


复旦大学数学科学学院
数学综合报告会

报告题目：Knowledge-Powered Causal Learning: From Modern AI Algorithms to Classic Wald-type Inference

报告人：王惠远 (University of Pennsylvania)

时间：2025-12-26 星期五 10:45-11:15

地点：腾讯会议ID：339 476 861；密码：251226

报告摘要：

Real-world data (RWD) presents a "double-edged sword" for casual inference: while it offers unprecedented scale and phenotypic richness, it lacks the design control of randomized trials, leaving evidence vulnerable to unmeasured confounding and credibility gaps. In this talk, I introduce a “Knowledge-Powered” framework that harnesses the phenotypic richness of RWD—specifically through Reference Outcomes (e.g., negative controls, historical outcomes)—to build reliable tools for real-world evidence generation. Central to this framework is the Transferability Assumption: the premise that confounding information captured by reference outcomes can be structurally transferred to the primary outcome. By bridging modern AI algorithms with exploratory real-world data analysis, we illustrate the utility of this assumption through three distinct methodological advancements.

First, assuming the bias functions of reference and primary outcomes share a common representation, we develop a debiased causal tree algorithm. Relying on the partial identification of the Conditional Average Treatment Effect on the Treated (CATT), this method employs a reference-outcome-informed splitting criterion to minimize the confounding bias for the primary outcome. Second, from a sensitivity analysis perspective, we show that partial identification also yields data-adaptive feasibility regions for the bias function, generalizing conventional marginal sensitivity models that rely on single global parameters. We identify sharp bounds for the ATT and utilize double machine learning to provide semiparametric efficient inference for these endpoints. Finally, through exploratory data analysis, we uncover an exchangeability structure among the biases of reference outcomes. Leveraging this structure allows us to bypass bound estimation entirely and directly recover valid, Wald-type inference for testing the null treatment effect of the primary outcome. The methodology is validated through a large-scale comparative effectiveness study of GLP-1RA vs. SGLT2i using the global TriNetX research network, demonstrating how phenotypic richness can be harnessed to automate and rigorously ground real-world evidence generation.

非线性数学模型与方法教育部重点实验室
中法应用数学国际联合实验室
上海市现代应用数学重点实验室
复旦大学数学研究所