Shuting Shen

Mailing Address: 11 Peabody Terrace, Cambridge, MA 02138, Unit 608

+1 8577563165 shs145@g.harvard.edu

Education Background

Department of Biostatistics, Harvard University, Boston, United States 08/2018 - Present

- PhD in Biostatistics, GPA: 3.931
- Courses: Inference I, Inference II, Method I, Data Structures and Algorithms, Probability II, Multivariate Statistical Analysis, Advanced Regression, Optimization, Discrete Probability, Bayesian Methodology in Biostatistics
- Awards: Robert B. Reed Prize (awarded each year to the student(s) receiving the highest grade on the Department's written qualifying exam); ICSA Student Paper Award; NESS Student Research Award; ICSA Junior Research Award; WNAR Best Student Paper Award; ASA SLDS Student Paper Award.

School of Mathematical Sciences, Peking University (PKU), Beijing, China 09/2015-07/2018

- Bachelor in Mathematics, Major GPA: 95.9/100, Ranking: top 1%
- Core courses: Mathematical Analysis (99/100), Advanced Algebra (90/100), Functions of Real Variables and Functional Analysis (99/100), Abstract Algebra (94/100), Probability (93/100), Theory of Functions of a Complex Variable (98/100), Statistics (97/100), Ordinary Differential Equations (97/100).

School of Foundational Education, Peking University Health Science Center, Beijing, China 09/2013-07/2018

- Bachelor in Biomedical English, Major GPA: 90.4/100, Ranking: 1/36
- Core courses: Biostatistics (96/100), Epidemiology (89/100), General Biology (96/100), Organic Chemistry (97/100), Advanced Mathematics (92/100), Biochemistry (95/100), Histology and Embryology (99/100), Physiology (97/100), Immunology (98/100).
- Awards: Special Grade Scholarship (top 2% in 2016); First Grade Scholarship (top 4% in 2015 and 2017); Merit Student Awards at PKU (top 10% in 2015, 2016 and 2017)

Selected Publications

- Shen, S., and Lu, J. (2020). 'Combinatorial-Probabilistic Trade-Off: Community Properties Test in the Stochastic Block Models', *arXiv preprint arXiv:2010.15063* in submission to *IEEE Transactions on Information Theory* (a short version accepted by ICLR 2023 as spotlight)
- Shen, S., Lu, J., and Lin, X. (2021). 'Fast Distributed Principal Component Analysis of Large-Scale Federated Data' in submission to *Annals of Statistics*
- Shen, S., Chen, X., Fang, E. X., and Lu, J. (2023). 'Combinatorial Inference on the Optimal Assortment in Multinomial Logit Models', *arXiv preprint arXiv: 2301.12254* in submission to *Management Science*

Research Experience

Doubly-Robust Sequence Kernel Association Test (SKAT) for High-Dimensional Data 09/2022 – Present

*Advisors: **Xihong Lin**, Professor, Department of Biostatistics, Harvard University; **Junwei Lu**, Assistant Professor, Department of Biostatistics, Harvard University*

- Generalized the Sequence Kernel Association Test (SKAT) to the high-dimensional setting with data dimension larger than sample size.

- Established theoretical guarantee for the distributional convergence of the test statistic.
- Proved that the test is doubly-robust.
- Conducting simulation studies to evaluate performance of the test.

Inference on Optimal Offer Set in the Multinomial Logit Model 10/2021 - Present

Advisors: **Junwei Lu**, *Assistant Professor, Department of Biostatistics, Harvard University;* **Ethan X. Fang**, *Assistant Professor, Department of Statistics and Department of Industrial and Manufacturing Engineering, Penn State University;* **Xi Chen**, *Associate Professor, Department of Technology, Operations, and Statistics, New York University.*

- Conducted theoretical analysis on the statistical rate of regularized MLE for the preference score parameter under the Multinomial Logit (MNL) Model.
- Derived asymptotic normality results related to the optimal offer set for expected total revenue.
- Constructed confidence interval with multiplier bootstrap for the smallest optimal offer set and provided theoretical guarantee for its validity.
- Conducted simulation studies to evaluate the performance of the proposed algorithm.
- Finished the manuscript to be submitted *to Management Science*.

Distributed Fast Principle Component Analysis for Large-scale Data 06/2020 – 12/2021

Advisors: **Xihong Lin**, *Professor, Department of Biostatistics, Harvard University;* **Junwei Lu**, *Assistant Professor, Department of Biostatistics, Harvard University*

- Develop a scalable distributed PCA method that leverages FAST PCA and distributed estimation of principal eigenspaces, for big data with both large sample size and high dimensionality.
- Derived theoretical error bound for the proposed method that shows the method enjoys the same error rate as traditional full sample PCA.
- Characterized the asymptotic normality of the estimator and its inferential applications.
- Conducted extensive simulations comparing the algorithm with existing methods, showing the high computational efficiency and statistical accuracy of the proposed method.
- Implemented the algorithm on the 1000 Genomes Data.
- Finished the manuscript in submission to *Annals of Statistics.*

Hypothesis Testing for Clustering Properties Based on Likelihood Ratio Test 01/2019 – 10/2020

Advisor: **Junwei Lu**, *Assistant Professor, Department of Biostatistics, Harvard University*

- Developed a general inference method for clustering property tests based on the log likelihood ratio statistic in Homogeneous Stochastic Block Model.
- Proved the validity of the test in controlling type 1 and type 2 error rates.
- Provided a general lower bound characterizing difficulty of the clustering property test using novel metrics with side results on exact recovery rate.
- Implemented numerical experiments on both the synthetic data and the protein interaction application to show the validity of our method.
- Finished the manuscript "Combinatorial-Probabilistic Trade-Off: Community Properties Test in the Stochastic Block Models" in submission to IEEE Transactions on Information Theory.
- A conference version of this paper is accepted by International Conference on Learning Representations (ICLR) 2023 and selected as spotlight (notable-top-25%).

Skills and Others

Computer skills: R language, Python, Latex, Matlab, C language, SAS, Stata