Home Assignment 1

Matthew Zhao

1/14/2023

Lecture 1 HW

Problem 1

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X_n})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2X_i\bar{X_n} + \bar{X_n}^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i^2) - 2(\frac{1}{n}\sum_{i=1}^{n} X_i\bar{X_n}) + \frac{1}{n}(\sum_{i=1}^{n}\bar{X_n})^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i^2) - 2(\frac{1}{n}\sum_{i=1}^{n} X_i)^2 + \frac{1}{n}(\sum_{i=1}^{n} X_i)^2$$

$$= \frac{1}{n}\sum_{i=1}^{n}(X_i^2) - (\frac{1}{n}\sum_{i=1}^{n} X_i)^2$$

By LLN, $\frac{1}{n}\sum_{i=1}^{n}(X_i^2) \xrightarrow{p} E_P(X^2)$ and $\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{p} E_P(X)$

By CMT, $\hat{\sigma}_n^2 \xrightarrow{p} E_P(X^2) - E_P(X)^2 = Var_P(X)$

Problem 2

Intuitively, as $n \to \infty$, the probability that $X_1,\ldots,X_n$ does not contain θ approaches 0 by definition of uniform distribution. Hence, the probability that our estimator for θ ($\max\{X_1, ..., X_n\}$) does not equal θ also goes to zero, thus $P(|\hat{\theta} - \theta| > \epsilon) \to 0$ as $n \to \infty$.

Formally, we can derive the CDF F_θ of our estimator and use it to show that as $n \to \infty$, our estimator converges in probability to θ. Specifically, we can use order statistics to show that $F_\theta(x) = F(x)^n$ where $F(x)$ is the CDF for each individual X_i drawn from $U[0, \theta]$:

$$F_\theta(x) = \mathbb{P}(X_1, ..., X_n \leq x)$$

$$= \mathbb{P}(X_1 \leq x)\mathbb{P}(X_2 \leq x)...\mathbb{P}(X_n \leq x)$$
$$= F(x)^n$$

Plugging this result in we show:

$$\mathbb{P}(|\hat{\theta}_n - \theta| > \epsilon) = \mathbb{P}(\theta - \epsilon < \hat{\theta}_n < \theta + \epsilon)$$

$$
\begin{aligned}
&= F_\theta(\theta+\epsilon) - F_\theta(\theta-\epsilon) \\
&= F(\theta+\epsilon)^n - F(\theta-\epsilon)^n \\
&= \frac{(\theta+\epsilon)^n}{\theta} - \frac{(\theta-\epsilon)^n}{\theta} \to 0
\end{aligned}
$$

as $\epsilon \to 0$, thus $\hat{\theta}_n \xrightarrow{p} \theta$

Problem 3

```
library(tidyverse)
library(haven)
library(stargazer)
library(xtable)
options(scipen=999)

make_ci <- function(xbar,var){
  c(xbar - 1.96*var, xbar + 1.96*var)
}

ci_coverage <- function(n){
  X <- rnorm(n)
  xbar <- mean(X)
  sd <- sd(X)

  known <- make_ci(xbar,1/sqrt(n))
  est <- make_ci(xbar,sd/sqrt(n))

  c(0 >= known[1] & 0 <= known[2],
    0 >= est[1] & 0 <= est[2])
}

known_results <- c()
est_results <- c()
N <- c(30,100,500)
for (n in N){
  known_temp <- c()
  est_temp <- c()
  for (x in 1:1000){
    known_temp <- c(known_temp,ci_coverage(n)[1])
    est_temp <- c(est_temp,ci_coverage(n)[2])
  }
  known_results <- c(known_results,mean(known_temp))
  est_results <- c(est_results,mean(est_temp))
}
data.frame(N,est_results,known_results)
```

```
##     N est_results known_results
## 1  30       0.938         0.948
## 2 100       0.948         0.948
## 3 500       0.942         0.950
```

For each N, the CIs have better coverage when using the actual variance than the estimated. This generally stays the same across values of N and different runs.

Lecture 2 HW

Problem 1

```r
df <- read.csv('penn.csv')

df_sub <- df %>%
  filter(tg == 0 | tg == 4)
Y <- log(df_sub$inuidur1)
D <- ifelse(df_sub$tg == 4,1,0)
W <- df_sub %>%
  select(female,black,hispanic,othrace,agelt35,agegt54)
```

(1),(2),(3)

```r
stargazer(lm(Y ~ ., data = data.frame(D)),type='text',digits=3,
          title='Model Estimates with Treatment')
```

(3)

```
##
## Model Estimates with Treatment
## ================================================
##                         Dependent variable:
##                     ---------------------------
##                                  Y
## ------------------------------------------------
## D                             -0.085**
##                                (0.036)
##
## Constant                      2.057***
##                                (0.021)
##
## ------------------------------------------------
## Observations                    5,099
## R2                              0.001
## Adjusted R2                     0.001
## Residual Std. Error      1.214 (df = 5097)
## F Statistic           5.686** (df = 1; 5097)
## ================================================
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

We interpret $\hat{\beta}_1$ as an average approximate decrease in an individual's first spell of unemployment (in weeks) of 8.5% when being offered a cash bonus for finding a job.

To interpret the coefficient as the causal effect of being offered a cash bonus for finding a job on initial unemployment duration, we would need to assume that the treatment was completely random, specifically that being offered the cash bonus (treatment) is independent of all other determinants of the duration of an individual's first spell of unemployment, formally $E[DU] = 0$.

```r
stargazer(lm(Y ~ .,data=cbind(D,W)),type='text',digits=3,
          title='Model Estimates with Controls for sex, race, and age')
```

(4)

```
##
## Model Estimates with Controls for sex, race, and age
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                  Y
## -----------------------------------------------
## D                             -0.080**
##                                (0.036)
##
## female                        0.094***
##                                (0.034)
##
## black                         -0.277***
##                                (0.052)
##
## hispanic                      -0.284***
##                                (0.095)
##
## othrace                       -0.469**
##                                (0.199)
##
## agelt35                       -0.171***
##                                (0.037)
##
## agegt54                       0.190***
##                                (0.058)
##
## Constant                      2.136***
##                                (0.035)
##
## -----------------------------------------------
## Observations                   5,099
## R2                             0.021
## Adjusted R2                    0.020
## Residual Std. Error      1.202 (df = 5091)
## F Statistic          15.934*** (df = 7; 5091)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

We interpret $\hat{\beta}_1$ as an average approximate decrease in an individual's first spell of unemployment (in weeks) of 8.0% when being offered a cash bonus for finding a job, holding gender, race, and age fixed.

If we do not believe that the treatment D is independent of all other variables, a highly plausible scenario given differential unemployment between groups of different race, gender, and age, we need different assumptions to interpret $\hat{\beta}_1$ as causal. Specifically, we could assume that the above factors are the only determinants of unemployment duration (a slightly more plausible assumption than random assignment) and control for them in a regression. If we believe our assumption holds (that D is independent of U|W where U is our error term including unobservables and W are our controls), we can make a causal interpretation of our coefficient via Selection on Observables.

The magnitude and significance level of the coefficient does not change much from the regression in the previous part, hence leading us to believe that the program was successful in reducing the length of unemployment.

```
stargazer(lm(Y ~ D*(.), data=cbind(D,W)),type='text',digits=3,
          title='Model Estimates with Controls and Treatment/Control Interactions')
```

(5)

```
##
## Model Estimates with Controls and Treatment/Control Interactions
## ===============================================
##                         Dependent variable:
##                     ---------------------------
##                                  Y
## -----------------------------------------------
## D                             -0.163**
##                               (0.070)
##
## female                         0.060
##                               (0.043)
##
## black                         -0.305***
##                               (0.064)
##
## hispanic                      -0.368***
##                               (0.113)
##
## othrace                        -0.449
##                               (0.277)
##
## agelt35                       -0.179***
##                               (0.045)
##
## agegt54                        0.134*
##                               (0.072)
##
## D:female                       0.097
##                               (0.073)
##
## D:black                        0.074
##                               (0.109)
##
## D:hispanic                     0.280
##                               (0.211)
##
## D:othrace                      -0.046
```

```
##                               (0.398)
##
## D:agelt35                      0.018
##                               (0.078)
##
## D:agegt54                      0.156
##                               (0.124)
##
## Constant                      2.167***
##                               (0.041)
##
## -----------------------------------------------
## Observations                   5,099
## R2                             0.022
## Adjusted R2                    0.020
## Residual Std. Error      1.203 (df = 5085)
## F Statistic          9.001*** (df = 13; 5085)
## ===============================================
## Note:               *p<0.1; **p<0.05; ***p<0.01
```

We interpret $\hat{\beta}_1$ as an average approximate decrease in an individual's first spell of unemployment (in weeks) of 16.3% when being offered a cash bonus for finding a job for a white male between the ages of 35 and 54.

We see that there is a significant difference between the coefficients from this regression compared to the model from part (3). Furthermore, this estimate is large and statistically significant. This means that if we can continue to believe that our prior assumptions about causal identification hold, it may be the case that our prior regression may not have accurately matched the true/underlying population model/been misspecified. Specifically, we are able to account for potential heterogeneity in treatment effects to some extent by including interactions (i.e. now looking at CATE). While we see that the coefficients for all of the interactions are insignificantly different from zero, a comparison of models using ANOVA below provides some evidence that the model with interactions is a better representation of the data (although causality is not guaranteed).

```r
base_model <- lm(Y ~ ., data = data.frame(D))
interactions_model <- lm(Y ~ D*(.), data=cbind(D,W))
stargazer(anova(base_model,interactions_model),
          type='text',digits=3,
          title='Comparison of Models')
```

```
##
## Comparison of Models
## ===================================================
## Statistic N   Mean    St. Dev.    Min       Max
## ---------------------------------------------------
## Res.Df    2 5,091.000  8.485     5,085     5,097
## RSS       2 7,433.929 113.732  7,353.509 7,514.350
## Df        1  12.000              12        12
## Sum of Sq 1  160.841            160.841   160.841
## F         1   9.269              9.269     9.269
## Pr(> F)   1   0.000                0         0
## ---------------------------------------------------
```

(6) We see that for other groups, our conclusions do change somewhat since the treatment effect for some groups, that is β_1, is insignificantly different from zero. This is more often the case for the regression from part 5 which includes interactions, with the coefficients for the first two regressions generally being fairly similar.

Problem 2

(0) Let $\beta = 3$

```
dgp <- function(n_obs){
  Z <- matrix(
      rnorm(n = n_obs*2, mean = 2, sd = 1),
      nrow = n_obs,
      ncol = 2
  )
  beta_instrument_vector <- matrix(5, nrow = 2, ncol = 1)
  confounder <- matrix(rnorm(n = n_obs, mean = 0, sd = sqrt((1:n_obs)^1.5)))
  X <- rnorm(n_obs, mean = 2, sd = 1) + confounder + Z %*% beta_instrument_vector
  Y <- 3*X + confounder
  as.matrix(data.frame(Y,X,Z))
}

beta <- function(data,num){
  if (num ==1){
    omega <- omega_1(data)
  } else if (num == 2){
    omega <- omega_2(data)
  } else{
    omega <- omega_3(data)
  }
  l <- solve(t(data[,2]) %*% data[,3:4] %*% omega %*% t(data[,3:4]) %*% data[,2])
  r <- t(data[,2]) %*% data[,3:4] %*% omega %*% t(data[,3:4]) %*% data[,1]
  l %*% r
}

omega_1 <- function(data){
  diag(2)
}

omega_2 <- function(data){
  prod <- matrix(0,2,2)
  for (i in 1:nrow(data)) {
    prod <- prod + data[i,3:4] %*% t(data[i,3:4])
  }
  solve(prod/nrow(data))
}

omega_3 <- function(data){
  beta_2sls <- beta(data,2)
  res <- data[,1] - data[,2] %*% beta_2sls

  prod <- matrix(0,2,2)
  for (i in 1:nrow(data)) {
```

```
    prod <- prod + data[i,3:4] %*% t(data[i,3:4]) * res[i]^2
  }
  solve(prod/nrow(data))
}
```

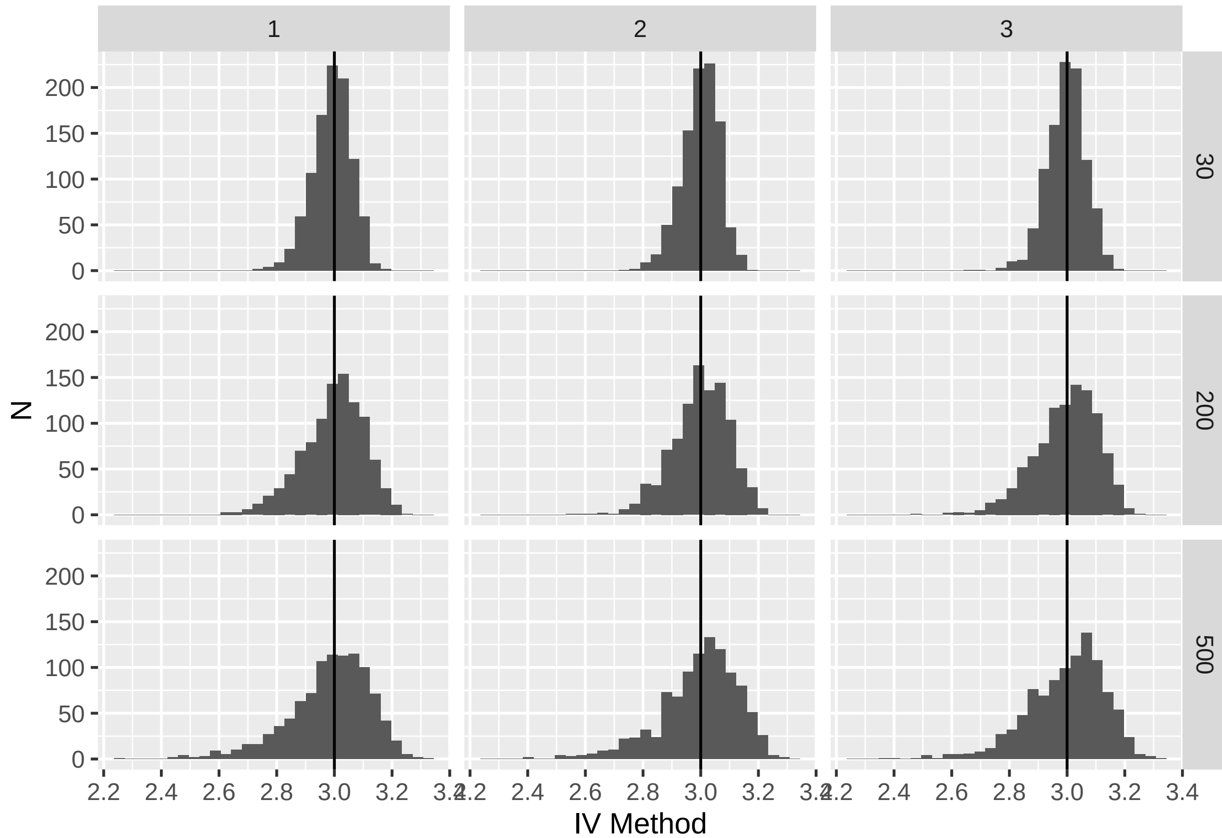
```
N <- c(30,200,500)
B <- 1:3
results <- data.frame(N = numeric(),
                      `IV method` = numeric(),
                      beta = numeric())

for (n in N) {
  for (b in B) {
    betas <- c()
    for (x in 1:1000){
      data <- dgp(n)
      temp <- beta(data,b)
      betas <- c(betas,temp)
    }
    new <- data.frame(rep(n,1000),
                      rep(b,1000),
                      betas)
    results <- rbind(results,new)
  }
}
```

```
colnames(results) <- c('N',"IV Method",'beta')
ggplot(data = results,mapping=aes(x=beta)) +
  geom_histogram() +
  facet_grid(N ~ `IV Method`) +
  labs(x = 'IV Method',y = 'N') +
  geom_vline(xintercept = 3)
```

(1),(2),(3)

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



From the above histograms we see that as n increases, our estimate of β approaches the true value. Additionally, we see that the second and third methods for finding Ω produce *beta*s that are more concentrated around the true value as n increases. This confirms their theoretical properties where these methods produce estimators with lower asymptotic variance.