Homework 4

Xinyi Lin

10/7/2019

Problem 1

Question a

p: the probability of the mice are obese at baseline is 0.2.

$theta0$: the probability of exposure to light for non-obese mice is 0.5.

$theta1$: the difference of probabilities of exposure to light between non-obese mice and obese mice is -0.2.

$beta0$: the population mean of glucose outcome for non-obese mice that are not exposed to light is 110.

$beta1$: the difference of glucose outcome population mean between non-obese mice and obese mice is 20.

$beta2$: the difference of glucose outcome population mean between mice exposed to light and mice not exposed to light is 5.

Question b

Marginal PACE:

$$
\begin{aligned}
E[Y_1] - E[Y_0] \quad &= \sum_C E\,(Y_1|C=c)Pr(C=c) - \sum_C E\,(Y_0|C=c)Pr(C=c)\,(IE) \\
&= \sum_C E\,(Y_1|A=1,C=c)Pr(C=c) - \sum_C E\,(Y_0|A=1,C=c)Pr(C=c)\,(RA) \\
&= \sum_{C,U} E\,(Y_1|A=1,C=c,U=u)Pr(C=c,U=u) - \sum_{C,U} E\,(Y_0|A=0,C=c,U=u)Pr(C=c,U=u)\,(IE+CA) \\
&= E(Y|A=1) - E(Y|A=0)
\end{aligned}
$$

Conditional PACE:

$$
\begin{aligned}
E[Y_1|V] - E[Y_0|V] \quad &= \sum_C E\,(Y_1|W=W,V=v)Pr(W=w|V=v) - \sum_C E\,(Y_0|W=w,V=v)Pr(W=w|V=v)\,(IE) \\
&= \sum_C E\,(Y_1|A=1,W=W,V=v)Pr(W=w|V=v) - \sum_C E\,(Y_0|A=1,W=w,V=v)Pr(W=w|V=v)\,(RA) \\
&= \sum_C E\,(Y|A=1,W=W,V=v)Pr(W=w|V=v) - \sum_C E\,(Y|A=1,W=w,V=v)Pr(W=w|V=v)\,(CA)
\end{aligned}
$$

In oder to let the marginal and conditional PACE are identified, following assumptions need to be satisfied: (1) no unobserved confounding assumption and (2) consistency assumption, the marginal and conditional PACE are identified.

Question c

g-formula in current homework:

$$
\begin{aligned}
E[Y_a] \quad &= E[E[Y_a|C]] \\
&= \sum_C E[Y_a|C = c]Pr[C = c] \\
&= \sum_C E[Y_a|A = a, C = c]Pr[C = c] \\
&= \sum_C E[Y|A = a, C = c]Pr[C = c]
\end{aligned}
$$

g-formula in previous homework:

$$
\begin{aligned}
&= \sum_C E[Y_a|C = c]Pr[C = c] \\
&= \sum_C E[Y_a|A = a, C = c]Pr[C = c] \\
&= \sum_C E[Y|A = a, C = c, U = u]Pr[C = c, U = u] \\
&= E[Y|A = a]
\end{aligned}
$$

The g-formula in previous homework is g-formula for randomized study, while g-formula in current homework is g-formula for observational study. Both of them yield the same causal effect, provided the covariates C that confound A have the same distribution in both types of studies. However, for observational study, the g-formula does not have a simple form as a conditional mean.

Question d

The estimate of $E[Y|A = 1] - E[Y|A = 0]$ is around -0.918.

Interpretation: the difference of glucose outcome mean between mice exposed to light and mice not exposed to light is around -0.918

Question e

The estimate of $E[Y_1] - E[Y_0]$ is around -5.714.

Interpretation: The estimator of average causal effect between mice exposed to light and mice not exposed to light is around -5.714 under no unmearsured confounder assumption and consistency assumption.

Difference between two estimator:

For estimator in question (d), we can estimate the value directly from data as data obtained from randomized study. For estimator in question (e), we need to consider probability of covariate(C) to get the estimator of average casual effect.

Question f

Assumptions:

1. no unobserved confounding

2. consistency

3. probabilistic

4. 10 covariates are baseline covariates

Problem 2

Question a

Before coming to P8122, I learnt confounding in P6400.

The definition of coufounder are: 1. Associate with treatment; 2. Associate with outcome; 2. Not lie on the casual pathway between treatment and outcome.

This is the tradional definition of confounder we discuss in class.

Question b

Based on Backdoor criterion, confounding bias arises when the treatment and outcome in view share a common cause. If one has observed enough variables to block all backdoor paths, that is if treatment and outcome are d-separated given the measured covariates in a graph in which the arrow out of treatment are removed, the treatment effect is identified. The variables that can be adjusted to block backdoor paths is confounder.

Question c

$1. Y = f(A, L, \epsilon_Y), A = f(L, \epsilon_A), L = f(\epsilon_L)$

$2. Y = f(A, U, \epsilon_Y), A = f(L, \epsilon_A), L = f(U, \epsilon_L), U = f(\epsilon_U)$

$3. Y = f(U, \epsilon_Y), L = f(A, U, \epsilon_L), A = f(\epsilon_A), U = f(\epsilon_U)$

$4. Y = f(A, L, \epsilon_Y), L = f(U, \epsilon_L), A = f(U, \epsilon_A), U = f(\epsilon_U)$

$5. Y = f(A, U_1, \epsilon_Y), L = f(U_1, U_2, \epsilon_L), A = f(U_2, \epsilon_A), U_1 = f(\epsilon_{U_1}), U_2 = f(\epsilon_{U_2})$

Question d
1. Yes; 2. Yes; 3. Yes; 4. Yes; 5. Yes.

Question e

1. Yes; 2. Yes; 3. No; 4. Yes; 5. No;

Question f

According to the tradiontional definition, L in DAG 3 and 5 are confounders and conditioning on it helps adjust for confounding. But according to backdoor criterion, L in DAG 3 and 5 are not confounders but colliders. So conditioning on L will make A and Y be d-connected instead of block the path. Adjusting for confounding is satisfied by blocking the path.

Appendix

```
knitr::opts_chunk$set(echo = FALSE)
set.seed(123)
n <- 16
p_C <- 1/5
C <- rbinom(n,1,p_C)
theta0 <- 1/2
theta1 <- -1/5
p_A <- theta0+theta1*C
A <- rbinom(n,1,p_A)
beta0 <- 110
beta1 <- 20
beta2 <- -5
sigma_Y <- 1
mu_Y <- beta0+beta1*C+beta2*A
Y <- rnorm(n,mu_Y, sigma_Y)
est_mean = mean(Y[A==1])-mean(Y[A==0])
est_mean2=(mean(Y[A==1 & C==1])*mean(C) + mean(Y[A==1 & C==0])*(1-mean(C)))-
(mean(Y[A==0 & C==1])*mean(C) + mean(Y[A==0 & C==0])*(1-mean(C)))
```