

## Education Background

PhD in Statistics and Data Science, Carnegie Mellon University  
Bachelor in Statistics, University of Science and Technology of China

GPA 4.08, 2018-(exp) May 2023  
GPA 3.86, July 2018

## Qualification

- 4+ Years research experience in statistical modeling and inference
- 4+ Years of programming experience, familiar with R, Python (Tensorflow, PyTorch), MATLAB, know of C, Git, SQL

## Work Experience

### Online experimentation in E-commerce with revenue and time constraints

June– Aug. 2020.

Applied Scientist Intern, AWS, Amazon. Supervisor: Lenon Minorics, Professor Guido Imbens.

- Constructed an new online experimentation algorithm that allows efficient trade-off between revenue constrain and time constrain in E-commerce utilizing recent advances in reinforcement learning and any-time valid inference.
- Proposed new algorithm to deal with unknown abrupt changes over time using Thompson Sampling and sequential change detector, which only cost  $O(1)$  for each updates, and has much lower regret comparing with other state-of-arts.

## Research Experience

My research interests take root in developing method and theory for high dimensional, nonparametric data analysis with the focus on data interactions, including **high-dim clustering**, **relational data modeling**, **online inference**, **graphical models**.

- Nonpara co-clustering in noisy high-dim mixture based on pairwise local dependency** Sept. 2020 –Present  
Under supervision of Professor Jing Lei & Kathryn Roeder.
  - Proposed to represent noisy high-dim mixture data as multi-layer network mixture, where each layer is the pairwise local dependence network among all dimensions in a sample. Mined the resulted multi-layer network mixture using tensor clustering, which gives clean sample clusters and sub-network clusters at the same time.
- Single-cell expression simulation incorporating gene interaction** ([bioRxiv](#), [code](#)) July. 2019 –May. 2020  
Supervisors: Professor Kathryn Roeder & Jiebiao Wang.
  - Constructed a new semiparametric scRNA-seq data simulation tool, which ensembles current state of art in first moment characterization, and fills in the gap of depicting higher order gene interaction using a copula model.
- Large-scale simultaneous (post-hoc) inference under dependence** Sept. 2019 – Present  
Supervisors: Professor Aaditya Ramdas & Gene Katsevich & Jelle Geoman.
  - Constructed a class of error control methods that allows for simultaneous inference and model selection under arbitrary dependence with only linear time computation.
  - Derived tighter calibration and the corresponding asymptotic power in our methods assuming a Gaussian dependence model. The theory suggests our algorithms class is rich enough to contain powerful strategy for various settings.
- Adaptive algorithms for online error control** Jan – Sept. 2019  
Supervisor: Professor Aaditya Ramdas.
  - Proposed new online FDR control algorithm that adapts to both signal proportion and conservative nulls, which outperforms current state of arts in terms of range as well as power. ([NeurIPS 2019](#), [code](#))
  - Extended the adaptive idea to develop new online FWER control methods. Formally proved substantial gains of power of the new methods, and derived closed form optima in a Gaussian sequence model. ([arxiv](#), [code](#))
- Recovering Graphical Structures with FDR control via Knockoffs** Oct. 2017– May. 2018  
Outstanding Graduation Thesis in undergrad. Supervisor: Professor Zemin Zheng.
  - Constructed new method for structure recovering in Gaussian graphical models with FDR control using knockoff filter. Extended the method to cases with additive measurement error using CoCoLasso.
- Brain task classification with Graph Neural Network guided by region connectivity** Jun.– Sept. 2017  
Undergraduate visiting summer research. Supervisor: Professor Jing Lei & Kehui Chen.
  - Proposed customized neural network models with regularization based on brain connectivity structures, to classified brains signals (MEG data), which is shown to reduce over-fitting and improve accuracy.

## Publication

- J. Tian, A. Ramdas. *ADDIS: an adaptive discarding algorithm for online FDR control with conservative nulls*. NeurIPS 2019.
- J. Tian, A. Ramdas. *Online control of the familywise error rate*. Statistical Methods in Medical Research 2021.
- J. Tian, J. Wang, K. Roeder. *ESCO: single-cell expression simulation incorporating gene co-expression*. Bioinformatics, minor revision.
- J. Tian, L. Minorics. *Flexible, Efficient and Robust online experimentation platform using MAB*. ALMC 2020, RL workshop.

## Coursework

- Statistics:** Measure Theory, Intermediate Statistics, Advanced Probability Theory, Reproducibility, Regression Analysis, Advanced Data Analysis, Stochastic Process, Time Series
- Machine Learning:** Introduction to Advanced ML, Statistical ML, Convex Optimization, Deep reinforcement learning
- Math:** Functional Analysis, Real Analysis, Complex Analysis, Mathematical Analysis, Linear Algebra
- Programming:** Statistical Computing, Data structure and Data base, C and Algorithm.