Causal Inference (Part 1)

Jae-kwang Kim [1]

[1]Some materials are taken from the lecture notes of professors Kosuke Imai and Marie Davidian

Causal Inference

- Interested in finding a causal relationship.
- We are interested in the effect of treatment over control on the outcome Y of interest.

Subject	Treatment (A)	Outcome (Y)
Clinical trial	New drug	Health outcome
Labor economics	Job training	Employment status
Politics	Canvassing	Vote turnout

- We assume that A is binary: $A = 1$ for treatment and $A = 0$ for control.
- Two potential outcomes for Y: $Y(0)$ for $A = 0$ and $Y(1)$ for $A = 1$
- The concept of potential outcomes was first introduced by Neyman (1923) in his PhD thesis. The particular chapter was translated to English and published later (Splawa-Neyman et al., 1990).

Randomized Experiment vs Observational Study

- Randomized experiment (e.g. clinical trial): the event for $A_i = 1$ is completely determined by a pure random mechanism.
- Observational study: Each unit i is assigned to $A_i = 0$ or $A_i = 1$ by other factors (such as physician's discretion).

Defining Causal Effects

- For each unit i, we observe (X_i, A_i, Y_i) where $Y_i = Y_i(A_i)$.
- We observe only one potential outcome for each unit. So, it is a missing data problem.
- In some literature, the unobserved potential outcome is called the counterfactual outcome. If $A_i = 1$, then we observe $Y_i(1)$ only and $Y_i(0)$ is the counterfactual outcome for unit i.
- Intuitively: if we could observe the counterfactual outcomes, then the difference $Y_i(1) - Y_i(0)$ is attributable to the treatments.
- Causal effect for unit i: $\tau_i = Y_i(1) - Y_i(0)$.

Average treatment effect

- Unit-level causal effects are difficult to estimate.
- We can average them over a sample of units.
  1. Sample average treatment effect:

$$\mathrm{SATE} = \frac{1}{n}\sum_{i=1}^{n}\{Y_i(1) - Y_i(0)\}$$

  2. Sample average treatment effect for the treated:

$$\mathrm{SATT} = \frac{1}{n_1}\sum_{i=1}^{n} A_i\{Y_i(1) - Y_i(0)\}$$

- Population average treatment effects:

$$\mathrm{PATE} = E(Y(1) - Y(0)) := \tau$$

$$\mathrm{PATT} = E(Y(1) \mid A = 1) - E(Y(0) \mid A = 1)$$

Conditional average treatment effect (CATE)

- The causal effect can be heterogeneous for different groups (age group, gender, race, etc).
- In this case, we may be interested in computing the conditional average treatment effect (CATE):

$$\tau(\mathbf{x}) = E\left(Y(1) - Y(0) \mid \mathbf{X} = \mathbf{x}\right)$$

- Applications to precision medicine and micro-targeting.

Formal causal problem

- Data: IID observed data

$$(X_i, A_i, Y_i), \quad i = 1, \ldots, n$$

- Goal: We wish to estimate the causal parameters (such as ATE) from the observed data.
- Problem: Under what conditions can we do this?
- Rubin (2005) wrote a nice review article on this topic.

Stable Unit Treatment Value Assumption (SUTVA)

- Definition of SUTVA:

$$Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i), \quad i = 1, \ldots, n \tag{1}$$

- First introduced by Rubin (1980).
- The outcome Y_i observed for individual i, who received treatment A_i, is the same as his potential outcome for that treatment regardless of the conditions under which he received that treatment.
- For example, the observed Y_i for treatment $A_i = 1$ in a randomized clinical trial is the same as the outcome he would have if the same treatment is obtained at the discretion of his physician.
- Implies no interference: Potential outcomes for an individual are unaffected by treatments received or potential outcomes of other individuals

Randomized studies

- Randomization ensures treatment assignment is independent of all other factors, including individual characteristics
- Thus, we have
$$\{Y(1), Y(0)\} \perp A \tag{2}$$
- **Main Result**: Under (1) and (2), we have
$$E(Y \mid A = a) = E(Y(a)) \tag{3}$$

for $a = 0, 1$.

Justification

Implication of (3)

- We can use
$$\hat{\tau} = \frac{1}{n_1}\sum_{i=1}^{n} A_i Y_i - \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i)Y_i \tag{4}$$
to estimate $\tau = E\{Y(1)\} - E\{Y(0)\}$, where $n_1 = \sum_{i=1}^{n} A_i$ and $n_0 = n_{n1}$. The estimator in (4) is called the difference-in-means estimator.
- Note that, by SUTVA,
$$\hat{\tau} = \frac{1}{n_1}\sum_{i=1}^{n} A_i Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}(1 - A_i)Y_i(0).$$
- If we treat $\mathcal{F}_n = \{Y_i(0), Y_i(1); i = 1, \ldots, n\}$ as a finite population, the randomization distribution due to treatment assignment is exactly equal to that of the simple random sampling (without replacement).

- Under the randomization distribution (treating $\mathcal{F}_n$ as fixed),

$$
\begin{aligned}
& E\left(\hat{\tau} \mid \mathcal{F}_n\right) \\
= \; & \frac{1}{n_1}\sum_{i=1}^{n} E(A_i \mid \mathcal{F}_n) Y_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\{1 - E(A_i \mid \mathcal{F}_n)\} Y_i(0) \\
= \; & \frac{1}{n}\sum_{i=1}^{n}\left(Y_i(1) - Y_i(0)\right) = \mathrm{SATE},
\end{aligned}
$$

where $E(A_i \mid \mathcal{F}_n) = n_1/n$.

- Randomness comes only from the treatment assignment.

The variance of the Difference-in-Means estimator

- Variance of $\hat{\tau}$: We can show that

$$V\left(\hat{\tau} \mid \mathcal{F}_n\right) = \frac{1}{n}\left(\frac{n_0}{n_1}S_1^2 + \frac{n_1}{n_0}S_0^2 + 2S_{01}\right) \tag{5}$$

  where, for $a = 0, 1$,

$$S_a^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i(a) - \overline{Y(a)}\right)^2$$

  and

$$S_{01} = \frac{1}{n-1}\sum_{i=1}^{n}\left(Y_i(0) - \overline{Y(0)}\right)\left(Y_i(1) - \overline{Y(1)}\right)$$

  with $\overline{Y(a)} = n^{-1}\sum_{i=1}^{n} Y_i(a)$.

- The variance is NOT identifiable.

Justification for (5)

- Since $\sum_{i=1}^{n} A_i = n_1$ and $E(A_i)$ are all equal, we have $E(A_i) = n_1/n$. Furthermore, we can show

$$E(A_i A_j) = \begin{cases} \frac{n_1(n_1-1)}{n(n-1)} & \text{if } i \neq j \\ n_1/n & \text{if } i = j \end{cases}$$

- Thus, writing $f_1 = n_1/n$, we obtain

$$Cov(A_i, A_j) = \begin{cases} -(n-1)^{-1} f_1(1-f_1) & \text{if } i \neq j \\ f_1(1-f_1) & \text{if } i = j \end{cases}$$

- Now,

$$
\begin{aligned}
& V\left\{n_1^{-1}\sum_{i=1}^{n} A_i Y_i(1) \mid \mathcal{F}_n\right\} \\
= \; & V\left\{n_1^{-1}\sum_{i=1}^{n} A_i (Y_i(1) - \overline{Y(1)}) \mid \mathcal{F}_n\right\} \\
= \; & n_1^{-2}\sum_{i=1}^{n} f_1(1-f_1)(Y_i(1) - \overline{Y(1)})^2 \\
- \; & n_1^{-2}(n-1)^{-1}\sum_{i=1}^{n}\sum_{j\neq i} f_1(1-f_1)(Y_i(1) - \overline{Y(1)})(Y_j(1) - \overline{Y(1)}) \\
= \; & n(n-1)^{-1} n_1^{-2}\sum_{i=1}^{n} f_1(1-f_1)(Y_i(1) - \overline{Y(1)})^2 \\
= \; & n^{-1}(n_0/n_1)S_1^2
\end{aligned}
$$

- Similarly, we can show

$$V\left\{n_0^{-1}\sum_{i=1}^{n}(1-A_i)Y_i(0)\mid \mathcal{F}_n\right\} \;=\; n^{-1}(n_1/n_0)S_0^2$$

- Now, wring $Z_i(a) = Y_i(a) - \overline{Y(a)}$

$$\begin{aligned}
& Cov\left\{n_1^{-1}\sum_{i=1}^{n}A_iY_i(1), n_0^{-1}\sum_{i=1}^{n}(1-A_i)Y_i(0)\mid \mathcal{F}_n\right\} \\
= & \; Cov\left\{n_1^{-1}\sum_{i=1}^{n}A_iZ_i(1)n_0^{-1}\sum_{i=1}^{n}(1-A_i)Z_i(0)\mid \mathcal{F}_n\right\} \\
= & \; n_1^{-1}n_0^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}Cov(A_i, 1-A_j)Z_i(1)Z_j(0) \\
= & \; -n^{-1}S_{01}
\end{aligned}$$

Bounds on the variance: Neyman (1923)'s idea

- Cauchy-Schwartz inequality:

$$-S_1 S_0 \le S_{01} \le S_1 S_0$$

- Thus, we obtain

$$V(\hat{\tau} \mid \mathcal{F}_n) \le \frac{n_0 n_1}{n}\left(\frac{S_1}{n_1} + \frac{S_0}{n_0}\right)^2 \le \frac{S_1^2}{n_1} + \frac{S_0^2}{n_0} = E\left(\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0} \mid \mathcal{F}_n\right), \tag{6}$$

where

$$\hat{\sigma}_a^2 = \frac{1}{n_a - 1}\sum_{i=1}^{n} \mathbb{I}(A_i = a)\left(Y_i - \overline{Y(a)}\right)^2, \quad a = 0, 1.$$

- The usual variance estimator is conservative on average.

Justification for (6)

Inference for population average treatment effect

- Assumption: the potential outcomes are IID from a superpopulation model

$$\begin{pmatrix} Y_i(0) \\ Y_i(1) \end{pmatrix} \sim \left[\begin{pmatrix} \mu_0 \\ \mu_1 \end{pmatrix}, \begin{pmatrix} \sigma_0^2 & \sigma_{01} \\ \sigma_{01} & \sigma_1^2 \end{pmatrix} \right] \tag{7}$$

- Two-phase sampling structure:
  1. Phase One: $(Y_i(0), Y_i(1))$ are generated from (7).
  2. Phase Two: Select the treatment group by SRS of size n_1. The others are in the control group.
- We are interested in $\tau = \mu_1 - \mu_0$, population average treatment effect.

Statistical Properties of $\hat{\tau}$ in (4)

- Unbiasedness of $\hat{\tau}$

$$E(\hat{\tau}) = \mu_1 - \mu_0$$

- Variance

$$V(\hat{\tau}) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_0^2}{n_0}$$

- Unbiased variance estimator is

$$\widehat{V}(\hat{\tau}) = \frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_0^2}{n_0}$$

- The CLT can be established.

Justification

REFERENCES

Rubin, D. B. (1980), 'Randomization analysis of experimental data: The fisher randomization test comment', *Journal of the American Statistical Association* **75**, 591–593.

Rubin, D. B. (2005), 'Causal inference using potential outcomes: Design, modeling, decisions', *Journal of the American Statistical Association* **100**, 322–331.

Splawa-Neyman, Jerzy, D. M. Dabrowska and T. P. Speed (1990), 'On the application of probability theory to agricultural experiments. essay on principles. section 9.', *Statistical Science* **5**, 465–472.