



Department of Quantitative
Theory and Methods

Minimax optimal subgroup identification

Quantifying treatment effect heterogeneity is a crucial task in many areas of causal inference, e.g. optimal treatment allocation and estimation of subgroup effects. We study the problem of estimating the level sets of the conditional average treatment effect (CATE). Given a user-specified threshold, the goal is to estimate the set of all units for whom the treatment effect exceeds that threshold. For example, if the cutoff is zero, the estimand is the set of all units who would benefit from receiving treatment. Assigning treatment just to this set represents the optimal treatment rule that maximises the mean population outcome. Being able to accurately estimate CATE level sets is therefore of great practical relevance. The level set estimator that we study follows the plug-in principle and consists of simply thresholding a good estimator of the CATE. While many CATE estimators have been recently proposed and analysed, how their properties relate to those of the corresponding level set estimators remains unclear. Our first goal is thus to fill this gap by deriving the asymptotic properties of level set estimators depending on which estimator of the CATE is used. Next, we identify a minimax optimal estimator in a model where the CATE, the propensity score and the outcome model are Holder-smooth of varying orders. We consider data generating processes that satisfy a margin condition governing the probability of observing units for whom the CATE is close to the threshold. We investigate the performance of the estimators in simulations and illustrate our methods on a dataset from REFLUX, a multi-center study that aimed to compare the effectiveness of surgery to treat Gastro-Oesophageal Reflux Disease.

Matteo Bonvini

Dept of Statistics & Data Science
Carnegie Mellon University

10.12.2022 | 1pm-2:15pm
PAIS 561

Fall 2022 QTM Speaker Series