samsi
NSF·Duke·NCSU·UNC

Causal Inference Program Opening Workshop
December 9-11, 2019

SPEAKER TITLES/ABSTRACTS

Mark van der Laan
University of California, Berkeley

"Targeted Learning for Causal Inference Based on Real World Data"

We discuss a general roadmap for generating causal inference based on observational studies used to general real world evidence. We review targeted minimum loss estimation (TMLE), which provides a general template for the construction of asymptotically efficient plug-in estimators of a target estimand for realistic (i.e, infinite dimensional) statistical models. TMLE is a two stage procedure that first involves using ensemble machine learning termed super-learning to estimate the relevant stochastic relations between the treatment, censoring, covariates and outcome of interest. The super-learner allows one to fully utilize all the advances in machine learning (in addition to more conventional parametric model based estimators) to build a single most powerful ensemble machine learning algorithm. We present Highly Adaptive Lasso as an important machine learning algorithm to include.

In the second step, the TMLE involves maximizing a parametric likelihood along a so-called least favorable parametric model through the super-learner fit of the relevant stochastic relations in the observed data. This second step bridges the state of the art in machine learning to estimators of target estimands for which statistical inference is available (i.e, confidence intervals, p-values etc). We also review recent advances in collaborative TMLE in which the fit of the treatment and censoring mechanism is tailored w.r.t. performance of TMLE. We also discuss asymptotically valid bootstrap based inference. Simulations and data analyses are provided as demonstrations.