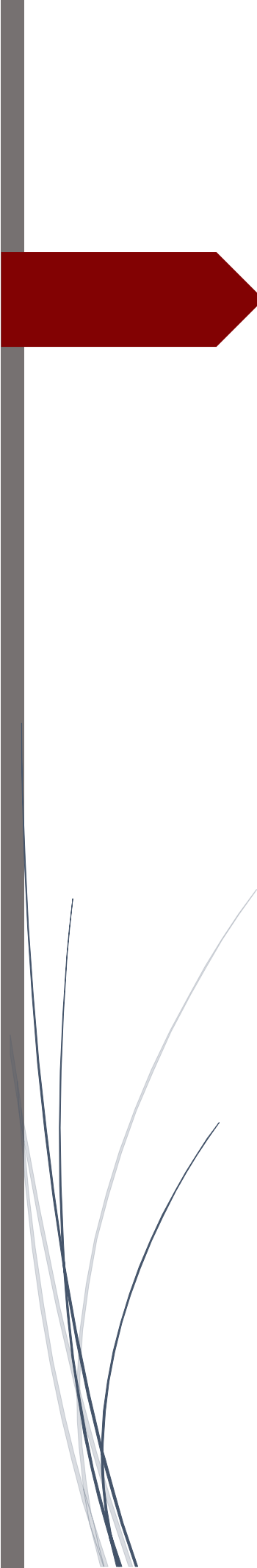




北京大学统计科学中心 2021 年度总结报告



北京大学统计科学中心

2022-2-17

北京大学统计科学中心 2021 年度总结报告

目 录

一、中心概况	1
二、要闻概览	3
三、学科建设	7
四、科研工作	9
(一) 科研项目	9
(二) 科研成果	14
(三) 学术交流	17
五、人才培养	19
(一) 学生获奖情况	19
(二) 毕业生及论文	19
(三) 毕业生去向	20
六、组织架构	21
(一) 指导委员会	21
(二) 行政团队	22
(三) 科学委员会	24
(四) 中心教员	26

一、中心概况

2021 年统计科学中心在统计学研究、博士生培养、师德师风建设以及推动统计学发展等方面开展工作。

在学科发展方面，2017 年北大统计学科入选“双一流”建设学科，在第四轮教育部学科评估中获评 A+；在软科公布的中国最好学科排名中最近 4 年连续排名第一（2018-2021）。北京大学统计学科排名的上升，为统计学科的未来发展提供了更加广阔的舞台。与此同时，北大统计科学中心的发展也面临新的机遇与挑战。

在教研团队方面，为建设一支专业结构合理、学术素养良好、适应能力强的统计学教师队伍，中心积极引进优秀人才。目前，中心现有教研人员 37 人，2021 年新增两人，分别为张瑞勋助理教授（统计科学中心与数学科学学院合聘）、占翔副教授（统计科学中心与公共卫生学院合聘）。北大统计学科业已形成具有国际一流水平的教研团队。

在国际合作交流方面，持续受全球新冠疫情的影响，2021 年中心的国内外学术交流活动依然受到严重影响。中心克服困难，在严格执行各级防疫政策的前提下，不断推进同国内外学者的学术交流活动，共举办学术报告 33 场，其中大多采用线上会议形式；共接待海内外访问学者 20 余人次，参与学术报告、短期课程以及合作项目等各类学术活动。2021 年中心共邀请到来自加州大学戴维斯分校 Shizhe Chen 教授、清华大学杨立坚教授、中国科学院大学洪永淼教授、中国人民大学马维教授、复旦大学夏寅教授、哥本哈根大学 Jonas Peters 教授等开展了 33 场学术报告活动。这些报告拓宽了中心教员和学生的学术视野，活跃了统计科学中心的学术氛围，促进了统计学领域的交流。此外，2021 年中心主办了第五届北大——清华统计论坛、Eric 教授海外名家讲学计划、“全国统计学优秀大学生”夏令营，承办了第二十一次全国统计科学讨论会，更广泛地促进了海内外统计学界师生及统计学从业者的交流与合作。

在行政队伍建设方面，中心为加强行政工作，于 2019 年底重新组建了一支办事高效、组织决策能力强的行政团队，姚方教授担任新一届中心主任。2021 年，中心在姚方主任的带领下，继续完善队伍建设，精进团队效能。目前管理团队包括周晓华和王汉生副主任、席瑞斌主任助理、穆慧娜综合办公室主任、张苑新行

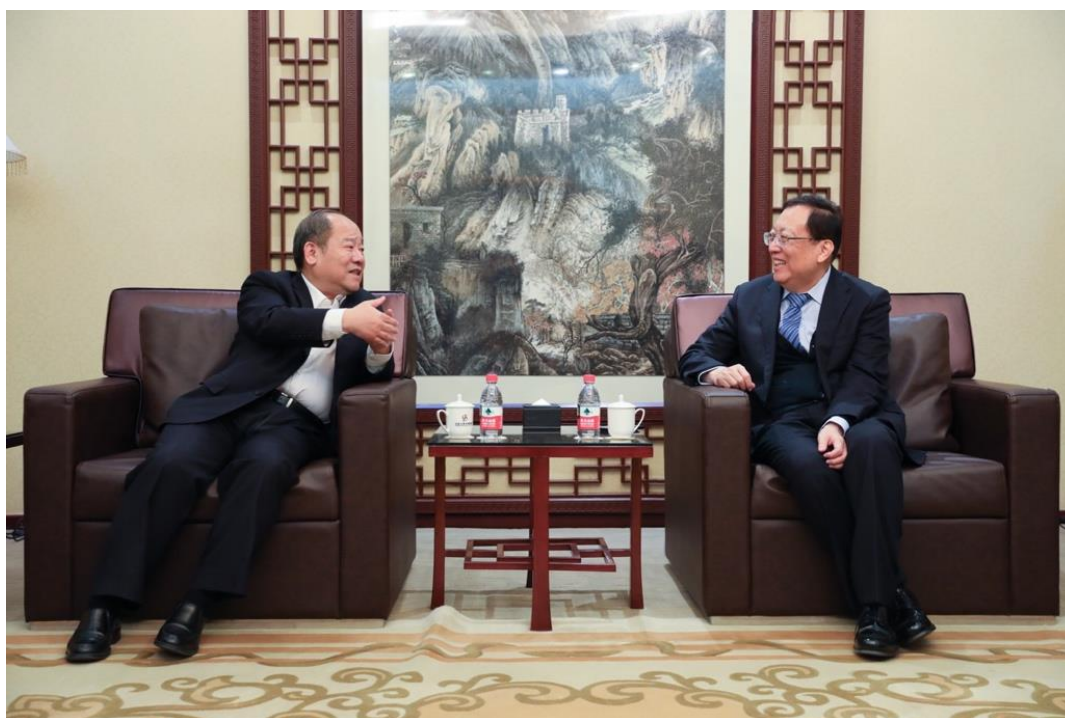
政秘书。2021 年，中心全面加强日常运营及拓展业务等事务的管理，各项工作稳中有进。2021 年，中心行政团队牵头，完成了五年一轮的学科评估（统计学）整体工作、新一轮“一流”学科建设方案编制工作，同时主办和承办了多场 200 人以上（线上线下结合）的学术会议。由于北大统计学科的特殊性，这些工作，都需要跨学部、院系协调沟通资源，中心在人员精减、资源紧张的情况下，顺利组织并完成上述多项复杂度极高的工作，充分展示了中心行政团队的业务能力。

在科学研究方面，2021 年中心教员累计发表学术论文近 150 篇，包含统计学及其交叉学科顶级期刊近 25 篇；2021 年在研项目 52 项，其中国家级科研项目 33 项；2021 年新获批项目 10 项。

二、要闻概览

1. 中心承办第二十一次全国统计科学讨论会，加强统计学同仁沟通与合作

2021 年 12 月 18-19 日，北京大学统计科学中心承办了第二十一次全国统计科学讨论会。会议由中国统计学会主办，参会人员来自全国统计系统、高等院校、科研院所，线上线下共计 300 余人参与了本次会议。本次会议促成了郝平校长与国家统计局宁吉喆局长一行的会面，双方就统计学科发展交换了意见。通过本次会议，国内外高校、全体统计学同仁进一步加强了联系与沟通，后续在统计相关问题研究上，将有更多广泛而深入的交流与合作，在“百年未有之大变局”下，共同回应时代的召唤，携手推进统计学事业的全面深入发展。



2. 高效推进统计学科发展，学科建设工作成效显著

中心作为北大统计学科建设的牵头单位，横向协调数学科学学院、光华管理学院、公共卫生学院、国家发展研究院等，于 2021 年 1 月中旬顺利完成第五轮学科评估的材料填报工作。同时，中心于 2021 年 9 月，顺利完成新一轮“一流学科”建设方案编纂工作，并邀请到中国科学院马志明院士、北京大学张平文副校长、中国人民大学统计学院王晓军院长等专家对建设方案进行评审，肯定了北大统计学科在上一轮“一流学科”建设中取得的成绩，同时表达了对北大统计新一轮“一流学科”方案的肯定与信心。

3. 陈松蹊教授当选中科院院士，持续为大气污染治理提供政策建议

2021 年 11 月 18 日，中国科学院、中国工程院公布 2021 年院士增选结果，统计科学中心陈松蹊教授当选为中国科学院院士。陈松蹊教授长期从事统计学与大气环境交叉学科研究，相关成果多次被生态环境部和中国环境监测总站采纳，有效的服务了国家大气污染防治的重大需求。截至 2021 年底，该团队共发布 8 份《空气质量评估报告》，每份报告均通过北大校办递交中办、国办、环保部和科技部，并引起广泛的媒体关注，如中新社、人民网、央视国际频道、《中国日报》、《北京日报》、《纽约时报》和《参考消息》等。在 *Journal of Geophysical Research* 等大气环境领域顶级杂志发表近 10 篇论文，获得一项空气质量评估方法的专利。

该团队提出客观评估大气污染的方法，去除监测数据中的气象混杂，从而度量背景排放的变化。团队建立了在时间和空间两个尺度上进行气象调整的方法，提出“人努力-天帮忙”指数。同时，团队运用统计学方法对冬季供暖效应、臭氧的辐射效应、重污染预警及 2021 年去除疫情影响后的空气质量评估等开展研究。



4. 北大统计学科在“2021 软科中国最好学科”获评第一

2021 年 10 月 25 日，高等教育评价专业机构软科中国基于对学科人才培养、科研项目、成果获奖、学术论文、高端人才五个指标的综合评分，发布“2021 软科中国最好学科排名”，北大统计学科排名第一。这已是北大统计学科连续第 4 年保持第一名（2018-2021）。

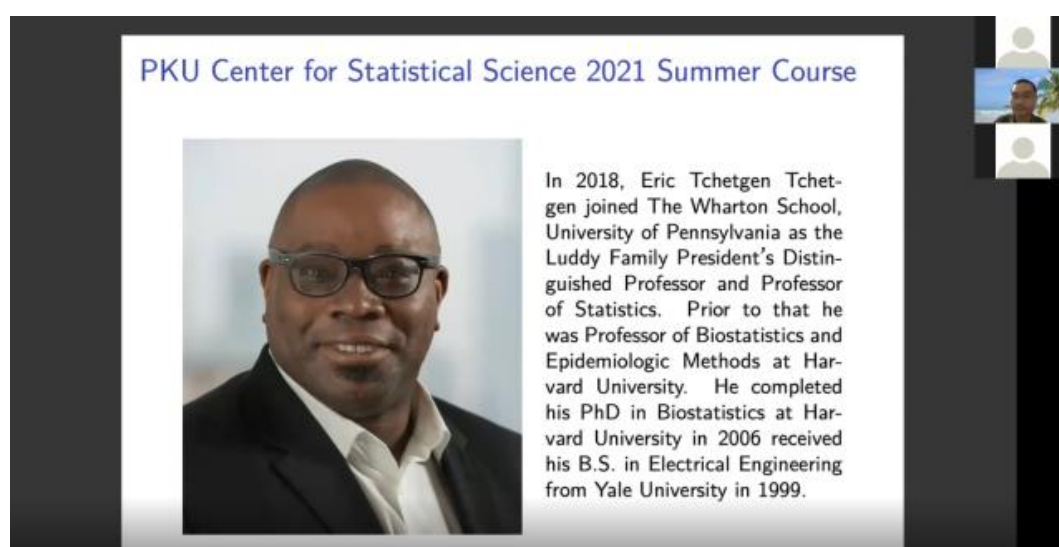
北大统计学科拥有雄厚的科研实力和光荣的历史底蕴，自 1940 年许宝騄先生从英国获统计学博士学位回国，首次在我国大学开设数理统计课程开始，几十

年筚路蓝缕，北京大学为中国统计学科人才培养、学科发展做出了巨大的贡献。2010 年在数理统计研究所的基础上成立统计科学中心，整合全校力量加强统计学科建设；2011 年设立数量经济与数理金融教育部重点实验室；2018 年成立跨学部的生物统计系。今天的北大统计站在了一个前所未有的高度，也必将不忘初心，砥砺前行，开创更辉煌的事业。

5. 北京大学海外名家讲学计划暨北京大学统计科学中心 2021 年本科生暑期学校成功举办

2021 年 7 月 8 日至 7 月 14 日，我院苗旺教授邀请沃顿商学院 Eric Tchetgen Tchetgen 教授，开设系列讲座暨北京大学统计科学中心 2021 年本科生暑期学校：Semiparametrics with Application to Causal Inference and Machine Learning。讲座通过线上 Zoom 会议形式成功举办。

Eric Tchetgen Tchetgen 教授通过 10 余小时的短课，带领大家充分了解了关于半参数统计大家普遍关心的问题，以及该理论的优势所在。通过其深入浅出的讲解，深刻的理论变得通俗易懂，这很好地帮助了北京大学数学学院和统计科学中心师生，以及 110 余名海内外其他高校相关领域的学生学习相关理论中的重点和难点，促进他们在科研上取得新的进展。



6. “全国统计学优秀大学生”夏令营成功举办

2021 年 6 月 4 日至 6 日，北京大学统计科学中心成功举办了“全国统计学优秀大学生”夏令营。夏令营旨在增进国内高校优秀本科生对统计学科研究的了解，并从中选拔有潜力者继续深造。来自全国各大高等院校的近 30 名优秀大学三年级学生从 300 余名申请者中脱颖而出，取得参营资格。营员在一系列活动中充分

了解了中心教师的研究方向，汲取了前沿的创新知识，并在交流中发现了自己的兴趣方向，充分展示了自己的风采，中心也从活动中选拔出优秀的统计学后备人才。



7. 第五届北大——清华统计论坛成功举办

2021年6月10日，北京大学统计科学中心成功举办第五届北大——清华统计论坛。该论坛目的在于发挥北大和清华两校的统计学科优势，互通有无，推动中国优秀统计青年人才的成长，助力中国统计学科的发展。论坛由北京大学统计科学中心和清华大学统计学研究中心联合举办，来自海内外多所高校的140余名师生通过线上线下结合的形式参加了本次论坛。



三、学科建设

北大统计学科按照国家“双一流”学科建设的战略部署和指导方案，稳中求进，继承创新，一直朝着建成亚太地区领先的统计学研究与人才培养基地的目标迈进。截至 2021 年底，北大统计学科“双一流”学科建设发展成效初显，完成“双一流”学科近期建设设定的目标，已达亚太地区一流水平。

作为我国重要的统计学研究和人才培养基地，北大统计学科充分发挥基础学科的支撑作用，为解决国家重大需求做贡献，同时积极参与多个学术组织、平台与会议，助力我国统计学学科发展。

利用专业知识建言献策，在国家卫生、维护国际声誉等方面发挥智库作用。耿直团队与国家疾控中心合作评估艾滋病人数，帮助卫健委依据事实有力驳斥反华势力散布的中国器官移植谣言，相关工作受到央视新闻的报道；自 2016 连续 4 年吴岚团队与国家农业农村部信息中心联合在北大举办“春耕论坛”，系国内首个聚焦农业大数据的产学研论坛，为发展现代农业和精准扶贫战略献策；周晓华任世界中医药联合会中医大数据产业分会会长，参与国家医疗器械审评中心的指南制定工作；房祥忠、周晓华在国家食品安全风险评估中心举办多场培训；艾明要参与国家体育总局科研所体育统计新指标体系的研讨定稿工作；张志华和华为等公司合作，针对业务需求，解决技术瓶颈。

不断推进跨学科交叉融合。2011 年设立北京大学数量经济与数理金融教育部重点实验室；2018 年成立跨学部的生物统计系，服务于精准医疗和生物医学数据挖掘的国家战略需求，对提高我国生物统计专业人才培养质量、促进统计学与生物医学的交叉和融合具有重要意义。张志华和林伟积极参与北京智源人工智能研究院的科研工作，葛颢承担北京大学数学生物医学前沿创新中心的研究工作，而生命科学学院李程则一直是统计科学中心成员。为更好服务于校内外科研工作，统计咨询中心于 2020 年正式成立。

积极为国内外学术共同体贡献力量。2016 年至今，房祥忠任中国现场统计研究会理事长，教育部高等学校统计学类教学指导委员会主任委员；陈大岳任中国数学会秘书长；耿直任国务院学位委员会统计学学科评议组成员。陈松蹊和姚方先后当选数理统计学会(IMS)理事会理事，陈松蹊任伯努利学会科学书记；陈大岳和姚方任 IMS 提名委员会委员；姚方担任 ASA 诺特奖评委会委员等。多人

在国际学术期刊任主编、副主编及编委，其中姚方任《加拿大统计学期刊》主编，张志华任 2020 国际机器学习大会、国际深度学习大会领域主席等。

持续精进教学与科研功底，取得多项教学与科研奖励。2021 年度，中心教员持续优化课程与教学方案，精进科研工作，在北京大学校内、北京地区及全国范围内，取得多项奖励和荣誉。现列举部分教员获奖情况如下：

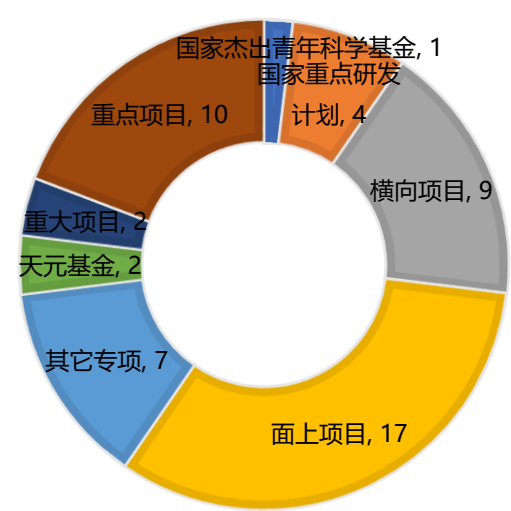
奖项名称	获奖等级/ 荣誉称号	成果完成人
教育部 2020 年度 “长江学者奖励计划”	青年学者	涂云东
北京大学教学成果奖	一等奖	吴岚、杨静平、黄海、 徐恺、何洋波、程雪、李东风
北京大学教学成果奖	一等奖	房祥忠、艾明要、 邓明华、耿直、李东风
北京大学教学成果奖	二等奖	章复熹、任艳霞、 陈大岳、蒋达权、葛颢
北京高校第十二届青年教师 教学基本功	三等奖	张原
北京高校第十二届青年教师 教学基本功比赛最受学生欢迎奖	理科类第一名	张原
北京高校第十二届青年教师 教学基本功比赛最佳教学反思奖	理科类第一名	张原
北京大学教学优秀奖-研究生	优秀奖	席瑞斌
中国统计学会统计科学技术进步奖	三等奖	席瑞斌

四、科研工作

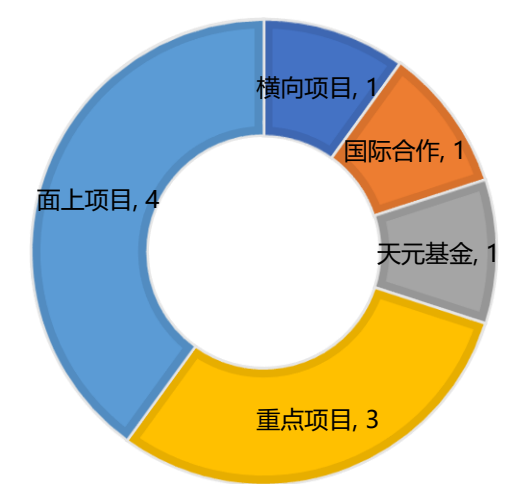
(一) 科研项目

统计科学中心教员目前在研项目累计 52 项，其中包括国家级科研项目 33 项，其他科研项目 19 项；2021 年新获批项目 10 项。

2021年在研项目（52项）分类统计



2021年新获批项目(10项)分类统计



2021 年度新获批项目						
序号	项目来源	项目类型	项目（课题）名称	负责人	开始时间	结题时间
1	国家自然科学基金	重点项目	大数据采样技术与统计设计理论研究	艾明要	202201	202612
2	国家自然科学基金	重点项目	表观修饰和染色质高级结构在细胞全能性建立中的关联调控及机制研究（项目负责人姚红杰）	李程	2021	2026
3	国家自然科学基金	数学天元基金	单细胞数据的多尺度统计建模与分析	邓明华	202201	202212
4	科技部	重点项目	面向药品现代化监管的智能化服务平台研发与应用	周晓华	202201	202412
5	国家自然科学基金	面上项目	FROC 框架下的诊断精确度评估的统计方法国家自然科学基金面上项目项目负责人按计划研究	周晓华	202201	202512
6	国家自然科学基金	面上项目	过参数化模型的正则化理论与方法	林伟	202201	202512
7	国家自然科学基金	面上项目	连续时间模型极大似然估计的大样本理论与应用	李辰旭	202201	202512
8	国家自然科学基金	面上项目	受限非平稳时间序列的计量经济模型：理论和应用	涂云东	202101	202412
9	NIH	国际合作	Fast and flexible Bayesian phylogenetics via modern machine learning	张成	202107	202606
10	企事业委托	横向项目-技术开发	肝癌的智能精准诊疗方案（项目负责人张宁）	席瑞斌	202112	202312

在研项目-国家级科研项目						
序号	项目来源	项目类型	项目（课题）名称	负责人	开始时间	结题时间
1	国家自然科学基金	重点项目	面向管理决策大数据分析的理论与方法	陈松蹊	202101	202212
2	国家自然科学基金	重点项目	复杂结构的函数型数据分析	姚方	202001	202412
3	国家自然科学基金	重点项目	多源异构数据的融合、特征提取与分析方法	王汉生	201901	202312

4	国家自然科学基金	重点项目	“非局部算子的随机分析及其应用”合作(项目负责人张希承)	刘勇	201801	202212
5	国家自然科学基金	重点项目	“非局部算子的随机分析及其应用”合作(项目负责人张希承)	任艳霞	201801	202212
6	国家社会科学基金	重点项目	人地协调与城乡协同的乡村规划建设研究	贾金柱	202101	202212
7	国家自然科学基金	数学天元基金项目	面向儿童脑发育障碍性疾病的神经机制建模与辅助诊疗算法	陈松蹊	202101	202412
8	国家自然科学基金	数学天元基金项目	面向儿童脑发育障碍性疾病的神经机制建模与辅助诊疗算法	周晓华	202101	202501
9	国家自然科学基金	国家杰出青年科学基金	三维基因组	李程	2020	2025
10	科技部	重点项目	大气污染对呼吸和心血管系统健康影响的早期识别技术	贾金柱	201701	202112
11	国家自然科学基金	面上项目	大数据分析中的试验设计和抽样技术研究	艾明要	202101	202412
12	国家自然科学基金	面上项目	树上随机游动中的多层分析问题	陈大岳	202001	202312
13	国家自然科学基金	面上项目	变系数流行病学模型的统计推断	陈松蹊	202101	202412
14	国家自然科学基金	面上项目	与时间变换有关的资产定价和最优执行策略相关问题研究	程雪	202001	202312
15	国家自然科学基金	面上项目	成分数据分析方法及其在宏基因组数据中的应用	邓明华	201901	202212
16	国家自然科学基金	面上项目	现代非平衡态热力学和统计物理中若干重要问题的随机数学理论	葛颢	202001	202312
17	国家自然科学基金	面上项目	因果推断方法研究及在食品安全与生物医学中的应用	耿直	201801	202112
18	国家自然科学基金	面上项目	非对称马氏过程的理论研究及应用	蒋达权	201901	202212
19	国家自然科学基金	面上项目	健康生态学视角下乳腺癌淋巴水肿风险预测与精准干预策略研究	梁宝生	202201	202412
20	国家自然科学基金	面上项目	关于无对照研究和中介分析的因果建模与推断	苗旺	202101	202412
21	国家自然科学基金	面上项目	测度值分枝马氏过程及相关 Fisher-KPP 方程性质研究	任艳霞	202101	202412

22	国家自然科学基金	面上项目	基于单细胞转录组测序的网络分析方法和理论	席瑞斌	202001	202312
23	国家自然科学基金	面上项目	基于风险相依性的风险组合的若干问题	杨静平	202101	202412
24	国家自然科学基金	面上项目	高维函数型数据的检验、回归与分类	姚方	201901	202212
25	国家自然科学基金	面上项目	大规模优化问题的近似牛顿方法：理论与实现	张志华	201801	202112
26	国家自然科学基金	面上项目	二维高斯场及其相关问题	章复熹	201801	202112
27	国家自然科学基金	面上项目	基于因果推断框架下的最优治疗方案选择的统计方法研究	周晓华	201801	202112
28	国家自然科学基金	其它专项	基于新型数学和统计模型的新型冠状病毒（2019-nCoV）发生和发展规律的研究	周晓华	202001	202112
29	科技部	重大项目-科技创新2030	部分可观测马尔科夫决策过程的随机规划算法（项目负责人杨晓光）	张志华	201912	202212
30	科技部	国家重点研发计划	粒子系统与渗流模型（项目负责人李增沪）	陈大岳	202012	202511
31	科技部	国家重点研发计划	食品安全风险分级评价指标体系与预警模型研究	房祥忠	201812	202112
32	科技部	国家重点研发计划	基于统计与数值模式的多污染物数据场构建	林伟	201701	202112
33	科技部	国家重点研发计划	金融风险的计量理论与方法（项目负责人陈增敬）	杨静平	201909	202408

在研项目-其他科研项目

序号	项目来源	项目类型	项目（课题）名称	负责人	开始时间	结题时间
1	北京市自然科学基金	重点项目	人工智能的统计理论与算法基础	林伟	201910	202310
2	国防科技委	重点项目	实时数据流的统计学习理论与方法	席瑞斌	201809	202109
3	国防科技委	重点项目	针对复杂军事系统的强化学习方法	张志华	201809	202109
4	军委科技委	重大项目	军事复杂数据中的因果推断	贾金柱	202009	202212
5	其他基金-北京大学生命科学学院逸夫发展基金	联合实验室	新型染色质构象捕获技术开发与应用	李程	2021	2022
6	其它部门	其它专项	基于深度学习的 ICD 编码预测	贾金柱	202106	202206
7	其它部门	其它专项	食品安全风险评估操作平台建设	贾金柱	202103	202203
8	其它部门	其它专项	基于删失数据建模和强化学习算法的肺癌放疗动态风险管理及自适应决策优化	梁宝生	202101	202212
9	其它部门	其它专项	基于同源重组基因全景突变检测和 Rubin 因果推断的乳腺癌遗传机制研究	梁宝生	202201	202312
10	其它部门	其它专项	基于新型数学和统计模型的新型冠状病毒（2019-nCoV）发生和发展规律的研究	张云俊	202006	202205
11	企事业委托	横向项目-技术开发	大数据背景下统计分析核心技术研究	艾明要	202009	202206
12	企事业委托	横向项目	深度学习相关的统计学理论研究	王汉生	202104	202304
13	企事业委托-北京智源人工智能研究院	横向项目-技术开发	过参数化模型与因果学习的统计理论	林伟	201906	202105
14	企事业委托-北京智源人工智能研究院	横向项目-技术服务	非参数贝叶斯机器学习方法与概率编程	张志华	201906	202105
15	企事业委托-比尔及梅琳达·盖茨基金会	横向项目	A study on the outbreak and pattern of development of 2019 Novel Coronavirus, based on new mathematical and statistical models	周晓华	202005	202205

16	企事业委托-方正证券股份有限公司	横向项目-技术咨询	量化投资策略和证券公司客户量化分析研究	吴岚	201801	202101
17	企事业委托-华为技术有限公司	横向项目-技术开发	Large Scale Causal Structure Learning	耿直	201910	202210
18	企事业委托-中国某某研究所	横向项目	典型统计模型及可靠性评估算法梳理及确认	房祥忠	202006	202112
19	企事业委托-中债金融估值中心有限公司	横向项目-技术咨询	北京大学——中债估值研究中心	杨静平	201912	202412

（二）科研成果

自 2021 年 1 月至 12 月，统计科学中心教员共发表论文 146 篇。以下为部分代表性论文（按第一作者姓氏字母排序排列）：

- [1] Aït-Sahalia, Yacine; **Li, Chenxu**; Li, Chen Xu(2021). Closed-Form Implied Volatility Surfaces for Stochastic Volatility Models with Jumps. *Journal of Econometrics*, 222(1), 364-392.
- [2] Aït-Sahalia, Yacine; **Li, Chenxu**; Li, Chen Xu(2021). Implied Stochastic Volatility Models. *Review of Financial Studies*, 34(1), 394-450.
- [3] Chang, Jinyuan, **Chen, Song Xi**; Tang, Chengyong; Wu, Tongtong(2021). High-dimensional empirical likelihood inference, *Biometrika*, 108, 127-147.
- [4] Chen, Hui; Ren, Haojie; **Yao, Fang**; Zou, Changliang(2021). Data-driven selection of the number of change-points via error rate control. *Journal of the American Statistical Association*, doi: 10.1080/01621459.2021.1999820.
- [5] **Chen, Song Xi**; Peng, Lihua(2021). Distributive statistical inference for massive data. *The Annals of Statistics*, 49(5), 2851-2869.
- [6] Deng, Yuhao; Chen, Fangyi; Li, Yang; Qian, Kaihua; Wang, Rui; **Zhou, Xiao-Hua**(2021). A powerful test for the maximum treatment effect in through QT.QTc studies. *Statistics in Medicine*, 40, 1947-1959.
- [7] Eviatar B.Procaccia; **Zhang, Yuan**(2021). On sets of zero stationary harmonic measure. *Stochastic Processes and their Applications*, 131, 236-252.
- [8] Gao, Suwei; Shi, Qiang; Zhang, Yifan; Liang, Guixian; Kang, Zhixin; Huang, Baofeng; Ma, Dongyuan; Wang, Lu; Jiao, Jianwei; Fang, Xiangdong; Xu, Cheng-Ran; Liu, Longqi; Xu, Xun; Göttgens, Berthold; **Li, Cheng**; Liu, Feng. Identification of HSC.MPP expansion units in fetal liver by single-cell spatiotemporal transcriptomics. *Cell Research*, 2021b, 32, 38-53.
- [9] Guo, Wenchuan; **Zhou, Xiao-Hua**; Ma, Shujie(2021). Estimation of Optimal Individualized Treatment Rules Using a Covariate-Specific Treatment Effect Curve with High-Dimensional Covariates. *Journal of the American Statistical Association*, 116(533), 309-321.
- [10] Han, Shasha; Cai, Jun; Yang, Juan; Zhang, Juanjuan; Wu, Qianhui; Zheng, Wen; Shi, Huilin; Marco Ajelli; **Zhou, Xiao-Hua**; Yu, Hongjie(2021). Time-varying optimization of COVID-19 vaccine prioritization in

the context of limited vaccination capacity. *Nature Communications*, 12(1), 4673.

- [11] Hattori Satoshi; **Zhou, Xiao-Hua**(2021). Summary concordance index for meta-analysis of prognosis studies with a survival outcome. *Statistics in Medicine*, 40, 5218-5236.
- [12] Li, Hongkai; Jia, Jinzhu; Yan, Ran; Xue, Fuzhong; **Geng, Zhi**(2021). A causal data fusion method for the general exposure and outcome. *Statistics in Medicine*, 41(2), 328-339.
- [13] Li, Shaomin; Liu, Rui; Wang, Shuai; **Chen, Song Xi**(2021). Radiative Effects of Particular Matters on Ozone Pollution in Six North China Cities. *Journal of Geophysical Research-Atmosphere*, doi:10.1029/2021JD035963.
- [14] Li, Tingting; Li, Ruifeng; Dong, Xuan; Shi, Lin; Lin, Miao; Peng, Ting; Wu, Pengze; Liu, Yuting; Li, Xiaoting; He, Xxuheng; Han, Xu; Kang, Bin; Wang, Yinan; Liu, Zhiheng; Chen, Qing; Shen, Yue; Feng, Mingxiang; Wang, Xiangdong; Wu, Duojiang; Wang, Jian; **Li, Cheng**(2021). Integrative Analysis of Genome, 3D Genome, and Transcriptome Alterations of Clinical Lung Cancer Samples. *Genomics Proteomics Bioinformatics*, S1672-0229(21)00096-6. doi:10.1016/j.gpb.2020.05.007
- [15] Li, Wei; **Geng, Zhi; Zhou, Xiao-Hua**(2021). Causal mediation analysis with sure outcomes of random events model. *Statistics in Medicine*, 40 (17), 3975-3989.
- [16] Li, Xiang; Chen, Kun; Wang, Shusen; **Zhang, Zhihua**(2021). Communication-efficient distributed SVD via local power iterations. *International Conference on Machine Learning*, 139(2640-3498), 139.
- [17] Li, Xinyu; **Miao, Wang**; Lu, Fang; **Zhou, Xiao-Hua**(2021). Improving efficiency of inference in clinical trials with external control data. *Biometrics*, doi: 10.1111/biom.13583.
- [18] Liang, Decai; Huang, Hui; Guan, Yongtao; **Yao, Fang**(2021). Test of weak separability for spatially stationary functional field. *Journal of the American Statistical Association*, doi: 0.1080/01621459.2021.2002156.
- [19] Lin, Dachao; Sun, Ruoyu; **Zhang, Zhihua**(2021). Faster directional convergence of linear neural networks under spherically symmetric data. *NeurIPS*. <https://openreview.net/forum?id=Q9hZdUBTC9S>.
- [20] Lin, Dachao; Ye, Haishan; **Zhang, Zhihua**(2021). Greedy and random quasi-newton methods with faster explicit superlinear convergence. *NeurIPS*. <https://arxiv.org/abs/2104.08764v2>.
- [21] Lin, Zhenhua; **Yao, Fang**(2021). Functional regression on the manifold with contamination. *Biometrika*, 108(1), 167-181.
- [22] Liu, Rongli; **Ren, Yanxia**; Song, Renming; Sun, Zhenyao(2021). Quasi-stationary distributions for subcritical superprocesses, *Stochastic Processes and Their Applications*, 132, 108-134.
- [23] Luo, Shanshan; Li, Wei; **He, Yangbo**(2021). Causal inference with outcomes truncated by death in multiarm studies. *Biometrics*. Advance online publication. doi:10.1111/biom.13554.
- [24] Luo, Xin; Liu, Yuting; Dang, Dachang; Hu, Ting; Hou, Yingping; Meng, Xiaoyu; Zhang Fengyun; Li, Tingting; Wang, Can; Li, Min; Wu, Haixu; Shen, Qiushuo; Hu, Yan; Zeng, Xuerui; He, Xiechao; Yan, Lanzhen; Zhang, Shihua; **Li, Cheng**; Su, Bing. 3D Genome of macaque fetal brain reveals evolutionary innovations during primate corticogenesis. *Cell*, 184(3), 723-740.e21.
- [25] Ma, Yingying; Guo, Shaojun; **Wang, Hansheng**(2021). Sparse spatio-temporal autoregressions by profiling and bagging. *Journal of Econometrics*, doi: [10.1016/j.jeconom.2020.10.010](https://doi.org/10.1016/j.jeconom.2020.10.010).
- [26] **Miao, Wang**; Hu, Wenjie; Ogburn, Elizabeth L.; **Zhou, Xiao-Hua**(2021). Identifying effects of multiple causes with an unobserved confounder. *Journal of the American Statistical Association*, <https://arxiv.org/pdf/2011.04504.pdf>.

- [27] **Miao, Wang**; Li, Wei; Hu, WenJie; Wang, Ruoyu; **Geng, Zhi**(2021). Invited commentary: estimation and bounds under data fusion. *American Journal of Epidemiology*, doi: [10.1093/aje/kwab194](https://doi.org/10.1093/aje/kwab194).
- [28] Qiu, Yumou; Tao, Jing; **Zhou, Xiao-Hua**(2021). Inference of heterogeneous treatment effects using observational data with high-dimensional covariates. *Journal of the Royal Statistical Society. Series B, Statistical methodology* (1369-7412), 83 (5), 1016.
- [29] Qiu, Yumou; **Zhou, Xiaohua**(2021). Inference on Multi-level Partial Correlations based on Multi-subject Time Series Data. *Journal of the American Statistical Association*, doi: [10.1080/01621459.2021.1917417](https://doi.org/10.1080/01621459.2021.1917417).
- [30] Qu, Molong; Xiong, Liang; Lyu, Yulin; Zhang, Xiannian; Shen, Jie; Guan, Jingyang; Chai, Ppeiuyan; Lin, Zhongqing; Nie, Boyao; **Li, Cheng**; Xu, Jun; Deng, Hongkui. Establishment of intestinal organoid cultures modeling injury-associated epithelial regeneration. *Cell Research*, 2021a. 31(3), 259-271
- [31] **Ren, Yanxia**; Song, Renming; Zhang, Rui(2021). The extremal process of super-Brownian motion. *Stochastic Processes and Their Applications*, 137, 1-34.
- [32] Srinivasan, Arun; Xue, Lingzhou; **Zhan, Xiang**(2021). Compositional knockoff filter for high-dimensional regression analysis of microbiome data. *Biometrics*, 77(3),984-995.
- [33] **Tu, Yundong**; Liang, Hanying; Wang, Qiying(2021). Nonparametric Inference in Quantile Cointegrations with Stationary Covariates. *Journal of Econometrics*, doi: [10.1016/j.jeconom.2021.06.002](https://doi.org/10.1016/j.jeconom.2021.06.002).
- [34] **Tu, Yundong**; Wang Ying(2021). Spurious Functional-coefficient Regression Models and Robust Inference with Marginal Integration. *Journal of Econometrics*, doi: [10.1016/j.jeconom.2020.12.010](https://doi.org/10.1016/j.jeconom.2020.12.010).
- [35] Wan, Yating; **Xu, Mingya**; Huang, Hui; **Chen, Song Xi**(2021). A spatio-temporal model for the analysis and prediction of fine particulate matter concentration in Beijing. *Environmetrics*, 32(1), e2648.
- [36] Wang, Feifei; **Zhang, Junni L.**; Li, Yichao; Deng, Ke; Liu, Jun S.(2021). Bayesian Text Classification and Summarization via A Class-Specified Topic Model. *Journal of Machine Learning Research*, 22(89), 1-48.
- [37] Wang, Tianying; Ling, Wodan; Plantinga Anna M; Michael C Wu; **Zhan, Xiang**(2021). Testing microbiome association using integrated quantile regression models. *Bioinformatics*, 38(2), 419-425.
- [38] Yang, Ying; **Yao, Fang**(2021). Online estimation for functional data. *Journal of the American Statistical Association*, doi: [10.1080/01621459.2021.2002158](https://doi.org/10.1080/01621459.2021.2002158).
- [39] Ye, Haishan; Luo, Luo; **Zhang, Zhihua**(2021). Approximate newton methods, *Journal of Machine Learning Research*, 22.
- [40] Yu, Jun; Wang, HaiYing; **Ai, Mingyao**; Zhang, Huiming(2021). Optimal Distributed Subsampling for Maximum Quasi-Likelihood Estimators with Massive Data. *Journal of the American Statistical Association*, doi: [10.1080/01621459.2020.1773832](https://doi.org/10.1080/01621459.2020.1773832).
- [41] **Zhang, Cheng**; Vu Dinh; Frederick A. Matsen IV(2021). Non-bifurcating phylogenetic tree inference via the adaptive LASSO. *Journal of the American Statistical Association*, 116(534), 858-873.
- [42] Zhao, Junlong; Liu, Xiumin; **Wang, Hansheng**; Leng, Chenlei(2021). Dimension reduction for covariates in network data. *Biometrika*, 20(10), 1-17.
- [43] Zheng, Zemin; Lv, Jinchu; **Lin, Wei**(2021). Nonsparse learning with latent variables. *Operations Research*, 69(1), 346-359.
- [44] 顾嘉, **陈松蹊**, 董倩, 邱宇谋(2021). 基于 vSEIdRm 模型的人口迁移以及离汉交通管控对新冠肺炎疫情发展的影响分析. *统计研究*, 2021, 38(9), 114-127.
- [45] **任艳霞**, 宋仁明(2021). 分枝马氏过程的脊柱分解与其应用. *中国科学: 数学*, 2021, 51(11), 1819-1844.

（三）学术交流

统计科学中心 2021 年通过线上线下相结合的形式，共组织学术报告 33 场，报告人来自国内外各高校和研究机构。其中来自海外的学者 15 场，国内的学者 18 场。

序号	时间	报告人	单位及职务	报告题目
1	2021/4/7	张涛	山东大学	生命历程视角下慢病纵向队列分析
2	2021/4/8	Shizhe Chen	加利福尼亚大学戴维斯分校	An Instrumental Variable Method for Point Processes with Application to Optogenetic Experiments
3	2021/4/15	侯琳	清华大学	Statistical Inference of Local Genetic Correlation in Human Complex Diseases
4	2021/4/21	杨立坚	清华大学	Oracally Efficient Estimation for Single-index Link Function with Simultaneous Confidence Band
5	2021/4/29	洪永淼	中国科学院	Estimation of High-dimensional Multivariate GARCH Models: A New Regularization Approach
6	2021/5/6	Stijn Vansteelandt	根特大学	Assumption-lean Inference for Generalised Linear Model Parameters
7	2021/5/13	林蔚	对外经济贸易大学	Seasonal adjustment of time series observed at mixed frequencies using singular value decomposition with wavelet thresholding
8	2021/5/20	孟澄	中国人民大学	Dimension Reduction Techniques in Optimal Transport Problems
9	2021/5/28	陈洛南	中科院生化细胞所	动力学的数据科学理论及应用
10	2021/6/3	Kenichi Nagasawa	华威大学	Treatment Effect Estimation with Noisy Conditioning Variables
11	2021/6/16	冯鑫	宽德投资联合创始人	“万花统”学术沙龙 量化投资那些事——行业发展、中外比较、统计方法的应用
12	2021/6/17	杨朋昆	清华大学	Learning Overparametrized Neural Networks and Statistical Models
13	2021/6/24	Ke Yuan	格拉斯哥大学	Modelling Intratumour Heterogeneity Across Genomes and Images with Machine Learning
14	2021/7/1	Xuan Wang	哈佛大学	Model Free Approach to Quantifying the Proportion of Treatment Effect Explained by a Surrogate Marker
15	2021/7/1	Xavier D'Haultfœuille	CREST-ENSAE	Fixed Effects Binary Choice Models with Three or More Periods

16	2021/7/6	Jae-kwang Kim	爱荷华州立大学	Statistical Inference with Neural Network Imputation for Item Nonresponse
17	2021/7/19	Yuhong Yang	明尼苏达大学	Goodness of Fit Assessment for Binary Classification Learning
18	2021/9/22	Kun Zhang	卡内基梅隆大学	Learning Causality and Using Causality for Learning
19	2021/9/29	刘汉中	清华大学	Regression analysis for covariate-adaptive randomization: a robust and efficient inference perspective
20	2021/10/14	王霞	中国人民大学	Specification Tests for Time-Varying Coefficient Models
21	2021/10/20	李伟	中国人民大学	Identification and estimation of nonignorable missing outcome mean without identifying the full data distribution
22	2021/10/21	张瑞勋	北京大学	Quantifying the Impact of Impact Investing
23	2021/10/28	Jonas Peters	哥本哈根大学	Exploiting Invariance: from Causal Discovery to Robust Prediction
24	2021/11/4	Theis Lange	哥本哈根大学	Dose finding with two active components
25	2021/11/11	占翔	北京大学	Statistical Methods for Microbiome Association Analysis
26	2021/11/17	Zhichao Jiang	马萨诸塞大学阿默斯特分校	Experimental Evaluation of Algorithm-Assisted Human Decision Making
27	2021/11/25	骆威	浙江大学	On efficient dimension reduction with respect to the interaction between two response variables
28	2021/12/2	刘玉坤	华东师范大学	Biased-sample empirical likelihood weighting: an alternative to inverse probability weighting
29	2021/12/8	Ying Ding	匹兹堡大学	New Statistical Development in Precision Medicine: from Targeted Treatment Development to Individualized Treatment Recommendation
30	2021/12/9	夏寅	复旦大学	A Locally Adaptive Weighting and Screening Approach To Spatial Multiple Testing
31	2021/12/15	马维	中国人民大学	A general theory of regression adjustment for covariate-adaptive randomization: OLS, Lasso, and beyond
32	2021/12/17	Per Johansson	乌普萨拉大学	Inference in experiments conditional on observed imbalances in covariates
33	2021/12/29	Anqi Zhao	新加坡国立大学	Reconciling Design-based and Model-based Causal Inferences for Split-plot Experiments

五、人才培养

（一）学生获奖情况

2021 年 1 月至 12 月，统计学科学生获奖颇丰，累计 36 人次获得国家、学校及学院的奖学金或荣誉称号奖励。其中，罗珊珊获得“国家奖学金”；邵凌轩、朱玉茹、李新宇等 12 人获得校级“三好学生”称号；李臻获得校级“优秀学生干部”称号；马辰辰、顾嘉、唐俊杰、李翔等 7 人获得“优秀科研奖”；罗霄荣获“学术创新奖”；另有金子捷、邓宇昊、赵朝熠、胡文杰等 14 人获得企业奖学金或院级奖学金。

（二）毕业生及论文

姓名	导师	论文题目
程元	吴岚	市场冲击模型与最优交易策略研究
传一健	吴岚	主动投资管理的概率统计建模及中国市场实证研究
杜孟坤	吴岚	多重检验的理论研究及其在资产定价模型中的应用
房庄颜	何洋波	因果图模型中的经验知识：表示、推理及应用
高家兴	蒋达权	单细胞转录组测序数据分析的随机建模
国乃心	房祥忠	关于深度学习模型对时间序列数据生成的研究
胡晓玉	姚方	多元与函数型数据分析中的非参数方法
李向阳	周晓华	基于深度学习的医疗信息抽取和医学图像识别
刘思雨	吴岚	订单簿价格模型和实证研究
王菲	艾明要	几类最优性准则下抽样与设计的构造问题
王涵	席瑞斌	拷贝数变异和突变指纹分析的统计方法
薛丁川	姚方	惩罚样条的理论性质及动态估计
张鑫	吴岚	基于排序学习的多因子多空策略研究

（三）毕业生去向

姓名	导师	研究方向	单位名称
程元	吴岚	金融统计	中银国际证券股份有限公司
传一健	吴岚	金融统计	中国工商银行股份有限公司
杜孟坤	吴岚	金融统计	圣托马斯大学
房庄颜	何洋波	金融统计、因果推断	小米通讯技术有限公司
高家兴	蒋达权	随机过程及其应用、非平衡统计物理、系统生物学	北京华为数字技术有限公司
国乃心	房祥忠	高维统计推断、因果推断	香港城市大学
胡晓玉	姚方	函数型数据分析，高维和复杂数据建模	新加坡国立大学
李向阳	周晓华	统计机器学习，互联网搜索	腾讯科技（北京）有限公司
刘思雨	吴岚	金融统计	上海知知智能科技有限公司
钱力	陈大岳	马氏过程与相互作用粒子系统	B 站 up 主
王菲	艾明要	试验设计与分析，大数据分析	中央民族大学
王涵	席瑞斌	生物信息学，生物统计	香港大学
薛丁川	姚方	因果推断及生物医学统计	香港中文大学（深圳）
张鑫	吴岚	金融统计	中信证券股份有限公司

六、组织架构

(一) 指导委员会



张平文

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



陈大岳

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



刘俏

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

（二）行政团队

1. 统计科学中心主任



姚方，统计科学中心主任、数学科学学院概率统计系主任，数理统计学会 IMS 理事、《加拿大统计期刊》主编；讲席教授、国家高层次专家；数理统计学会与北美统计学会会士；至今任 9 个国际统计学期刊编委，包括 JASA、AOS 等。

2. 统计科学中心副主任



周晓华，统计科学中心副主任，现任国际生物统计学会中国分会会长；中国现场统计研究会生物医学统计分会会长；讲席教授、国家高层次专家；美国科学促进会、数理统计学会、北美统计学会会士；中国数学会医学数学专业委员会主任委员。



王汉生，统计科学中心副主任，，光华管理学院商务统计与经济计量系教授，博导。北京大学商务智能研究中心、主任。美国统计学会会士，现为国际统计研究所、美国数理统计研究所、英国皇家统计协会以及泛华统计学会会员。

3. 统计科学中心主任助理



席瑞斌，统计科学中心主任助理，数学科学学院长聘副教授、研究员。2009年毕业于美国圣路易斯华盛顿大学，同年以助理研究员身份加入哈佛大学医学院从事生物医学信息学方面的研究。

4. 统计科学中心综合办公室



穆慧娜，统计科学中心综合办公室主任，2015年毕业于武汉大学经济与管理学院，2019年9月加入中心。



张苑新，统计科学中心秘书，2017年毕业于华中农业大学，2021年10月加入中心。

(三) 科学委员会



陈松蹊

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



李程

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



任艳霞

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



王汉生

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



杨静平

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



姚方

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



虞吉海

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



张志华

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



周晓华

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

(四) 中心教员

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

姓名	职务	研究方向	学校院系
涂云东	中心教员	金融计量理论、理论计量经济学、应用计量经济学	光华管理学院 商务统计与经济计量系
林 伟	中心教员	高维统计、统计机器学习	数学科学学院 概率统计系
宋晓军	中心教员	计量经济学理论及应用计量经济学	光华管理学院 商务统计与经济计量系
邓明华	中心教员	生物信息学、计算生物学	数学科学学院 概率统计系
张俊妮	中心教员	因果推断、贝叶斯分析、蒙特卡洛方法、数据挖掘及其应用	光华管理学院 商务统计与经济计量系
房祥忠	中心教员	生存分析、可靠性、纵向数据、基尼系数、空间统计等	数学科学学院 概率统计系
张 原	中心教员	互相作用粒子系统、随机游动的覆盖问题以及 DLA 研究上提出新的理论	数学科学学院 概率统计系
苗 旺	中心教员	大数据的因果推断方法及其在经济学、环境科学中的应用	光华管理学院 商务统计与经济计量系
张 成	中心教员	统计学习、贝叶斯推断、计算生物学	数学科学学院 概率统计系
占 翔	中心教员	生物统计与生物信息学、统计遗传、基于希尔伯特再生核空间的统计建模、高维分子组学数据分析与统计推断	北京国际数学研究中心 公共卫生学院生物统计系
张瑞勋	中心教员	金融数学、金融工程、机器学习在金融、科技中的应用，行为金融学和适应性市场理论	数学科学学院 金融数学系
陈 嵘	特聘教授	非线性和多变量时间序列分析、蒙特卡洛方法、统计计算和贝叶斯分析及其应用	Rutgers University
金加顺	特聘教授	大规模多元假设检验、癌症分类、变量选择、谱聚类和主成分分析（PCA）、图论和精度矩阵、网络分析、随机矩阵理论	Carnegie Mellon University