



THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY

Department of Mathematics

SEMINAR ON STATISTICS

Harnessing AI-Predicted Data for Trustworthy Statistical Inference

By

Prof. Jiwei ZHAO

University of Wisconsin

Abstract

The rapid advancement of AI/ML has made it increasingly feasible to generate predicted data at scale using black-box models, including deep learning, large language models, and generative AI. These predictions offer a promising way to augment scarce gold-standard data and boost the power of downstream scientific analyses. However, naively substituting predictions for true outcomes in statistical inference can introduce bias and inflate false positive rates, undermining the trustworthiness of scientific conclusions. In this talk, I will present a series of methods for safely and efficiently integrating AI-predicted data into rigorous statistical inference. I will first introduce an assumption-lean and data-adaptive framework that guarantees valid inference without assumptions on the ML model while adaptively leveraging accurate predictions for efficiency gains. I will then extend these ideas to settings with covariate shift arising from automated computational phenotypes in electronic health records, and to SADA, a method for safely aggregating multiple black-box predictions of uncertain quality. Throughout, I will highlight the semiparametric efficiency theory underpinning these methods and demonstrate their practical impact through applications in precision medicine.

bio: Jiwei Zhao is currently an Associate Professor at the University of Wisconsin-Madison. His research interests include semiparametric statistics, the trade-off between efficiency and robustness, domain adaptation and transfer learning, missing data analysis, and causal inference. He also works on developing trustworthy statistical inference methods for AI-predicted data and, more broadly, for synthetic data. His application areas include patient-reported outcomes, clinical trials, real-world evidence and real-world data, survey data, aging, mental health, and cancer. His work has been published in top-tier statistical journals as well as in leading machine learning conferences. His research has been consistently supported by the US National Science Foundation and the National Institutes of Health. Jiwei is now Associate/Action Editor for journals in both statistics and machine learning. He also serves as Area Chair for NeurIPS, ICML and AISTATS. He is an Elected Fellow of the American Statistical Association and an Elected Member of the International Statistical Institute.

Date : 18 June 2026 (Thursday)

Time : 3:00p.m.

Venue : Room 4502 (near Lift 25/26)

All are welcome!