PH C240B: HW 2

Yue You

Problem 2

Let (W, A, Y) be the observed data with distribution, $P_0 \in M$ nonparametric. Define the following parameter mapping for $P \in M : \Psi(P) = E_P[(1, 1, W)\beta]$ where $\beta = \Psi^1(P) = \text{argmin}_\gamma E_P(Y - (1, A, W)\gamma)^2$.

(a) The empirical distribution, $\mathbf{P}_n$, is the NPMLE for the true distribution. We use $\mathbf{P}_n$ as a plug-in estimator, $\Psi(\mathbf{P}_n)$, for the true parameter, $\Psi(P_0)$. This is called the NPMLE for $\Psi(P_0)$. Derive this estimator's influence curve. You may derive the efficient influence curve, $D^*_\Psi(P)$ first and then your answer is $D^*_\Psi(\mathbf{P}_n)$ This is a valid approach but not the only approach.

Because $\Psi(P)$ is a function of $\Psi^1(P)$ we first derive the efficient influence curve for $\Psi^1(P) = \beta = \text{argmin}_\gamma E_P(Y - (1, A, W)\gamma)^2$. First we need to compute the pathwise derivative (also let's assume that W is d-dimensional:

We can use the delta method to derive the efficient influence curve for this parameter mapping:

$D^*_\Psi(P) = \frac{d\Psi}{d\beta} D^*_\beta(P)(O)$
$= E_P(1, 1, W)D^*_\beta(P)(O)$

Next we compute $D^*_\beta(P)(O)$: $\because \Psi^{'}(P) = \beta = \text{argmin}_\gamma E_P(Y - (1, A, W)\gamma)^2$
$\therefore \beta\ satisfies\ \frac{d}{d\beta}\text{argmin}_\gamma E_P(Y - (1, A, W)\beta)^2 = 0_{d\times 1}$
$\Rightarrow 2[E_P(Y - (1, A, W)\beta)(1, A, W)^T] = 0_{d\times 1}$
$\Rightarrow E_P(Y - (1, A, W)\beta)(1, A, W)^T = 0_{d\times 1}$.....(1)

And $\because \Psi^{'}(P_\epsilon) = \beta = \text{argmin}_\gamma E_{P_\epsilon}(Y - (1, A, W)\gamma)^2$
$\therefore \beta_\epsilon\ satisfies\ \frac{d}{d\beta_\epsilon}\text{argmin}_\gamma E_{P_\epsilon}(Y - (1, A, W)\beta_\epsilon)^2 = 0_{d\times 1}$
$\Rightarrow 2[E_{P_\epsilon}(Y - (1, A, W)\beta_\epsilon)(1, A, W)^T] = 0_{d\times 1}$
$\Rightarrow E_{P_\epsilon}(Y - (1, A, W)\beta_\epsilon)(1, A, W)^T = 0_{d\times 1}$.....(2)
from (1)(2)$\therefore E_P(Y - (1, A, W)\beta)(1, A, W)^T - E_{P_\epsilon}(Y - (1, A, W)\beta_\epsilon)(1, A, W)^T = 0_{d\times 1}$
$\Rightarrow E_P(Y - (1, A, W)\beta)(1, A, W)^T - E_{P_\epsilon}(Y - (1, A, W)\beta)(1, A, W)^T$
$+E_{P_\epsilon}(Y - (1, A, W)\beta)(1, A, W)^T - E_{P_\epsilon}(Y - (1, A, W)\beta_\epsilon)(1, A, W)^T = 0_{d\times 1}$
$\therefore E_{P_\epsilon}(Y - (1, A, W)\beta_\epsilon)(1, A, W)^T - E_{P_\epsilon}(Y - (1, A, W)\beta)(1, A, W)^T$.....(3)
$= E_P(Y - (1, A, W)\beta)(1, A, W)^T - E_{P_\epsilon}(Y - (1, A, W)\beta)(1, A, W)^T$.....(4)
$\because lim_{\epsilon\to 0}\frac{(4)}{\epsilon}$
$= E_{P_\epsilon}(Y - (1, A, W)\frac{\beta_\epsilon - \beta}{\epsilon})(1, A, W)^T$
$= lim_{\epsilon\to 0}E_{P_\epsilon}(1, A, W)^T(1, A, W)\frac{\beta_\epsilon - \beta}{\epsilon})$
$= E_{P_\epsilon}(1, A, W)^T(1, A, W)\frac{d\Psi^{'}(P_\epsilon)}{d\epsilon}|_{\epsilon=0}$.....(5)
And $\because lim_{\epsilon\to 0}\frac{(3)}{\epsilon}$
$= lim_{\epsilon\to 0}\int(Y - (1, A, W))\beta(1, A, W)^T\frac{P_\epsilon(O) - P(O)}{\epsilon}dv(O)(\because P_\epsilon(O) = (1 + \epsilon h)P(O))$
$= \int(Y - (1, A, W))\beta(1, A, W)^T S(O)P(O)dv(O)$
$= E_p[(Y - (1, A, W))\beta(1, A, W)^T S(O)]$.....(6)
$\therefore (5) = (6)$
$\Rightarrow E_{P_\epsilon}(1, A, W)^T(1, A, W)\frac{d\Psi^{'}(P_\epsilon)}{d\epsilon}|_{\epsilon=0}$
$= E_p[(Y - (1, A, W))\beta(1, A, W)^T S(O)]$
$\Rightarrow \frac{d\Psi^{'}(P_\epsilon)}{d\epsilon}|_{\epsilon=0}$
$= [E_{P_\epsilon}(1, A, W)^T(1, A, W)]^{-1}E_p[(Y - (1, A, W))\beta(1, A, W)^T S(O)]$

$\Rightarrow D^*_\beta(P)(O) = [E_{P_\epsilon}(1,A,W)^T(1,A,W)]^{-1}E_p[(Y-(1,A,W))\beta(1,A,W)^T]$
$\Rightarrow IC = E_{P_n}(1,1,W)[E_{P_{n,\epsilon}}(1,A,W)^T(1,A,W)]^{-1}E_{p_n}[(Y-(1,A,W))\beta(1,A,W)^T] - \Psi(P)$

(b) Consider the following data generating process: W_1, W_2 are standard normals, $Pr(A=1) = expit(1 + 0.4W_1 - 0.4W_2W_1)$, $Y = (W_1+W_2)^2 + \epsilon$, $\epsilon \sim N[0,1]$. What is the true parameter value, $\Psi(P_0)$, for this data generating process?

To determine the true parameter value, $\Psi(P_0)$, we ran a simulation:

```
expit <- function(x){
  1/(1+exp(-x))
}

# generate the data
n <- 10000
W1 <- rnorm(n, 0, 1)
W2 <- rnorm(n, 0, 1)
A <- rbinom(n, 1, expit(1 + .4 * W1 - .4 * W2 * W1))
epsilon <- rnorm(n, 0 ,1)
Y <- (W1 + W2)^2 + epsilon

df <- data.frame(W1, W2, A, Y)

beta_hat <- lm(Y ~ ., data = df)

mean(beta_hat$fitted.values)
```

```
## [1] 1.975642
```

```
# pseudo-counterfactual dataframe forcing the treatment to 1
cf_df <- df
cf_df$A <- 1
cf_df <- cf_df[c("W1", "W2", "A")]

pred <- predict(beta_hat, cf_df)

mean(pred)
```

```
## [1] 1.729106
```

(c) Now consider the parameter, $\Psi^2(P) = E_P E_P[Y|A=1,W]$ or the true treatment specific mean for our non-parametric model. What is the true parameter value for the data generating process in part (b)?

The treatment specific mean is $E_{U,X}[Y(A=1)]$. Since A is fixed, Y is not directly affected by A so we can simply solve the mean of our Y values to find the treatment specific mean. Note that Y is indirectly affected by A through $W1$ and $W2$. We have that

$E_{U,X}[Y(A=1)] = EE[Y|A=1,W]$

When we condition on both A and W, we are making the assumption that $Y \perp A|W$.

```
n <- 10000
W1 <- rnorm(n, 0, 1)
W2 <- rnorm(n, 0, 1)
A <- 1
epsilon <- rnorm(n, 0 ,1)
Y <- (W1 + W2)^2 + epsilon

df <- data.frame(W1, W2, A, Y)
```

```r
mean(df$Y)
```

```
## [1] 1.996325
```

(d) Run a simulation where you take 1000 draws of $n = 1000$ from the data generating process in part (b). Using your NPMLE from part (a) and its influence curve to form 95% Wald confidence intervals, check coverage of the true Ψ and Ψ^2 for each draw. Report the coverage percentage for each parameter. Considering that estimating Ψ is often used to estimate Ψ^2, comment on your results.

```r
sim <- function(n) {
W1 <- rnorm(n, 0, 1)
W2 <- rnorm(n, 0, 1)
A <- rbinom(n, 1, expit(1 + .4 * W1 - .4 * W2 * W1))
epsilon <- rnorm(n, 0 ,1)
Y <- (W1 + W2)^2 + epsilon
df <- data.frame(W1, W2, A, Y)
cf_df <- df
cf_df$A <- 1
pred <- predict(beta_hat, cf_df)

mean(pred)
# true Psi_1 from part (b)
Psi_1 <- 1.75
# true Psi_2 from part (c)
Psi_2 <- 2

}
sim(1000)

registerDoParallel(cores = detectCores())
getDoParWorkers()
```

```
## [1] 4
```

```r
B <- 1000
n <- 1000
ALL <- foreach(i=1:B,.packages=c("mvtnorm"), .errorhandling = "remove") %dopar% {sim(n)}
res <- do.call(rbind, ALL)

colMeans(res)
```

```
## [1] 2
```

Problem 3

Bonus question: What is the remainder term for your estimator in part (a)? What remainder term conditions need to be satisfied for the estimator to be asymptotically linear?

We assume the NPMLE is well-behaved in the sense that it estimates the parameters in the second-order remainder at a rate faster than $n^{-1/4}$ (i.e. consistent at rate $< n^{-1/4}$). Then, $R_2(\mathbf{P}_n, P_0) = o_P(n^{-1/2})$ and $\Psi(\mathbf{P}_n) - \Psi(P_0) = (P_n - P_0)D^*(\mathbf{P}_n) + o_P(n^{-1/2})$. We had to make this assumption to derive the efficient influece curve for the NPMLE and the remainder term in part (a) is

$$R_2(\mathbf{P}_n, P_0) = \Psi(\mathbf{P}_n) - \Psi(P_0) - (P_n - P_0)D^*_\Psi(\mathbf{P}_n)$$

where $\Psi(\mathbf{P}_n) = \frac{1}{n}\sum_{i=1}^{n}(1, 1, W_i)\hat{\beta}$, $\Psi(P_0) = E_P[(1, 1, W_i)\beta]$, P_n is the empirical measure, P_0 is the true data distribution, and $D^*_\Psi(\mathbf{P}_n)$ is defined in part (a). Note that $(P_n - P_0)D^*_\Psi(\mathbf{P}_n)$ is the empirical processes applied to random functions, not a sum of i.i.d random variables because the random variable is defined by the data.

Collaborators & Resources

Rachael Philips, Chris Kennedy, Tommy Carpenito