Localized Debiased Machine Learning: Efficient Inference on Quantile Treatment Effects and Beyond

Nathan Kallus, Assistant Professor at Cornell Tech

Abstract

We consider the efficient estimation of a low-dimensional parameter in an estimating equation involving high-dimensional nuisances that depend on the parameter of interest. An important example is the (local) quantile treatment effect ((L)QTE) in causal inference, for which the efficient estimating equation involves as a nuisance the covariate-conditional cumulative distribution function evaluated at the quantile to be estimated. Debiased machine learning (DML) is a data-splitting approach to address the need to estimate nuisances using flexible machine learning methods that may not satisfy strong metric entropy conditions, but applying it to problems with parameter-dependent nuisances is impractical. For (L)QTE estimation, DML requires we learn the whole conditional cumulative distribution function, conditioned on potentially high-dimensional covariates, which is far more challenging than the standard supervised regression task in machine learning. We instead propose localized debiased machine learning (LDML), a new data-splitting approach that avoids this burdensome step and needs only estimate the nuisances at a single initial rough guess for the parameter. For (L)QTE estimation, this involves just learning two binary regression (i.e., classification) models, for which many standard, time-tested machine learning methods exist, and the initial rough guess may be given by inverse propensity weighting. We prove that under lax rate conditions on nuisances, our estimator has the same favorable asymptotic behavior as the infeasible oracle estimator that solves the estimating equation with the unknown true nuisance functions. Thus, our proposed approach uniquely enables practically-feasible and theoretically-grounded efficient estimation of important quantities in causal inference such as (L)QTEs and in other coarsened data settings.