

团队概况

团队荣誉

科研成果

研究项目

团队记事

未来可期

陈松蹊团队 年度总结

**Song Xi Chen's Research Group
Annual Summary**

www.songxichen.com



20
25

团队概况

团队荣誉

科研成果

研究项目

团队记事

未来可期

第一部分

团队概况

Brief Introduction to the Group



ONE

指导老师

陈松蹊（清华大学统计与数据科学系，兴华讲席教授，中国科学院院士）

邱宇谋（北京大学数学学院概率统计系及统计科学中心，长聘副教授）

郑小谷（上海张江数学研究院、新西兰大气和水文研究所资深科学家）

团队成员

本科生 清华大学：李浩铖、赵烽洵

研究生 北京大学：贾尧、钱磊、冯宇宁

博士生 北京大学：苏武、丘竞昆、李天宇、丁心如、李和、肖泽霖、詹皓翔、陆启隆、边明远、陈涵玥、王昺皓；清华大学：于丁一、陈新佑、刘睿、朱梓芸

博士后 北京大学：孙杨、李虹霖；清华大学：李鑫桐

毕业生去向 孙浩轩（博士）哈尔滨工业大学助理教授 彭阿晓（硕士）江西省纪委监委
童培峰（博士）清华大学航空航天学院博士后
闫晗（博士）伦敦政治经济大学博士后
杨巧诗（本科）宾夕法尼亚大学博士在读

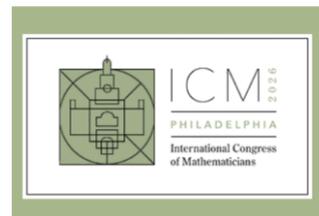


2025 学术活动

- 陈松蹊教授获颁北师香港浸会大学 (BNBU) 荣誉院士荣衔
- 受邀在2026国际数学家大会 (ICM) 作45分钟报告
- 连续入选2025年Stanford University全球2%顶尖科学家榜单
- 入选全球著名出版集团爱思唯尔 (Elsevier) 2025年 '中国高被引学者' (Highly Cited Chinese Researchers) 榜单



Song-Xi Chen
17 - Statistics, ML
University Chair Professor
Tsinghua University



2026国际数学家大会
Philadelphia, USA



北京生物医学统计与数据管理学会暨
第九届学术交流研讨会



香港中文大学 (深圳) 人工智能学
院成立仪式暨国际人工智能论坛



西太平洋多圈层相互作用数据集成
项目2025年度学术研讨会



第八届数字中国建设峰会
“数字人才培养分论坛”



全国数学、统计学及交叉学科的
博士后海河学术交流

一月

二月

四月

五月

七月

八月

九月

十月

十一月



第三届全国统计与数据科学
联合会议



2025中国国际大数据产业博览会
“大数据统计”交流活动



首届中国人民大学政府统计
国际研讨会



中国现场统计研究会数字孪生
与不确定性量化分会成立大会
暨UQ中心十周年研讨会



管理科学与工程学会
2025年年会

交叉
研究

推动数据科学交叉研究发展



9月17日，清华大学数据科学交叉研究院正式成立
(Institute for Interdisciplinary Data Science, Tsinghua University, THU IIDS)
陈松蹊教授担任院长



首届海洋统计与数据科学研讨会·厦门
(2026.1.11)



首届数据要素与智慧统计研讨会·赣州
(2026.1.18)

学科
建设

统计学“101计划”课程与教材建设



《高等学校统计学类专业人才培养
战略研究报告暨核心课程体系》
(简称“白皮书”)

发展历程

- 统计学“101计划”工作启动会 (清华大学)
- ABCD组分别召开统计学“101计划”课程体系线上研讨会
- 统计学“101计划”教材建设推进会 (江西景德镇)
- 统计学“101计划”核心课程骨干教师研修班 (云南昆明)
- 统计学“101计划”核心课程教材建设研讨会 (福建厦门)

2024年11月19日

2024年12月—2025年1月

2025年4月15日

2025年7月28-31日

2026年1月9-11日

统计学“101计划”工作启动会

教育部基础学科系列统计学“101”计划教材建设推进会
景德镇 2025-04-12

教育部统计学“101计划”核心课程骨干教师研修班

人才培养

数据科学人才培养与实践教学

为各专业本科生搭建跨学科数据分析基础

025年秋季学期，面向全校本科生开设《数据分析引论》课程，
并同步上线抖音平台作为公开课开放。

成为抖音平台**上榜播放量top10高校公开课**

清华大学统计与数据科学系
Department of Statistics and Data Science, Tsinghua University

数据分析引论

INTRODUCTORY DATA ANALYSIS

自9月16日，每周二 09:50-12:15
地点：清华三教 3200 主讲人：陈松蹊教授

非参数回归 数据科学 机器学习 深度学习 神经网络 数据挖掘 数据可视化 数据清洗 数据预处理 数据集成 数据融合 数据隐私 数据安全 数据伦理 数据治理 数据质量 数据生命周期 数据驱动决策 数据赋能 数据创新 数据生态 数据文化 数据素养 数据思维 数据意识 数据自信 数据自强 数据自立 数据自新 数据自胜 数据自爱 数据自敬 数据自尊 数据自重 数据自爱 数据自敬 数据自尊 数据自重

清华公开课-数据分析引论
427.5万 播放·更新至4集

第1集 跟着院士学数据分析！这门清华数据分析课在家就能上！
112:23 5.9万

第2集 跟着院士学数据分析！这门清华数据分析课在家就能上！
104:53 2.6万

第3集 跟着院士学数据分析！这门清华数据分析课在家就能上！
102:32 2.8万

第4集 跟着院士学数据分析！这门清华数据分析课在家就能上！
108:45 2.2万

前沿AI计算机课程最受欢迎

抖音精选公开课覆盖14个学科大类
86个一级学科
计算机科学与技术为最受欢迎的一级学科

播放数TOP10高校公开课

- Python语言基础与应用 (陈斌, 北京大学)
- 计算机组成原理 (刘宏伟, 哈尔滨工业大学)
- 大数据分析中的算法 (文再文, 北京大学)
- 实用Python程序设计 (郭炜, 北京大学)
- 数据分析引论 (陈松蹊, 清华大学)
- 流行病学基础 (吴涛, 北京大学)
- 离散数学概论 (陈斌, 北京大学)
- 高级语言程序设计 (车万翔, 哈尔滨工业大学)
- 积极心理学 (彭凯平, 清华大学)
- 人工智能与Python程序设计 (胡迪, 中国人民大学)

科普
传播

走进“统计与数据科学”主题演讲



启蒙数据分析思维，建立统计直观认知

4月

- 包头第九中学
- 北京师范大学数学科学学院
- 长沙第一中学

5月

- 福州第三中学

11月

- 中南财经政法大学

12月

- 南昌大学
- 北京十一中学

26年
1月

- 西南大学三亚中学

数智
感言

探讨AI时代统计学发展新方向

统计学不仅是工具，更是一种科学思维

【数据之上】陈松蹊院士：统计学如何“丈量万物”

黄大年茶思屋科技网



“统计学为数据治理与AI发展筑基”



“中国急需加速构建高质量的科学数据集”

“人工智能时代统计学将绽放异彩”



谈“统计学人才培养”的问题

政协提案

以统计分析能力赋能社会服务建设

运用数据分析解决实际问题，助力社会进步

建议加强仿制药上市后监管，丰富医保药种类



全国政协委员陈松璞：建议加强仿制药上市后监管，丰富医保药种类

时间：2025-03-06 来源：新京报



“我国仿制药上市的审评标准已同国际接轨，对药品上市前监管相对较严，但上市后的监管部门分散，影响监管效果。”今年全国两会，除了带来关于放宽人口生育约束、构建婚育激励体系的建议，全国政协委员、中国科学院院士陈松璞还关注仿制药监管。

他建议，加强药品上市后监管，推进药品再评价工作，尤其是进入医保集采的药品。同时，丰富医保药品种类的选择，在医院同时提供原研药和仿制药，设置不同的报销比例，视患者病情合理选择药品。

仿制药上市后监管部门分散

新京报：近期，关于国产仿制药的疗效问题受到广泛关注。此次，你提出了仿制药监管的建议。在调研中，你发现了什么问题？

陈松璞：一般来说，仿制药可以基本达到原研药的效果，同时降低医疗成本，减少百姓的医疗负担。目前国际上通用的仿制药评价是与相对应的原研药进行生物等效性试验，基于两种药品的动力学参数数据，构造等效性的统计学90%置信区间。我国仿制药上市的审评标准已同国际接轨，对药品上市前监管相对较严，但上市后的监管部门分散，影响监管效果。

仿制药被许可上市后，企业可能会出于各种原因，对产品进行变更，包括药品原料药、辅料、包装材料来源以及生产工艺、地点等多个方面。尽管国内外对变更管理均有相应的指导原则，但从管理事权的角度讲，存在差异。

具体来说，我国将药品审评划归国家药品监督管理局下属的国家药品审评中心，但备案和年报类的变更在省级药品监管部门。监管分离会导致企业将产品的变更拆分成很多个小变更，这些小变更更加之后，可能会影响产品质量，继而影响既往的生物等效性试验结果，造成变更后的产品无法再与原研药品生物等效。

加之现有药品监管数据库不健全，国家与地方、地方与地方之间对变更评价标准存在差异等问题同时存在。因此，始终生产与进行生物等效性试验批次相同质量的仿制药，应该是上市后监管工作的重点。

在药物监管能力方面，以统计团队为例，我国药监局生物统计团队只有20余人，而2024年批准新药39种，承担新药上市申请审评高达500余项。同时，相关工作人员还要承担新药临床试验方案沟通、指导原则起草和修订等工作，任务重、压力大，人员存在较大的流失风险。其他专业人员也存在力量不足问题。



21财经

打开APP

全国政协委员陈松璞：建议取消家庭生育数量限制，推行生育激励和提升人口素质 | 2025全国两会前瞻

21世纪经济报道 21财经APP 高谦荣 北京报道 2025-02-24 13:48

关联内容

全国政协委员陈松璞：建议加强药品上市后药效监管，给集采药品企业适度利润空间 | 2025全国两会前瞻

21世纪经济报道记者周潇泉 北京报道 2025年全国两会召开前夕，21世纪经济报道记者从全国政协委员、中国科学院院士陈松璞获得其今年全国两会拟提交的提案，其中一份是《关于放宽人口生育约束、构建婚育激励体系的建议》。

陈松璞指出，2022-2024年，中国年出生人口连续三年维持在900多万的水平。过去三年，我国总和生育率（TFR）介于1.02至1.08之间，仅为替代水平生育率的一半。出生人数的急剧下降将导致人口年龄结构失衡和劳动人口萎缩，加剧人口老龄化，也将削弱国家的生产、消费能力，进而影响整体经济动能。

近期，陈松璞团队采用人口统计的队列要素法，在假定未来人口出生性别比和分年龄死亡率保持在当前水平的基础上，测算了三个总和生育率方案下到本世纪末的中国总人口数、年出生人数、年龄中位数、劳动人口规模以及老龄化率，并对外推出了《中国2025-2100年人口预测与政策建议》。三个生育情景分别为：总和生育率保持在2022年的1.05（中等生育方案），10年线性下降到韩国2023年的0.72（超低生育方案），以及10年线性上升到1.31（高生育方案）。我国人口将面临出生人数骤减、劳动人口萎缩和严重超高龄化的挑战。

为避免超低生育模式发生，减缓中国人口将面临的严重结构失衡，陈松璞团队提出如下建议：

第一，修改人口与计划生育法，取消对家庭生育数量的限制。2021年修订的《中华人民共和国人口与计划生育法》仅明确一对夫妻可以生育三个子女。建议将“人口与计划生育法”更名为“人口与生育法”，全面放开生育，以适应新时期人口发展的迫切需要。

第二，降低法定结婚年龄，提高生育人口基数。目前，我国男22岁、女20岁的法定结婚年龄，是计划生育时代的产物。中国的法定婚龄为世界最高，国际上普遍的法定婚龄不超过18岁。建议将法定婚龄调至18岁，以提高生育人口基数，释放生育潜能。

第三，推行生育激励政策，抓住2025-2035黄金窗口期。鉴于目前鼓励城市人口生育措施的效率低于预期，建议在全国（包括农村地区）推行更多样、更大力度的激励政策，如每月对每个儿童提供现金补助和医疗支持到一定年龄等。有些政策在城市效果有限，但在农村有望产生效果。2025-2035年，我国育龄女性数量将稳定在2.9亿人-3.1亿人，建议抓住这一黄金窗口期，开展激励政策试点、生育意愿统计调查与数据分析，制定针对不同人群的生育鼓励方案。

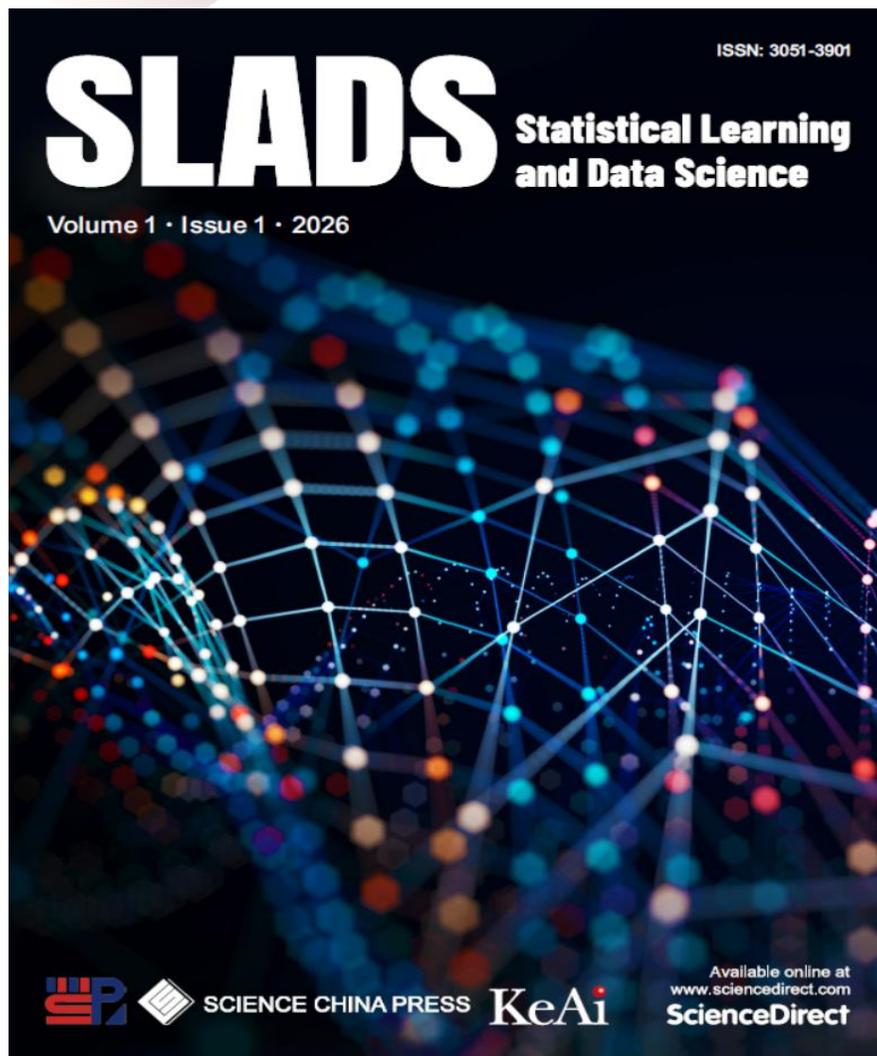


建议取消家庭生育数量限制，推行生育激励和提升人口素质

学术
共享

领航国内统计期刊，接轨国际学术前沿

搭建中外统计学界合作桥梁



SLADS是由陈松蹊教授和金加顺教授主编、中国科学院主办的高度国际化期刊，也是《中国科学》系列的第一本统计学期刊。**该刊成功入选"卓越行动计划"高起点新刊项目**，获得国内统一连续出版物号（CN号），**符合国家自然科学基金项目结题的国内期刊要求。**

Young Researchers Workshop 为面向统计与数据科学领域青年学者的小型学术研讨会，旨在为2020年后入职的高校青年教师和博士后（含2026年博士毕业生）提供一个专属的学术交流平台。





2025学术活动

学术报告



邱宇谋, 北京大学数学学院概率统计系及统计科学中心, 长聘副教授

论文发表

1. 中国运筹学会数学规划分会第十五届全国数学优化学术会议 (MOS 2025)
 - 时间: 2025年5月17-18日, 上海
 - 身份: 青年邀请报告人
 - 报告题目: Inference for Local Conditional Quantile Treatment Effect Curve with High-Dimensional Covariates "
2. 天元数学国际交流中心研讨会-大数据背景下统计前沿与交叉研讨会
 - 2025年2月20日
 - 报告题目: "Minimax Optimal Clustering and Signal Recovery Under Block Signals and Computational Constraint.
3. Versatile Differentially Private Learning for General Loss Functions, invited talk at Statistical Network and Beyond (SNAB) 2025, Workshop, Tokyo, Japan, June, 2025.
4. Multiple Robust Inference for Average Treatment Effects under High-dimensional Data, invited talk at Korean Statistical Society Summer Conference, Gyeongju, Korean, June, 2025
5. JCSDS 2025会议
 - 时间: 2025年7月12日
 - 分会场: IS008: Recent Developments of Prediction and Inference for High-dimensional Models
 - 报告题目: Minimax Detection Boundary and Sharp Optimal Test for Gaussian Graphical Model.
6. Physics-informed Statistical Data Fusion for Reconstructing 3D Current Fields of Oceanic Eddies, invited talk at Foundation Models and their Biomedical Applications, Banff, Canada, August, 2025.

1. Kwon, Y., Kim, J. K. and **Qiu, Y.*** (2025). Debiased calibration estimation using generalized entropy in survey sampling. Accepted by *The Journal of the American Statistical Association*.
2. Lu, Q., Chen, S. X. and **Qiu, Y.*** (2025). Versatile Differentially Private Learning for General Loss Functions. Accepted by *The Annals of Statistics*.
3. Sun, B., Yao, Y., **Qiu, Y.*** and Zhang, K. (2025). A Sample Efficient Conditional Independence Test in the Presence of Discretization. *ICML 2025*.
4. Sun, B., Yao, Y., Hao, H.-Y., **Qiu, Y.*** and Zhang, K. (2025). A Conditional Independence Test in the Presence of Discretization. *ICLR 2025*.
5. Cho, S., Kim, J. K. and **Qiu, Y.** (2025). Multiple Bias Calibration for Valid Statistical Inference under Nonignorable Nonresponse. *Biometrics* 81 ujaf044.
6. **Qiu, Y.*** and Gu, Y. (2025). F Statistics for High-Dimensional Inference of Linear Model. *The Bernoulli Journal* 31 2434-2458.
7. Zhang, S., Chen, S. X. and **Qiu, Y.** (2025). Mean Tests for High-Dimensional Time Series. *Statistica Sinica* 35 171-201.

团队概况

团队荣誉

科研成果

研究项目

团队记事

未来可期

第二部分

团队荣誉

The Group Awards and Honors



TWO

北大光华两名博士生荣获国际数理统计学会(IMS)旅行奖项



闫晗

北京大学光华管理学院
2020级统计学博士生

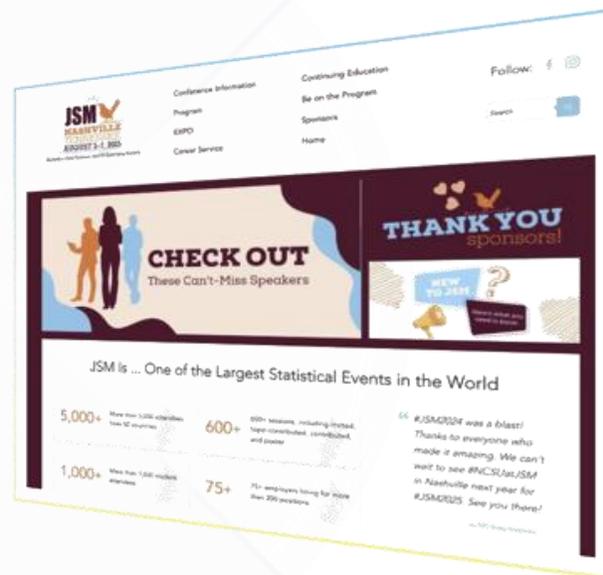
新研究员旅行奖 (New
Researcher Travel Award)



丘竞昆

北京大学光华管理学院
2021级统计学博士生

Hannan研究生旅行奖
(Hannan Graduate
Student Travel Award)



IMS资助他们赴美参加2025年Joint Statistical Meetings (JSM) 进行学术交流

国际数理统计协会 (IMS) 是全球三个权威统计学学术组织之一, 旨在发展和推广统计与概率的理论及应用, 出版发行高质量学术期刊, 如Annals of Statistics, Annals of Probability等。IMS颁发的荣誉奖项, 如IMS Fellow和各类Travel Awards, 都被国际统计学界高度。



国际顶级学术会议NeurIPS 2025海报展示



Versatile Differentially Private Learning for General Loss Functions

QiLong Lu, lu_qilong@stu.pku.edu.cn
 Song Xi Chen, sxchen@tsinghua.edu.cn
 YuMou Qiu, qiuyumou@math.pku.edu.cn

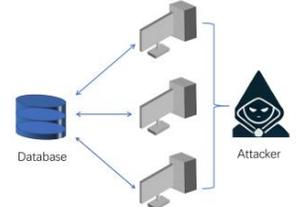
NeurIPS (全称Conference on Neural Information Processing Systems, 神经信息处理系统大会) 是全球人工智能与机器学习领域最具影响力的顶级学术会议之一, 与 ICML、ICLR 并称为机器学习领域“三大顶会”。



作者: 陆启隆

北京大学光华管理学院
2023级博士生

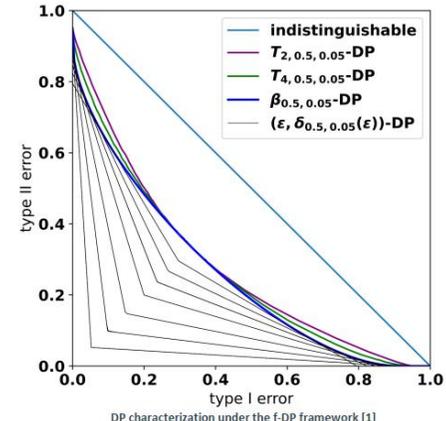
Non-versatility of Noisy-SGD



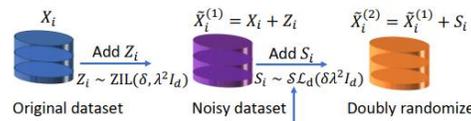
- Many hyperparameters need to be adjusted based on training results, such as epochs, learning rate, batch size, and the initial point.
- Different models need Noisy-SGD to be rerun, and the hyperparameters need to be reselected.
- **Therefore, the DP budget may be violated when Noisy-SGD is applied multiple times.**

Adding noise to data

Data: $X_1, \dots, X_n \in \mathbb{R}^d$
 Noise: $Z_1, \dots, Z_n, Z_i \in \mathbb{R}^d$ follows zero-inflated multivariate Laplace (ZIL)
 Published noisy data: $\tilde{X}_1^{(1)} = X_1 + Z_1, \dots, \tilde{X}_n^{(1)} = X_n + Z_n$
 What is the DP characterization for this mechanism?

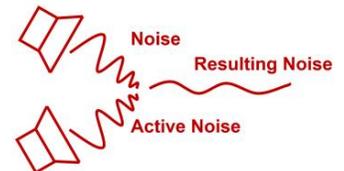


Versatile DP learning by double random(DR) corrected loss



Doubly random corrected loss: $\ell^{DR}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}; \theta; \delta) = (1 - \delta^{-1}) \ell(\tilde{X}_i^{(2)}, \theta) + \delta^{-1} \ell(\tilde{X}_i^{(1)}, \theta)$

The form of the DR corrected loss only involves basic arithmetic!



Unbiasedness: $\mathbb{E} \ell^{DR}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}; \theta; \delta) = \mathbb{E} \ell(X_i, \theta)$. Even discontinuities of ℓ are permissible!

- Why DR works?
- $Z \sim \text{ZIL}(\delta, \Sigma)$
 - $S \sim \mathcal{SL}_d(\delta \Sigma)$
 - Z and S are independent.
 - A function f .
 - $\varphi_f(t) = \int f(x) e^{it^T x} dx$. (Fubini's theorem)
 - $\varphi_S(t) = \frac{1}{1 + \frac{\delta t^T \Sigma t}{2}} = \frac{1}{(2\pi)^d} \int \varphi_f(t) e^{-it^T x} \left[\frac{1 - 1/\delta}{1 + \frac{t^T \Sigma t}{2}} + 1/\delta \cdot \frac{1 + \delta \frac{t^T \Sigma t}{2}}{1 + \frac{t^T \Sigma t}{2}} \right] dt$ (Fourier inversion)
 - $\varphi_Z(t) = \frac{1 + \frac{\delta t^T \Sigma t}{2}}{1 + \frac{t^T \Sigma t}{2}} = \frac{1}{(2\pi)^d} \int \varphi_f(t) e^{-it^T x} dt = f(x)$ (Fourier inversion)
 - $\varphi_{Z+S}(t) = \frac{1}{1 + \frac{t^T \Sigma t}{2}}$

Alternative approaches for smooth loss

$$\ell^{SDR}(\tilde{X}_i^{(1)}, \tilde{X}_i^{(2)}; \theta; \delta, \lambda) = \ell(\tilde{X}_i^{(1)}, \theta) - \frac{(1 - \delta)\lambda^2}{2} \sum_{k=1}^d \frac{\partial^2}{\partial x_k^2} \ell(\tilde{X}_i^{(2)}, \theta)$$

$$\ell^{SL}(\tilde{X}_i^{(2)}; \theta; \lambda) = \ell(\tilde{X}_i^{(2)}, \theta) - \frac{\lambda^2}{2} \sum_{k=1}^d \frac{\partial^2}{\partial x_k^2} \ell(\tilde{X}_i^{(2)}, \theta)$$

Simulation

Objective: Demonstrate the DR estimator's performance for M-estimation with non-smooth loss functions.

Loss Functions:

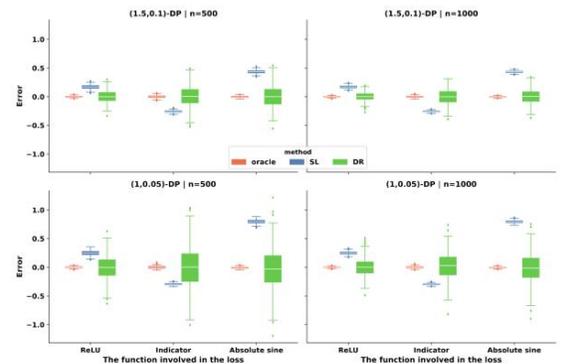
- ▶ $\ell_1(x, \theta) = (\theta - \text{ReLU}(x))^2$
- ▶ $\ell_2(x, \theta) = (\theta - \mathbf{1}_{[0.5, 1]}(x))^2$
- ▶ $\ell_3(x, \theta) = (\theta - |\sin 2\pi x|)^2$

Data: $X_i \sim U(0, 1)$, noisy and doubly randomized data generated with privacy parameters $(\delta, \lambda) = (0.1, 0.94)$ and $(0.05, 1.4)$.

True Parameters:

- ▶ $\theta_{0,1} = 0.5, \theta_{0,2} = 0.5, \theta_{0,3} = \frac{2}{\pi}$

Sample Sizes: $n = 500, 1000$.



1. DR: Always works. 2. SL: Fails, even if the second derivative of the loss function fails at only one point.

References

[1] Dong, J., Roth, A. and Su, W. J. (2022) Gaussian Differential Privacy. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 84, 3–37.
 [2] Avella-Medina, M., Bradshaw, C. and Loh, P.-L. (2023) Differentially private inference via noisy optimization. *Annals of Statistics*, 51, 2067–2092.
 [3] Koltz, S., Kozubowski, T. J. and Podgórski, K. (2001) The Laplace distribution and generalizations. Birkhäuser Boston, Inc., Boston, MA.
 [4] Wang, H. J., Stefanski, L. A. and Zhu, Z. (2012) Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika*, 99, 405–21.



团队概况

学生荣誉

科研成果

研究项目

团队记事

未来可期

第三部分

科研成果

Research Achievements

THREE

论文

丘竞昆等关于自正则化高斯逼近的研究被《统计年刊》接受发表



丘竞昆，第一作者，北京大学光华管理学院2021级博士生。其研究得到国家自然科学基金项目No.12292983和No.12071013的部分支持。

丘竞昆与陈松蹊教授、邵启满教授合作的研究论文《高斯逼近的自正则化Cramér中偏差定理》(Self-Normalized Cramér Type Moderate Deviation Theorem for Gaussian Approximation) 被《统计年刊》(The Annals of Statistics) 接受发表。

SELF-NORMALIZED CRAMÉR TYPE MODERATE DEVIATION THEOREM FOR GAUSSIAN APPROXIMATION

BY JINGKUN QIU^{1,a}, SONG XI CHEN^{2,b} AND QI-MAN SHAO^{3,c}

¹Guanghua School of Management, Peking University, ^ajkqiu@stu.pku.edu.cn

²Department of Statistics and Data Science, Tsinghua University, ^bsxchen@tsinghua.edu.cn

³Department of Statistics and Data Science, Shenzhen International Center for Mathematics, National Center for Applied Mathematics Shenzhen, Southern University of Science and Technology, ^cshaoqm@sustech.edu.cn

高斯逼近 (Gaussian Approximation) 刻画了感兴趣的高维统计量的分布与其二阶矩相对应的正态分布之间的逼近关系。经过Chernozhukov等十余年的发展，该方法已被广泛用于高维统计理论和复杂数据分析领域，包括但不限于假设检验、变点检测、回归模型、因子模型和图模型等。

具体而言，记 X_1, \dots, X_n 是独立同分布的零均值且方差有限的 p 维随机向量， $S_{n,j} = \sum_{i=1}^n X_{ij}$ 是第 j 个分量的求和， $\sigma_{n,j}^2 = \sum_{i=1}^n \mathbb{E}X_{ij}^2$ 是第 j 个分量的方差和。现有的高斯逼近结果通常可以推出，当 X_1 服从次指数或次高斯条件，且 $\log p = o(n^\epsilon)$ 时，

$$\left| \mathbb{P}\left(\max_{1 \leq j \leq p} S_{n,j}/\sigma_{n,j} > x\right) - \mathbb{P}\left(\max_{1 \leq j \leq p} Z_j > x\right) \right| \rightarrow 0$$

对 $x \in \mathbb{R}$ 一致成立，其中 (Z_1, \dots, Z_p) 是正态随机向量，且协方差矩阵与 X_1 的相关系数矩阵相同。然而在实践中，上述方法面临如下挑战：(1)用于标准化的方差 $\sigma_{n,j}^2$ 通常不可知，使得 $S_{n,j}/\sigma_{n,j}$ 无法直接使用；(2)次指数或次高斯假设对于一些重尾数据过于严苛；(3)当 x 较大时， $\mathbb{P}(\max_{1 \leq j \leq p} Z_j > x)$ 会以极快的速度趋于0，使得上式的绝对误差收敛结果不够精确。

为了解决上述三点挑战，文章提出使用自正则化的求和 $S_{n,j}/V_{n,j}$ 来替代标准化的求和 $S_{n,j}/\sigma_{n,j}$ ，并给出相对误差收敛结果，其中 $V_{n,j} = \sum_{i=1}^n X_{ij}^2$ 是第 j 个分量的平方和。具体而言，文章证明了当 X_1 的相关系数矩阵是单位阵，且三阶矩有限时，

$$\frac{\mathbb{P}(\max_{1 \leq j \leq p} S_{n,j}/V_{n,j} > x)}{\mathbb{P}(\max_{1 \leq j \leq p} Z_j > x)} \rightarrow 1$$

对 $0 \leq x \leq o(n^{1/6})$ 和 $p \geq 1$ 一致成立。文章内亦具体讨论了对于一般相关系数矩阵和仅有 $2 + \tau$ 阶矩有限时的类似收敛结果。

基于上述结果，文章提出一个新的高维 t 检验，并发现相较于现有的基于最大值的高维均值检验方法，在同等的高维情形下，新检验方法可以将以往所需的次高斯分布条件减弱为 $8/3$ 阶矩有限，由此提升检验的适用范围。

论文

陆启隆等关于普适差分隐私统计学习的研究被《统计年刊》接受发表



陆启隆，第一作者，北京大学光华管理学院2023级博士生。本研究得到了国家自然科学基金项目No.12292980和No.92358303的资助。

陆启隆与陈松蹊教授、邱宇谋副教授合作完成的研究论文《适用于一般损失函数的普适差分隐私统计学习》（Versatile Differentially Private Learning for General Loss Functions）被《统计年刊》（*The Annals of Statistics*）接受发表。另外，该论文亦被 *The Annals of Statistics* 推荐入选其首次设立的“期刊—会议交流通道”（journal-to-conference track），在本年度12月份举行的人工智能与机器学习领域的国际顶级学术会议NeurIPS 2025上进行海报展示（Poster Presentation）。

VERSATILE DIFFERENTIALLY PRIVATE LEARNING FOR GENERAL LOSS FUNCTIONS

BY QILONG LU^{1,a}, SONG XI CHEN^{2,b} AND YUMOU QIU^{3,c}

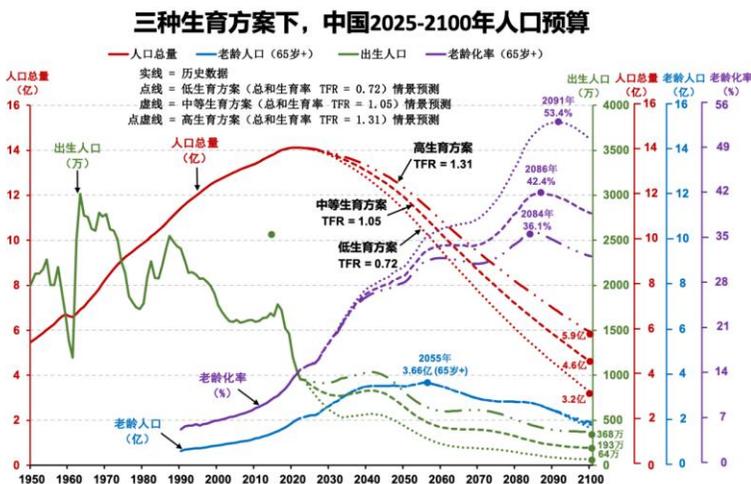
¹Guanghua School of Management, Peking University, ^alu_qilong@stu.pku.edu.cn

本文证明了 DR 修正损失估计量的一致性与渐近正态性，并给出了可用于推断的方差估计。与经典的高斯机制和拉普拉斯机制相比，该方法避免了积分或求导计算，因而在实现上更为简洁，且适用范围更广。进一步地，当研究对象限于二阶光滑损失函数时，本文提出了平滑双重随机（SDR）程序，其估计误差在一定意义下可与 Duchi 和 Ruan（2024）的结果相媲美。最后，文章系统讨论了所提估计量在隐私保护水平与估计效率之间的权衡关系。

中国人口研究专题报告

中国 2025-2100 年人口预测与政策建议

2025 年 1 月



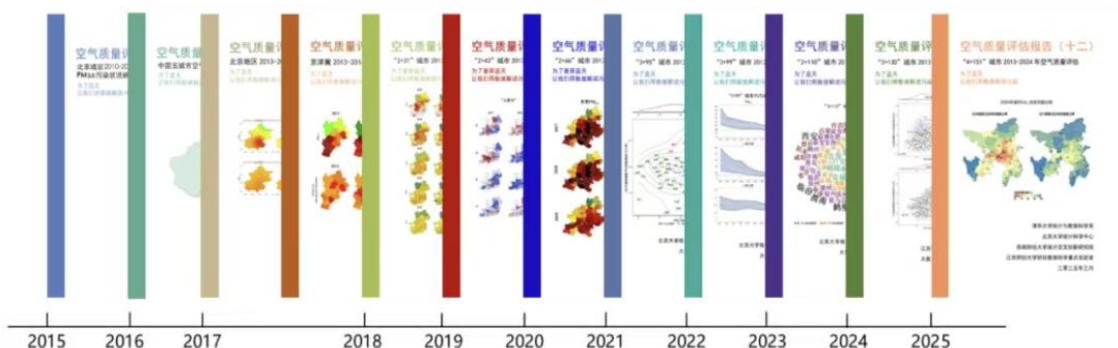
西南财经大学统计交叉创新研究院
西南财经大学数据科学与商务智能联合实验室
西南财经大学统计学院
清华大学统计与数据科学系



1月16日，由西南财经大学统计交叉创新研究院主办的《中国2025-2100年人口预测与政策建议》研究报告发布会暨中国人口现状与趋势分析研讨会在四川成都召开

陈松蹊院士正式发布了《中国2025-2100年人口预测与政策建议》研究报告，对我国2025年至2100年的人口总量和结构进行预测，从人口总量、出生人口、人口年龄结构、老龄化和劳动力等方面深入分析我国人口到本世纪末的变化趋势。

陈松蹊团队系列空气质量报告



为了重获蓝天，我们用数据解读污染！

空气质量评估报告（十二）

“4+151”城市 2013-2024 年空气质量评估

为了蓝天
让我们用数据解读污染

- 该报告首次采用结构化统计决策结合大语言模型自动生成，研究区域拓展至川渝地区，覆盖十六省市155城，涵盖全国57.2%人口。

一、后疫情期首现多数污染物大范围改善

二、PM_{2.5}尚未退场 臭氧挤占C位

三、“十四五”收官 多省空气质量“毕业”压力大

四、治“气”12年 污染红黑榜洗牌

空气质量评估报告系统



团队开发的「空气质量自定义评估报告系统」为空气质量评估系列报告的衍生产品，基于2013年起持续更新的小时国控空气质量监测数据和自2011年起的小时气象数据，采用统计学方法剔除气象因素影响，提供科学、可比的的城市空气质量评估服务。

系统支持自由选择城市+年份组合，获取专属分析报告。目前支持全国4直辖市+151城市查询，占全国地级及以上城市总数（339城市）的45.7%，常住人口占全国人口的57.2%（2020年第七次人口普查数据，不含港澳台）。定制化报告涵盖6类常规空气污染物浓度年际变化趋势、空间分布特征、区域排名、多污染物协同控制建议、气象影响分析等核心内容。

课题组成员

- 李虹霖 博士后
北京大学数学科学学院
- 刘奔 博士研究生
西南财经大学统计交叉创新研究院
- 刘俊 讲师、王佳 讲师
江西财经大学财经数据科学重点实验室
- 詹皓翔 博士研究生
北京大学数学科学学院
- 郭斌 教授
西南财经大学统计交叉创新研究院
- 陈松蹊 教授 课题负责人
清华大学统计与数据科学系

详情请参阅：
<https://www.songxichen.com/AQAssess/#/home>

发表论文

已接收的论文

- Wu Su, Xiaoyuan E, Zhao Jing, Song Xi Chen.(2026). Glider Path Design and Control for Reconstructing Three-Dimensional Structures of Oceanic Mesoscale Eddies. Proceedings of the Royal Society A. To appear.
- Song Xi Chen(2026) Statistical Signal Detection in High Dimension. Proceeding of International Meeting of Mathematicians, to appear.

- Chen, HY, Chen SX and JK Qiu (2025) Comments on 'Consistent climate fingerprinting' by McKittrick (2025). *Climate Dynamics* 63, 261 (2025).
- Pei Feng Tong, Haoran Yang, Xinru Ding, Yuchuan Ding, Xiaokun Geng, Shan An, Guoxin Wang, and Song Xi Chen.(2025) *Debiased Estimation and Inference for Spatial–Temporal EEG/MEG Source Imaging*. *IEEE TRANSACTIONS ON MEDICAL IMAGING*, VOL. 44, NO. 3
- Jingkun Qiu, Song Xi Chen, Qiman Shao. (2025) *Self-Normalized Cramér Type Moderate Deviation Theorem for Gaussian Approximation*. *Ann. Statist.* 53 (3) 1319 - 1346, June 2025.
- Zheng, Xiangyu and Chen, S.X. (2025) *Segmented Linear Regression Trees*, *Acta Mathematica Sinica*, Feb., 2025, Vol. 41, No. 2, pp. 498–521.
- Tong, P.F., Chen, S.X. and Tang, C.Y. (2025) *Multivariate calibrations with auxiliary information*, *Statistica Sinica*, 35, 1-18.
- Zhang, SY, S.X. Chen and Yumou Qiu (2025) *Mean Tests For High-dimensional Time Series*, *Statistica Sinica* 35, 171-201
- Kwon, Y., Kim, J. K. and Qiu, Y.* (2025). *Debiased calibration estimation using generalized entropy in survey sampling*. Accepted by *The Journal of the American Statistical Association*.
- Lu, Q., Chen, S. X. and Qiu, Y.* (2025). *Versatile Differentially Private Learning for General Loss Functions*. Accepted by *The Annals of Statistics*.
- Sun, B., Yao, Y., Qiu, Y.* and Zhang, K. (2025). *A Sample Efficient Conditional Independence Test in the Presence of Discretization*. *ICML 2025*.
- Sun, B., Yao, Y., Hao, H.-Y., Qiu, Y.* and Zhang, K. (2025). *A Conditional Independence Test in the Presence of Discretization*. *ICLR 2025*.
- Cho, S., Kim, J. K. and Qiu, Y. (2025). *Multiple Bias Calibration for Valid Statistical Inference under Nonignorable Nonresponse*. *Biometrics* 81 uja044.
- Qiu, Y.* and Gu, Y. (2025). *F Statistics for High-Dimensional Inference of Linear Model*. *The Bernoulli Journal* 31 2434-2458.

团队概况

团队荣誉

科研成果

研究项目

团队记事

未来可期

第四部分

研究项目

Research Project

FOUR

科研项目

国家自然科学基金重大研究计划

“西太平洋地球系统多圈层相互作用”的集成项目

西太平洋多圈层相互作用数据集成研究

(2024 - 2026)



西太平洋多圈层相互作用数据集成项目
2025年度学术研讨会 2025.4.10

国家自然科学基金重大项目

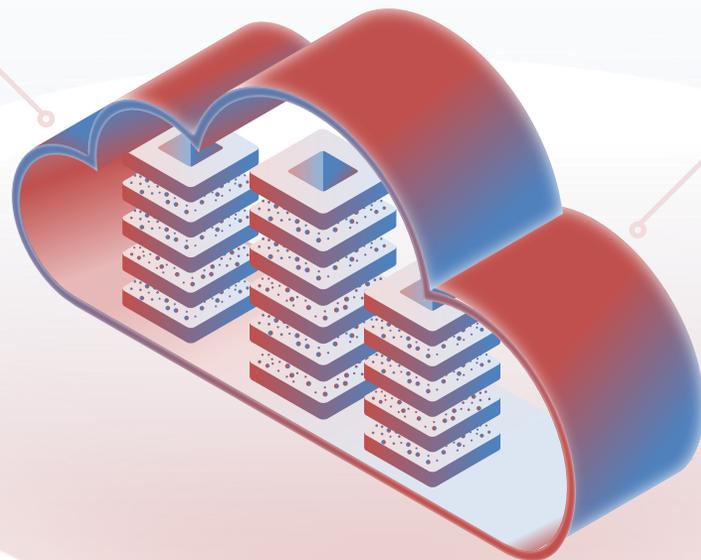
融汇海量观测数据的大气系统建模
与预报中的关键数学问题与算法

(2023 - 2027)

中国海洋大学物理海洋教育部重点实验室

西太平洋多圈层相互作用综合数据产品研制

(2025 - 2026)



专利项目



公开(公告)号:
CN114838845B

海洋涡旋温度场重建方法、装置、设备、介质和程序产品

发明人: 陈松蹊, 鄂潇原

本发明公开了一种海洋涡旋温度场重建方法。通过获取滑翔机在中尺度涡旋中采集的各涡旋位点坐标及温度信息, 计算插值函数, 进而推算其他位点的温度数据, 最终构建完整的涡旋温度场。该方法实现了海洋涡旋温度场的精确重建。

一种立体定向脑电图的电极触点标定方法及装置

发明人: 陈松蹊, 童培峰

本专利通过三维点云处理、广度优先搜索与主轴分析技术, 实现电极触点全自动化精准标定, 适配非理想电极形态。应用于癫痫诊疗全流程, 可提升诊疗效率与精度, 降低操作门槛。技术优势在于自动化程度高、减少医疗人力成本, 推动神经调控领域精准医疗发展。



公开(公告)号:
CN202411799829.2

团队概况

团队荣誉

科研成果

研究项目

团队记事

未来可期

第五部分

团队记事

The Group Activities



FIVE



5月 三位博士学位答辩：孙浩轩、童培峰、闫晗



生日庆祝



11月 秋游访古登居庸关长城，师生豪情入胸怀

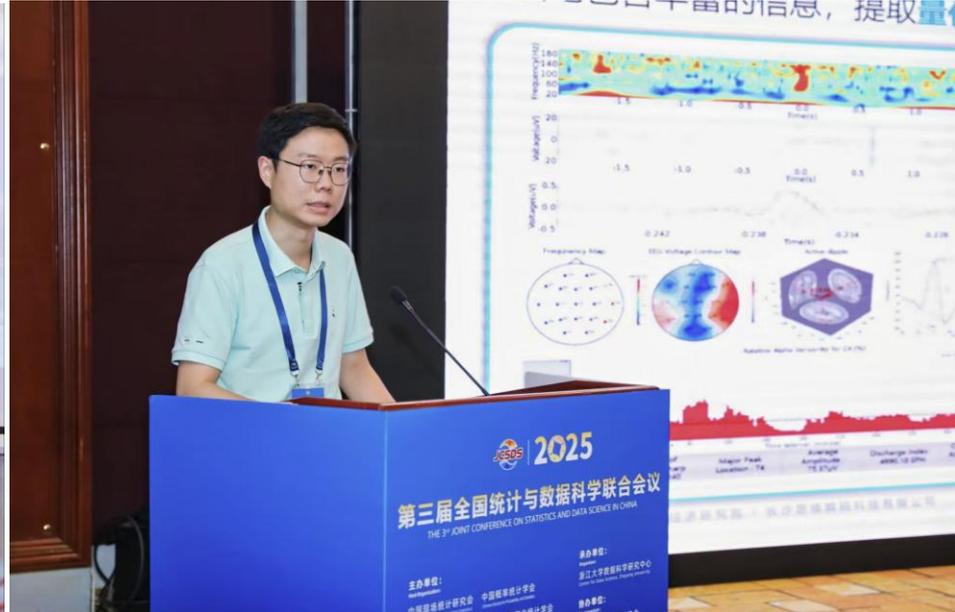




6月 北大-清华论坛上的钱磊



7月 JCSDS上的陈涵玥



7月 JCSDS上的童培峰



4月 西太项目年度会议上的王守霞



2026年1月 海洋统计研讨会上的朱梓芸和赵烽河



2026年1月 海洋统计研讨会上的苏武



7月 JCSDS上的导师郑小谷教授



陈松蹊教授与2025级本科生分享交流



陈松蹊 院士
清华大学统计与数据科学系系主任

陈松蹊教授获颁
北师大香港浸会大学 (BNBU)
荣誉院士荣衔



团队概况

团队荣誉

科研成果

研究项目

团队记事

未来可期



2026

新岁启幕，征程再扬！

愿诸君实心任事，不畏繁难，

驰而不息追真理，跃马扬鞭赴新程。

祝大家新春大吉，学术精进，硕果盈枝！