Xinyi Wang

Mobile: (708) 5363707 | Email: xinyiwang@uchicago.edu

EDUCATION

The University of Chicago, Chicago   Sep 2023 – Present
MS in Statistics, expected in June 2025

The Chinese University of Hong Kong, Hong Kong   Sep 2019 – Jul 2023
B.Sc. with honors (first class) in Risk Management Science, Minor in Statistics
Stream: Risk Analytics, Overall GPA: 3.82/4.00 (Rank 1/30), Major GPA: 3.88/4.00
- Academic honours: Dean's Honors List (2019 – 2020, 2020 – 2021, 2021 – 2022, 2022 – 2023)
- Scholarships: Bang How Memorial Scholarships (2020 – 2021, 2 out of 800); Chung Chi College Class Scholarship (2020 – 2021, 2021 – 2022); Department of Statistics Scholarships (2021 – 2022, 2022 – 2023, 2 out of 40); Q.W. Lee Scholarships (2022 – 2023, 2 out of 800)

PUBLICATION

Molei Liu*, **Xinyi Wang***, and Chuan Hong. A Semiparametric Approach for Robust and Efficient Learning with Biobank Data. preprint at arXiv. https://arxiv.org/abs/2404.01191 (*: equal contribution).

RESEARCH EXPERIENCE

Robust Empirical Bayesian Hierarchical Model for Covariate Adjustment in Stratified Randomized Experiments
Research Assistant Supervised by Prof. Xinran Li, University of Chicago   Oct 2023 – Present
- Literature review on general forms of finite population central limit theorems and several covariate adjustment approaches designed specifically for small and large strata sizes
- Developed the covariate adjustment method based on Bayesian hierarchical model motivated by the difficulty in distinguishing small and large strata sizes
- Proved the asymptotic properties of our adjusted difference-in-means estimator and ran simulation studies for four settings
- Ran comprehensive simulation studies and decreased estimation variance compared to the previous methods
- Preparing the corresponding manuscript as the first author to submit

Federated Learning Adjusting for Hidden Confounding
Research Assistant Supervised by Prof. Molei Liu, Columbia University   Mar 2024 – Present
- Literature review on factor models adjusting for hidden confounding and federated learning methods
- Inspired by projected principal component analysis in factor models introduced by Fan et al. (2016), developed methods that adjust for site information (e.g. measure of healthcare utilization) and hidden confounding
- Conducted numerical experiments by R under generalized linear models for different covariate dimensions

A Semiparametric Approach for Robust and Efficient Learning with Biobank Data
Research Assistant Supervised by Prof. Molei Liu, Columbia University   Oct 2022 – Apr 2024
- Literature review on large sample sieve estimation of semi-nonparametric models and semi-supervised validation of multiple surrogate outcomes with application to electronic medical records phenotyping
- Motivated by the low robustness under nonlinear genetic effects in the previous semi-supervised validation method
- Proved the consistency, convergence rate and the asymptotic normality of the proposed estimators in classification and validation of the EHR features which contains step functions referring to asymptotic statistics by van der Vaart
- Ran simulation studies for three settings and increased accuracy of parameters in two nonlinear settings while maintaining comparable accuracy in one linear setting compared to the previous method without sieves
- Prepared the corresponding manuscript as a co-first author

An Adaptive Test on Cox Model
Research Assistant Supervised by Prof. Tony Sit, the Chinese University of Hong Kong   Jul 2022 – Oct 2022
- Literature review on an adaptive test that yields high statistical power under various high-dimensional scenarios, and social network dependence with time to event data modelled by the latent Cox model
- Proved the convergence to zero of the third moment of Cox score function under i.i.d. condition

Transformed Dynamic Quantile Regression under Biased Sampling.
Research Assistant Supervised by Prof. Tony Sit, the Chinese University of Hong Kong   Jul 2021 – Oct 2022
- Literature review on biased sampling and quantile regression with censored data and dynamic Box-Cox transformation
- Developed a transformed dynamic estimation method of quantile regression under biased sampling
- Implemented estimation algorithms via R and Linux server, designed numerical studies to verify the estimation accuracy and overcame inefficient computation by parallel computing
- Proved asymptotic properties of the proposed estimators and the wild bootstrap method using product integral