Causal Inference: What If

Chapter 1 to 3

Kunwoong Kim

Seoul National University

July 22, 2021

Table of Contents

Introduction

- We want to answer the following questions.
  - Does cigarette smoking causes lung cancer?
  - Does the obesity increases mortality?
- Measures of causal effect
- From *randomized experiments* to *observational studies*

Table of Contents

1. A definition of causal effect

Let $Y \in \{0, 1\}$ (e.g., lung cancer) as an outcome of an input X. Denote Y^a as the outcome under the action $A = a \in \{0, 1\}$ (e.g., smoking).

Definition (Counterfactual outcome)

The variables $Y^{a=1}$ and $Y^{a=0}$ are called as *counterfactual outcomes.*

Definition (Causal effect for an individual)

The treatment A has a *causal effect* on an individual's outcome Y if $Y^{a=1} \neq Y^{a=0}$ for the individual.

Definition (Consistency)

If $A = a$, then $Y^a = Y^A = Y$.

- Individual causal effects cannot be identified: we have missing data. We cannot observe the counterfactual world.

1. A definition of causal effect

Thus we provide another definition of causal effect: *average causal effect*. We call an *average causal effect* of treatment A on an outcome Y is present if

$$\mathsf{Pr}(Y^{a=1}=1) \neq \mathsf{Pr}(Y^{a=0}=1)$$

or equivalently,

$$\mathsf{E}(Y^{a=1}) \neq \mathsf{E}(Y^{a=0}).$$

1. A definition of causal effect

We compute the average causal effects by the following three measures.

1. Causal risk difference

$$\mathsf{Pr}(Y^{a=1} = 1) - \mathsf{Pr}(Y^{a=0} = 1) = 0$$

2. Causal risk ratio

$$\mathsf{Pr}(Y^{a=1} = 1)/\mathsf{Pr}(Y^{a=0} = 1) = 1$$

3. Causal odds ratio

$$\frac{\mathsf{Pr}(Y^{a=1} = 1)/\mathsf{Pr}(Y^{a=1} = 0)}{\mathsf{Pr}(Y^{a=0} = 1)/\mathsf{Pr}(Y^{a=0} = 0)} = 1$$

1. A definition of causal effect

We say that treatment A and outcome Y are dependent (associated) if $\Pr(Y = 1|A = 1) - \Pr(Y = 1|A = 0) \neq 0$.

1. Associational risk difference

$$\Pr(Y = 1|A = 1) - \Pr(Y = 1|A = 0) = 0$$

2. Associational risk ratio

$$\Pr(Y = 1|A = 1)/\Pr(Y = 1|A = 0) = 1$$

3. Associational odds ratio

$$\frac{\Pr(Y = 1|A = 1)/\Pr(Y = 0|A = 1)}{\Pr(Y = 1|A = 0)/\Pr(Y = 0|A = 0)} = 1$$

1. A definition of causal effect

Association is not causation.

- Two disjoint subsets determined by actual treatment vs. Population under two different treatment values

$$\Pr(Y^a) \neq \Pr(Y|A = a)$$

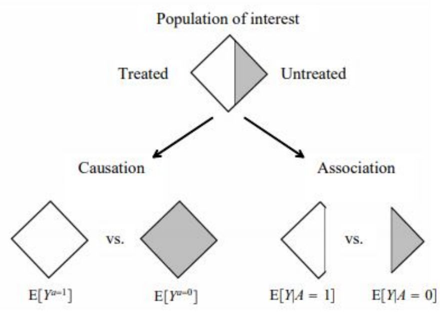

Table 2.1

	A	Y	Y^0	Y^1
Rheia	0	0	0	?
Kronos	0	1	1	?
Demeter	0	0	0	?
Hades	0	0	0	?
Hestia	1	0	?	0
Poseidon	1	0	?	0
Hera	1	0	?	0
Zeus	1	1	?	1
Artemis	0	1	1	?
Apollo	0	1	1	?
Leto	0	0	0	?
Ares	1	1	?	1
Athena	1	1	?	1
Hephaestus	1	1	?	1
Aphrodite	1	1	?	1
Cyclope	1	1	?	1
Persephone	1	1	?	1
Hermes	1	0	?	0
Hebe	1	0	?	0
Dionysus	1	0	?	0

Figure: 1.1

Table of Contents

2. Randomized experiments

- Treat an input X as $A = 1$ with a fair coin!
- Randomized experiments generate data with missing values of counterfactual outcomes.
- Then, association is causation.

$$\mathsf{E}(Y^a) = \mathsf{E}(Y^a | A = a) = \mathsf{E}(Y | A = a)$$

since $Y^a \perp A$ and $Y^a = Y$.

2. Randomized experiments

- What about the case when we do not treat individuals randomly but conditionally random?

e.g. $A = 1$ if X received a transplant, $Y = 1$ if X died, and $L = 1$ if X was in a critical condition (measured before treatment was assigned). Assume that doctors treated individuals with $A = 1$ with probability 0.75 if $L = 1$ (with prob 0.5 otherwise).

- The treatment A and the critical condition L are dependent.
- How to compute causal effects in this situation?

2. Randomized experiments

The standardization technique helps us to compute the causal risk ratio

$$\frac{\text{Pr}(Y^{a=1}=1)}{\text{Pr}(Y^{a=0}=1)} = \frac{\sum_l \text{Pr}(Y^{a=1}=1|L=l)\text{Pr}(L=l)}{\sum_l \text{Pr}(Y^{a=0}=1|L=l)\text{Pr}(L=l)}$$

$$= \frac{\sum_l \text{Pr}(Y=1|L=l,A=1)\text{Pr}(L=l)}{\sum_l \text{Pr}(Y=1|L=l,A=0)\text{Pr}(L=l)}$$

since $\text{Pr}(Y^a=1|L=l) = \text{Pr}(Y=1|L=l,A=a)$ for all l by the conditional exchangeability.

- That is, we can compute the causal risk ratio in a conditionally randomized experiment via standardization.

2. Randomized experiments

- Inverse probability (IP) weighting is an equivalent to the standardization technique.
- It holds by the conditional exchangeability, that is, we create pseudo-population.

2. Randomized experiments

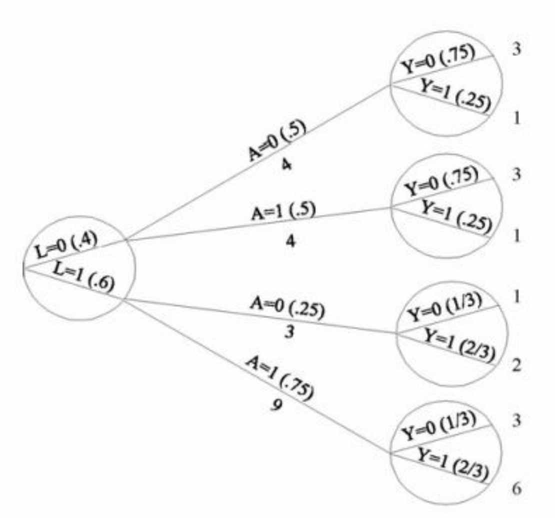

Figure: 2.1

2. Randomized experiments

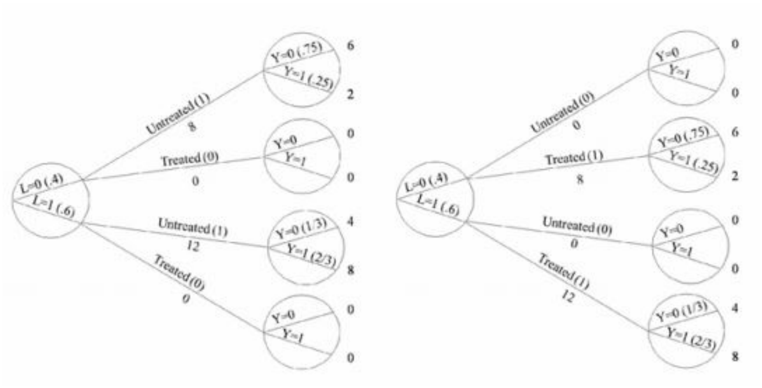

Figure: 2.2

2. Randomized experiments

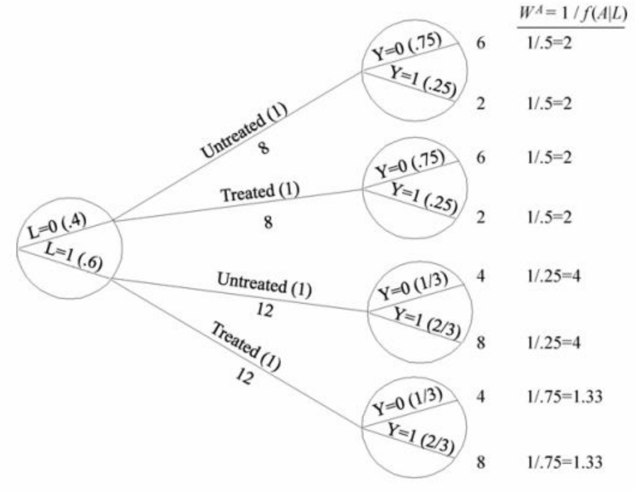

Figure: 2.3

2. Randomized experiments

- Then, we can always compute the causal risks through the two calculation techniques **if** we can conduct *(conditionally) randomized experiments*.

Q. **Can we always conduct randomized experiments?** What about the case when A is the heart transplant treatment and Y indicates death? Doctors assign individuals who are more likely to benefit from the transplant, rather than assigning randomly.

- However, *randomized experiments* can be impractical in many cases.
- Thus we conduct an *observational study* as the least bad option.

Table of Contents

3. Observational studies

- Investigators observe and record.
- From the observed data, how can we compute causal effects?
- We link observational study to conditionally randomized experiment.

What we need are:

1. Exchangeability
2. Positivity
3. Consistency

If the above three conditions hold (actually, we assume.), then we can compute causal effects using observed data.

3. Observational studies

1. Exchangeability
   - We “assume” the exchangeability.
   - L should be the only variable that is unequally distributed between the treated and the untreated.

3. Observational studies

e.g., heart transplants:
(Case 1) Doctors assign to individuals with low probability of rejecting the transplant (i.e., possessing HLA genes). HLA is not a predictor of Y. Thus the heart transplanting is random within levels of L.
(Case 2) Doctors prefer to transplant hearts into nonsmokers ($U = 0$), which is not known to the investigators. Then, X with $U = 1$ has a lower probability of receiving $A = 1$. But the doctors should have randomly treating individuals independent to U.

- The investigator should use their expert knowledge to measure sufficiently many Ls, and we should trust the experts' knowledge.

3. Observational studies

2. Positivity
   - What if doctors always transplant a heart to individuals in critical condition $L = 1$? Then, $Pr(A = 0|L = 1) = 0$.
   - One cannot compute the causal effects through the standardization or IP weighting.
   - We assume the following condition to avoid it.

Positivity:

$$Pr(A = a|L = l) > 0$$

for all l with $Pr(L = l) \neq 0$.

3. Observational studies

3. Consistency
   - We should avoid defining ill-defined counterfactual outcomes.

e.g., Ill-defined counterfactual outcome Y^a
The causal effect of obesity A at age 40 on the risk of mortality Y by age 50. X was not obese at 40 but would have died by age 50 because of an accident.
We should define A more precisely, then probabilities of miscommunications reduce which leads to ill-defined counterfactuals.

3. Observational studies

Summary: how can we use observational data in computing causal effects?

- The study should satisfy three conditions (1), (2) and (3).

Note: We can replace (1) and (2) by other conditions (Chapter 16) and extrapolations via modeling (Chapter 14), respectively. (3) should be satisfied.