Shuting Shen

Mailing Address: 2211 Hillsborough RD, Unit 2074, Durham, NC 27705, United States

+1 8577563165 ss1446@duke.edu

Current Position

Duke University, Durham, NC, United States 07/2023 – Present

- Postdoctoral Associate (jointly at Fuqua School of Business and Department of Biostatistics & Bioinformatics)
- Advisors: Dr. Ethan Xingyuan Fang (Department of Biostatistics & Bioinformatics) and Dr. Alexandre Belloni (Fuqua School of Business)
- Research Interest: Choice Model Asymptotic Properties, Operations Research Theories, Applied Probability.

Education Background

Department of Biostatistics, Harvard University, Boston, MA, United States 08/2018 – 05/2023

- PhD in Biostatistics, GPA: 3.931
- Advisors: Dr. Xihong Lin and Dr. Junwei Lu.
- 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

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 English (Medicine Track), 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

- **Shuting Shen** and Junwei Lu. (2020). 'Combinatorial-Probabilistic Trade-Off: Community Properties Test in the Stochastic Block Models', *arXiv preprint arXiv:2010.15063* accepted by *IEEE Transactions on Information Theory (short version is ICLR 2023 spotlight paper).*
- **Shuting Shen**, Junwei Lu, and Xihong Lin (2021). 'Fast Distributed Principal Component Analysis of Large-Scale Federated Data' under revision at *Journal of the American Statistical Association, Theory and Methods Section.*

- **Shuting Shen**, Xi Chen, Ethan X. Fang, and Junwei Lu (2023). 'Combinatorial Inference on the Optimal Assortment in Multinomial Logit Models', under revision at *Operations Research (abstract at EC'23).*
- Zhiwei Xu, Ziming Gan, Doudou Zhou, **Shuting Shen**, Junwei Lu and Tianxi Cai (2023). 'Inference of Dependency Knowledge Graph for Electronic Health Records', reject and resubmit at *Journal of the Royal Statistical Society, Series B.*

Awards and Honors

- ASA SLDS Student Paper Award, 2023
- WNAR Best Student Paper Award, 2022
- ICSA Junior Research Award, 2022
- ICSA Student Paper Award, 2022
- NESS Student Research Award, 2022
- Robert B. Reed Prize (awarded each year to the student(s) receiving the highest grade on the Department's written qualifying exam), 2020
- 1st Prize (Provincial), National High School Math League, 2012

Selected Research Projects

Anti-Concentration Inequalities for the Difference of Maxima of Gaussian Random Vectors 09/2023 – 08/2024

*Advisors: **Alexandre Belloni**, Professor, Fuqua School of Business, Duke University; **Ethan X. Fang**, Associate Professor, Department of Biostatistics & Bioinformatics, Duke University*

- Developed anti-concentration inequalities for the difference between the maxima of two Gaussian random variables.
- Employed new smoothing techniques that utilizes the derivative of a conditional probability for Gaussian random vectors as an intermediate bound.
- Unlike existing literature, theoretical conditions depend only on pairwise correlations rather than minimum eigenvalue of covariance matrix.
- Applied anti-concentration bounds to develop the central limit theorem for the maximizer of empirical processes indexed by a discrete set.
- Manuscript available upon request.

Inference on Optimal Offer Set in the Multinomial Logit Model 10/2021 – 01/2023

*Advisors: **Junwei Lu**, Assistant Professor, Department of Biostatistics, Harvard University; **Ethan X. Fang**, Associate Professor, Department of Biostatistics & Bioinformatics, Duke 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.

Distributed Fast Principal 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.
- 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.
- Conducted simulations comparing the algorithm with existing methods, showing the high efficiency and statistical accuracy of the proposed method.

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.