Variable Selection in Causal Inference Using Penalization

Brief Paper Summary

Denis Ostroushko

December 12, 2023

Motivation

- In small and moderate sample sizes common variable selection methods will lead to common 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

- What if I told you there is a good method that addresses two common issues!
- Proposed method aims to find non-ignorable confounders and predictors of outcome
- Intuitively, we shrink or remove the parameters 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 with the same set of predictors. Variable selection is done using penalized joint maximum likelihood
2. Penalty for each coefficient depends on the levels of contribution to both the outcome and treatment models. Authors call this a booster parameter
    - For example, when a covariate that barely predicts the outcome and treatment it has a stronger penalty on the associated parameter.
    - When a covariate barely predicts the outcome and is strongly related to treatment, penalty is less strict.
    - Idea is similar to adaptive lasso

Estimation Continued

3. Select final predictors with effects above a certain threshold: $\frac{1}{\sqrt{n}}$
4. Define $S_i = A_i - \pi(\mathbf{X}_i)$
5. 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

Questions

NO QUESTIONS