Yale SCHOOL OF PUBLIC HEALTH
Biostatistics

Optimal Doubly Robust Estimation of Heterogeneous Causal Effects

Edward Kennedy, PhD
Assistant Professor
Department of Statistics & Data Science
Carnegie Mellon University

12:00 Noon Eastern time, Tuesday, November 3, 2020

Virtual seminar via Zoom

ABSTRACT

Heterogeneous effect estimation plays a crucial role in causal inference, with applications across medicine and social science. Many methods for estimating conditional average treatment effects (CATEs) have been proposed in recent years, but there are important theoretical gaps in understanding if and when such methods are optimal. This is especially true when the CATE has nontrivial structure (e.g., smoothness or sparsity). Our work contributes in several main ways. First, we study a two-stage doubly robust CATE estimator and give a generic model-free error bound, which, despite its generality, yields sharper results than those in the current literature. We apply the bound to derive error rates in nonparametric models with smoothness or sparsity, and give sufficient conditions for oracle efficiency. Underlying our error bound is a general oracle inequality for regression with estimated or imputed outcomes, which is of independent interest; this is the second main contribution. The third contribution is aimed at understanding the fundamental statistical limits of CATE estimation. To that end, we propose and study a local polynomial adaptation of double-residual regression. We show that this estimator can be oracle efficient under even weaker conditions, if used with a specialized form of sample splitting and careful choices of tuning parameters. These are the weakest conditions currently found in the literature, and we conjecture that they are minimal in a minimax sense. We go on to give error bounds in the non-trivial regime where oracle rates cannot be achieved. Some finite-sample properties are explored with simulations.