Stat 111 Homework 3

Kojin Oshiba

1. Randomized Control Trials

(a) Yi(1)

So as not to clutter the document, I am only showing the first 10 rows of the treatment group.

```
df <- read.csv(file="RCT.csv", header=TRUE, sep=",")
treatment <- df[df$group == 1,]
treatment[1:10,]
```

```
##      X  id age sex migraine f1   pk1 chronicity group    f5 pf5
## 3    6 112  45   1        1 15  9.25         27     1  6.25  13
## 4    7 113  45   1        1 25 42.50         30     1 51.25  27
## 5    8 114  49   1        1 14 24.25         49     1 25.25  13
## 7   10 130  46   1        1 11 21.75          3     1  1.00   2
## 8   11 131  64   1        1  6 14.50         23     1  2.50   2
## 10  13 137  53   1        1  8 11.75         32     1 13.50   9
## 11  16 141  37   1        0  9 15.50          7     1  2.75   1
## 15  20 149  23   1        1 25 49.25         23     1 19.50  12
## 16  21 150  59   1        1  9  9.75         54     1 21.50  10
## 17  25 161  32   1        1 25 59.75          5     1 38.00  22
##    withdrawal_reason completer
## 3                 NA         1
## 4                 NA         1
## 5                 NA         1
## 7                 NA         1
## 8                 NA         1
## 10                NA         1
## 11                NA         1
## 15                NA         1
## 16                NA         1
## 17                NA         1
```

(b) Proxy for p

A good proxy for p can be the fraction of people in the experiment who were assgined to the treatment group.

```
n = nrow(df)
cat('proxy for p:', nrow(df[df$group == 1,]) / n)
```

```
## proxy for p: 0.5348837
```

(c) Estimate the average causal treatment effect

```
p = 1/2
cat('estimate of the average causal treatment effect',1/n*p * sum(df[df$group == 1,]$pf5) - 1/n*(1-p) *
```

```
## estimate of the average causal treatment effect -0.1228073
```

2. Joint PDF/PMF

(a)

$$
\begin{aligned}
f_{Y_1,...,Y_n}(y_1,...,y_n) &= f_{Y_1}(y_1)...f_{Y_n}(y_n) \\
&= \prod_{i=1}^{n}(1-p)^{y_i-1}p \\
&= (1-p)^{\sum_{i=1}^{n} y_i - n} p^n
\end{aligned}
$$

(b)

$$
\begin{aligned}
f_{Y_1,...,Y_n}(y_1,...,y_n) &= f_{Y_1}(y_1)...f_{Y_n}(y_n) \\
&= \prod_{i=1}^{n} \lambda e^{-\lambda y_i} \\
&= \lambda^n e^{-\lambda \sum_{i=1}^{n} y_i}
\end{aligned}
$$

3. Linear Regression Model

(a)

$$
\begin{aligned}
E(\hat{\theta}_1) &= E(\frac{\sum_{i=1}^{n} Y_i}{\sum_{i=1}^{n} x_i}) \\
&= \frac{1}{\sum_{i=1}^{n} x_i} \sum_{i=1}^{n} E(Y_i) \\
&= \frac{\sum_{i=1}^{n} \theta x_i}{\sum_{i=1}^{n} x_i} \\
&= \theta
\end{aligned}
$$

$$
\begin{aligned}
E(\hat{\theta}_2) &= E(\frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{x_i}) \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{1}{x_i} E(Y_i) \\
&= \frac{1}{n} \sum_{i=1}^{n} \frac{\theta x_i}{x_i} \\
&= \theta
\end{aligned}
$$

(b)

$$
\begin{aligned}
mse(\theta, \hat{\theta}_1) &= Var(\hat{\theta}_1) + bias(\theta, \hat{\theta}_1)^2 \\
&= Var(\hat{\theta}_1) \\
&= Var(\frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n x_i}) \\
&= \frac{1}{(\sum_{i=1}^n x_i)^2} \sum_{i=1}^n Var(Y_i) \\
&= \frac{n\sigma^2}{(\sum_{i=1}^n x_i)^2}
\end{aligned}
$$

$$
\begin{aligned}
mse(\theta, \hat{\theta}_2) &= Var(\hat{\theta}_2) + bias(\theta, \hat{\theta}_2)^2 \\
&= Var(\hat{\theta}_2) \\
&= Var(\frac{1}{n} \sum_{i=1}^n \frac{Y_i}{x_i}) \\
&= \frac{1}{n^2} \sum_{i=1}^n \frac{1}{x_i^2} Var(Y_i) \\
&= \frac{1}{n^2} \sum_{i=1}^n \frac{1}{x_i^2} \sigma^2 \\
&= \frac{\sigma^2}{n^2 \sum_{i=1}^n x_i^2}
\end{aligned}
$$

(c)

From lecture note (3.15),

$$
\begin{aligned}
mse(\theta, \hat{\theta}) &= Var(\hat{\theta}) + bias(\theta, \hat{\theta})^2 \\
&= Var(\hat{\theta}) \\
&= \frac{\sigma^2}{\sum_{i=1}^n x_i^2}
\end{aligned}
$$

Now, since

$$
\begin{aligned}
mse(\theta, \hat{\theta}) &= \frac{\sigma^2}{\sum_{i=1}^n x_i^2} \\
&= \frac{n}{\sum_{i=1}^n x_i^2} \frac{\sigma^2}{n} \\
mse(\theta, \hat{\theta}_1) &= \frac{n\sigma^2}{(\sum_{i=1}^n x_i)^2} \\
&= (\frac{n}{\sum_{i=1}^n x_i)})^2 \frac{\sigma^2}{n} \\
mse(\theta, \hat{\theta}_2) &= \frac{\sigma^2}{n^2 \sum_{i=1}^n x_i^2} \\
&= \frac{1}{n \sum_{i=1}^n x_i^2} \frac{\sigma^2}{n}
\end{aligned}
$$

we can compare the part that doesn't involve $\frac{\sigma^2}{n}$.

Recall Cauchy-Schwartz inequality:

$$
(\sum_{i=1}^n x_i b_i) \leq (\sum_{i=1}^n x_i^2)(\sum_{i=1}^n b_i^2)
$$

Set $b_i = 1$. Then we have

$$
\frac{n}{\sum_{i=1}^n x_i^2} \leq (\frac{n}{\sum_{i=1}^n x_i)})^2
$$

Hence,

$$
mse(\theta, \hat{\theta}) \leq mse(\theta, \hat{\theta}_1)
$$

Set $b_i = \frac{1}{a_i}$. Then we have

$$
n^2 \leq (\sum_{i=1}^n x_i^2)(\frac{1}{\sum_{i=1}^n x_i^2})
$$

Hence,

$$
mse(\theta, \hat{\theta}) \leq mse(\theta, \hat{\theta}_2)
$$

Thus, we proved that $\hat{\theta}$ has the lowest mse.

4. Auto MPG

a. Estimate θ

```
df <- read.table(paste0("https://archive.ics.uci.edu/ml",
                        "/machine-learning-databases/auto-mpg/auto-mpg.data"))
colnames <- c('mpg','cylinders','displacement','horsepower',
              'weight','acceleration','model year','origin','car name')
colnames(df) <- colnames

Y <- df$mpg
X <- df$weight
n <- nrow(df)

calc_thetahat <- function (X,Y) {
```

```r
  sum(Y * X)/sum(X**2)
}

calc_thetahat1 <- function (X,Y) {
  sum(Y)/sum(X)
}

calc_thetahat2 <- function (X,Y) {
  1/n * sum(Y / X)
}

thetahat  <- calc_thetahat(X,Y)
thetahat1 <- calc_thetahat1(X,Y)
thetahat2 <- calc_thetahat2(X,Y)

cat("Theta Hat:  ", thetahat, '\n')
```

```
## Theta Hat:   0.006746881
```

```r
cat("Theta Hat 1:", thetahat1, '\n')
```

```
## Theta Hat 1: 0.007916233
```

```r
cat("Theta Hat 2:", thetahat2)
```

```
## Theta Hat 2: 0.009210333
```

```r
require(reshape2)
```

```
## Loading required package: reshape2
```

```r
pop <- data.frame(X,Y)
R <- 1000

thetahats <- replicate(R, {
                            sample <- pop[sample(nrow(pop), n, replace=TRUE), ]
                            calc_thetahat(sample$X,sample$Y)
                          })
thetahat1s <- replicate(R, {
                            sample <- pop[sample(nrow(pop), n, replace=TRUE), ]
                            calc_thetahat1(sample$X,sample$Y)
                          })
thetahat2s <- replicate(R, {
                            sample <- pop[sample(nrow(pop), n, replace=TRUE), ]
                            calc_thetahat2(sample$X,sample$Y)
                          })

cat("Variance of Theta Hat:  ", sd(thetahats), '\n')
```

```
## Variance of Theta Hat:   0.0002067892
```

```r
cat("Variance of Theta Hat 1:", sd(thetahat1s), '\n')
```

```
## Variance of Theta Hat 1: 0.0002316594
```

```r
cat("Variance of Theta Hat 2:", sd(thetahat2s))
```

```
## Variance of Theta Hat 2: 0.0002549126
```

We observe that the variance of $th\hat{e}ta$ is the smallest, which is consistent with the result from the previous problem.

5. Unbiased Estimates

(a)

$$\begin{aligned} E(e^{-3X}) - e^{-3\lambda} &= M_X(-3) - e^{-3\lambda} \\ &= e^{\lambda(e^{-3}-1)} - e^{-3\lambda} \end{aligned}$$

Hencem it is biased.

(b)

$$\begin{aligned} E((-2)^X) - e^{-3\lambda} &= \sum_{x=0}^{\infty} (-2)^x \frac{\lambda^x e^{-\lambda}}{x!} - e^{-3\lambda} \\ &= e^{-\lambda} \sum_{x=0}^{\infty} \frac{(-2\lambda)^x}{x!} - e^{-3\lambda} \\ &= e^{-\lambda} e^{-2\lambda} - e^{-3\lambda} \ (\because \text{Maclaurin series of } e^{-2\lambda}) \\ &= 0 \end{aligned}$$

Hence it is biased.

(c)

θ is always positive. On the other hand, $g(X)$ is negative half of the time. Hence, it is silly that we are using an estimator whose value is nonsensical for estimating the estimand half of the time. A better estimator would be $h(X) = max(0, g(X))$ since $h(X)$ is always non negative.