Segment 5: Analyzing Observational Studies

Section 05: Inverse Probability Weighting

Inverse Probability of Treatment Weighting

One way we've seen to "restructure" the data:

1. Estimate the propensity score
2. Weight the sample $\rightarrow$ pseudopopulation to approximate the design of a randomized trial
    - Weights will depend on the estimated propensity score
    - Treatment assignment will be unconfounded in the pseudopopulation

There are many methods for estimating causal effects that rely on this type of *Inverse Probability of Treatment* (IPTW) weighting

IPTW: Basic Intuition

Goal: Weight the observed sample to create a *pseudopopulation* that is balanced with respect to covariates (i.e., as though "randomized")

- The less likely a unit is to have received treatment $Z_i = z$, the larger representation it should get in the pseudopopulation
- Up-weight "weirdos": units whose treatment assignment is very different from what would be suggested by the propensity score
  - Unit $Z_i = 1$ and $e_i(X_i) = 0.05$ gets a weight of $\frac{1}{0.05} = 20$
  - Unit with $Z_i = 1$ and $e_i(X_i) = 0.9$ gets weight of $\frac{1}{0.9} = 1.11$
- Weights adjust for confounding by creating a pseudiopopulation where all individuals have the same probability of receiving $Z = 1$ and $Z = 0$
- Rooted in ideas of survey sampling

Example: IPTW Intuition

(adapted from Hernan and Robins textbook)

	X_i	Z_i	Y_i^{obs}
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	0
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1
16	1	1	1
17	1	1	1
18	1	1	0
19	1	1	0
20	1	1	0

- $X_i \in (0,1)$ for not severe, severe injury
- $Z_i \in (0,1)$ for control, surgery
- $Y_i^{obs} \in (0,1)$ for no complication, complication
- 69% $Z = 1$ have $X_i = 1$
- 43% $Z = 0$ have $X_i = 1$
- Assume unconfoundedness conditional on X
  - $Pr(Z_i = 1 | X_i = 1) = 0.75$
  - $Pr(Z_i = 1 | X_i = 0) = 0.50$

IPTW: Example

	X_i	Z_i	Y_i^{obs}
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	0
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1
16	1	1	1
17	1	1	1
18	1	1	0
19	1	1	0
20	1	1	0

- What is $Pr(Y_i(0) = 1)$: Prob of complication if everyone had received control?

IPTW: Example

	X_i	Z_i	Y_i^{obs}
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	0
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1
16	1	1	1
17	1	1	1
18	1	1	0
19	1	1	0
20	1	1	0

- What is $Pr(Y_i(0) = 1)$: Prob of complication if everyone had received control?
- Of the 8 patients with $X_i = 0$, 4 received $Z_i = 0$ and 1 had $Y_i^{obs} = 1$
  - How many complications would have occurred if all 8 of these units had had $Z_i = 0$?
  - 2: if the # of units is doubled, so is the number of complications
- Of the 12 with $X_i = 1$, 3 had $Z_i = 0$ and 2 had $Y^{obs} = 1$
  - How many complications would have occurred if all 12 of these had had $Z_i = 0$?
  - 8 (4 $\times$ as many units, 4$\times$ as many Y)

Inverse Probability Weighting

	X_i	Z_i	Y_i^{obs}
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	0
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1
16	1	1	1
17	1	1	1
18	1	1	0
19	1	1	0
20	1	1	0

- What is $Pr(Y_i(0) = 1)$: Prob of complication if everyone had received control?
- If all 20 patients had received $Z_i = 0$
  - 2 (not 1) would have complications among the $X_i = 0$ patients
  - 8 (not 2) would have complications among the $X_i = 1$ patients
- $\Rightarrow Pr(Y_i(0) = 1) = \frac{10}{20} = 0.5$
- Analogously, $Pr(Y_i(1)) = 1) = 0.5$

(Why can we assume that the rate of complication would have been the same in each level of X_i?)

Inverse Probability Weighting: Intuition

- Think about two hypothetical scenarios where all are treated or untreated
- **Key Idea:** Combine those two hypothetical situations to create a hypothetical population in which every unit appears as *both* treated and untreated
  - A *pseudopopulation* larger than the observed data
- With conditionally unconfounded treatment assignment, the pseuodopopulation will have *unconditionally* unconfounded assignment
  - I.e., the covariate distribution will be the same in the treated and control group
- $\Rightarrow$ The treated/control contrast in the pseudopopluation represents a causal effect (without further adjustment in the pseudopopulation)

Inverse Probability Weighting: Intuition

	X_i	Z_i	Y_i^{obs}
1	0	0	0
2	0	0	1
3	0	0	0
4	0	0	0
5	0	1	0
6	0	1	0
7	0	1	0
8	0	1	1
9	1	0	1
10	1	0	1
11	1	0	0
12	1	1	1
13	1	1	1
14	1	1	1
15	1	1	1
16	1	1	1
17	1	1	1
18	1	1	0
19	1	1	0
20	1	1	0

- Consider the 4 units with $Z_i = 0$ and $X_i = 0$
  - Used to create 8 units in the pseudopopulation
  - Each receives weight, $w_i = 2 = \frac{1}{0.5}$
  - $Pr(Z_i = 0|X_i = 0) = 0.5$
- Consider the 9 units with $Z_i = 1$ and $X_i = 1$
  - Used to create 12 members of the pseudopopulation
  - Each receives weight, $w_i = 1.33 = \frac{1}{0.75}$
  - $Pr(Z_i = 1|X_i = 1) = 0.75$

Inverse Probability Weighting

In general, for unit i with $X_i = x$ and $Z_i = z$

$$w_i = \frac{1}{Pr(Z_i = z | X_i = x)}$$

More generally,

$$w_i = \frac{1}{f(Z_i | X_i)} = \frac{Z_i}{e_i(X_i)} + \frac{1 - Z_i}{1 - e_i(X_i)}$$

Inverse of probability of receiving the treatment actually received.

For estimating causal effects:

$$E[Y(z)] = E\left[\frac{I(Z = z)Y^{obs}}{f(Z|X)}\right]$$

IP Weights Based on Propensity Scores

- In simple cases, could estimate $Pr(Z|X)$ nonparametrically
  - Ensure exact balance on X in the pseudopopulation
- We have already discussed why this may not be feasible and motivated the propensity score $e(X_i)$ (and its estimate) as a quantity that can
- The idea for IPW with the estimated propensity score extends in a straightforward way
  - Where $\hat{e}_i(X_i)$ comes from a parametric model
  - Should create balance on *propensity score* in pseudopopulation
  - Should still check balance on individual covariates in the pseudopopulation

$$w_i = \frac{Z_i}{\hat{e}(X_i)} + \frac{(1 - Z_i)}{(1 - \hat{e}(X_i))}$$

IPTW Intuition

Based on propensity scores

- $Z_i = 1, \hat{e}(X_i) = 0.05$
  - $\Rightarrow$ Unit i has X_i that look different from other $Z = 1$ units
  - $\Rightarrow$ Units with this X_i are underrepresented in the $Z = 1$ group
  - $\Rightarrow$ Upweight them by $\frac{1}{0.05}$
  - $\Rightarrow$ Represent 20 observations in the pseudopopulation

- $Z_i = 1, \hat{e}(X_i) = 0.90$
  - $\Rightarrow$ Unit i has X_i that look similar to other $Z = 1$ units
  - $\Rightarrow$ Units with this X_i are common in the $Z = 1$ group
  - $\Rightarrow$ Upweight them by $\frac{1}{0.90}$
  - $\Rightarrow$ Represent 1.11 observations in the pseudopopulation

Estimating the ATE with IPW

Difference in the mean of the weighted outcomes is an unbiased estimate of the ATE:

$$
\begin{aligned}
\hat{\tau}_{ipw,1} =& \frac{1}{N}\Big(\sum_{i=1}^{N}\frac{Y_iZ_i}{e(X_i)} - \sum_{i=1}^{N}\frac{Y_i(1-Z_i)}{1-e(X_i)}\Big) \\
=& \frac{1}{N}\sum_{i=1}^{N}\big(Y_iZ_iw_1(X_i) - Y_i(1-Z_i)w_0(X_i)\big)
\end{aligned}
$$

where $w_1(X_i) = \frac{1}{e(X_i)}$ and $w_0(X_i) = \frac{1}{1-e(X_i)}$ are weights for the treated, untreated, respectively

In practice, replace $e(X_i)$ with $\hat{e}(X_i)$, which actually improves efficiency

In practice, frequently accomplished with the `library(survey)` package in `R`

IPW Notes

- Advantages
  - Elegant theoretical foundation
  - Improved efficiency (relative to matching or coarser stratification)
  - Extension to more complex settings (e.g., time-varying treatments)
  - Can be thought of as a continuation of stratification as the # of strata goes up and size of strata goes down
- Disadvantages
  - More sensitive to propensity score model specification
    - E.g., vs. stratification, which "shrinks" or "smooths" weights to the average within the stratum
  - $\hat{e}(X_i)$ near 0 or 1 $\rightarrow$ very extreme weights
    - $e(X_i) \rightarrow 1$: $w_1(X_i) \rightarrow$
  - Erratic finite sample performance
  - Results can be dominated by a few observations