High Dimensional Inference: Semiparametrics, Counterfactuals, and Heterogeneity

Ying Zhu

Abstract

Semiparametric regressions enjoy the flexibility of nonparametric models as well as the interpretability of linear models. These advantages can be further leveraged with recent advance in high dimensional statistics. This talk begins with a simple partially linear model, $Y_i = \mathbf{X}_i\boldsymbol{\beta}^* + g^*(\mathbf{Z}_i) + \varepsilon_i$, where the parameter vector of interest, $\boldsymbol{\beta}^*$, is high dimensional but sufficiently sparse, and g^* is an unknown nuisance function. In spite of its simple form, this high dimensional partially linear model plays a crucial role in counterfactual studies of heterogeneous treatment effects. In the first half of this talk, I present an inference procedure for *any* subvector (regardless of its dimension) of the high dimensional $\boldsymbol{\beta}^*$. This method does not require the "beta-min" condition and also works when the vector of covariates, $\mathbf{Z}_i$, is high dimensional, provided that the function classes $\mathbb{E}(X_{ij}|\mathbf{Z}_i)$s and $\mathbb{E}(Y_i|\mathbf{Z}_i)$ belong to exhibit certain sparsity features, e.g., a sparse additive decomposition structure. In the second half of this talk, I discuss the connections between semiparametric modeling and Rubin's Causal Framework, as well as the applications of various methods (including the one from the first half of this talk and those from my other papers) in counterfactual studies that are enriched by "big data".