Potential Outcome Model

Hans Jarett Ong

September 8, 2022

Contents

1. The Potential Outcome Model
2. Average Treatment Effects
3. Assumptions in Estimating Treatment Effect
4. Confounders
5. References

The Potential Outcome Model

The Potential Outcome Model

- This framework stipulates the existence of **Potential Outcome Random Variables** usually written as

$$Y^{D=d}$$

  where Y is the *outcome* and D is the *exposure* variable (a.k.a. *treatment*).

- Furthermore, this Potential Outcome is also defined on an **individual level**.

Individual-Level Treatment Effects

If D were a binary treatment, we define the individual-level causal effect as

$$\delta_i = y_i^1 - y_i^0$$

There are other ways to represent effects (e.g. ratio, odds ratio, etc.), but this is the most common way.

- y_i^1 is the outcome of individual i *had they taken* the treatment
- y_i^0 is the outcome of individual i *had they NOT taken* the treatment

Fundamental Problem of Causal Inference

Group	Y^1	Y^0
Treatment group $(D = 1)$	Observable as Y	Counterfactual
Control group $(D = 0)$	Counterfactual	Observable as Y

- It is impossible to measure δ_i because we could only measure either y_i^1 or y_i^0 but not both.
- The unobserved or the "what if" (alternate universe) outcome is known as the **Counterfactual**.
- For example, if $d_i = 0$ then $y_i^0 = y_i$ but the counterfactual y_i^1 is unknown and unobserved.

Average Treatment Effects

Average Treatment Effects (ATE)

Although individual-level treatment effects are impossible to measure, we could still measure group-level treatment effects. One of the most common measures of group treatment effects if the **Average Treatment Effect (ATE)**:

$$E[\delta] = E[Y^1 - Y^0] = E[Y^1] - E[Y^0]$$

An alternative group-level treatment effect measure is the *causal risk ratio*:

$$\frac{Pr[Y^1 = 0]}{Pr[Y^0 = 1]}$$

which is mostly used in epidemiology and the health sciences.

"Correlation is not Causation"

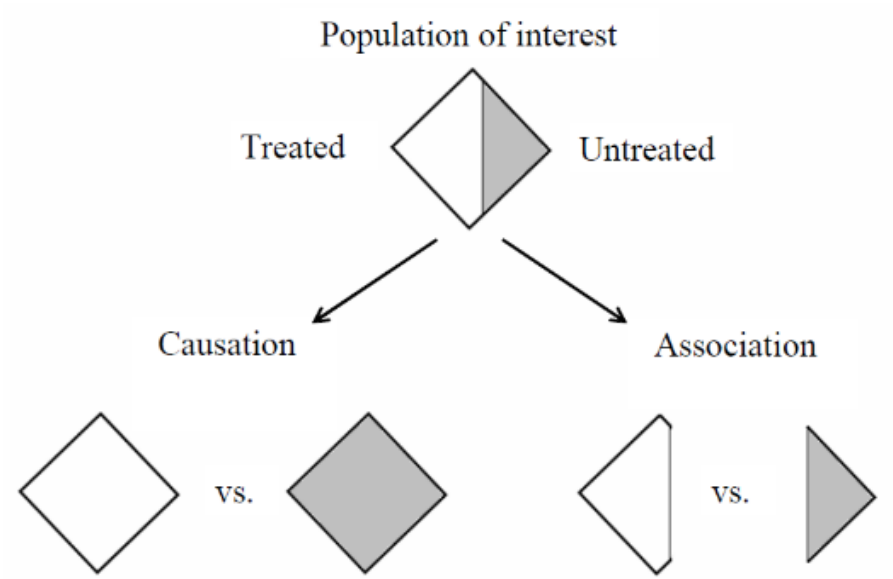

Image Source: *Hernan and Robins 2018 - Causal Inference: What If*

$$E[Y|D=1] - E[Y|D=0] \neq E[Y^1] - E[Y^0]$$

Naive group difference contains bias.

$$
\begin{aligned}
E[Y|D=1] - E[Y|D=0] &= E[Y^1|D=1] - E[Y^0|D=0] \\
&= E[Y^1|D=1] - E[Y^0|D=1] + E[Y^0|D=1] - E[Y^0|D=0] \\
&= E[Y^1 - Y^0|D=1] + \{E[Y^0|D=1] - E[Y^0|D=0]\}
\end{aligned}
$$

- The first term is the *ATT* (ATE of the treatment group).
- The second term is *bias*. This bias captures the difference due to treatment assignment, i.e. the inherent difference between the groups not caused by the treatment.

Why experiments work

Remember that

$$\begin{aligned} &E[Y|D=1] - E[Y|D=0] \\ &\quad = E[Y^1 - Y^0|D=1] + \{E[Y^0|D=1] - E[Y^0|D=0]\} \end{aligned}$$

- For randomized experiments, **treatment is assigned at random** making it independent to potential outcomes, i.e. $(Y^1, Y^0) \perp\!\!\!\perp D$
- This implies that $E[Y^0|D=1] = E[Y^0|D=0]$ and $E[Y^1 - Y^0|D=1] = E[Y^1 - Y^0]$
- Therefore, for randomized experiments

$$\text{ATE} = E[Y^1 - Y^0] = E[Y|D=1] - E[Y|D=0]$$

Assumptions in Estimating Treatment Effect

Assumption 1: SUTVA

Stable Unit Treatment Value Assumption (SUTVA)

- *The potential outcomes for any unit do not vary with the treatment assigned to other units, and, for each unit, there are no different forms or versions of each treatment level, which lead to different potential outcomes.* In other words:
  1. **There should be no interactions between units.** The potential outcomes of one unit shouldn't depend on the treatment assignment of other units.
  2. **There should only be a single version of the treatment.** For example, different dosages of the treatment should be counted as different treatments.
- SUTVA is usually violated when the intervention is large enough to saturate the system. Some examples are:
  - Effect of Vaccinations on Recovery/Transmission.
  - Effect of Training on Salary.

Assumption 2: Ignorability

$$(Y^0, Y^1) \perp\!\!\!\perp D|X$$

- *Given background variable X, potential outcomes are independent to the treatment assignment.*
- a.k.a. the *Unconfoundedness Assumption* – for units with the same background, treatment assignment can be viewed as random.
- Another way of saying it is "the treatment assignment mechanism D is *ignorable* given background variable X"
- This means that, assuming no unmeasured confounders, we can measure the Conditional Average Treatment Effect (CATE).

$$CATE = E[Y^1 - Y^0|X = x]$$

Assumption 3: Positivity

$$P(D = d|X = x) > 0 \quad \forall\ d, x$$

- *For any value of X, treatment assignment is not deterministic.*
- In other words, there should be no segment x in which every unit gets the same treatment. Any way to put it is that there should be treatment variance within each segment.
- Violating this assumption is problematic because it prevents us from estimating the potential outcomes. If everyone in a segment takes the same treatment, then we won't have information on the other potential outcomes (had they taken the other treatments).

Confounders

Confounders

- *Confounders are the variables that affect both the treatment assignment and the outcome.*
- Confounders are eliminated in randomized experiments but are common in observational studies.
- For example, a common confounder in medicine is *age*. Younger patients are more likely to recover regardless of the actual effect of the treatment, and older patients are more likely to receive the treatment.
- In Machine Learning, this is often called *covariate shift* which is a reason why it is hard for some models to generalize. e.g. You can't apply a model trained on an older population to a younger population.

Simpson's Paradox

We've seen this earlier, but here's another example:

	Treatment A	Treatment B
Young	$234/270 = 87\%$	$81/87 = \mathbf{92}\%$
Older	$55/80 = 69\%$	$192/263 = \mathbf{73}\%$
Overall	$289/350 = \mathbf{83}\%$	$273/350 = 78\%$

Taken from *Yao 2021. A Survey of Causal Inference.*

This paradox is caused by *confounders*.

Adjusting for Confounders

Adjusting for confounders involves adjusting the covariate distribution to make the distribution the same for both treatment and control groups. This can be done with the following methods:

- Matching
- Re-weighting
- Regression

We will discuss these in future lectures!

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.