释网络结构数据。

24、26日，由 Rutgers University 的张存惠教授主讲的“高维数据的统计推断”课程开讲。张存惠教授是 Rutgers University 统计和生物统计系特聘教授，他是 IMS 和美国统计协会的研究员，主要研究方向为：高维数据、经验贝叶斯方法、半参数与非参数方法、生存分析、网络数据等。张教授在课程中首先向大家介绍了高维数据广泛的应用领域，包括生物信息学、互联网、资本市场、信号过程、神经成像等学科都是产生高维数据的来源。高维数据



的统计推断作为数据科学的核心理论基础，在大数据研究中占据着不可或缺的地位，但目前高维数据带给研究人员的主要挑战是，数据维数的急剧增长甚至增长超过了样本数量，如果继续沿用传统统计方法到高维领域将不能带来有效的推断结论，张教授在课程中向大家介绍了包括补偿最小二乘、变量选择、偏差矫正、线性回归中的 bootstrap 方法、图模型、copula 模型等在高维领域被广泛应用的统计方法，使大家充分了解了高维分析中最前沿的统计方法。

本次短期课程吸引了百余名研究生、博士后、青年教师及数据科学从业人员参加，大家在为期 4 天的学习中与两位教授在交流想法，碰撞观点，收获了高维统计与网络分析领域国际最前沿的理论与方法，为自己的科研与工作提供了帮助。

3. 2016 北京大学统计科学中心冬季短期课程



2016 年 12 月 12 日至 2017 年 01 月 05 日韩放教授在统计科学中心开设基础的机器（统计）学习理论。课程内容包括 PAC learning model, concentration inequalities, VC dimension, generalization bounds, 以及如何运用这些理论工具分析一些重要的机器学习方法（SVM, kernel method, boosting, etc.）。

韩放教授是华盛顿大学的助理教授是他的主要研究兴趣包括统计学，机器学习以及他们在复杂数据分析上的应用。他所开发的研究工具属于高维统计学，鲁棒性统计学，非参数和半参数回归模型（nonparametric 以及时间序列分析（time series analysis））。

该课程吸引了来自国内的百余位师生参加，大家在学习中积极与韩放教授交流想法，收获了机器学习的理论与方法，为自己的科研与工作提供了帮助。

4. 哥本哈根大学副教授 Theis Lange 到访统计科学中心



2016.1.1 日至 2016.12.31 Theis Lange 博士在北大统计科学中心进行访问。Theis Lange 博士 2008 年毕业于哥本哈根大学，此后从事生物统计、流行病学、计量经济学等领域的研究工作，有许多优秀的研究成果。自 2008 年至今共发表学术论文 60 余篇，现为哥本哈根大学公共卫生学院副教授。Theis Lange 博士曾于 2015 年 12 月到访统计中心并做学术报告。

5. 新西兰统计局资深研究员 John Bryant 访问统计科学中心

2016 年 10 月 2 日至 10 月 29 日，新西兰统计局资深研究员 John Bryant 访问了统计科学中心。John Bryant 是人口学博士，在新西兰统计局推广用贝叶斯方法估计和预测小区域特征，在人口学界具有国际知名度。



小区域是指按照性别、年龄和地区等因素划分的子群体。估计和预测小区域特征而不仅仅是总体特征，能更切合政策制定者和行业决策者的需求。对于总体特征而言，样本通常很大因而可以进行直接估计；但是对于比较细致的小区域，样本量就比较小，所以无法直接采用属于这些小区域的样本对其特征进行精确的估计。需要通过统计建模从其他区域或时期的数据借力才能更好地估计小区域特征。

John Bryant 之前曾多次访问统计科学中心并与张俊妮教授合作。他们发展了一套小区域估计和预测的贝叶斯框架，能够以一致连贯的方式应对不确定性和缺失值，很好地结合不同来源的数据，并很好地引入来自专家判断的信息。此次访问中，John Bryant 与张俊妮教授合作，详细讨论了他们正在合作撰写的《Bayesian Demographic Estimation and Forecasting》一书（将由 CRC Press 出版）的整体框架，并进行了一些章节的写作。他们还与统计科学中心的博士生王菲菲合作，研究了如何使用贝叶斯非参数模型刻画小区域划分因素之间复杂并随时间变化的交互效应，并完成了相关程序。

John Bryant 的访问进一步加深了他与统计科学中心的合作，也体现了统计科学中心对国际合作研究的大力支持。

6. 统计中心 2016 年来访人员名单

2016 年统计科学中心共接待海内外访问学者 56 次，参与学术报告，短期课程以及合作项目等各类学术活动。

来访人员	时间
郭小波博士, 中山大学	2016-01-13/15
刘荣丽副教授, 南京大学	2016-03-07
Haiqiang Chen, Xiamen University	2016-03-23/25
薛晓峰博士, 中国科学院大学	2016-03-28
Yanchu Liu, Sun Yat-Sen University	2016-03-31/04-1
Ke Wu, Renmin University of China	2016-04-05
林正炎教授, 浙江大学	2016-04-07
林正炎教授, 浙江大学	2016-04-11
Qiwei Yao, London School of Economics	2016-04-14
Prof. V. Vatutin, Steklov Mathematical Institute, Moscow, Russia	2016-04-18
范青亮, 厦门大学	2016-04-28
Alan Welsh, The Australian National University	2016-04-30/05-13
Haoxi Yang, Nankai University	2016-05-05
向绪言教授, 湖南文理学院	2016-05-09
郭旭, 南京航空航天大学	2016-05-11/13
庄额嘉, Wuhan University	2016-05-18/20
Xiaogu Zheng, Beijing Normal University	2016-05-19
刘伟教授, 江苏师范大学	2016-05-22/24
唐继军教授, 天津大学计算机学院	2016-05-25
Ping Yu, The University of Hong Kong	2016-05-25/27
Bin Yu, Chancellor's Professor of Statistics and EECS, UC Berkeley	2016-05-26
陈昆, 康涅狄格大学	2016-05-27
Ji-Liang Shiu, Renmin University of China	2016-05-30
Nianqing Liu, Shanghai University of Finance and Econom	2016-06-01/03
Marc G. Genton, King Abdullah University of Science and Technology (KAUST)	2016-06-02
Ying Sun, King Abdullah University of Science and Technology (KAUST)	2016-06-02

Myoung-jae Lee, Korea University	2016-06-03
王一早博士, University of Cincinnati	2016-06-06
Lijun Bo, University of Science and Technology of China	2016-06-07/09
Hans-Juergen Lomp, Boehringer Ingelheim	2016-06-17
Jeff Wu, Georgia Institute of Technology	2016-06-22
Hongzhe Li, University of Pennsylvania Perelman School of Medicine	2016-06-28
Prof. Iuliana Ionita-Laza, Columbia University	2016-06-30
Mike Baiocchi, Stanford School of Medicine	2016-06-30
Peter M. Robinson, London School of Economics	2016-07-01
Prof. Lawrence D. Brown, University of Pennsylvania	2016-07-06
Qingyang Zhang, University of Arkansas	2016-07-07
Yichuan Zhao, Georgia State University	2016-07-07
Shujie Ma, University of California, Riverside	2016-09-08
Cheng-Der Fuh, National Central University	2016-09-14
Zongming Ma, University of Pennsylvania	2016-10-19
Jintai Lin, School of Physics, PKU	2016-10-20
Chun-houh Chen, Institute of Statistical Science, Academia Sinica	2016-10-20
John Bryant, Senior Researcher, Statistics New Zealand	2016-10-02/29
Per Johansson, Professor in statistics, Department of statistics, Uppsala University	2016-10-28
Diane Donovan, The University of Queensland	2016-11-01/03
Xingguo Luo, Zhejiang University	2016-11-02/04
薛原, 对外经济贸易大学	2016-11-17
Hongbiao Zhao, Xiamen University	2016-11-23/25
Ping Ma, University of Georgia	2016-12-14
Fang Han, University of Washington	2016-12-12/2016-01-17
方方, 华东师范大学	2016-12-14/15
Peng Ding, University of California, Berkeley	2016-12-15
高健博士, 中国环境科学研究院	2016-12-23
Johannes Schmidt-Hieber, Leiden University	2016-12-28/30
Jingshu Wang, University of Pennsylvania	2016-12-29/31

五、论文发表

自 2016 年 1 月至 12 月, 统计科学中心教员共发表论文 57 篇,按第一作者姓氏字母排序排列:

[1] **Ai, Mingyao**; Kong, Xiangshun; Li, Kang (2016). A general theory for orthogonal array based Latin hypercube designs. *Statistica Sinica*, 26(2), 761-777.

[2] **Ai, Mingyao**; Liang, Jiajuan; Tang, Man-Lai (2016). A generalized T3 graph for testing highdimensional normality. *Frontiers of Mathematics in China*, 11(6), 1363-1378.

[3] Bryant, J., **Zhang, Junni** (2016). "Bayesian Forecasting of Demographic Rates for Small Areas: Emigration Rates by Age, Sex, and Region in New Zealand, 2014-2038," *Statistica Sinica*, 26, 1337-1363.

[4] **Chen Song Xi**, Lei Lihua and **Tu Yundong** (2016). "Functional Coefficient Moving Average Model with Applications to Forecasting Chinese CPI," *Statistica Sinica*, 26, 1649-1672, October 2016.

[5] Chen, H., Ding, P., **Geng, Z.** and Zhou, X. H. (2016). Semiparametric Inference of the Complier Average Causal Effect with Nonignorable Missing Outcomes. *ACM Transactions on Intelligent Systems and Technology*, 7(2), 19:1-19:15.

[6] **Chen, S.X.** (2016). Peter Hall's Contribution to the Bootstrap, *The Annals of Statistics*, 44, No. 5, 1821–1836.

[7] **Chen, S.X.**, Lei, L.-H. and **Tu, Y-D** (2016). Functional Coefficient Moving Average Models with applications to forecasting Chinese CPI, *Statistica Sinica*, 26, 1649-1672.

[8] Chen, Zhen-Qing; **Ren, Yan-Xia**; Yang, Ting (2016). Boundary Harnack principle and gradient estimates for fractional Laplacian perturbed by non-local operators. *Potential Anal.* 45, no. 3, 509–537.

[9] Chen, Zhijin; **Yang, Jingping**; Wang, Xiaoqian (2016). Pricing k(th) realization derivatives and collateralized debt obligation with multivariate Frchet copula. *Frontiers Of Mathematics In China*, Volume: 11 Issue: 6 Pages: 1419-1450.

[10] Chu, Y., Dai, L., Qi, S., Smith, M. L., **Huang, H.**, Li, Y. and Shen, Y. (2016). Challenges from variation across regions in cost effectiveness analysis in multi-regional clinical trials, *Pharmaceutical Medicine and Outcomes Research*, 7:371.

- [11] Engländer, János; **Ren, Yan-Xia**; Song, Renming (2016). Weak extinction versus global exponential growth of total mass for superdiffusions. *Ann. Inst. Henri Poincaré Probab. Stat.* 52, no. 1, 448–482.
- [12] Guo, B. and **Chen S.X.** (2016). Tests for High Dimensional Generalized Linear Models. *Journal of the Royal Statistical Society, Series B.* 78, 1079 – 1102.
- [13] He, J. and **Chen S. X.** (2016). Testing Super-Diagonal Structure in High Dimensional Covariance Matrices, *Journal of Econometrics*, 194, 283-297.
- [14] He, P., Wu, Z. G., Zhang, X. H. and **Geng, Z.** (2016). Identification of causal mediation models with an unobserved pre-treatment confounder. *Statistical Causal Inferences and Their Applications in Public Health Research*. Ed by He, H., Wu, P., Chen, D. G. et al. *Springer Science.* 241-262.
- [15] Huang, D., Yin, J., Shi, T., and **Wang, H.** (2016). A statistical model for social network labeling, *Journal of Business and Economics Statistics*, 34(3), 368-374.
- [16] Jiang, Z. C., Ding, P. and **Geng, Z.** (2016). Principal causal effect identification and surrogate endpoint evaluation by multiple trials. *J Royal Statist. Soc. B.* 78, 829-848.
- [17] Lan, W., Zhong, P. S., Li, R., **Wang, H.** and Tsai, C. L. (2016). Testing a Single Regression Coefficient in High Dimensional Model, *Journal of Econometrics*, 195(1), 154-168.
- [18] Lee, L.F. and **Yu J.**, (2016). Identification of spatial panel Durbin models. *Journal of Applied Econometrics* 31, 133-162.
- [19] **Li, Chenxu** (2016). Bessel Processes, Stochastic Volatility, and Timer Options. *Mathematical Finance* 26(1), 122–148.
- [20] **Li, Chenxu** and Chen D. (2016). Estimating Jump-Diffusions Using Closed-form Likelihood Expansions. *Journal of Econometrics*, 195(1), 51–70.
- [21] **Li, Chenxu**, An Y., Chen D., Lin Q., and Si N. (2016). Efficient Computation of Likelihood Expansions for Diffusion Models. *IIE Transactions*, 48(12), 1156–1171.
- [22] Li, R, Liu, Y, Li, T and **Li, C.** (2016). 3Disease Browser: A Web server for integrating 3D genome and disease-associated chromosome rearrangement data, *Scientific Reports*, 2016 Oct 13;6:34651.
- [23] Li, Shuo and **Tu, Yundong** (2016). Root-n Consistent Density Estimation in Semiparametric Regression Models, *Computational Statistics and Data Analysis*, 104,

91-109, December 2016.

[24] Li, Shuo and **Tu, Yundong** (2016). On Estimating Nonparametric Multiplicative Error Models, *Economics Letters*, 143, 66-68, June 2016.

[25] Liang, X., Li, S., Zhang, SY, **Huang, H.** and **Chen S.X.** (2016). PM2.5 Data Reliability, Consistency and Air Quality Assessment in Five Chinese Cities, *Journal of Geophysical Research—Atmosphere*, 121, 10,220–10,236.

[26] **Liu L.**, Guo Z. and Duan X. (2016). Population size estimation with covariate values missing non-ignorable. *Acta Mathematicae Applicatae Sinica*, 32, 659-668.

[27] Liu M, Li Y, Liu A, Li R, Su Y, Du J, **Li C**, Zhu AJ. (2016).The exon junction complex regulates the splicing of cell polarity gene *dlg1* to control Wingless signaling in development, *Elife*. 2016 Aug 18;5. pii: e17200.

[28] **Liu, Yong**; Zhai, Jianliang (2016). Time regularity of generalized Ornstein-Uhlenbeck processes with Lévy noises in Hilbert spaces.J. *Theoret. Probab.*29 (2016), no. 3,843–866.

[29] Lu, X., Zhao, J., Chen, Y., and **Wang, H.** (2016).A Choice Model with a Diverging Choice Set for POI Data Analysis, *Statistics and its Interface*, 9, 355-363.

[30] Luo, P. and **Geng, Z.** (2016). Bounds on direct and indirect effects of treatment on a continuous endpoint. *ACM Transactions on Intelligent Systems and Technology*. 7 (2):1-18.

[31] Luo, P. and **Geng, Z.** (2016).Causal mediation analysis for survival outcome with unobserved Mediator-outcome confounders. *Comp. Statist Data Analy.* 93, 336-347.

[32] Miao, W., Ding, P. and **Geng, Z.** (2016).Identifiability of normal and normal mixture models with nonignorable missing data. *Journal of the American Statistical Association*, 111, 1673-1683.

[33] Pan, R., **Wang, H.**, and Li, R. (2016).Ultrahigh Dimensional Multi-Class Linear Discriminant Analysis by Pairwise Sure Independence Screening. *Journal of the American Statistical Association*, 2016, 111(513), 169-179.

[34] Peng, LH, **Chen S.X.** and Zhou W. (2016). More Powerful Tests for Sparse High-Dimensional Covariances Matrices, *Journal of Multivariate Analysis*, 149, 124-143.

[35] Ping, Z. Xia, Y., Shen, T., Parekh, V., Siegal, G.P. Eltoum, I.E., He, J.,Chen, D. Deng, M., **Xi, R.**, and Shen, D. * (2016).A microscopic landscape of the invasive breast cancer genome, *Scientific Reports*, 6, 27545.

- [36] Qian Junyang and **Jia Jinzhu** (2016). On Stepwise Pattern Recovery of Fused Lasso. *Computational Statistics and Data Analysis* 94 (2016), pp 221-237.
- [37] Shen, Y., **Huang, H.** and Guan, Y. (2016). A Conditional Estimating Equation Approach for Recurrent Event Data with Additional Longitudinal Information, *Statistics in Medicine*, 35, 4306-4319.
- [38] Shi, J., Xing, Q., Duan, M., Wang, Z., Yang, L., Zhao, Y., Wang, Xiao, Liu, Y., Deng, M., Ding, Z., Ke, A., Zhou, J., Fan, J., Cao, Y., Wang, J.,*, **Xi, R.**,* and Gao, Q.* (2016). Inferring the progression of multifocal liver cancer from spatial and temporal genomic heterogeneity, *Oncotarget*, 7(3):2867-77.
- [39] Strazza, M., Pirrone, V., Wigdahl, B., Dampier, W., **Lin, W.**, Feng, R., Maubert, M. E., Weksler, B., Romero, I. A., Couraud, P.-O. and Nonnemacher, M. R. (2016). Prolonged morphine exposure induces increased firm adhesion in an in vitro model of the blood-brain barrier. *International Journal of Molecular Sciences*, 17, 916.
- [40] Wang, Y., **Tu, Y-D** and **Chen S. X.** (2016). Improving inflation prediction with the quantity theory. *Economics Letters*, 149, 112-115.
- [41] Wang, Yaping; **Ai, Mingyao** (2016). Optimal designs for multiple treatments with unequal variances. *Journal of Statistical Planning and Inference*, 171, 175-183.
- [42] Wang, Yaping; **Ai, Mingyao**; Li, Kang (2016). Optimality of pairwise blocked definitive screening designs. *Annals of the Institute of Statistical Mathematics*, 68(3), 659-671.
- [43] **Xi, R.** and Lin, N.* (2016). Direct Regression Modelling of High-order Moments in Big Data, *Statistics and Its Interface*, 9(4), 445-452.
- [44] **Xi, R.***, Lee, S., Xia, Y., Kim, T. and Park, P.* (2016). Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants, *Nucleic Acids Research*, 44(13):6274-86.
- [45] **Xi, R.***, Li, Y. and Hu, Y. (2016). Bayesian Quantile Regression Based on the Empirical Likelihood with spike and slab priors, *Bayesian Analysis*, 11(3): 821-855.
- [46] **Xu, Minya**, Ju Dong, Qin Xin and Dorenzo S. Marco (2016). Boundary Conditions of the Emotional Exhaustion-Unsafe Behavior Link: The Dark Side of Group Norms and Personal Control. *Asia Pacific Journal of Management*, 33(1), 113-140
- [47] **Xu, Minya**, Zhong, Ping-Shou and Wang Wei (2016). Detecting Variance Change-Points for Blocked Time Series and Dependent Panel Data. *Journal of Business & Economic Statistics*, 34(2): 213-226.

[48] 徐敏亚, 刘海洋、刘圣明、王辉(2016). “领导与下属权力距离的一致性对下属工作绩效的影响及其机制”, 《南开管理评论》, 19(5), 55-65。

[49] Yan, T., Qin, H., and **Wang, H.** (2016). Asymptotics in undirected random graph models parameterized by the strengths of vertices, *Statistica Sinica*, 26, 273-293.

[50] Yang, Z, **Yu J.** and Liu S.F., (2016). Bias Correction for Fixed Effects Spatial Panel Data Models. *Regional Science and Urban Economics* 61, 52-72.

[51] **Yu, J.**, L.A. Zhou and G. Zhu, (2016). Strategic Interaction in Political Competition: Evidence from Spatial Effects across Chinese Cities. *Regional Science and Urban Economics* 57, 23-37.

[52] **Zhang, Junni**, Härdle, W. H., Chen, C. Y., Bommers, E. (2016). Distillation of News Flow into Analysis of Stock Reactions, *Journal of Business and Economic Statistics*, 34 (4), 547-563.

[53] Zhu, X., Huang, D., and Pan, R., and **Wang, H.** (2016). An EM Algorithm for Click Fraud Detection, *Statistics and its Interface*, 9, 389-394.



六、研究报告

- [1] He Jing and **Chen Song Xi**
Testing Super-Diagonal Structure in High Dimensional Covariance Matrices
- [2] **Tu Yundong**
Efficient Estimation of Nonparametric Simultaneous Equations Models
- [3] **Li C.**
Estimating Jump-Diffusions Using Closed-form Likelihood Expansions
- [4] Li Shuo and **Tu Yundong**
On Estimating the Nonparametric Multiplicative Error Models
- [5] **Li C.**, Tang C. Y. Tang, and Yan J.
On the Prevalence of Leverage Effect Puzzle: Evidence from Multifactor Stochastic Volatility Models with High Frequency Financial Data
- [6] An Y., Chen D., **Li C.**, Lin Q., and Si Nian
Efficient Computation of Likelihood Expansions for Diffusion Models
- [7] Li Shuo and **Tu Yundong**
 \sqrt{n} -consistent Density Estimation in Semiparametric Regression Models
- [8] **Tu Yundong** and Yi Yanping
Forecasting Cointegrated Nonstationary Time Series with Time-varying Variance
- [9] **Song Xiaojun**, Taamouti Abderrahim
Measuring Nonlinear Granger Causality in Mean
- [10] Wang Ying, **Tu Yundong** and **Chen Song Xi**
Adaptive Quantity Theory of Money
- [11] **Chen Song Xi**
PETER HALL'S CONTRIBUTIONS TO THE BOOTSTRAP
- [12] Liang Xuan, Li Shuo, Zhang Shuyi, **Huang Hui**, and **Chen Song Xi**
PM2.5 Data Reliability, Consistency and Air Quality Assessment
- [13] Huang Yimin, **Ai Mingyao**, Kong Xiangshun
Optimal designs for the Lasso in sparse linear models

- [14] Liang Xuan, Li Shuo, Zhang Shuyi, **Huang Hui**, and **Chen Song Xi**
中国五城市空气污染状况之统计学分析
- [15] Kong Xiangshun, **Ai Mingyao**
Flexible sliced designs for computer experiments
- [16] Kong Xiangshun, **Ai Mingyao**
Design for sequential follow-up experiments in computer emulations
- [17] **Xu Minya**, Zhong Ping-Shou & Wang Wei
Detecting Variance Change-Points for Blocked Time Series and Dependent Panel Data
- [18] Ju Dong, Qin Xin, **Xu Minya**, DiRenzo S. Marco
Boundary conditions of the emotional exhaustion-unsafe behavior link: The dark side of group norms and personal control
- [19] **Li Chenxu** and Chen Dachuan
Estimating jump-diffusions using closed-form likelihood expansions
- [20] **Xi Ruibin**, Lee Semin, Xia Yuchao, Kim Tae-Min, and Park J. Peter
Copy number analysis of whole-genome data using BIC-seq2 and its application to detection of cancer susceptibility variants
- [21] Qu Xi, Lee Lung-fei, **Yu Jihai**
QML Estimation of spatial dynamic panel data models
- [22] **Ren Yan-Xia**, Song Renming and Yang Ting
Spine decomposition and $L \log L$ criterion for superprocesses with non-local branching mechanisms
- [23] **Ren Yan-Xia**, Song Renming Song and Zhang Rui
Williams decomposition for superprocesses
- [24] Wang Yaping, **Ai Mingyao**
De nitive Screening Designs with Extreme Numbers of Level Changes
- [25] Yu Jun, **Ai Mingyao**, Wang Yaping
Optimal designs for linear models with Fredholm-type errors

七、学术讲座

北京大学统计科学中心 2016 年共组织学术报告 56 场，报告人来自美国、英国、德国、澳大利亚、沙特、俄罗斯、韩国、瑞典、新西兰、荷兰、香港和台湾以及国内各高校、研究机构和知名企业。其中来自海外的学者 31 场，国内的学者 25 场。学术报告拓宽了老师和学生的学术视野，促进了统计学领域的学术交流，活跃了统计中心的学术氛围。

[1] Statistical modeling problems in the Guangzhou Twin Eye Study

郭小波博士, 中山大学

2016-01-14 14:00-15:00

理科一号楼 1479 教室

[2] Multivariate Extension of the Ballot Theorem and Its Applications in Multitype branching forests

刘荣丽副教授, 南京大学

2016-03-07 15:00-16:00

理科一号楼 1303 教室

[3] A Regime Shift Model with Nonparametric Switching Mechanism

Haiqiang Chen, Xiamen University

2016-03-24 14:00-15:30

光华管理学院新楼 217 教室

[4] Some results about critical values for contact processes in random environments

薛晓峰博士, 中国科学院大学

2016-03-28 15:00-16:00

理科一号楼 1303 教室

[5] A New Smoothed Perturbation Analysis Estimation of Greeks for Financial Options with Discontinuous Payoffs

Yanchu Liu, Sun Yat-Sen University

2016-03-31 14:00-15:00

光华管理学院新楼 217 教室

[6] Asymmetry in Stock Comovements: An Entropy Measure

Ke Wu, Renmin University of China

2016-04-05 14:00-15:30

光华管理学院新楼 216 教室

[7] 弱收敛 (2016 年北京大學許寶騮講座)

林正炎教授, 浙江大學

2016-04-07 15:00-16:00

理科一号楼 1114 教室

[8] On weak convergence of stochastic processes to stochastic integrals

林正炎教授, 浙江大学

2016-04-11 15:00-16:00

理科一号楼 1303 教室

[9] Kriging Over Space and Time Based on a Latent Reduced Rank Structure

Qiwei Yao, London School of Economics

2016-04-14 14:00-15:00

光华管理学院新楼 217 教室

[10] Some limit theorems for subcritical branching processes in random environment

Prof. V. Vatutin, Steklov Mathematical Institute, Moscow, Russia

2016-04-18 16:00-17:00

理科一号楼 1303 教室

[11] Penalized Generalized Method of Moments with Many Weak Instrumental Variables

范青亮, 厦门大学

2016-04-28 14:00-15:00

理科一号楼 1114 教室

[12] Fitting Misspecified Linear Mixed Models

Alan Welsh, The Australian National University

2016-05-03 14:00-15:00

理科一号楼 1114 教室

[13] Implications of Return Predictability across Horizons for Asset Pricing Models

Haoxi Yang, Nankai University

2016-05-05 14:00-15:00

光华管理学院新楼 217 教室

[14] Statistical Identification of Markov Chain

向绪言教授, 湖南文理学院

2016-05-09 15:00-16:00

理科一号楼 1303 教室

[15] Model checking for parametric single-index models: A dimension reduction model-adaptive approach

郭旭, 南京航空航天大学

2016-05-12 14:00-15:00

光华管理学院新楼 217 教室

[16] Intergenerational Top Income Mobility in Taiwan

庄额嘉, Wuhan University

2016-05-19 14:00-15:00

光华管理学院新楼 217 教室

[17] Global Carbon Assimilation System using a modified Ensemble Kalman filter

Xiaogu Zheng, Beijing Normal University

2016-05-19 15:00-16:00

光华管理学院新楼 217 教室

[18] Variational Approach for Stochastic Partial Differential Equations

刘伟教授, 江苏师范大学

2016-05-23 15:00-16:00

理科一号楼 1303 教室

[19] 基于全基因组数据的系统发育分析

唐继军教授, 天津大学计算机学院

2016-05-25 12:30-13:30

北京大学镜春园全斋 29

[20] Treatment Effects Estimators Under Misspecification

Ping Yu, The University of Hong Kong

2016-05-26 14:00-15:30

光华管理学院新楼 217 教室

[21] 撩基因空间表达神秘面纱, 展交叉学科合作研究之密

Bin Yu, Chancellor's Professor of Statistics and EECS, UC Berkeley

2016-05-26 19:00-21:00

第二教学楼 107 教室

[22] A tailored robust multivariate clustering approach via mean-shift penalization

陈昆, 康涅狄格大学

2016-05-27 14:00-15:00

理科一号楼 1418 教室

[23] Internally Consistent Estimation of Nonlinear Panel Data Models with Correlated Random Effects

Ji-Liang Shiu, Renmin University of China

2016-05-30 14:00-15:30

光华管理学院旧楼 219 教室

[24] Nonparametric Test of Monotonicity of Bidding Strategy in First Price Auctions

Nianqing Liu, Shanghai University of Finance and Economics

2016-06-02 14:00-15:30

光华管理学院新楼 217 教室

[25] Tukey g-and-h Random Fields

Marc G. Genton, King Abdullah University of Science and Technology (KAUST)

2016-06-02 15:30-16:30

理科一号楼 1114 教室

[26] Approximating Likelihoods for Large Environmental Datasets

Ying Sun, King Abdullah University of Science and Technology (KAUST)

2016-06-02 16:30-17:30

理科一号楼 1114 教室

[27] Regression Discontinuity with Integer Running Variables: Effect of Education on Fertility

Myoung-jae Lee, Korea University

2016-06-03 14:00-15:30

光华管理学院旧楼 219 教室

[28] From infinite urn schemes to decompositions of self-similar Gaussian processes

王一早博士, University of Cincinnati

2016-06-06 15:00-16:00

理科一号楼 1303 教室

[29] Optimal Credit Investment with Borrowing Costs

Lijun Bo, University of Science and Technology of China

2016-06-08 14:00-15:00

光华管理学院旧楼 115 教室

[30] Statistical challenges and opportunities in drug development – the Boehringer-Ingelheim perspective

Hans-Juergen Lomp, Boehringer Ingelheim

2016-06-17 14:00-15:00

理科一号楼 1114 教室

[31] Computer experiments and uncertainty quantification: an overview

Jeff Wu, Georgia Institute of Technology

2016-06-22 16:00-17:00

理科一号楼 1114 教室

[32] Sparse Simultaneous Signal Detection and Its Applications in Genomics

Hongzhe Li, University of Pennsylvania Perelman School of Medicine

2016-06-28 14:00-15:00

理科一号楼 1114 教室

[33] Statistical approaches for predicting the functional effect of

Prof. Iuliana Ionita-Laza, Columbia University

2016-06-30 12:30-13:30

北京大学镜春园全斋 29

[34] Targeted Error Correction: a framework for making cheap data useful

Mike Baiocchi, Stanford School of Medicine

2016-06-30 14:00-15:00

理科一号楼 1114 教室

[35] Semiparametric Inference on Trending Panel Data with Spatial Dependence

Peter M. Robinson, London School of Economics

2016-07-01 10:00-11:30

光华管理学院新楼 217 教室

[36] Empirical Bayes Prediction under Check Loss

Prof. Lawrence D. Brown, University of Pennsylvania

2016-07-06 10:00-11:00

光华管理学院旧楼 215 教室

[37] Duration Estimation in Hidden Markov Model

Qingyang Zhang, University of Arkansas

2016-07-07 14:00-15:00

理科一号楼 1114 教室

[38] A nonparametric approach for partial areas under ROC curves

Yichuan Zhao, Georgia State University

2016-07-07 15:00-16:00

理科一号楼 1114 教室

[39] Estimating Subgroup-Specific Treatment Effects via Concave Fusion

Shujie Ma, University of California, Riverside

2016-09-08 14:00-15:00

光华管理学院新楼 217 教室

[40] Model Selection in Hidden Markov Models via Marginal Likelihood

Cheng-Der Fuh, National Central University

2016-09-14 14:00-15:00

光华管理学院新楼 216 教室

[41] Optimal Community Detection in Degree-Corrected Stochastic Block Models

Zongming Ma, University of Pennsylvania

2016-10-19 14:00-15:00

理科一号楼 1418 教室

[42] Economic trade and transboundary air pollution

Jintai Lin, School of Physics, PKU

2016-10-20 13:00-14:00

光华酒店 K02

[43] Matrix Visualization: New Generation of Exploratory Data Analysis

Chun-houh Chen, Institute of Statistical Science, Academia Sinica

2016-10-20 14:00-15:00

理科一号楼 1303 教室

[44] The world's first fully Bayesian official statistics

John Bryant, Senior Researcher, Statistics New Zealand

2016-10-27 14:00-15:00

光华管理学院新楼 217 教室

[45] Economic Status, Air Quality, and Child Health: Evidence from Inversion Episodes

Per Johansson, Professor in statistics, Department of statistics, Uppsala University

2016-10-28 14:00-15:30

光华管理学院新楼 217 教室

[46] Studies in Orthogonal Designs with Application to Experimental Design

Diane Donovan, The University of Queensland

2016-11-01 14:00-15:00

理科一号楼 1114 教室

[47] Instantaneous Squared VIX and VIX Derivatives

Xingguo Luo, Zhejiang University

2016-11-03 14:00-15:00

光华管理学院新楼 217 教室

[48] Multiple Changepoints Detection via Deep Learning

薛原, 对外经济贸易大学

2016-11-17 14:00-15:00

理科一号楼 1303 教室

[49] A Beautiful Breakdown of CIR Intensity: Exact Simulation and Financial Application

Hongbiao Zhao, Xiamen University

2016-11-24 14:00-15:30

光华管理学院新楼 217 教室

[50] Leveraging methods for massive data logistic regression

Ping Ma, University of Georgia

2016-12-14 14:00-15:00

光华管理学院新楼 217 教室

[51] Minimax Risk Bounds for Piecewise Constant Models

Fang Han, University of Washington

2016-12-15 09:00-10:00

理科一号楼 1303 教室

[52] Model Averaging for Prediction with Fragmentary Data

方方, 华东师范大学

2016-12-15 10:00-11:00

理科一号楼 1303 教室

[53] Asymptotic Theory of Rerandomization in Treatment-Control Experiments

Peng Ding, University of California, Berkeley

2016-12-15 11:00-12:00

理科一号楼 1303 教室

[54] 颗粒物监测及源解析新技术发展与应用

高健博士, 中国环境科学研究院

2016-12-23 13:15-14:15

光华酒店 K01

[55] Frequentist Analysis of the Posterior for High-Dimensional Models

Johannes Schmidt-Hieber, Leiden University

2016-12-29 14:00-15:00

理科一号楼 1303 教室

[56] Adaptive Filtering Multiple Testing Procedures for Partial Conjunction Hypotheses

Jingshu Wang, University of Pennsylvania

2016-12-30 14:00-15:00

理科一号楼 1418 教室



八、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。



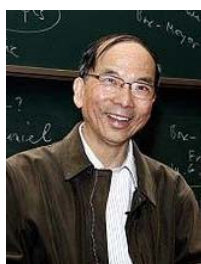
Peter G. Hall(2016年去世)

委员，墨尔本大学和加州大学戴维斯分校的统计学教授，英国皇家学会院士，澳大利亚科学院院士，美国科学院外籍院士。



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、美国统计学会和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



郁彬

主席，加州大学伯克利分校统计系教授，2013 年当选 AAAS，现任 IMS President，千人计划学者，长江学者讲席教授，是 2006 年的古根海姆学者，她曾经还是美国瑟姆西国家咨询委员会（统计和应用数学科学研究所）联席主席。



陈大岳

委员，北京大学统计科学中心副主任，曾获国家自然科学基金委“杰出青年”等多项资助，现任北京大学数学科学学院副院长、中国数学会副理事长和《数学进展》主编等职。



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系主任, 国家首批“千人计划”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, 国家杰出青年科学基金获得者。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、流行病学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。



刘军

委员, 哈佛大学统计系教授, 曾任北京大学长江学者讲座教授。国际数理统计学会会士, 美国统计学会会士, 曾获得过统计学界最高奖“考普斯会长奖”以及华人数学最高奖“晨兴数学金奖”。



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”。



蔡洪滨

委员，北京大学光华管理学院院长，2006年获教育部新世纪优秀人才称号，2007年获国家自然科学基金杰出青年科学基金，2008年被聘为教育部长江学者特聘教授，2011年当选为世界计量经济学会会士(Fellow)，并于2012年被选为理事会理事。




田刚

委员，北京大学数学科学学院院长，美国普林斯顿大学数学系教授，北京国际数学研究中心主任，中国科学院院士，全国政协常委，中国民主同盟中央副主席。在微分几何和数学物理领域做出了重大的贡献，曾经在2002年国际数学家大会上被邀请作大会报告。



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。曾获国家自然科学基金二等奖、教育部高等学校自然科学一等奖、国家杰出青年基金、冯康科学计算奖、教育部“长江学者”特聘教授、“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。



Tel: 0086-010-62760736

Fax: 0086-010-62760736

Mail: stat-center@pku.edu.cn

Http: www.stat-center.pku.cn

Add: 北京大学燕东园 32 号小楼,100871