

Covariate selection with group lasso and doubly robust estimation of causal effects

Julian Wolfson *

julianw@umn.edu

The efficiency of doubly robust estimators of the average causal effect (ACE) can be improved by including in the treatment and outcome models only those covariates which are related to both treatment and outcome (i.e., confounders) or related only to the outcome. However, it is often challenging to identify such covariates if a large number are measured in a study. We propose GLiDeR (Group Lasso and Doubly Robust Estimation), a novel variable selection technique for identifying confounders and predictors of outcome using an adaptive group lasso approach that simultaneously performs coefficient selection, regularization, and estimation across the treatment and outcome models. The selected variables and corresponding coefficient estimates are used in a standard doubly robust estimator to estimate the ACE. We provide asymptotic results showing that, for a broad class of data generating mechanisms, GLiDeR yields a consistent estimator of the ACE when either the outcome or treatment model is correctly specified. A comprehensive simulation study shows that GLiDeR is more efficient than doubly robust methods using standard variable selection techniques, and has substantial computational advantages over a recently proposed doubly robust Bayesian model averaging method. We illustrate our method by estimating the causal treatment effect of bilateral versus single-lung transplant on forced expiratory volume in one year after transplant using an observational registry.

This is joint work with Ph.D. candidate Brandon Koch and Prof. David Vock.

*Division of Biostatistics, School of Public Health, University of Minnesota, A460 Mayo Building MMC 303, 420 Delaware St SE, Minneapolis, MN 55455, USA