Causal Analysis

Impact Evaluation and Causal Machine Learning with Applications in R

Chapter 2: Causality and No Causality

Table of Contents

Fundamental Problem of Causal Inference

- Causal effects are fundamentally unidentifiable at the individual level (Holland, 1986).
- Reason: The world cannot be observed simultaneously with and without a particular intervention at the same point in time.

Example

- Aim to compare an individual's wage with and without training participation.
- At any point in time, an individual has either participated or not participated in the training, but never both.
- Therefore, causal effect (e.g. difference in wage with and without training) cannot be directly observed for any individual.

Notation and Potential Outcome Framework

- Statistical notation:
  - Capital letters for random variables.
  - Lowercase letters for specific values of random variables.
- Aim to evaluate the causal effect of D on Y.
  - D is the intervention/treatment; $D = 1$ if someone receives the treatment, $D = 0$ if not.
  - Y is the observed outcome.

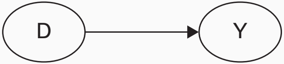


Figure 2.1: The causal effect of the treatment on the outcome in a causal graph (Pearl, 2000)

- Following Neyman (1923) and Rubin (1974), $Y(1)$ and $Y(0)$ denote the potential outcomes hypothetically realized if treatment D were set to 1 (treatment) and 0 (nontreatment), respectively.

Stable Unit Treatment Value Assumption (SUTVA)

Stable unit treatment value assumption (SUTVA)

The potential outcomes of an individual are not affected by the treatment status of others (Cox, 1958, Rubin, 1980).

- SUTVA implies, for any individual i in the population, that:

$$Y_i(d_i, \mathbf{d}_{-i}) = Y_i(d_i) \text{ for } d_i \in \{0, 1\} \text{ and any assignment } \mathbf{d}_{-i}, \tag{2.1}$$

where $Y_i(d_i, \mathbf{d}_{-i})$ is the potential outcome under individual i's treatment assignment $D_i = d_i$ and all others' assignments $\mathcal{D}_{-i} = \mathbf{d}_{-i}$.

  - Subject i's potential outcomes are only a function of own treatment, d_i.
  - Rules out spillover or interference effects from the treatment of others to subject i's outcome.

Assessing the Plausibility of the SUTVA

- The plausibility of the SUTVA needs to be assessed in the empirical context at hand.

Example

- Whether other labor market participants are trained may have an impact on one's own labor market outcomes.
    - If more individuals obtain training, supply of a certain skill in the labor market increases.
    - Companies can choose among a larger pool of trained individuals.
    - This may negatively affect the wage of an individual independent of own training participation.
- If number of trained individuals is small relative to total supply of that skill, the SUTVA might approximately hold.

Observed and Potential Outcomes

- Association between potential outcomes $Y(1)$ and $Y(0)$ and observed outcome Y:

$$Y = Y(1) \cdot D + Y(0) \cdot (1 - D) \text{ for } D \text{ being either 1 or 0} \tag{2.2}$$

- Equivalent equation:

$$Y = Y(0) + \underbrace{(Y(1) - Y(0))}_{\text{causal effect}} \cdot D \tag{2.3}$$

- It is impossible to observe both potential outcomes at the same time for any individual:
  - $Y = Y(1)$ for treated individuals ($D = 1$).
  - $Y = Y(0)$ for nontreated individuals ($D = 0$).
- As either $Y(1)$ or $Y(0)$ is not observed, individual causal effects are never directly identified.

Average Causal Effects

Specific assumptions permit evaluating aggregated causal effects:

ATE, ATET, and ATENT

Average treatment effect of D on Y (ATE):

$$\Delta = E[Y(1)] - E[Y(0)] \tag{2.4}$$

Average treatment effect on the treated (ATET):

$$\Delta_{D=1} = E[Y(1)|D=1] - E[Y(0)|D=1] \tag{2.5}$$

Average treatment effect on the nontreated (ATENT):

$$\Delta_{D=0} = E[Y(1)|D=0] - E[Y(0)|D=0] \tag{2.6}$$

- Effects refer to total population, or groups of treated (treatment group) and nontreated individuals (control group).

Treatment Selection Bias (1)

Can we identify average causal effects by simply comparing the outcomes of treated and nontreated individuals?

- In general, no: treatment and control groups may differ in background characteristics that affect the outcome.
- Differences in Y across groups may not only reflect the treatment effect, but also the effect of background characteristics.

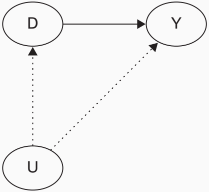
Example

- Assume that individuals participating in training are on average more motivated than those who do not.
- If motivation affects wages, comparing the average wages of participants and non-participants will not correspond to the causal effect of training.
- Instead, it will provide a mixture of the effects of training and motivation.

Treatment Selection Bias (2)

- Generally, cannot infer causality from patterns observed in data.
- We might observe a statistical association (correlation) between Y and D, but this may not correspond to the causal effect of D.

Figure 2.2: Treatment selection bias



- U: Characteristics (e.g., motivation) affecting both D and Y.
- Dotted arrows indicate that U is unobserved: its causal effects on D and Y cannot be assessed.

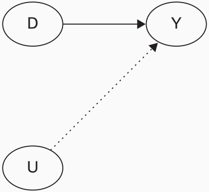
Ceteris Paribus Condition (1)

For a proper evaluation of the causal effect of D on Y, the ceteris paribus condition must be satisfied.

Ceteris paribus condition

The treated and control groups must be comparable in background characteristics U that affect Y ("everything else equal" apart from D).

Figure 2.3: No treatment selection bias



- Graphically, the ceteris paribus condition implies that if U affects Y, it must not affect D.

Ceteris Paribus Condition (2)

Example

- If motivation affects wages (Y), it must not influence training participation (D).
- Ceteris paribus condition implies that motivation does not systematically differ across treatment and control groups.
- Differences in Y across groups with $D = 1$ and $D = 0$ can be attributed to differences in the treatment alone.
- Therefore, the differences correspond to the causal effect of D.

Conditions for Identifying the ATE (1)

- Under which condition does the mean difference in the observed outcomes across groups correspond to the ATE?
- Formally: Under which condition is

$$\Delta = E[Y(1)] - E[Y(0)] = E[Y|D=1] - E[Y|D=0]\,?$$

- Because $Y = Y(1)$ for the treatment group, while $Y = Y(0)$ for the control group (see slide 7), we first note that:

$$E[Y|D=1] = E[Y(1)|D=1], \quad E[Y|D=0] = E[Y(0)|D=0]$$

- Therefore,

$$E[Y|D=1] - E[Y|D=0] = E[Y(1)|D=1] - E[Y(0)|D=0]$$

Conditions for Identifying the ATE (2)

$E[Y(1)|D=1] - E[Y(0)|D=0]$ equals $\Delta = E[Y(1)] - E[Y(0)]$ if the following conditions hold:

(i)	$E[Y(1)\|D=1] = E[Y(1)]$	The average of $Y(1)$ in the treatment group and in the total population are the same.
(ii)	$E[Y(0)\|D=0] = E[Y(0)]$	The average of $Y(0)$ in the control group and in the total population are the same.

- To satisfy these conditions, the treatment and control groups must be comparable with regard to U (at least on average).
- If these conditions do not hold, the potential outcomes of treated/controls are not representative of the total population.

Selection Bias

- From $E[Y|D=1] \neq E[Y(1)]$ and/or $E[Y|D=0] \neq E[Y(0)]$, it generally follows that:

$$
\begin{aligned}
& E[Y|D=1] - E[Y|D=0] \neq E[Y(1)] - E[Y(0)], \\
\text{such that } \quad & \underbrace{E[Y|D=1] - E[Y(1)] - E[Y|D=0] + E[Y(0)]}_{\text{selection bias when assessing } E[Y(1)]-E[Y(0)]} \neq 0 \qquad (2.8)
\end{aligned}
$$

- The mean difference in the observed outcomes across groups does not correspond to the ATE, but includes selection bias.
- Note that the size (or even the existence) of the treatment selection bias can in general not be verified in the data.