Homework 1 Data Analysis

Jimmy Hickey

2020/09/03

```
library(modelObj)
library(DynTxRegime)
library(ggplot2)
source("estimators.R")
source("value_estimators.R")
```

1

```
data <- read.csv(file = file.choose(), header = TRUE, sep = ",")
data$A= data$trt

y = data$chol0 - data$chol6
```

i. naive

```
delta_N_se(y = y, A = data$A)
```

```
## $EY
##         0         1
## -11.43539  33.05634
##
## $deltaHat
## [1] 44.49173
##
## $sigmaHat
## [1] 2.118561
```

ii. outcome regression

```
lm = buildModelObj(model = ~A + exercise + wt + smoke + trig0 + age + gender +
                     A:exercise + A:wt + A:smoke + A:trig0 + A:age + A:gender,
                   solver.method = "lm",
                   predict.method = "predict.lm",
                   predict.args = list("type"="response"))
delta_OR_se(moOR = lm, data = data, y = y)
```

```
## $fitOR
##
## Call:
## lm(formula = YinternalY ~ A + exercise + wt + smoke + trig0 +
##     age + gender + A:exercise + A:wt + A:smoke + A:trig0 + A:age +
##     A:gender, data = data)
##
## Coefficients:
## (Intercept)            A     exercise           wt        smoke        trig0
##   3.246e+01   -2.615e+02    2.058e+01   -2.393e-01    2.908e+00   -1.671e-02
##         age       gender   A:exercise         A:wt      A:smoke      A:trig0
##   9.491e-03    5.239e-01   -2.108e+01    1.621e+00   -5.081e+00    3.512e-02
##       A:age     A:gender
##   2.488e-02    9.056e-01
##
##
## $EY
##         0         1
## -11.77170  28.78214
##
## $deltaHat
## [1] 40.55384
##
## $sigmaHat
## [1] 1.727727
```

iii. propensity score stratification

```
propensity <- modelObj::buildModelObj(model = ~ age + wt + gender + exercise + smoke + trig0 + chol0,
                              solver.method = 'glm',
                              solver.args = list(family='binomial'),
                              predict.method = 'predict.glm',
                              predict.args = list(type='response'))
delta_S_se(moPS = propensity, data = data, y = y, K = 5)
```

```
## $fitPS
##
## Call:  glm(formula = YinternalY ~ age + wt + gender + exercise + smoke +
##     trig0 + chol0, family = "binomial", data = data)
##
## Coefficients:
## (Intercept)          age           wt       gender     exercise        smoke
##  -2.9404887    0.0009917    0.0083114   -0.0929438    0.3816924   -0.0976649
##       trig0        chol0
##  -0.0009232    0.0061492
##
## Degrees of Freedom: 999 Total (i.e. Null);  992 Residual
## Null Deviance:       1386
## Residual Deviance: 1366  AIC: 1382
##
## $cj
##        0%       20%       40%       60%       80%      100%
```

```
## 0.0000000 0.4364824 0.4737117 0.5097436 0.5554141 1.0000000
##
## $grp
## bin
##   1   2   3   4   5
## 200 200 200 200 200
##
## $EY
##               0         1
## [1,] -13.610390  64.08943
## [2,] -16.576471  51.91304
## [3,] -12.481818  31.41111
## [4,] -10.486486  12.59551
## [5,]  -6.316667 -17.15000
##
## $deltaHatj
##         1         2         3         4         5
##  77.69982  68.48951  43.89293  23.08199 -10.83333
##
## $deltaHat
## [1] 40.46618
##
## $sigmaHat
## [1] 1.713776
```

iv. IPW

```
delta_IPW_se(moPS = propensity, data = data, y = y)
```

```
## $fitPS
##
## Call:  glm(formula = YinternalY ~ age + wt + gender + exercise + smoke +
##     trig0 + chol0, family = "binomial", data = data)
##
## Coefficients:
## (Intercept)          age           wt       gender     exercise        smoke
##  -2.9404887    0.0009917    0.0083114   -0.0929438    0.3816924   -0.0976649
##       trig0        chol0
##  -0.0009232    0.0061492
##
## Degrees of Freedom: 999 Total (i.e. Null);  992 Residual
## Null Deviance:       1386
## Residual Deviance: 1366  AIC: 1382
##
## $EY
##         0         1
## -11.84030  29.01366
##
## $deltaHat
## [1] 40.85396
##
## $sigmaHat
```

```
## [1] 2.22835
```

v. double robust AIPW

```
delta_DR_se(moOR = lm, moPS = propensity, data = data, y = y)
```

```
## $fitOR
##
## Call:
## lm(formula = YinternalY ~ A + exercise + wt + smoke + trig0 +
##     age + gender + A:exercise + A:wt + A:smoke + A:trig0 + A:age +
##     A:gender, data = data)
##
## Coefficients:
## (Intercept)            A     exercise           wt        smoke        trig0
##   3.246e+01   -2.615e+02    2.058e+01   -2.393e-01    2.908e+00   -1.671e-02
##         age       gender   A:exercise         A:wt      A:smoke      A:trig0
##   9.491e-03    5.239e-01   -2.108e+01    1.621e+00   -5.081e+00    3.512e-02
##       A:age     A:gender
##   2.488e-02    9.056e-01
##
##
## $fitPS
##
## Call:  glm(formula = YinternalY ~ age + wt + gender + exercise + smoke +
##     trig0 + chol0, family = "binomial", data = data)
##
## Coefficients:
## (Intercept)          age           wt       gender     exercise        smoke
##  -2.9404887    0.0009917    0.0083114   -0.0929438    0.3816924   -0.0976649
##       trig0        chol0
##  -0.0009232    0.0061492
##
## Degrees of Freedom: 999 Total (i.e. Null);  992 Residual
## Null Deviance:       1386
## Residual Deviance: 1366  AIC: 1382
##
## $deltaHat
## [1] 40.39392
##
## $EY
##         0         1
## -18.06398  22.32993
##
## $sigmaHat
## [1] 1.750613
```

2

a

```
regime = as.integer({data$chol0 > 280})
```

i. outcome regression

```
value_OR_se(moOR = lm, data = data, y = y, regime = regime, txName = 'A')
```

```
## $fitOR
##
## Call:
## lm(formula = YinternalY ~ A + exercise + wt + smoke + trig0 +
##     age + gender + A:exercise + A:wt + A:smoke + A:trig0 + A:age +
##     A:gender, data = data)
##
## Coefficients:
## (Intercept)            A     exercise           wt        smoke        trig0
##   3.246e+01   -2.615e+02    2.058e+01   -2.393e-01    2.908e+00   -1.671e-02
##         age       gender   A:exercise         A:wt      A:smoke      A:trig0
##   9.491e-03    5.239e-01   -2.108e+01    1.621e+00   -5.081e+00    3.512e-02
##       A:age     A:gender
##   2.488e-02    9.056e-01
##
##
## $EY
##         0         1
## -7.285852 10.518170
##
## $valueHat
## [1] 3.232318
##
## $sigmaHat
## [1] 1.130432
```

ii. IPW

```
value_IPW_se(moPS = propensity, data = data,  y = y, regime = regime, txName = 'A')
```

```
## $fitPS
##
## Call:  glm(formula = YinternalY ~ age + wt + gender + exercise + smoke +
##     trig0 + chol0, family = "binomial", data = data)
##
## Coefficients:
## (Intercept)          age           wt       gender     exercise        smoke
```

```
##  -2.9404887    0.0009917    0.0083114   -0.0929438    0.3816924   -0.0976649
##       trig0        chol0
##  -0.0009232    0.0061492
##
## Degrees of Freedom: 999 Total (i.e. Null);  992 Residual
## Null Deviance:       1386
## Residual Deviance: 1366  AIC: 1382
##
## $EY
##         0         1
## -7.342378 10.596387
##
## $valueHat
## [1] 3.254009
##
## $sigmaHat
## [1] 1.685697
```

iii. alternative IPW

```
value_AIPW_se(moPS = propensity, moOR = lm, data = data, y = y, regime =regime, txName = 'A')
```

```
## $fitOR
##
## Call:
## lm(formula = YinternalY ~ A + exercise + wt + smoke + trig0 +
##     age + gender + A:exercise + A:wt + A:smoke + A:trig0 + A:age +
##     A:gender, data = data)
##
## Coefficients:
## (Intercept)            A     exercise           wt        smoke        trig0
##   3.246e+01   -2.615e+02    2.058e+01   -2.393e-01    2.908e+00   -1.671e-02
##         age       gender   A:exercise         A:wt      A:smoke      A:trig0
##   9.491e-03    5.239e-01   -2.108e+01    1.621e+00   -5.081e+00    3.512e-02
##       A:age     A:gender
##   2.488e-02    9.056e-01
##
##
## $fitPS
##
## Call:  glm(formula = YinternalY ~ age + wt + gender + exercise + smoke +
##     trig0 + chol0, family = "binomial", data = data)
##
## Coefficients:
## (Intercept)          age           wt       gender     exercise        smoke
##  -2.9404887    0.0009917    0.0083114   -0.0929438    0.3816924   -0.0976649
##       trig0        chol0
##  -0.0009232    0.0061492
##
## Degrees of Freedom: 999 Total (i.e. Null);  992 Residual
## Null Deviance:       1386
## Residual Deviance: 1366  AIC: 1382
```

```
##
## $EY
##         0         1
## 0.4863589 2.6758700
##
## $valueHat
## [1] 3.162229
##
## $sigmaHat
## [1] 1.185867
```

iv. optimal AIPW

```
qLearn(moMain = lm,
       moCont = lm,
       data = data,
       response = y,
       txName = 'A',
       iter = 0L,
       verbose = FALSE)
```

```
## Q-Learning: step 1
## Outcome Regression Analysis
## Combined
##
## Call:
## lm(formula = YinternalY ~ A + exercise + wt + smoke + trig0 +
##     age + gender + A:exercise + A:wt + A:smoke + A:trig0 + A:age +
##     A:gender, data = data)
##
## Coefficients:
## (Intercept)            A     exercise           wt        smoke        trig0
##   3.246e+01   -2.615e+02    2.058e+01   -2.393e-01    2.908e+00   -1.671e-02
##         age       gender   A:exercise         A:wt      A:smoke      A:trig0
##   9.491e-03    5.239e-01   -2.108e+01    1.621e+00   -5.081e+00    3.512e-02
##       A:age     A:gender
##   2.488e-02    9.056e-01
##
## Recommended Treatments:
##   0   1
## 211 789
##
## Estimated value: 33.75671
```

b

```
w = seq(110,284, length.out = 100)

regimes <- sapply(X = w, FUN = function(x,y){{y > x}*1L}, y = data$chol0)
```

```r
### i. Outcome regression
lm <- buildModelObj(model = ~A+age+wt+gender+exercise+smoke+trig0+
                      A*age+A*gender+wt*gender+wt*exercise+smoke*age+chol0,
                    solver.method = "lm",
                    predict.method = "predict.lm",
                    predict.args = list("type"="response"))

w = seq(110,284, length.out = 100)

regimes <- sapply(X = w, FUN = function(x,y){{y > x}*1L}, y = data$wt)




or_vals <- apply(X = regimes,
                 MARGIN = 2L,
                 FUN = function(x, moOR, data, y) {
                   temp <- value_OR_se(moOR = moOR,
                                       data = data,
                                       y = y,
                                       regime = x,
                                       txName = 'A')
                   return( "valueHat" = temp$valueHat)
                 },
                 moOR = lm,
                 data = data,
                 y = y)


### ii. IPW
ipw_vals <- apply(X = regimes,
         MARGIN = 2L,
         FUN = function(x, moPS, data, y) {
           temp <- value_IPW_se(moPS = moPS,
                                data = data,
                                 y = y,
                                regime = x,
                                txName = 'A')
           return( "valueHat" = temp$valueHat )
         },
         moPS = propensity,
         data = data,
         y = y)


### iii. optimal AIPW
aipw_vals <- apply(X = regimes,
         MARGIN = 2L,
         FUN = function(x, moPS, moOR, data, y) {
           temp <- value_AIPW_se(moPS = moPS,
                                 moOR = moOR,
                                data = data,
                                y = y,
```

```
                                regime = x,
                                txName = 'A')
          return( "valueHat" = temp$valueHat )
        },
        moPS = propensity,
        moOR = lm,
        data = data,
        y = y)

plot_data = data.frame(or = or_vals, ipw = ipw_vals, aipw= aipw_vals, w = w)


# an unnecessarily complicated way to find the max
max_w = plot_data$w[(which.max(c(plot_data$or, plot_data$ipw,plot_data$aipw))) %% length(w)]


ggplot(data = plot_data) +
  geom_line(aes(x = w, y = or), color = "red", linetype = "dashed") +
  geom_line(aes(x = w, y = ipw), color = "blue", linetype = "dotted") +
  geom_line(aes(x = w, y = aipw), color = "black", linetype = "solid") +
  geom_vline(xintercept = max_w)+
  ylab("V(d)") +
  xlab("weight")
```