BST258_pset3

Problem 1

Let $\mathcal{P}_t = t\widetilde{\mathcal{P}} + (1-t)\mathcal{P}$ and by definition,

$$\Psi(\mathcal{P}) = E_{\mathcal{P}}\left[\{Y - E_{\mathcal{P}}(Y)\}\{X - E_{\mathcal{P}}(X)\}\right] = E_{\mathcal{P}}(g(O, \mathcal{P}))$$

where $g(O, \mathcal{P}) = \{Y - E_{\mathcal{P}}(Y)\}\{A - E_{\mathcal{P}}(A)\}$ and $O = (A, Y)$. With the hint,

$$\begin{aligned}
\frac{d\Psi(\mathcal{P}_t)}{dt}|_{t=0} &= E_{\mathcal{P}}[\frac{d}{dt}g(O, \mathcal{P}_t)]|_{t=0} + g(\tilde{o}, \mathcal{P}) - E_{\mathcal{P}}[g(O, \mathcal{P})] \\
&= E_{\mathcal{P}}[\frac{d}{dt}(Y - E_{\mathcal{P}_t}Y)(A - E_{\mathcal{P}_t}A)]|_{t=0} + (\tilde{y} - E_{\mathcal{P}}Y)(\tilde{a} - E_{\mathcal{P}}A) - E_{\mathcal{P}}[(Y - E_{\mathcal{P}}Y)(A - E_{\mathcal{P}}A)]
\end{aligned}$$

The first term can be written as,

$$E_{\mathcal{P}}[\frac{d}{dt}(Y - E_{\mathcal{P}_t}Y)(A - E_{\mathcal{P}_t}A)]|_{t=0} = E_{\mathcal{P}}\{-\frac{d}{dt}E_{\mathcal{P}_t}Y|_{t=0}(A - E_{\mathcal{P}}A) - (Y - E_{\mathcal{P}}Y)\frac{d}{dt}E_{\mathcal{P}_t}A|_{t=0}\}$$

Similarly, we can have $E_{\mathcal{P}_t}Y = E_{\mathcal{P}_t}(f(Y, \mathcal{P}_t))$, where $f(Y, \mathcal{P}_t) = Y$, and using the hint, we have

$$\frac{d}{dt}E_{\mathcal{P}_t}Y|_{t=0} = E_{\mathcal{P}}[\frac{dY}{dt}]|_{t=0} + \tilde{y} - E_{\mathcal{P}}(Y) = \tilde{y} - E_{\mathcal{P}}(Y)$$

Then we have

$$\begin{aligned}
E_{\mathcal{P}}[\frac{d}{dt}(Y - E_{\mathcal{P}_t}Y)(A - E_{\mathcal{P}_t}A)]|_{t=0} &= E_{\mathcal{P}}\{-(\tilde{y} - E_{\mathcal{P}}(Y))(A - E_{\mathcal{P}}A) - (\tilde{A} - E_{\mathcal{P}}(A))(Y - E_{\mathcal{P}}Y)\} \\
&= -(\tilde{y} - E_{\mathcal{P}}(Y))E_{\mathcal{P}}(A - E_{\mathcal{P}}A) - (\tilde{A} - E_{\mathcal{P}}(A))E_{\mathcal{P}}(Y - E_{\mathcal{P}}Y) \\
&= 0
\end{aligned}$$

Thus,

$$\frac{d\Psi(\mathcal{P}_t)}{dt}|_{t=0} = (\tilde{y} - E_{\mathcal{P}}Y)(\tilde{a} - E_{\mathcal{P}}A) - E_{\mathcal{P}}[(Y - E_{\mathcal{P}}Y)(A - E_{\mathcal{P}}A)]$$

The efficient influence function is,

$$\phi(O, \mathcal{P}) = (Y - E_{\mathcal{P}}Y)(A - E_{\mathcal{P}}A) - E_{\mathcal{P}}[(Y - E_{\mathcal{P}}Y)(A - E_{\mathcal{P}}A)]$$

Problem 2

The derivation follows the scratch by Hines et al. (2012). Let's define the expected conditional covariance on L.

$$\mathrm{cov}_t(X,Y|L=l) = \mathbb{E}_{\mathcal{P}_t}\left[(Y-\mathbb{E}_{\mathcal{P}_t}(Y|L))(X-\mathbb{E}_{\mathcal{P}_t}(X|L))|L=l\right]$$

By the law of total variance,

$$\Psi(\mathcal{P}_t) = \mathbb{E}_{\mathcal{P}_t}(\mathrm{cov}_t(X,Y|L))$$

Then by the hint in Q1, we have

$$\begin{aligned}
\frac{d}{dt}\Psi\left(\mathcal{P}_t\right) &= \frac{d}{dt}\mathbb{E}_{\mathcal{P}_t}\left[\mathrm{cov}_t(X,Y|L)\right] \\
&= \mathbb{E}_{\mathcal{P}_t}\left[\frac{d}{dt}\mathrm{cov}_t(X,Y|L=l)\right]|_{t=0} + \mathrm{cov}_{\mathcal{P}}(X,Y|\tilde{l}) - \mathbb{E}_{\mathcal{P}}(\mathrm{cov}_{\mathcal{P}}(X,Y|L))
\end{aligned}$$

For any l, with the hint, we can have

$$\begin{aligned}
\frac{d}{dt}\mathrm{cov}_t(X,Y|L=l)|_{t=0} &= \frac{d}{dt}\mathbb{E}_{\mathcal{P}_t}\left[(Y-\mathbb{E}_{\mathcal{P}_t}(Y|L))(X-\mathbb{E}_{\mathcal{P}_t}(X|L))|L=l\right]|_{t=0} \\
&= \frac{\mathbb{1}_{\tilde{l}}(l)}{f(l)}\left[(\tilde{y}-\mathbb{E}_{\mathcal{P}}(Y|\tilde{l}))(\tilde{x}-\mathbb{E}_{\mathcal{P}}(X|\tilde{l})) - \mathbb{E}_{\mathcal{P}}((Y-\mathbb{E}_{\mathcal{P}}(Y|L))(X-\mathbb{E}_{\mathcal{P}}(X|L))|L=l)\right] \\
&+ \mathbb{E}_{\mathcal{P}}\left[\frac{d}{dt}(Y-\mathbb{E}_{\mathcal{P}_t}(Y|L))(X-\mathbb{E}_{\mathcal{P}_t}(X|L)|L=l\;|_{t=0}\right]
\end{aligned}$$

Similar to Q1, we have

$$\mathbb{E}_{\mathcal{P}}\left[\frac{d}{dt}(Y-\mathbb{E}_{\mathcal{P}_t}(Y|L))(X-\mathbb{E}_{\mathcal{P}_t}(X|L)|L=l\;|_{t=0}\right] = 0$$

Then, we take the expectation for the first term,

$$\begin{aligned}
&\mathbb{E}_{\mathcal{P}}\left[\frac{\mathbb{1}_{\tilde{l}}(l)}{f(l)}\left[(\tilde{y}-\mathbb{E}_{\mathcal{P}}(Y|\tilde{l}))(\tilde{x}-\mathbb{E}_{\mathcal{P}}(X|\tilde{l})) - \mathbb{E}_{\mathcal{P}}((Y-\mathbb{E}_{\mathcal{P}}(Y|L))(X-\mathbb{E}_{\mathcal{P}}(X|L))|L=l)\right]\right] \\
&= \int\int\int\frac{\mathbb{1}_{\tilde{l}}(l)}{f(l)}\left[(\tilde{y}-\mathbb{E}_{\mathcal{P}}(Y|\tilde{l}))(\tilde{x}-\mathbb{E}_{\mathcal{P}}(X|\tilde{l}))f(\tilde{x},\tilde{y}|\tilde{l})f(\tilde{l})d\tilde{l}d\tilde{x}d\tilde{y}\right] - \mathbb{E}_{\mathcal{P}}((Y-\mathbb{E}_{\mathcal{P}}(Y|L))(X-\mathbb{E}_{\mathcal{P}}(X|L))|L=l) \\
&= \int\int\left[(\tilde{y}-\mathbb{E}_{\mathcal{P}}(Y|\tilde{l}=l))(\tilde{x}-\mathbb{E}_{\mathcal{P}}(X|\tilde{l}=l))f(\tilde{x},\tilde{y}|\tilde{l}=l)d\tilde{x}d\tilde{y}\right] - \mathbb{E}_{\mathcal{P}}((Y-\mathbb{E}_{\mathcal{P}}(Y|L))(X-\mathbb{E}_{\mathcal{P}}(X|L))|L=l) \\
&= 0
\end{aligned}$$

Thus, we have $\frac{d}{dt}\mathbb{E}_{\mathcal{P}_t}[\text{cov}_t(X, Y|L = l)|_{t=0}] = \text{cov}_{\mathcal{P}}(X, Y|\tilde{l}) - \mathbb{E}_{\mathcal{P}}(\text{cov}_{\mathcal{P}}(X, Y|L))$. The efficient influence function is

$$\phi(O, \mathcal{P}) = (Y - \mathbb{E}_{\mathcal{P}}(Y|L))(A - \mathbb{E}_{\mathcal{P}}(A|L)) - \Psi(\mathcal{P})$$

Problem 3

The one-step estimator is

$$\Psi\left(\hat{\mathcal{P}}_n\right) + \frac{1}{n}\sum_{i=1}^{n}\phi\left(O_i, \hat{\mathcal{P}}_n\right) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \mathbb{E}_{\hat{\mathcal{P}}_n}(Y|L_i))(A_i - \mathbb{E}_{\hat{\mathcal{P}}_n}(A|L_i))$$

I will conduct a simple simulation where

- $L_i \sim N(1, 3)$
- $A_i \sim N(-2L_i + 1, \sigma^2 = 1)$
- $Y_i \sim N(L_i + 5L_i^2 + 3, \sigma^2 = 1)$

Thus, conditional on $L = l$, the covariance between Y and A is 0 for all $l = (0, 1)$, and the expected conditional covariance is 0.

a.

```r
simulation_function <- function(sample_size) {
  L <- rnorm(sample_size, 1, 3)
  A <- rnorm(sample_size, mean = -2*L + 1, sd = 1)
  Y <- rnorm(sample_size, mean = L + 5*L^2 + 3 , sd = 1)
  return(data.frame(L, Y, A))
}

glm_estimate <- function(data) {
  model_Y <- glm(Y ~ L + I(L^2), data = data, family = gaussian)
  model_A <- glm(A ~ L, data = data, family = gaussian)

  Y_pred <- predict(model_Y, data)
  A_pred <- predict(model_A, data)

  # One-step estimator with a glm
  est_psi <- mean((data$Y - Y_pred) * (data$A - A_pred))

  return(data.frame(true_psi = 0, est_psi = est_psi))
}


library(furrr)
```

```
Loading required package: future
```

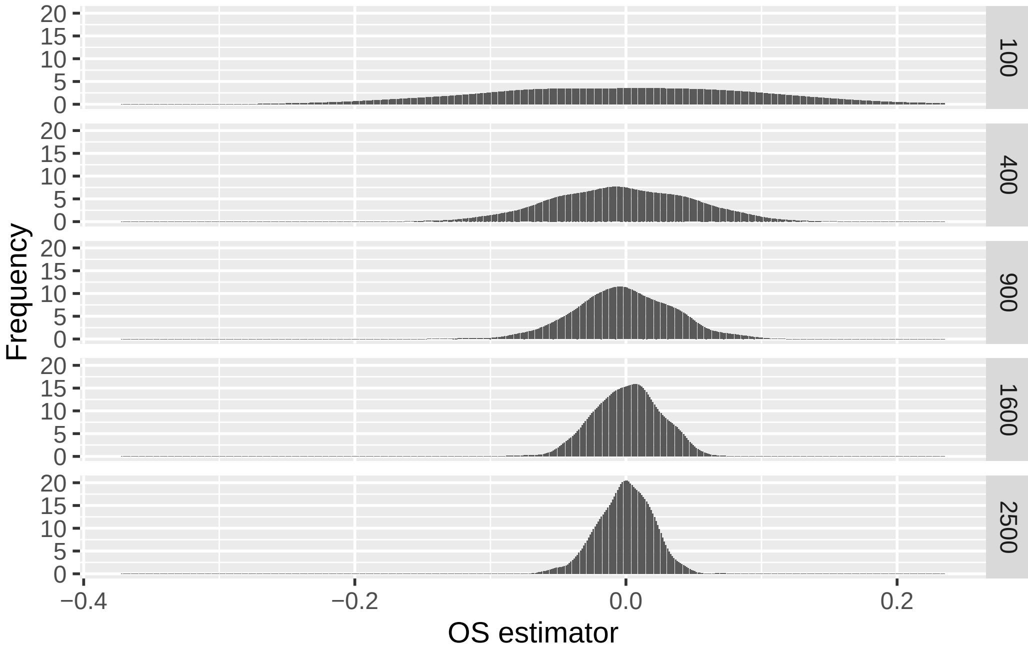
```r
# let's do 500 simulations
parallel_simulation <- function(sample_size) {
  future_map_dfr(1:500, ~glm_estimate(simulation_function(sample_size)),
                 .options = furrr_options(seed = TRUE))
}

# n = 100
set.seed(123)
plan(multisession)
glm_100 <- parallel_simulation(100)
glm_100$sample_size <- 100
# n = 400
glm_400 <- parallel_simulation(400)
glm_400$sample_size <- 400

# n= 900
glm_900 <- parallel_simulation(900)
glm_900$sample_size <- 900

#n=1600
glm_1600 <- parallel_simulation(1600)
glm_1600$sample_size <- 1600


# n = 2500
glm_2500 <- parallel_simulation(2500)
glm_2500$sample_size <- 2500

library(tidyverse)
```

```
-- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
v dplyr     1.1.4     v readr     2.1.5
v forcats   1.0.0     v stringr   1.5.1
v ggplot2   3.4.4     v tibble    3.2.1
v lubridate 1.9.3     v tidyr     1.3.1
v purrr     1.0.2

-- Conflicts ------------------------------------------ tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()    masks stats::lag()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
```

```r
ggplot(rbind(glm_100,glm_400, glm_900, glm_1600, glm_2500)) +
  geom_histogram(aes( x = est_psi),  stat = 'density') +
  facet_grid(sample_size ~ .) +
  labs(x = "OS estimator", y = "Frequency")
```

```
Warning in geom_histogram(aes(x = est_psi), stat = "density"): Ignoring unknown
parameters: `binwidth`, `bins`, and `pad`
```



- The OS estimators follow symmetric distributions, centered at 0. The dispersion decreases as the sample size increases.
- Sample size 100: average = -0.0058927, sd = 0.1006725.
- Sample size 400: average = -0.0033602, sd = 0.0502757.
- Sample size 900: average = -0.0015261, sd = 0.0353352.
- Sample size 1600: average = 9.7034424×10^{-4}, sd = 0.0240561.
- Sample size 2500: average = -5.752864×10^{-4}, sd = 0.0197859.
- The trend of standard deviation suggests a $\frac{1}{\sqrt{n}}$ convergence rate.

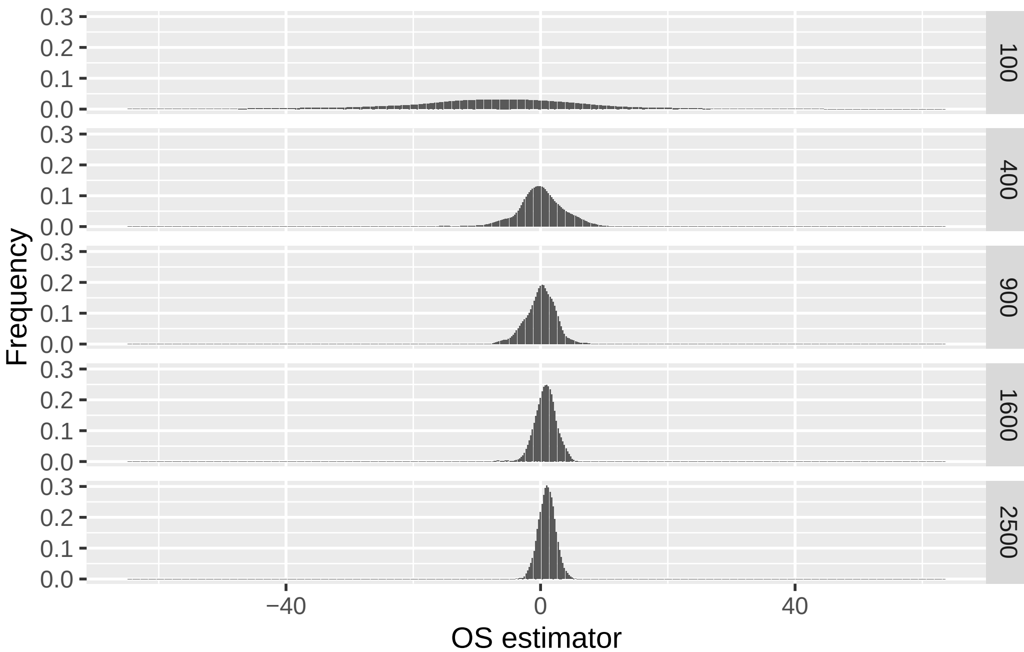
b.

I will use a simple random forest as the nonparametric regression approach

```
library(rpart)
RF_estimate <- function(data) {
  model_Y <- rpart(Y ~ L, data = data)
  model_A <- rpart(A ~ L, data = data)

  Y_pred <- predict(model_Y, data)
  A_pred <- predict(model_A, data)

  # One-step estimator with a glm
  est_psi <- mean((data$Y - Y_pred) * (data$A - A_pred))

  return(data.frame(true_psi = 0, est_psi = est_psi))
}

# let's do 500 simulations
parallel_simulation <- function(sample_size) {
  future_map_dfr(1:500, ~RF_estimate(simulation_function(sample_size)),
                 .options = furrr_options(seed = TRUE))
}

# n = 100
set.seed(123)
plan(multisession)
rf_100 <- parallel_simulation(100)
rf_100$sample_size <- 100
# n = 400
rf_400 <- parallel_simulation(400)
rf_400$sample_size <- 400

# n= 900
rf_900 <- parallel_simulation(900)
rf_900$sample_size <- 900

#n=1600
rf_1600 <- parallel_simulation(1600)
rf_1600$sample_size <- 1600


# n = 2500
```

```r
rf_2500 <- parallel_simulation(2500)
rf_2500$sample_size <- 2500

#  n = 4000

library(tidyverse)
ggplot(rbind(rf_100, rf_400, rf_900, rf_1600, rf_2500)) +
  geom_histogram(aes( x = est_psi), stat = 'density') +
  facet_grid(sample_size ~ .) +
  labs(x = "OS estimator", y = "Frequency")
```

```
Warning in geom_histogram(aes(x = est_psi), stat = "density"): Ignoring unknown
parameters: `binwidth`, `bins`, and `pad`
```



- The OS estimators follow symmetric distributions, centered at 0. The dispersion decreases as the sample size increases.
- Sample size 100: average = -6.9433792, sd = 16.0353327.
- Sample size 400: average = -0.1103702, sd = 4.2296345.
- Sample size 900: average = 0.0069962, sd = 2.3509404.
- Sample size 1600: average = 0.7813882, sd = 1.6810417.
- Sample size 2500: average = 0.9970777, sd = 1.2957094.

c

- With a larger sample size, the parametric approach has averages very close to 0, but the non-parametric approach still has non-zero averages, indicating asymptotic bias. Recall $\hat{\Psi}(\hat{\mathcal{P}}_n) - \Psi(\mathcal{P}) = \mathrm{P}_n(\hat{f}) - \mathrm{P}(f)$ can be decomposed into three terms, where $\Psi(\mathcal{P}) = \mathrm{P}(f) = 0$ in this case. Among them, $\mathrm{P}(\hat{f} - f)$ is asymptotic bias for non-parametric approaches but will vanish for parametric approaches.
- The variance of the non-parametric approach is larger than the parametric approach. In this simulation, we correctly specified the parametric models, which will be asymptotically efficient, and have smaller variance compared to the non-parametric approach.

Problem 4

a. Derive the equation

$$
\begin{aligned}
&\operatorname{cov}(Y, A \mid L) \\
&= \mathbb{E}(Y \mid L, A=1)\mathbb{E}(A \mid L) - \mathbb{E}(A \mid L) \cdot \mathbb{E}(Yl) \\
&= \mathbb{E}(Y \mid L, A=1)\mathbb{E}(A \mid L) - \mathbb{E}^2(A \mid L)\mathbb{E}(Y \mid L, A=1) - \mathbb{E}(A \mid L)(1-\mathbb{E}(A \mid L))\mathbb{E}(Y \mid L, A=0) \\
&= \mathbb{E}(A \mid L)[\mathbb{E}(Y \mid L, A=1) - \mathbb{E}(Y \mid L, A=0)] - \mathbb{E}^2(A \mid L)[\mathbb{E}(Y \mid L, A=1) - \mathbb{E}(Y \mid L, A=0) \\
&= [\mathbb{E}(A \mid L) - \mathbb{E}^2(A \mid L)][\mathbb{E}(Y \mid L, A=1) - \mathbb{E}(Y \mid L, A=0)] \\
&= \operatorname{Var}(A|L)[\mathbb{E}(Y^1 \mid L) - \mathbb{E}(Y^0 \mid L)]
\end{aligned}
$$

Then,

$$
\frac{\mathbb{E}\{\operatorname{Cov}(Y, A \mid L)\}}{\mathbb{E}\{\mathbb{V}(A \mid L)\}} = \mathbb{E}\{\frac{\operatorname{Var}(A|L)}{\mathbb{E}\{\mathbb{V}(A \mid L)\}}[\mathbb{E}(Y^1 \mid L) - \mathbb{E}(Y^0 \mid L)]\}
$$

b

We have derived the efficient influence function for $\mathbb{E}_{\mathcal{P}}(\operatorname{cov}_{\mathcal{P}}(Y, A|L))$, which we denote by $\phi_1(O, \mathcal{P})$. With the hint, we can derive the efficient influence function $\phi_2(O, \mathcal{P})$ for $\mathbb{E}_{\mathcal{P}}(\mathrm{V}_{\mathcal{P}}(A|L))$ as,

$$
\phi_2(O, \mathcal{P}) = (A - \mathbb{E}_{\mathcal{P}}(A|L))^2 - \mathbb{E}_{\mathcal{P}}(\mathrm{V}_{\mathcal{P}}(A|L))
$$

With the chain rule, the efficient influence function for $\frac{\mathbb{E}\{\operatorname{Cov}(Y,A|L)\}}{\mathbb{E}\{\mathbb{V}(A|L)\}}$ is

$$
\begin{aligned}
&\frac{d}{dt}\frac{\mathbb{E}_{\mathcal{P}_t}\{\operatorname{Cov}_{\mathcal{P}_t}(Y, A \mid L)\}}{\mathbb{E}_{\mathcal{P}_t}\{\mathbb{V}_{\mathcal{P}_t}(A \mid L)\}}\Big|_{t=0} = \frac{1}{\mathbb{E}_{\mathcal{P}}\{\mathbb{V}_{\mathcal{P}}(A \mid L)\}}\phi_1 - \frac{\mathbb{E}_{\mathcal{P}}\{\operatorname{Cov}_{\mathcal{P}}(Y, A \mid L)\}}{\mathbb{E}^2_{\mathcal{P}}\{\mathbb{V}_{\mathcal{P}}(A \mid L)\}}\phi_2 \\
&= \frac{1}{\mathbb{E}_{\mathcal{P}}\{\mathbb{V}_{\mathcal{P}}(A \mid L)\}}[(A - \mathbb{E}_{\mathcal{P}}(A|L))(Y - \mathbb{E}_{\mathcal{P}}(Y|L)) - \operatorname{Cov}_{\mathcal{P}}(Y, A|L)] \\
&- \frac{\mathbb{E}_{\mathcal{P}}\{\operatorname{Cov}_{\mathcal{P}}(Y, A \mid L)\}}{\mathbb{E}^2_{\mathcal{P}}\{\mathbb{V}_{\mathcal{P}}(A \mid L)\}}[(A - \mathbb{E}_{\mathcal{P}}(A|L))^2 - \mathbb{E}_{\mathcal{P}}(\mathrm{V}_{\mathcal{P}}(A|L))] \\
&= \frac{[A - \mathbb{E}_{\mathcal{P}}(A|L)][Y - \mathbb{E}_{\mathcal{P}}(Y|L)]\mathbb{E}_{\mathcal{P}}\{\mathbb{V}_{\mathcal{P}}(A \mid L)\} - \mathbb{E}_{\mathcal{P}}\{\operatorname{Cov}_{\mathcal{P}}(Y, A \mid L)\}(A - \mathbb{E}_{\mathcal{P}}(A|L))^2}{\mathbb{E}^2_{\mathcal{P}}\{\mathbb{V}_{\mathcal{P}}(A \mid L)\}}
\end{aligned}
$$

C

The one-step estimators is

$$
\begin{aligned}
\hat{\Psi} &= \frac{\mathbb{E}_{\hat{\mathcal{P}}}\{\mathrm{Cov}(Y, A \mid L)\}}{\mathbb{E}_{\hat{\mathcal{P}}}\{\mathbb{V}(A \mid L)\}} + \frac{1}{n}\sum_{i=n}^{n} \frac{\mathrm{cov}_{\hat{\mathcal{P}}}(A, Y \mid L)\mathbb{E}_{\hat{\mathcal{P}}}(\mathbb{V}_{\hat{\mathcal{P}}}(A|L)) - \mathbb{V}_{\hat{\mathcal{P}}}(A \mid L)\mathbb{E}_{\hat{\mathcal{P}}}(\mathrm{cov}_{\mathcal{P}}(A, Y|L))}{\mathbb{E}^2_{\hat{\mathcal{P}}}(\mathbb{V}_{\hat{\mathcal{P}}}(A|L))} \\
&= \frac{1}{n}\sum_{i=1}^{n}\{ \frac{[A - \mathbb{E}_{\hat{\mathcal{P}}}(A|L)][Y - \mathbb{E}_{\hat{\mathcal{P}}}(Y|L)]\mathbb{E}_{\hat{\mathcal{P}}}\{\mathbb{V}_{\mathcal{P}}(A \mid L)\} - [A - \mathbb{E}_{\hat{\mathcal{P}}}(A|L)]^2\mathbb{E}_{\hat{\mathcal{P}}}(\mathrm{cov}_{\mathcal{P}}(A, Y|L))}{\mathbb{E}^2_{\hat{\mathcal{P}}}(\mathbb{V}_{\hat{\mathcal{P}}}(A|L))} \\
&+ \frac{\mathbb{E}_{\hat{\mathcal{P}}}\{\mathrm{Cov}(Y, A \mid L)\}}{\mathbb{E}_{\hat{\mathcal{P}}}\{\mathbb{V}(A \mid L)\}}\}
\end{aligned}
$$

Problem 5

a.

I will parametric models to fit the conditional distribution of A and Y. Assuming the three identification assumptions, I assume the following model holds:

1. $A|L \sim Bernoulli(\theta(l))$, where

$$\text{logit}(\theta(l)) = \alpha^T l$$

2. $Y|L \sim N(\mu_l, \sigma^2)$, where

$$\mu_l = \beta^T L$$

I use the same logistic regression as in Pset 2 to fit the model for $A|L$. I use the same outcome regression model as in Pset2 but exclude the treatment (qsmk) and all its interactions terms. I use the empirical average to estimate the expectation.

```
dta <- read.csv("nhefs.csv")
dta <- dta %>% filter(!is.na(wt82))

os_estimator <- function(data) {
  fit_qsmk <- glm(qsmk ~ sex + age + I(age^2) + as.factor(race) + as.factor(education) + smol

theta <- predict(fit_qsmk, data = data, type = "response")


fit_outcome <- lm(wt82_71  ~ sex + age + I(age^2) + as.factor(race) + as.factor(education) +




A_predict <- predict(fit_qsmk, data = data, type = "response")
Y_predict <- predict(fit_outcome, data = data)

cov_AY <- (data$wt82_71 -Y_predict)*(data$qsmk - A_predict)
E_cov_AY <- mean(cov_AY)

Var_A <-(data$qsmk-A_predict)^2
E_Var_A <- mean(Var_A)

ic <- (cov_AY*E_Var_A - Var_A*E_cov_AY)/(E_Var_A^2)
```

```r
os <- mean(ic) + E_cov_AY/E_Var_A

# estiamte the analytic variance

analytic_sd <-sd(ic)/sqrt(nrow(data))

return(data.frame(os_estimate = os,
                  analytic_sd = analytic_sd))
}


os_estimator(dta)
```

```
  os_estimate analytic_sd
1     3.45932   0.4670395
```

The estimated variance weighted treatment effect for smoking cessation on weight change is 3.46.

b.

As before, the one-step estimator can be decomposed into three parts

$$\hat{\Psi} - \Psi = T^* + T_1 + T_2$$

where $T^* = (P_n - P)(\phi(O, P))$, T_1 is the empirical process term, and T_2 is the reminder bias term. Since we are using parametric models, $T_2 = o_p(1/\sqrt{n})$, then by the CLT,

$$\sqrt{n}(\hat{\Psi} - \Psi) \rightsquigarrow N(0, \mathbb{V}(\phi(O, P)))$$

I will use sample variance to estimate $\mathbb{V}(\phi(O, P)))$. See the output above, the analytic variance is 0.467. Then I will conduct a bootstrap to estimate the variance of the one-step estimator.

```r
set.seed(123)
library(future)
library(furrr)
plan(multisession)

boot_df <- furrr::future_map_dfr(1:1000,
                      .f = function(boot_id){
  boot_data <- dta[sample(1:nrow(dta), replace = TRUE),]
```

```r
  os_estimator(boot_data) %>% return()},
.options = furrr_options(seed = TRUE))



sd(boot_df$os_estimate)
```

```
[1] 0.462666
```

The estimated standard error by bootstrap is 0.462, very close to the analytic one.