Variable Selection in Causal Inference Using Penalization

Brief Paper Summary

Denis Ostroushko

December 11, 2023

Motivation

- Variable selection for causal inference in using common variable selection methods can lead to serious model misspecification issues
- Inclusion of spurious predictors can lead to inflated variance of the estimator
- Omission of important confounders may lead to a biased estimation of ATE
- Doubly robust estimators employ propensity score and outcome models to provide more credible estimates of ATE
- Under the conditions that both types are correctly specified, variance of such estimator is also minimized

Overview

- The three types of predictor are:
  - Related to the outcome only
  - Related to the treatment assignment only
  - A confounder: related to both!
- Proposed method selects those which are true confounders, and allows to select weak outcome related, but strong treatment related predictors
- We shrink or remove the parameter based on how much the predictor contributes to the outcome model

Estimation

1. Consider all types of variables as predictors, construct outcome and treatment models
2. Apply penalization to coefficients in the treatment model. Penalty for each coefficient depends on the levels of contribution to both the outcome and treatment models. Obtain selected variables and propensity scores.
3. Define $S_i = A_i - \pi(\mathbf{X}_i)$
4. Doubly robust estimator: $E[Y_i|S_i, \mathbf{X}_i] = \theta * S_i + g(\mathbf{X}_i, \beta)$

- this one is similar to propensity score regression
- θ is the average treatment effect

Results

Table 1. Performance of the proposed method when $r_2 > n$ and in the presence of a weak confounder. S.D^{emp}: empirical standard error; S.D^{tb}: sandwich standard error.

Method	Bias	S.D^{emp}	S.D^{tb}	MSE	Bias	S.D^{emp}	S.D^{tb}	MSE
Scenario 1.		$n = 300$			$n = 500$			
SCAD	0.010	0.515	0.502	0.266	0.012	0.386	0.381	0.149
LASSO	0.067	0.522	0.509	0.277	0.057	0.425	0.421	0.184
PS-fit	0.164	5.575	–	31.104	0.101	4.295	–	18.453
Oracle	0.017	0.510	–	0.260	0.007	0.373	–	0.139
Scenario 2.		$n = 300$			$n = 500$			
SCAD	0.062	0.606	0.592	0.372	0.019	0.483	0.456	0.234
LASSO	0.037	0.612	0.593	0.375	0.012	0.481	0.460	0.232
Y-fit	0.710	0.598	–	0.862	0.818	0.453	–	0.875
PS-fit	0.381	6.722	–	45.326	0.094	5.117	–	26.189
Oracle	0.045	0.638	–	0.409	0.018	0.459	–	0.211

- ATE compared to IPW and Outcome Regression with varying number of weak confounders
- SCAD and LASSO used for considered method