Jinjin Tian Email: jinjint@andrew.cmu.edu | Phone: (412)-616-6460 | Github: JINJINT

Education Background

PhD in Statistics and Data Science, Carnegie Mellon University *GPA 4.08, 2018-(exp) May 2023*
Bachelor in Statistics, University of Science and Technology of China *GPA 3.86, July 2018*

Qualification

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

Research Experience

My research interests take root in methods and theory for high dimensional, nonparametric data analysis with special focus on the interaction structure and temporal nature, specifically **high-dim clustering**, **relational data modeling**, **graphical structure recovery**, and **post-hoc**, **online inference**.

- **Signals recovery in noisy high-dim mixture via local structure learning** *Sept. 2020 –Present*
  Under supervision of Professor Jing Lei & Kathryn Roeder.
  - Proposed a statistics to capture local structure between a pair of features, which will only dominate iff the corresponding sample comes from a non-noise mixture component, in which the feature pairs are both relevant features.
  - Proved exact recovery of signals in a single non-noise mixture setting using a non-parametric estimation, under even impossible scenarios for canonical methods like sparse PCA. Working on proofs for multiple non-noise mixture cases.
- **Large-scale post-hoc model selection and inference under dependence** *Sept. 2019 – May. 2020*
  Supervisors: Professor Aaditya Ramdas & Eugene 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 in various settings.
- **Single-cell gene expression data modeling incorporating gene interaction** (bioRxiv, package)
  Supervisors: Professor Kathryn Roeder & Jiebiao Wang. *July. 2019 –May. 2020*
  - Modeled the single cell gene expression data to not only characterize various type of heterogeneity through a hierarchical model, but also depict higher order gene interaction using a copula model. Wrapped up the model as a simulation software which also allows mimicking real data via fitting the model using a moment matching method.
- **Adaptive algorithms for online error control** *Jan – Sept. 2019*
  Supervisor: Professor Aaditya Ramdas.
  - Constructed new algorithm for online false discovery rate control via adapting to both signal proportion and noise level, which outperforms current state of arts in terms of applicable range as well as power. (NeurIPS 2019, code)
  - Extended the adaptive idea for more stringent familywise error rate control. Formally proved substantial gains of power for the new methods, and derived closed form optima of hyperparameters in a Gaussian sequence model. (Statistical Methods in Medical Research 2021, code)
- **Recovering Graphical Structures with FDR control via Knockoffs** *Oct. 2017– May. 2018*
  Outstanding 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**
  Undergrad visiting summer research. Supervisors: Professor Jing Lei & Kehui Chen. *Jun.– Sept. 2017*
  - Proposed graphical neural network models with regularization based on brain connectivity structures, which is shown to reduce over-fitting and improve accuracy in classification of brains signals (MEG data).

Work Experience

Online experimentation in E-commerce with revenue and time constraints (package) *June– Aug. 2020.*
Applied Scientist Intern, AWS, Amazon. Supervisor: Lenon Minorics, Professor Guido Imbens.

- Constructed a 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.
- Accepted and presented at the RL workshop of the internal Amazon Machine Learning Conference 2020.

Publications

- **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, to appear.**
- **J. Tian**, L. Minorics. *Flexible, Efficient and Robust online experimentation platform using MAB.* **AMLC 2020, RL workshop.**
- **J. Tian**, A. Ramdas. *ADDIS: an adaptive discarding algorithm for online FDR control with conservative nulls.* **NeurIPS 2019.**