Qi Xu

Postdoctoral researcher / Carnegie Mellon University

qixu@andrew.cmu.edu | (217) 305-1708 | qi-xu.com

Research interests

In general, I am interested in developing principled statistical methodologies for **large-scale** and **heterogeneous** data with complex structures such as *block-wise missingness*, *distribution shift* and *high-dimensionality*. Recently, I am also interested in integrating **pre-trained AI** models in the loop of statistical analysis. I also collaborate with domain scientists in *genomics*, *maternal health*, etc.

Tools: semiparametric theory, representation learning, high-dimensional statistics, machine learning

Academic appointment

2024 - Present	**Carnegie Mellon University**, Pittsburgh, PA *Postdoctoral researcher* Host: Dr. Kathryn Roeder and Dr. Jing Lei

Education

2020 - 2024	**University of California, Irvine**, Irvine, CA *Ph.D. in Statistics* Advisor: Dr. Annie Qu
2018 - 2019	**University of Illinois Urbana Champaign**, Urbana, IL *M.S. in Statistics*
2014 - 2018	**Tongji University**, Shanghai, China *B.S. in Statistics*

Preprints

1. B.Bateni, Y.Yuan, **Q.Xu**, A.Qu. Distributional Treatment Effect Estimation for Heterogeneous Populations via Optimal Transport. Under review, **Journal of the Royal Statistical Society, Series B**

2. **Q.Xu**, L.Testa, J.Lei, K.Roeder. Blockwise Missing meets AI: A Tractable Solution for Semiparametric Inference. Under review, **Journal of the Royal Statistical Society, Series B**

3. **Q.Xu**, J.Lei, K.Roeder. Sparse Group Principal Component Analysis via Double Thresholding with Applications in Multi-Cellular Programs. Manuscript

4. L.Testa, **Q.Xu**, J.Lei, K.Roeder. Semiparametric Semi-supervised Learning for General Targets under Distribution Shift and Decaying Overlap. Reject & Resubmit, **Biometrika**

5. Y.Sui, **Q.Xu**, T.Li, Y.Bai, A.Qu. Partially-shared Imaging Regression on Integrating Heterogeneous Brain-Cognition Associations across Alzheimer's Diagnoses. Reject & Resubmit, **Journal of the American Statistical Association, Case Study & Application**

6. **Q.Xu**, A.Qu. Representation Retrieval Learning for Heterogeneous Data Integration. Major revision, **Journal of the American Statistical Association, Theory & Method**. *An earlier version was presented in the IMS Medallion Lecture by Annie Qu*

7. Y.Sui, **Q.Xu**, Y.Bai, A.Qu. Multi-task Learning for Heterogeneous Data via Integrating Shared and Task-Specific Encoding. Under review, **Journal of Machine Learning Research**

8. Y.Sui, **Q.Xu**, Y.Bai, A.Qu. Multi-task Learning for Heterogeneous Multi-source Block-Wise Missing Data. Minor revision, **Journal of Computational and Graphical Statistics**

Peer-reviewed publications

1. K.Wei, F.Xue, **Q.Xu**, Y.Yuan, Y.Zhang, G.Qin, A.Wani, A.Aiello, D.Wildman, M.Uddin, A.Qu. (2025+) Time-varying Mediation Analysis for Incomplete Data with Application to DNA Methylation Study for PTSD. **Annals of Applied Statistics**, just accepted

2. J.Rim, **Q.Xu**, X.Tang, Y.Guo, A.Qu. (2025) Individualized Time-varying Nonparametric Model with An Application in Mobile Health. **Statistics in Medicine**, 44: e70005.

3. **Q.Xu**, X.Cao, G.Chen, H.Zeng, H.Fu, A.Qu. (2024) Multi-Label Residual Weighted Learning for Individualized Combination Treatment Rule. **Electronic Journal of Statistics** 18(1), 1517-1548.

4. Y.Zhang, **Q.Xu**, N.Tang, A.Qu. (2024) Differential Private Data Release for Mixed-type Data via Latent Factor Model. **Journal of Machine Learning Research**, 25(116):1–37.

5. J.Zhang, F.Xue, **Q.Xu**, J.Lee, A.Qu. (2024) Individualized Dynamic Model for Multi-resolutional Data. **Biometrika**, 111(4): 1257-1275.

6. **Q.Xu**, H.Fu, A.Qu. (2024) Optimal Individualized Treatment Rules for Combination Treatments Under Budget Constraints. **Journal of the Royal Statistical Society, Series B**, 86(3), 714-741.

7. **Q.Xu**, Y.Yuan, J.Wang, A.Qu. (2024) Crowdsourcing Utilizing Subgroup Structure of Latent Factor Modeling. **Journal of the American Statistical Association, Theory & Method**, 119(546), 1192–1204.

Collaborative research

1. Y.Guo, **Q.Xu**, P.Kehoe, A.Qu. Daily Changes in Sleep Stages and Associated Cardiovascular Parameters During Pregnancy: Using a Wearable Device. **Frontiers in Global Women's Health**, accepted

2. J.Rim, **Q.Xu**, X.Tang, T.Jimah, Y.Guo, A.Qu. (2025+) Heterogeneous effects of physical activity on physiological stress during pregnancy. **PLOS Digital Health**, accepted

3. Y.Yuan, **Q.Xu**, W.Agaz, D.Janelle, C.Wang, W.Derek, U.Monica, A. Qu. (2024) Differentially Expressed Heterogeneous Overdispersion Genes Testing for Count Data. **PLOS One**, 19(7), e0300565.

4. Y.Guo, **Q.Xu**, N.Dutt, P.Kehoe, A.Qu. (2023) Longitudinal Objective Changes in Sleep Structure During Pregnancy. **Women's Health**, 2023; 19.

5. Y.Guo, **Q.Xu**, P.Kehoe, A.Qu. (2023) The Longitudinal Relationship between Objective Sleep and Cardiovascular Parameters During Pregnancy. **Sleep**, Volume 46, Issue Supplement1, Page A388

Ongoing projects

1. **Q.Xu**, K.Roeder, et al., Gene regulatory network discovery via post-selection inference with overlapped group lasso.

2. **Q.Xu**, L.Testa, J.Lei, K.Roeder. Nonparameteric regression under block-wise missingness

3. Y.Zhang, **Q.Xu**, A.Qu. Truth-inflated mechanism for differential privacy.

Teaching Experience

2024, 2022, 2021	**STAT295, Statistical Learning in Data Science**, University of California, Irvine *Teaching assistant* Graduate-level topic course. Lecture programming skills for data science and machine learning beginner, prepare homework, lead discussion.
2022	**Data296P/297P, Capstone Project**, University of California, Irvine *Teaching assistant* Master program capstone project. lead group discussion, host Q & A session for capstone project.
2021	**STAT120A, Introduction to Probability and Statistics I**, University of California, Irvine *Teaching assistant* Undergraduate-level probability course. Grade homework, host Q & A session.

Honors & Awards

2024 | **NSF-Simons Fellowship Award**, NSF-Simons Center for Multiscale Cell Fate Research

2024 | **UCI Statistics Fellowship Award for Methodology Research**, University of California, Irvine

2023 | **IMS Hannan Graduate Student Travel Award**, Institute of Mathematical Statistics

2023 | **ASA Student Paper Award Finalist**, ASA, Nonparametric Statistics Section

2022 | **ASA Student Paper Award Winner**, ASA, Statistical Learning and Data Science Section

Talks & presentations

1. (2025) Sparse Group Principal Component Analysis via Double Thresholding with Applications in Multi-Cellular Programs, *contributed talk*, **JSM**, Nashville, USA

2. (2025) Sparse Group Principal Component Analysis via Double Thresholding with Applications in Multi-Cellular Programs, *contributed talk*, **SSGG**, Minneapolis, USA

3. (2024) Mutli-Label Residual Weighted Learning for Individualized Combination Treatment Rule, *invited talk*, **JCSDS**, Kunming, China

4. (2024) Mutli-Label Residual Weighted Learning for Individualized Combination Treatment Rule, *invited talk*, **ICSA Applied Statistics Symposium**, Nashville, USA

5. (2024) Mutli-Label Residual Weighted Learning for Individualized Combination Treatment Rule, *contributed talk*, **ENAR**, Baltimore, USA

6. (2023) Estimation of Individualized Combination Treatment Rule, *invited talk*, **IMS Young Mathematical Scientist Forum**, Singapore

7. (2023) Trustworthy Machine Learning, *discussant*, **JSM**, Toronto, Canada

8. (2023) Optimal Individualized Treatment Rule for Combination Treatments, *contributed talk*, **JSM**, Toronto, Canada

9. (2023) Optimal Individualized Treatment Rule for Combination Treatments, *invited talk*, **North America Machine Learning, Optimization and Statistics Symposium**, Vancouver, Canada

10. (2022) Crowdsourcing Utilizing the Subgroup Structures in Latent Factor Modeling, *contributed talk*, **JSM**, Washington DC, USA

11. (2021) Multi-Categorical Crowdsourcing Using Subgroup Latent Factor Modeling, *invited talk*, **ICSA Applied Statistics Symposium**, Online

12. (2020) Multi-Categorical Crowdsourcing Using Subgroup Latent Factor Modeling, *contributed talk*, **JSM**, Online

Mentoring experience

2022-2024 | **UCInsipire summary research program**
Haoqun Cao, Jiayi Sun, Xieqing Yu, Xiaoke Cao, Geping Chen, Hanqi Zeng, Yuchen Li
Mentor undergraduate students for summary research, produce an EJS paper.

2023 | **Individual study**
Ziqi Zhang, Kevin Li
Mentor master students enrolled for individual study with my PhD advisor.

Professional services

2025 | **JSM session "Harnessing the power of large-scale and heterogeneous data with integrative analysis"**
Organizer

2022 | **IMS International Conference on Statistics and Data Science (ICSDS)**
Secretary, website manager

2021	**Brown Bag Seminar Series, Department of Statistics, UCI**
	Organizer

Reviewer experience

Journal: JASA, JRSSB, JCGS, EJS, JMLR, Statistics Sinica, Biometrics, Nature Communication, Journal of Nonparametric Statistics, Stat, Computational Statistics & Data Analysis, IEEE Transaction on Consumer Electronics

Conference: AISTAT, ICLR, AAAI