STA640 Homework 4

Fan Bu

10/20/2020

First creat a data frame object `dat` to represent the data.

```
library(tidyverse)
dat = data.frame(Z = c(rep(0,185+123+9+41), rep(1, 37+20+26+96)),
                 W = c(rep(0,185+123), rep(1, 9+41),
                       rep(0,37+20), rep(1, 26+96)),
                 Y = c(rep(0,185), rep(1,123), rep(0,9), rep(1,41),
                       rep(0,37), rep(1,20), rep(0,26), rep(1,96)))
```

Problem 1

The ITT estimate is simply

$$\frac{\sum_{i=1}^n Z_i Y_i}{\sum_{i=1}^n Z_i} - \frac{\sum_{i=1}^n (1-Z_i) Y_i}{\sum_{i=1}^n (1-Z_i)},$$

and I'll estimate the standard error via bootstrap.

```
# estimate
ITTy = mean(dat$Y[dat$Z == 1]) - mean(dat$Y[dat$Z == 0])

# SE via bootstrap
set.seed(42)
B = 1000
N = nrow(dat)
ITT.boot = sapply(1:B, function(i){
  dat.boot = dat %>% slice(sample(1:N, N, replace = T))
  mean(dat.boot$Y[dat.boot$Z == 1]) - mean(dat.boot$Y[dat.boot$Z == 0])
})

cat('ITT estimate:', ITTy, '\nStandard error:', sd(ITT.boot), '\n')
```

```
## ITT estimate: 0.1899441
## Standard error: 0.04495113
```

Problem 2

The 4 possible pre-assignment groups are

- $W(0) = 0, W(1) = 0$, "never-takers": people who would not take physiotherapy no matter whether or not they receive the discount.
- $W(0) = 0, W(1) = 1$, "compliers"; people who would take physiotherapy if offered a discount, but wouldn't otherwise.
- $W(0) = 1, W(1) = 0$, "defiers"; people who would take physiotherapy if **not** offered a discount, but wouldn't do so if offered a discount.
- $W(0) = 1, W(1) = 1$, "always-takers"; people who would always take physiotherapy no matter whether or not they receive the discount.

Problem 3

1. Existence of compliers (Z has direct effect on received treatment)

$$Pr(W_i = 1 \mid Z_i = 1) > 0 \text{ and } Pr(W_i = 0 \mid Z_i = 0) > 0 \quad \text{for all } i.$$

   This assumption is plausible and this should be true in almost all randomized trials (we would expect at least some people to comply)

2. Exclusion Restriction for non-compliers (Z doesn't have direct effects on the outcomes).

$$Y_i(0) = Y_i(1), \quad \text{for all } i \in S_i = n, a.$$

   This is plausible, as for never-takers and always-takers, their treatment assignment doesn't affect if they actually receive the treatment, and so the potential outcomes would be the same.

Problem 4

We need the "random assignment" assumption (which is implicitly assumed in **problem 3**), and also the **monotonicity** assumption, which is assumption 1 above plus $W_i(1) \geq W_i(0)$ (no defiers).

Since we assume there is no defiers, we only need to estimate π_c, π_n, π_a as follows:

$$\begin{aligned}
\hat{\pi}_a &= \frac{\sum_i W_i(1 - Z_i)}{\sum_i (1 - Z_i)}; \\
\hat{\pi}_n &= \frac{\sum_i (1 - W_i) Z_i}{\sum_i Z_i} = 1 - \frac{\sum_i W_i Z_i}{\sum_i Z_i}; \\
\hat{\pi}_c &= 1 - \hat{\pi}_a - \hat{\pi}_n = \frac{\sum_i W_i Z_i}{\sum_i Z_i} - \frac{\sum_i W_i(1 - Z_i)}{\sum_i (1 - Z_i)}.
\end{aligned}$$

Again, standard errors are estimated via bootstrap.

```r
get_prop_est <- function(d){
  est = numeric(3)
  # pi_a:
  est[1] = mean(d$W[d$Z==0])
  # pi_n:
  est[2] = 1 - mean(d$W[d$Z==1])
  # pi_c:
  est[3] = 1 - est[1] - est[2]

  est
}


# bootstrap
set.seed(42)
B = 1000
N = nrow(dat)
props.boot = sapply(1:B, function(i){
  dat.boot = dat %>% slice(sample(1:N, N, replace = T))
  get_prop_est(dat.boot)
})
```

```r
Est = get_prop_est(dat)
SEs = apply(props.boot, 1, sd)
```

```r
res = data.frame(Group = c('always-taker','never-taker','complier'),
                 Est.prop = Est,
                 SE = SEs)
knitr::kable(res, digits = 4)
```

Group	Est.prop	SE
always-taker	0.1397	0.0187
never-taker	0.3184	0.0355
complier	0.5419	0.0401

Problem 5

The intent-to-treat effect for true compliers is defined as

$$\begin{aligned} ITT_c &= \mathbb{E}(Y_i(1) - Y_i(0) \mid S_i = c) \\ &= \mathbb{E}(Y_i(1) - Y_i(0) \mid W_i(0) = 0, W_i(1) = 1). \end{aligned}$$

Under the assumptions in problems 3 (and 4), we have

$$ITT = \pi_c ITT_c,$$

which means we can estimate ITT_c by

$$\hat{\tau}_c = \hat{\tau}_y / \hat{\pi}_c,$$

where $\hat{\tau}_y$ is the esimate for ITT and $\hat{\pi}_c$ is the estimated proportion. Same as before, I'll use bootstrap to estimate the standard error.

```r
get_ITTc_est <- function(d){
  # 1. ITT estimate
  ITTy = mean(d$Y[d$Z == 1]) - mean(d$Y[d$Z == 0])

  # 2. pi_c
  pi_c = mean(d$W[d$Z==1]) - mean(d$W[d$Z==0])

  # 3. ITT_c
  ITTc = ITTy/pi_c

  ITTc
}


# SE via bootstrap
set.seed(42)
B = 1000
N = nrow(dat)
ITTc.boot = sapply(1:B, function(i){
  dat.boot = dat %>% slice(sample(1:N, replace = T))
  get_ITTc_est(dat.boot)
})
```

```r
cat('For true compliers,\nITT estimate:', get_ITTc_est(dat),
    '\nStandard error:', sd(ITTc.boot), '\n')
```

```
## For true compliers,
## ITT estimate: 0.3505155
## Standard error: 0.07652557
```

Problem 6

The estimate in **(5)** is really about the clinical **efficacy** of the treatment, as it represents the effect of actually taking physiotherapy. This estimate provides more meaningful clinical implications.

However, the estimate in **(1)** is more about the **effectiveness** of the treatment for a general population, as it represents the effect of "promoting physiotherapy". This estimate provides more meaningful healthy policy implications.

Problem 7

It's about "selective missing" (patient lost to follow-up depends on "compliance").

Since whether or not a patient is still reachable at six months after surgery is a post-treatment confounder - it is influenced by the treatment and also determines if we can access their outcome at all. Moreover, for those patients who are lost to follow-up, their ITT effects may not be well-defined (e.g., patients who die or have severe issues within six months cannot even have potential outcomes at six months).

Problem 8

Let the parameters θ consist of

$$\begin{aligned} p_{s,z} &= Pr(Y_i = 1 \mid Z_i = z, S_i = s) \\ \pi_s &= Pr(S_i = s) \end{aligned}$$

for $s \in \{a, n, c, d\}$ and $z \in \{0, 1\}$.

Then the likelihood contribution of the ith individual (assuming we know S_i) can be written as

$$\begin{aligned} & p(Y_i^{obs}, S_i \mid Z_i, \theta) \\ =& p(Y_i^{obs} \mid S_i, Z_i, \theta) p(S_i \mid \theta) \\ =& p_{S_i,Z_i}^{Y_i^{obs}} (1 - p_{S_i,Z_i})^{1-Y_i^{obs}} \pi_{S_i}. \end{aligned}$$

And then the likelihood would be

$$\begin{aligned} & \Pi_{i=1}^n p(Y_i^{obs}, S_i \mid Z_i, \theta) \\ =& \Pi_{i=1}^n [p_{S_i,Z_i}^{Y_i^{obs}} (1 - p_{S_i,Z_i})^{1-Y_i^{obs}} \pi_{S_i}] \\ =& \Pi_{i:S_i=a,Z_i=1} [p_{a,1}^{Y_i^{obs}} (1 - p_{a,1})^{1-Y_i^{obs}} \pi_a] \\ & \times \Pi_{i:S_i=a,Z_i=0} [p_{a,0}^{Y_i^{obs}} (1 - p_{a,0})^{1-Y_i^{obs}} \pi_a] \\ & \times \Pi_{i:S_i=n,Z_i=1} [p_{n,1}^{Y_i^{obs}} (1 - p_{n,1})^{1-Y_i^{obs}} \pi_n] \\ & \times \Pi_{i:S_i=n,Z_i=0} [p_{n,0}^{Y_i^{obs}} (1 - p_{n,0})^{1-Y_i^{obs}} \pi_n] \\ & \times \Pi_{i:S_i=c,Z_i=1} [p_{c,1}^{Y_i^{obs}} (1 - p_{c,1})^{1-Y_i^{obs}} \pi_c] \\ & \times \Pi_{i:S_i=c,Z_i=0} [p_{c,0}^{Y_i^{obs}} (1 - p_{c,0})^{1-Y_i^{obs}} \pi_c] \\ & \times \Pi_{i:S_i=d,Z_i=1} [p_{d,1}^{Y_i^{obs}} (1 - p_{d,1})^{1-Y_i^{obs}} \pi_d] \\ & \times \Pi_{i:S_i=d,Z_i=0} [p_{d,0}^{Y_i^{obs}} (1 - p_{d,0})^{1-Y_i^{obs}} \pi_d]. \end{aligned}$$

Problem 9

Under the assumptions in **problem 3**, we have $\pi_d = 0$ and that $p_{n,1} = p_{n,0} = p_n$ and $p_{a,1} = p_{a,0} = p_a$.

Our estimand is

$$ITT_c = p_{c,1} - p_{c,0}.$$

Then the likelihood is simplified to

$$\begin{aligned}
&\Pi_{i=1}^n p(Y_i^{obs}, S_i \mid Z_i, \theta)\\
=&\Pi_{i:S_i=a}[p_a^{Y_i^{obs}}(1-p_a)^{1-Y_i^{obs}}\pi_a]\\
&\times \Pi_{i:S_i=n}[p_n^{Y_i^{obs}}(1-p_n)^{1-Y_i^{obs}}\pi_n]\\
&\times \Pi_{i:S_i=c,Z_i=1}[p_{c,1}^{Y_i^{obs}}(1-p_{c,1})^{1-Y_i^{obs}}\pi_c]\\
&\times \Pi_{i:S_i=c,Z_i=0}[p_{c,0}^{Y_i^{obs}}(1-p_{c,0})^{1-Y_i^{obs}}\pi_c].
\end{aligned}$$

Note that in reality we don't observe the true S_i but only $W_i^{obs} = W_i(Z_i)$. Adopting flat priors for the parameters θ, we can draw samples of S_i and $p_a, p_n, p_{c,1}, p_{c,0}, \pi_a, \pi_n, \pi_c$ via the following Gibbs sampler:

For $r = 1 : R$, do:

1. Draw $S_i^{(r)}$ for each person i:

- if $Z_i = 0, W_i^{obs} = 1$, i is definitely an always-taker, so set $S_i^{(r)} = a$ (no need to update or redraw);
- if $Z_i = 1, W_i^{obs} = 0$, i is definitely a never-taker, so set $S_i^{(r)} = n$ (no need to update or redraw);
- if $Z_i = 0, W_i^{obs} = 0$, draw $S_i^{(r)} \in \{n, c\}$ where

$$\begin{aligned}
Pr(S_i = n \mid \text{everything else}) &\propto {p_n^{(r-1)}}^{Y_i^{obs}}(1-p_n^{(r-1)})^{1-Y_i^{obs}}\pi_n^{(r-1)}\\
Pr(S_i = c \mid \text{everything else}) &\propto {p_{c,0}^{(r-1)}}^{Y_i^{obs}}(1-p_{c,0}^{(r-1)})^{1-Y_i^{obs}}\pi_n^{(r-1)}.
\end{aligned}$$

- if $Z_i = 1, W_i^{obs} = 1$, draw $S_i^{(r)} \in \{a, c\}$ where

$$\begin{aligned}
Pr(S_i = n \mid \text{everything else}) &\propto {p_a^{(r-1)}}^{Y_i^{obs}}(1-p_a^{(r-1)})^{1-Y_i^{obs}}\pi_a^{(r-1)}\\
Pr(S_i = c \mid \text{everything else}) &\propto {p_{c,1}^{(r-1)}}^{Y_i^{obs}}(1-p_{c,1}^{(r-1)})^{1-Y_i^{obs}}\pi_n^{(r-1)}.
\end{aligned}$$

2. Draw $\pi_a^{(r)}, \pi_n^{(r)}, \pi_c^{(r)}$ from

$$(\pi_a^{(r)}, \pi_n^{(r)}, \pi_c^{(r)}) \sim Dir((N_a^{(r)}, N_n^{(r)}, N_c^{(r)})),$$

where $N_s^{(r)} = \sum_{i=1}^n \mathbf{1}(S_i^{(r)} = s)$ for each $s \in \{a, n, c\}$.

3. Draw $p_a^{(r)}, p_n^{(r)}, p_{c,1}^{(r)}, p_{c,0}^{(r)}$ via

$$\begin{aligned}
p_a^{(r)} &\sim Beta(\sum_{i:S_i^{(r)}=a} Y_i^{obs}, N_a^{(r)} - \sum_{i:S_i^{(r)}=a} Y_i^{obs})\\
p_n^{(r)} &\sim Beta(\sum_{i:S_i^{(r)}=n} Y_i^{obs}, N_n^{(r)} - \sum_{i:S_i^{(r)}=n} Y_i^{obs})\\
p_{c,1}^{(r)} &\sim Beta(\sum_{i:S_i^{(r)}=c,Z_i=1} Y_i^{obs}, \sum_{i:S_i^{(r)}=c} Z_i - \sum_{i:S_i^{(r)}=c,Z_i=1} Y_i^{obs})\\
p_{c,0}^{(r)} &\sim Beta(\sum_{i:S_i^{(r)}=c,Z_i=0} Y_i^{obs}, \sum_{i:S_i^{(r)}=c} (1-Z_i) - \sum_{i:S_i^{(r)}=c,Z_i=0} Y_i^{obs}).
\end{aligned}$$

Then after running the Gibbs sampler to obtain R samples of S_i's and θ, the posterior distribution of ITT_c is approximated by the set of samples

$$\{p_{c,1}^{(r)} - p_{c,0}^{(r)}\}_{r=1}^{R}.$$