Zhenghao Zeng

Home page: tigerzhzeng.com

Email : zhenghaz@andrew.cmu.edu
323, 5000 Forbes Ave, Pittsburgh, PA 15213

Education

- **Carnegie Mellon University** — Pittsburgh, PA
  PhD in Statistics; GPA: 4.14/4.30; Advisor: Edward H. Kennedy — *Aug 2020 - May 2025*
- **Carnegie Mellon University** — Pittsburgh, PA
  Master in Statistics — *Aug 2020 - May 2021*
- **University of Science and Technology of China** — Hefei, Anhui
  Bachelor of Statistics; GPA: 4.18/4.30 ; ***summa cum laude (Guo Moruo Scholarship 1/76)*** — *Aug 2016 - June 2020*

Skills Summary

- **Languages**: 5+ Years of programming experience, familiar with Python, R, MATLAB
- **Experience**: 5+ Years of research experience in statistics and machine learning

Work Experience

- **Research Scientist/Engineer Intern, Data Science Lab, Adobe** — San Jose, CA
  Joint work with Prof. Avi Feller (Berkeley) and David Arbour (Adobe Research) — *May 2023 - Aug 2023*
  - **Goal**: Propose a doubly robust estimator for continuous treatment effects/ dose response function with the help of surrogate outcomes.
  - **Identification**: Provided sufficient conditions to identify continuous treatment effects and formalized the problem as a nonparametric regression problem via doubly robust mapping.
  - **Estimation Guarantee**: Established theoretical results, including oracle estimation theory and asymptotic normality, for the estimator proposed.
  - **Real Application (ongoing)**: Applied proposed method to real data.

Research Experience

- **Causal Inference with High-dimensional Discrete Covariates** — Pittsburgh, PA
  Research Assistant - Prof. Edward H. Kennedy and Prof. Sivaraman Balakrishnan — *Nov 2022 - Current*
  - **Upper Bounds**: Characterized sufficient and necessary conditions for conventional plug-in style causal effect estimators to be consistent in high-dimensional discrete setting. Proposed additional structures under which the estimators enjoy faster rate and can be consistent in non-classic high-dimensional regime.
  - **Lower Bounds (ongoing)**: Examined minimax lower bound in high-dimensional regime and obtained partial answer. Working on tightening the bounds.
- **Flexibly Estimating and Interpreting Heterogenous Effects of Laparoscopic Surgery** — Pittsburgh, PA
  Research Assistant - Prof. Edward H. Kennedy and Prof. Luke Keele (University of Pennsylvania) — *Nov 2022 - Current*
  - **Methods**: Proposed different ways to characterize and interpret effect heterogeneity with doubly robust estimation and inferential tools.
  - **Application**: Applied state-of-the-art causal inference techniques to estimating and visualizing the heterogeneous effects of laparoscopic Surgery for Cholecystitis Patients.
- **Covariate-assisted Bounds on Causal Effects with Instrumental Variables** — Pittsburgh, PA
  Research Assistant - Prof. Edward H. Kennedy; Manuscript — *Aug 2022 - Apr 2023*
  - **Derived Bounds**: Applied linear programming to obtain bounds on average treatment effects under a valid instrument design given covariates.
  - **Methods to Estimate the Bounds**: Developed two methods to efficiently estimate the bounds as non-smooth functionals: Either approximated the bounds with smooth functionals and evaluated approximation error or assumed a margin condition and applied standard doubly robust estimation based on influence functions.
  - **Extension**: Extended the bounds to continuous response case.
- **Efficient Generalization and Transportation** — Pittsburgh, PA
  Research Assistant - Prof. Edward H. Kennedy; Manuscript — *May 2021 - July 2022*
  - **Goal**: Develop methods to generalize/transport the causal effects from a randomized trial/observational study to the target population.
  - **Efficiency Theory**: Formalized the problem as estimating statistical functionals. Derived the first-order influence functions and doubly robust estimators. Established the asymptotical normality of doubly robust estimator under additional assumptions.
  - **Minimax Optimality**: Derived the minimax lower bounds of the target functionals. Further proposed a high-order estimator that can achieve the minimax rate in a broad regime.

- **Application**: Applied the doubly robust estimator to transport the causal effects of dietary intake on adverse pregnancy outcomes from an observational study to the whole U.S. female population.
- **Award**: **2023 ENAR Distinguished Student Paper Award**

- **A Tensor-EM Method for Large-Scale Latent Class Analysis with Binary Responses** Ann Arbor, MI
  Research Assistant - Prof. Gongjun Xu; Published in Psychometrika *July 2019 - March 2021*
  - **Tensor-EM Algorithm**: Derived the tensor structure of cross-moments in latent class model. Applied tensor power method to the cross-moments and obtained an moment-based initial estimator for EM algorithm.
  - **Consistency**: Established the clustering consistency (i.e. consistency of estimating each individual's latent class membership) and consistency of item parameters for joint Maximum Likelihood Estimator (MLE).
  - **Application**: Applied Tensor-EM method to an educational assessment dataset. Clustered the students into three classes and evaluated the performance of students in each class.
- **On the Phase Transition of Wilks' Phenomenon** Ann Arbor, MI
  Research Assistant - Prof. Gongjun Xu; Published in Biometrika *July 2019 - May 2020*
  - **Chi-square Approximation**: Developed sufficient and necessary conditions for Wilks' phenomenon (i.e. twice the negative loglikelihood ratio asymptotically approaches a Chi-square distribution) under popular tests on multivariate mean and covariance structures in high dimensional setting.
  - **Asymptotic Bias**: Analyzed the accuracy of chi-squared approximations by deriving the asymptotic biases when the dimension is large and Wilks' theorem fails to hold. Provided helpful insights into the use of chi-squared approximations in scientific practices.

Academic Service

- **Reviewer**: Journal of American Statistical Association, Electronic Journal of Statistics, American Journal of Epidemiology

Coursework

- **Mathematics and Probability**: Mathematical Analysis(A+), Linear Algebra(A+), Real Analysis(A+), Functional Analysis(A+), Advanced Probability Theory(A+), Stochastic Process(A+), Probability Limiting Theory(A+)
- **Statistics**: Mathematical Statistics(A+), Regression Analysis(A+), Multivariate Analysis(A+), Bayesian Analysis(A+), Nonparametric Statistics(A+), Advanced Statistical Theory(A)
- **Machine Learning**: Advanced Machine Learning(A+), Convex Optimization(A+), Probabilistic Graphical Models(A+), Deep Learning(A+)
- **Programming**: Statistical Computing(A+), Deep Learning System(A), Foundations of Algorithms(A+)