Propensity Score Weighting

using machine learning

Young Geun Kim
ygeunkim.github.io

2019711358, Department of Statistics

24 Nov, 2020

Introduction

Propensity Score Estimation

Evaluation

Related Contents

Introduction

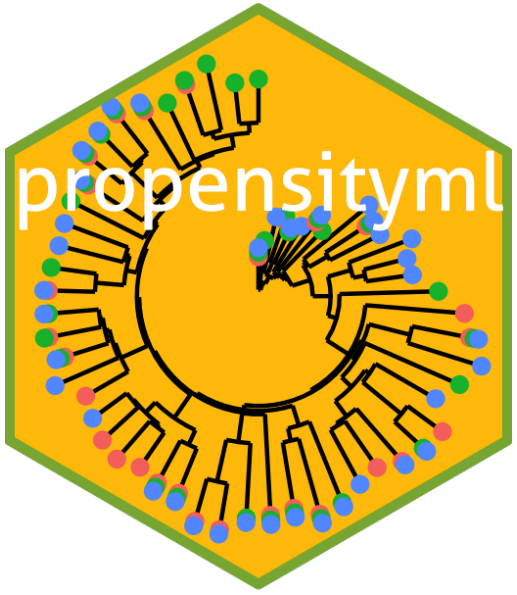
Reviewed Paper

Reviewd and apply Lee et al. (2010): estimate propensity score using

- Logistic regression
- Random forests
- CART
- SVM

Custom Package

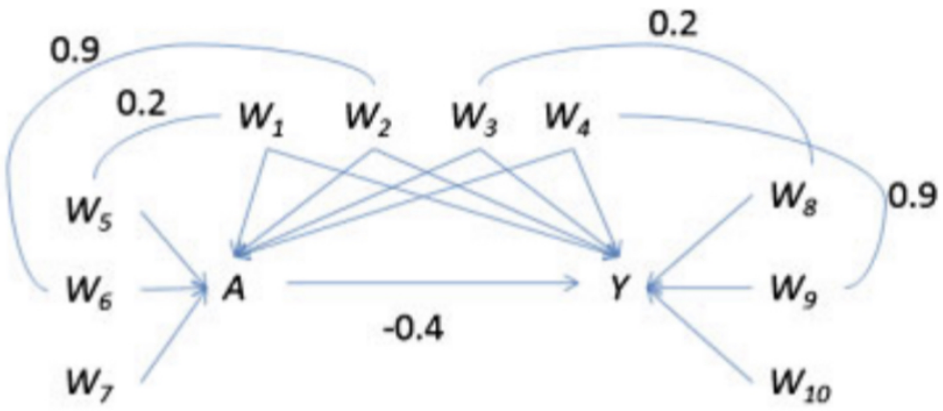
```
# remotes::install_github("ygeunkim/propensityml")
library(propensityml)
```



Simulation study

Simulation setting by Setoguchi et al. (2008):

- 10 covariates: confounders, exposure predictors, outcome predictors
- Treatment
- Outcome probability



A: exposure
Y: outcome
W_1-W_4: confounders
W_5-W_7: exposure predictors
W_8-W_{10}: outcome predictors

Binary variables: A, W_1, W_3, W_6, W_8, W_9
Continuous variables: Y, W_2, W_4, W_7, W_{10}

Scenarios

1. Additivity and linearity
2. Mild non-linearity: *1 quadratic term*
3. Moderate non-linearity: *3 quadratic term*
4. Mild non-additivity: *3 two-way interaction terms*
5. Mild non-additivity and non-linearity: *3 two-way interaction terms and 1 quadratic term*
6. Moderate non-linearity: *10 two-way interaction terms*
7. Moderate non-additivity and non-linearity: *10 two-way interaction terms and 3 quadratic terms*

Function to reproduce Setoguchi et al. (2008)

```
sim_outcome(n = 1000, covmat = build_covariate()) %>% str()
#> Classes 'data.table' and 'data.frame':   1000 obs. of  12 variables:
#>  $ w1           : Factor w/ 2 levels "0","1": 1 2 2 2 1 2 2 2 1 1 ...
#>  $ w2           : num  -0.28 0.306 0.633 -0.307 -0.59 ...
#>  $ w3           : Factor w/ 2 levels "0","1": 1 1 1 2 2 2 2 2 2 2 ...
#>  $ w4           : num  1.657 -1.44 -1.94 0.539 0.412 ...
#>  $ w5           : Factor w/ 2 levels "0","1": 2 2 2 1 1 2 1 1 2 2 ...
#>  $ w6           : Factor w/ 2 levels "0","1": 1 2 2 1 1 2 2 1 2 2 ...
#>  $ w7           : num  0.4874 -0.0162 -0.1558 -0.3943 0.3646 ...
#>  $ w8           : Factor w/ 2 levels "0","1": 2 2 1 1 2 1 2 2 1 1 ...
#>  $ w9           : Factor w/ 2 levels "0","1": 2 1 1 2 2 1 2 1 1 1 ...
#>  $ w10          : num  -0.305 0.594 0.418 0.763 0.881 ...
#>  $ exposure     : Factor w/ 2 levels "0","1": 2 2 2 2 2 1 2 2 2 2 ...
#>  $ outcome_prob: num  5.95e-53 7.20e-01 2.73e-23 1.85e-33 7.78e-03 ...
#>  - attr(*, ".internal.selfref")=<externalptr>
```

Propensity Score Estimation

Sample Size

For simulation, 1000 replicates

Small: 500
with 7 scenarios

Medium: 1000
with 7 scenarios

Large: 2000
with 7 scenarios

Covariate Balance

For example,

```
compute_balance(
  small_list[mcname == 1 & scenario == "A", -c("mcname", "scenario")],
  treatment = "exposure", trt_indicator = 1, outcome = "outcome_prob"
)
#>     variable  balance
#>  1:       w1 -0.05540
#>  2:       w2 -0.03770
#>  3:       w3 -0.09556
#>  4:       w4  0.09143
#>  5:       w5 -0.11176
#>  6:       w6  0.03223
#>  7:       w7 -0.06150
#>  8:       w8 -0.09707
#>  9:       w9  0.01704
#> 10:      w10  0.00309
```

Average(`balance`) = **Average standardized absolute mean distance** (ASAM).

Average standardized absolute mean distance (ASAM)

- Evaluation
- After applying weighting

Logistic Regression

Evaluation

Related Contents

About this project

Project repository

https://github.com/ygeunkim/psweighting-ml

Project package

https://github.com/ygeunkim/propensityml

References I

Lee, B. K., Lessler, J., and Stuart, E. A. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, 29(3):337–346.

Setoguchi, S., Schneeweiss, S., Brookhart, M. A., Glynn, R. J., and Cook, E. F. (2008). Evaluating uses of data mining techniques in propensity score estimation: a simulation study. *Pharmacoepidemiology and Drug Safety*, 17(6):546–555.