Regression and Reweighing Methods

Hans Jarett Ong

September 8, 2022

Contents

The Unreasonable Effectiveness of Linear Regression

Linear Regression for Estimating ATE

$$E[Y|X,D] = \beta_0 + \delta D + \beta_1 X$$

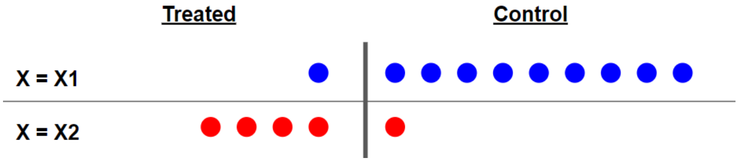
Assuming linearity,

$$\begin{aligned} E[Y|X,D=1] &= \beta_0 + \delta + \beta_1 X \\ E[Y|X,D=0] &= \beta_0 + \beta_1 X \\ E[Y|X,D=1] - E[Y|X,D=0] &= \delta = \text{ATE} \end{aligned}$$

- We can estimate ATE from the coefficient of the treatment variable D!
- Intuition: δ is the effect of the treatment on the outcome after accounting for the effect of the other variables in the model.

Inverse Probability of Treatment Weighting (IPTW)

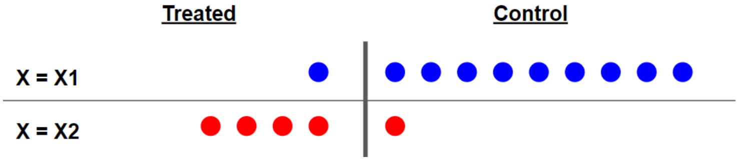
Intuition for IPTW



Say these 2 groups have the following propensity scores:

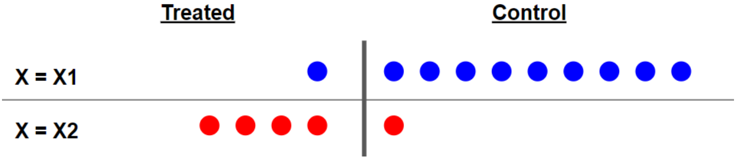
- $P(D = 1|X = x_1) = 0.1$
- $P(D = 1|X = x_2) = 0.8$

Intuition for IPTW



- We need to "balance" the treated and controlled groups per segment X.
- Segments with low probability of treatment should weigh more and vice versa.
- We can do this using the inverse of the probability of receiving treatment (hence IPTW).

Intuition for IPTW



$$\text{Weight} = \frac{1}{P(D|X)}$$

For $X = x_1$:

- Weight for Treated $= \frac{1}{P(D=1|X=x_1)} = \frac{1}{0.1} = 10$
- Weight for Control $= \frac{1}{P(D=0|X=x_1)} = \frac{1}{0.9} = \frac{10}{9} = 1.1$
- Each subject from the treatment group will weigh more compared to a subject from the control group.

ATE Estimation using IPTW

- More formally, the IPTW of an individual (w_i) is defined as

$$w_i = \frac{D_i}{\pi_i} + \frac{1 - D_i}{1 - \pi_i}$$

- ATE can be estimated using

$$\text{ATE} = \frac{1}{n} \sum_{i=1}^{n} \frac{D_i Y_i}{\pi_i} - \frac{(1 - D_i) Y_i}{1 - \pi_i}$$

- Note that π_i should be greater than 0 – this is the *positivity assumption*.
- Like in Propensity Score Matching, we usually have to use the estimated $\hat{\pi}_i$ in observational studies (e.g. using logistic regression)

Standard Errors in IPTW

- If we know the actual propensity scores π (if we're the one designing the experiment), then we can use the weighted average of the variance to get the standard error:

$$\sigma_w^2 = \frac{\sum_{i=1}^{n} w_i (y_i - \hat{\mu})^2}{\sum_{i=1}^{n} w_i}$$

- But if we're using the estimated propensity scores $\hat{\pi}_i$, then we can use **bootstrap estimation** to get the standard errors.

Doubly Robust Estimators

Doubly Robust Estimators

- a.k.a. augmented IPTW
- Doubly Robust Estimators combines Regression and IPTW estimators.
- It will be unbiased if at least one of the IPTW or Regression estimators are correctly specified.
- Think of this as diversification. Even if one of your estimators is wrong, you still have a chance to get a good estimate.
- Composed of 2 estimators (hence "double"):
  1. Outcome Estimation: $\hat{\mu_d}(x) = E[Y|X, D = d]$
  2. Propensity Score Estimation: $\hat{\pi}(x) = P(D = 1|X)$

Putting it together

$$\hat{E}[Y^1] = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{D_i(Y_i - \hat{\mu_1}(X_i))}{\hat{\pi}(X_i)} + \hat{\mu_1}(X_i)\right)$$

$$\hat{E}[Y^0] = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{(1-D_i)(Y_i - \hat{\mu_0}(X_i))}{1-\hat{\pi}(X_i)} + \hat{\mu_0}(X_i)\right)$$

$$\hat{\mathsf{ATE}} = \hat{E}[Y^1] - \hat{E}[Y^0]$$

Why it Works

Consider the first term:

$$\hat{E}[Y^1] = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{D_i(Y_i - \hat{\mu_1}(X_i))}{\hat{\pi}(X_i)} + \hat{\mu_1}(X_i)\right)$$

- If $\hat{\mu_1}(x)$ is correct, then $E[D_i(Y_i - \hat{\mu_1}(X_i))] = 0$.
- You will then be left with the correct outcome estimator $\hat{\mu_1}(X)$

The estimator above can be rearranged into the following form:

$$\hat{E}[Y^1] = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{D_iY_i}{\hat{\pi}(X_i)} - \frac{(D_i - \hat{\pi}(X_i))\hat{\mu_1}(X_i)}{\hat{\pi}(X_i)}\right)$$

- If $\hat{\pi}(X_i)$ is correct, then $E[D_i - \hat{\pi}(X_i)] = 0$.
- You will then be left with your regular IPTW estimator.

References

References

1. Hernan, Miquel A., and James M. Robins. Causal Inference. CRC Press, 2019.
2. Roy, Jason. “A Crash Course in Causality: Inferring Causal Effects from Observational Data — Coursera.” Coursera, https://www.coursera.org/learn/crash-course-in-causality. Accessed 15 Aug. 2022.
3. Matheus, Facure. “Causal Inference for The Brave and True.” Matheus Facure, https://matheusfacure.github.io/python-causality-handbook/landing-page.html. Accessed 15 Aug. 2022.
4. Morgan, Stephen L., and Christopher Winship. Counterfactuals and Causal Inference. Cambridge University Press, 2014.
5. Pearl, Judea, et al. Causal Inference in Statistics. John Wiley & Sons, 2016.
6. Yao, Liuyi, et al. “A Survey on Causal Inference.” ACM Transactions on Knowledge Discovery from Data, no. 5, Association for Computing Machinery (ACM), Oct. 2021, pp. 1–46. Crossref, doi:10.1145/3444944.