

Module 4

Wednesday, October 23, 2024 1:11 PM

Inverse Probability of Treatment Weighting (IPTW)

Another approach to this problem is, rather than matching, we could use all the data but down-weight some and up-weight others. This approach is called **Inverse Probability of Treatment Weighting**, and it is accomplished by weighting people by the inverse of their probability of treatment received.

- For treated subjects, weight by the inverse of $P(A = 1|X)$
- For control subjects, weight by the inverse of $P(A = 0|X)$

Suppose there is a single binary confounder X and also:

$$P(A = 1|X = 1) = 0.1$$

- Among people with $X=1$, only 10% will receive treatment (i.e. the propensity score for people with $X=1$ is 0.1)
- Note that among people with $X=1$, this means 90% do not receive treatment ($P(X = 0|X = 1) = 1 - P(A = 1|X = 1) = 1 - 0.1 = 0.9$)

$$P(A = 1|X = 0) = 0.8$$

- Among people with $X=0$, 80% will receive treatment (i.e. the propensity score for people with $X=0$ is 0.8)

Thus among people with $X=1$, there will be 1 person in the treatment group for every 9 such people in the control group. If we were to do Propensity Score matching to balance the groups, the one $X=1$ person in the treated group would represent 1 person but the one selected representative in the control group would be representing 9 people.

In the example above, the single treated subject with $X=1$ would be weighted like this: $\frac{1}{P(A=1|X=1)} = \frac{1}{0.1} = 10$

And the control subjects with $X=1$ would each be weighted like so: $\frac{1}{P(A=0|X=1)} = \frac{1}{0.9} = \frac{10}{9}$

Among subjects with $X=0$, there are four subjects in the treated group for every one subject in the control group.

So the treated subjects with $X=0$ would be weighted like this: $\frac{1}{P(A=1|X=0)} = \frac{1}{0.8} = \frac{5}{4}$

And the control subjects with $X=0$ would be weighted like this: $\frac{1}{P(A=0|X=0)} = \frac{1}{0.2} = 5$

In observational studies, certain groups are oversampled relative to the hypothetical sample from a randomized trial. There is confounding in the original population, but since IPTW creates a pseudo-population where treatment assignment no longer depends on X , there is no confounding in the pseudo-population.

In surveys, it is common to oversample some groups relative to the population. To estimate the population mean, we can weigh the data to account for the oversample (**Horvitz-Thompson estimator**). Under the assumption of exchangeability and positivity, we can estimate a potential outcome (let's say outcome=1) as the average value of Y if everyone had been treated. In the estimate below, we divide the sum of the outcomes in the treated pseudo-population divided by the number of subjects in the treated pseudo-population:

$$\frac{\sum_{i=1}^n I(A_i = 1) \frac{Y_i}{\pi_i}}{\sum_{i=1}^n \frac{I(A_i = 1)}{\pi_i}}$$

Where π_i is the propensity score.

Marginal Structural Models

IPTW estimation methods can be used more generally to estimate causal effect parameters from models.

A **Marginal Structural Model (MSM)** is a model for the mean of the potential outcomes.

- Marginal: model is not conditional on the confounders
- Structural: model for potential outcomes, not observed outcomes

Linear MSM:

$$E(Y^a) = \varphi_0 + \varphi_1 a \quad a=0,1$$

- $E(Y^0) = \varphi_0$
- $E(Y^1) = \varphi_0 + \varphi_1$

Thus φ_1 is the **average causal effect** $E(Y^1) - E(Y^0)$

Logistic MSM:

$$\text{logit}\{E(Y^a)\} = \varphi_0 + \varphi_1 a \quad a=0,1$$

Thus $\exp(\varphi_1)$ is the **causal odds ratio**:

$$\frac{\text{Odds that } Y^1 = 1}{\text{Odds that } Y^0 = 1} = \frac{\frac{P(Y^1 = 1)}{1 - P(Y^1 = 1)}}{\frac{P(Y^0 = 1)}{1 - P(Y^0 = 1)}}$$

Suppose V is a variable that modifies the effect of A ; then a **Linear MSM with effect modification**:

$$E(Y^a|V) = \varphi_0 + \varphi_1 a + \varphi_3 V + \varphi_4 aV \quad a=0,1$$

So the **average causal effect**: $E(Y^1|V) - E(Y^0|V) = \varphi_1 + \varphi_4 V$

The **general MSM** formula is as follows:

$$g\{E(Y^a|V)\} = h(a, V; \varphi)$$

- $g()$ is a link function
- $h()$ is a function specifying parametric form of a and V (typically additive, linear)

IPTW Estimation

Consider estimation of parameters from a linear regression model $Y = X\beta + \varepsilon$ where estimation involves solving:

$$\sum_{i=1}^n X_i(Y_i - X_i^T \hat{\beta}) = 0$$

For $\hat{\beta}$, the estimated value that minimizes the sum of squared deviations (the least squares estimator).

The same concept applies to other types of models for IPTW estimation. Since the pseudo-population obtained via IPTW is free from confounding (assuming ignorability and positivity), we can therefore estimate MSM parameters by solving estimating equations for the observed data of the pseudo-population. The generalized steps are as follows:

1. Estimate the propensity score
2. Create weights
 - a. 1 divided by the propensity score for treated subjects
 - b. 1 divided by (1 minus the propensity score for control subjects)
3. Specify the MSM of interest
4. Use software to fit a weighted generalized linear model
5. Use asymptotic (sandwich) variance estimator (or bootstrapping)
 - a. This accounts for the fact that the pseudo-population might be larger than the sample size

Assessing Balance

Covariate balance can be checked on the weighted sample using standardized differences, either in a Table 1 or in a plot. Recall that the standardized difference is the difference in means between groups divided by the pooled standard deviation:

$$smd = \frac{\bar{X}_{treatment} - \bar{X}_{control}}{\sqrt{s_{treatment}^2 + s_{control}^2} / 2}$$

The same concept applies after weighting, except the SMD is calculated on weighted means and weighted variances, stratified on treatment group.

Practice Quiz

1. Someone who was likely to be treated, given their covariates, but wasn't:
 - a. Will have a large weight
2. Marginal Structural Models are models of:
 - a. The mean of the potential outcome as a function of possible values of treatment
3. IPTW estimation works because:
 - a. It creates an unconfounded pseudo-population

Distribution of Weights

One downside to IPTW is that using larger weights can lead to large standard errors and noisier estimate of causal effects. If one data point is heavily weighted and especially if the outcome is binary, whether they have the event or not could have a big impact on the parameter estimate. In other words, if one person's outcome data can greatly affect the parameter estimate, the standard error will be large.

Consider bootstrapping, which is one way to estimate standard errors. This process involves randomly sampling from the original sample and estimating the parameters, and repeating this process many times. However, one person has a very large weight; whether or not they are included in the sample will have a big impact on the parameter estimates and therefore much of the variability of the estimate is caused by this one person.

Remedies for Large Weights

A good first step is to look into why the weights are large. Identify the subjects who have large weights and identify: what is unusual about them? Is there a problem with their data? Is there a problem with the propensity score model?

One way to handle large weights is, prior to doing the IPTW, consider **trimming the tails** of the propensity score distribution. This can eliminate some subjects who have extreme values of the propensity score (a common rule of thumb is to trim those whose scores are above the 98th or below the 2nd percentile), though keep in mind that trimming the tails changes the population.

Another option is known as **weight truncation**:

- Determine a maximum allowable weight (which could be a specific value or based off of a percentile)
- If a weight is greater than the maximum allowable, set it to the maximum allowable value.

Whether or not to truncate weights involves a variance-bias trade-off:

- Truncation: some bias, but smaller variance
- No truncation: unbiased, but larger variance

Truncating extremely large weights can result in estimators with lower MSE.

Doubly Robust Estimators

Recall that with IPTW, the weighting is the inverse of the estimated propensity score and if the propensity score is correctly specific, the estimator is unbiased.

Doubly Robust Estimation is also known as Augmented IPTW and it is unbiased if either the propensity score model or the outcome regression model are correctly specified. In general AIPTW should be more efficient than regular IPTW estimators.

- Suppose the propensity score is correctly specified, but the outcome regression model is not. Then the Augmentation portion of the estimation (where the regression model term m_1 is) ends up equaling zero and doesn't impact the estimation. So the estimation is the average of the propensity score.

$$\frac{1}{n} \sum_{i=1}^n \left\{ \underbrace{\frac{A_i Y_i}{\pi_i(X_i)}}_{\text{IPTW}} - \underbrace{\frac{A_i - \pi_i(X_i)}{\pi_i(X_i)} m_1(X_i)}_{\text{Augmentation}} \right\}$$

- Suppose the propensity score is incorrectly specific, but the outcome regression model is correctly specified. Then the difference in the numerator goes to zero, leaving only the portion with the model term. Thus the estimation is the expected value of Y^1 . This is easier to see after rearranging the estimation formula like below.

$$= \frac{1}{n} \sum_{i=1}^n \left\{ \frac{A_i (Y_i - m_1(X_i))}{\pi_i(X_i)} + m_1(X_i) \right\}$$

IPTW Quiz

1. Inverse Probability of Treatment Weighting involves weighting the data by:
 - a. 1 divided by the propensity score for treated subjects and 1 divided by 1 minus the propensity score for control subjects.
2. Among control subjects, would someone with a high value of the propensity score get more or less weight than someone with a low value of the propensity score?
 - a. More weight
3. Marginal Structural models are used to model:
 - a. Average causal effects
4. The pseudo-population refers to the population:
 - a. After weighting
5. Compared with situations where weights are small, large weights lead to causal effect estimates that are:
 - a. More variable
6. Near violation of the positivity assumption occurs when there are some weights that are very large:
 - a. True
7. Weight truncation is the same as trimming the tails.
 - a. False
8. Weight truncation, compared with no weight truncation, will likely lead to causal effect estimates:
 - a. That have more bias, but less variance
9. Doubly robust estimators require that:
 - a. Either the propensity score model or outcome model are correctly specified