香港中文大學統計學系
Department of Statistics
THE CHINESE UNIVERSITY OF HONG KONG

SEMINAR

DEPARTMENT OF STATISTICS
THE CHINESE UNIVERSITY OF HONG KONG

BART-enhanced models for causal inference and missing data

INVITED SPEAKER

Liangyuan Hu
Associate Professor
Department of Biostatistics and Epidemiology
Rutgers University

TIME

October 22, 2024 (Tue) · 10:30 am - 11:15 am

VENUE

SC L3 · Science Centre L3 · CUHK

ABSTRACT

In this talk, we present novel, flexible methods that leverage Bayesian Additive Regression Trees (BART) for causal inference and handling missing data. First, we introduce a random-intercept accelerated failure time model (riAFT-BART), designed to estimate the causal effects of multiple treatments on clustered survival outcomes. This model offers a robust approach for analyzing survival data with a hierarchical structure. Next, we introduce a new sensitivity analysis framework for assessing the impact of unmeasured confounding on causal inferences, providing a practical tool to quantify the robustness of causal conclusions. Moving into longitudinal data settings, we present a Bayesian non-parametric longitudinal sequential imputation method for handling missing-at-random longitudinal covariates. Our approach features two key advancements. First, we propose the Bayesian Trees Mixed-Effects Model (BMTrees), along with its variants, which incorporate non-parametric priors to accommodate complex, non-linear, and non-normal relationships in longitudinal data. Second, we employ the sequential imputation framework using these Bayesian non-parametric models to efficiently address missing covariate data. To mitigate computational challenges, our approach utilizes a fitting-with-imputing strategy, allowing for simultaneous model fitting and imputation sampling. We demonstrate the practical utility of these methods through two real-world applications. The riAFT-BART model and sensitivity analysis are applied to the National Cancer Database to compare the effects of three treatment approaches on survival outcomes for patients with high-risk localized prostate cancer. Additionally, our missing data methodology is applied to the Multi-Ethnic Study of Atherosclerosis dataset, where we conduct an integrative analysis to estimate the causal effects of various antihypertensive treatment initiation strategies.