Causal Inference

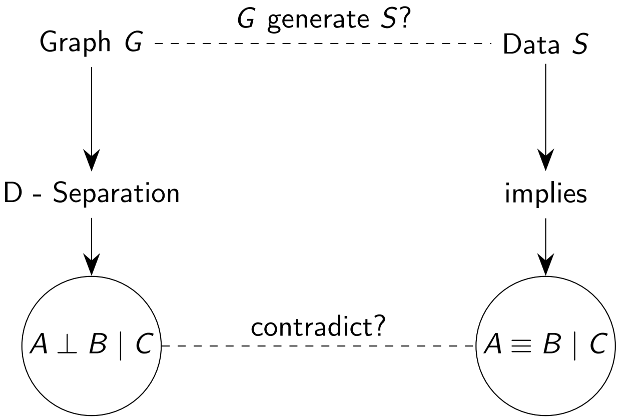
Ritsumeikan Beamer Theme

Qiunuo Chen

Fuzhou University

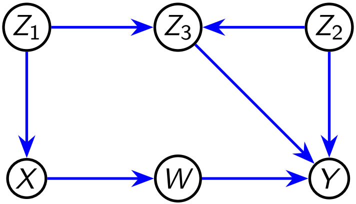
March, 2025

Causal Inference Problem



先假设一个因果关系，然后对这个假设的因果模型进行统计学上的数据的分析，可以来判断本身有因果关系的两个变量之间是否真的有因果关系。

Causal Graph Example



Causal Inference Example I

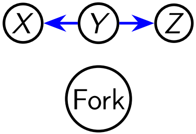
- Given Graph G implies $W \perp Z_1 \mid X$
- For Data S, we perform regression:

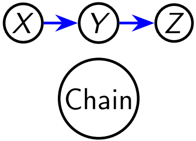
$$W = \beta_0 + \beta_{Z_1} Z_1 + \beta_X X + \epsilon$$

- If the result shows that $\beta_{Z_1} \neq 0$, then

$W \not\perp Z_1 \mid X$ (statistical correlation implies statistical dependence)

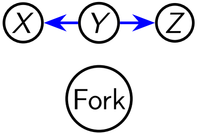
Dependence in Chain and Fork Structures

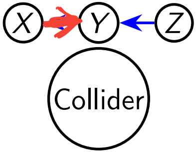




Dependence	Chain	Fork
$X \equiv Y$	✓	✓
$X \equiv Z$	✓	✓
$Y \equiv Z$	✓	✓
$X \equiv Y \mid Z$	✓	✓
$Y \equiv Z \mid X$	✓	✓
$X \perp Z \mid Y$	✓	✓

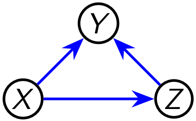
Dependence in Chain and Collider Structures



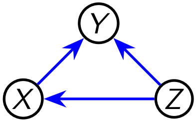


Dependence	Fork	Collider
$X \equiv Y$	✓	✓
$X \equiv Z$	✓	NO
$Y \equiv Z$	✓	✓
$X \equiv Y \mid Z$	✓	✓
$Y \equiv Z \mid X$	✓	✓
$X \perp Z \mid Y$	✓	NO

Dependence in Graph Structures



G1



G2

3

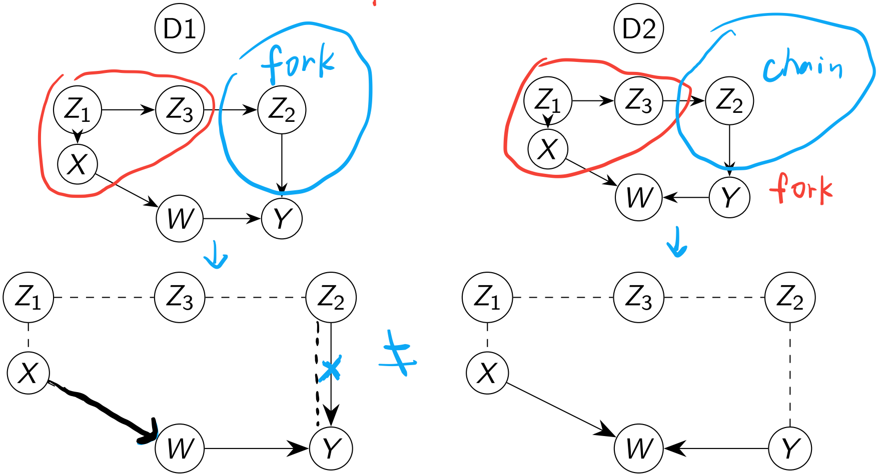
Dependence	G1	G2
$X \equiv Y$	✓	✓
$X \equiv Z$	✓	✓
$Y \equiv Z$	✓	✓
$X \equiv Y \mid Z$	✓	✓
$Y \equiv Z \mid X$	✓	✓
$X \equiv Z \mid Y$	✓	✓

Summary

1. Chain and Fork are indistinguishable.
2. Fork/chain and Collider are distinguishable.
3. Colliders with adjacent parents are indistinguishable.

Example Diagrams

去掉等价类



For any SCM model whose corresponding graph is not acyclic, the joint distribution of the variable in the model is given by

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \text{pa}(X_i))$$

联合分布

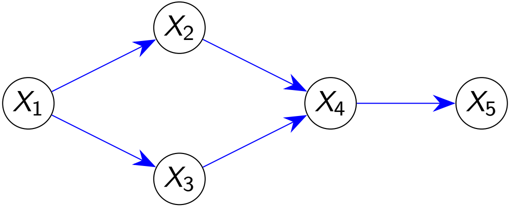
where $\text{pa}(X_i)$ stands for the values of parents of variable X_i.

$\Downarrow G$

化简

父结点

Example



$$P(X_1, X_2, X_3, X_4, X_5) = \prod_{i=1}^{5} P(X_i \mid \mathrm{Pa}(X_i))$$

$$= P(X_1)P(X_2 \mid X_1)P(X_3 \mid X_1)P(X_4 \mid X_2, X_3)P(X_5 \mid X_4)$$

Example proof

$$P(X_1,X_2,X_3,X_4,X_5) = \overbrace{P(X_5 \mid X_4,X_3,X_2,X_1)}^{A'}\overbrace{P(X_4 \mid X_3,X_2,X_1)}^{B'}$$

$$\overbrace{P(X_3 \mid X_2,X_1)}^{C'}P(X_2 \mid X_1)P(X_1)$$

$$= P(X_1)P(X_2 \mid X_1)\overbrace{P(X_3 \mid X_1)}^{C}\overbrace{P(X_4 \mid X_2,X_3)}^{B}$$

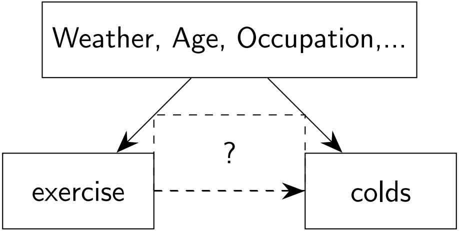
$$\overbrace{P(X_5 \mid X_4)}^{A}$$

D-sep

根据condition法则可以将条件部分进行精简。

$$where A' = A, B' = B, C' = C.$$

Relationship Exploration



Treatment and Outcome Variables

Definition

Treatment variable is the one that may cause the difference in the outcome variable.

- △ Do taller people make more money?
  - Treatment: Height
  - Outcome: Income
- △ Do magnets help relieve pain?
  - Treatment: Magnets
  - Outcome: pain

Confounding Variable (Confounder)

Definition

Confounding variable (confounder)

- △ has an effect on the outcome variable
- △ has an effect on the treatment variable

Example

Does exercise prevent cold?

- △ For a sample of subjects, record the amount of exercise per week, and the number of colds over a year. Suppose you find that exercise is correlated with fewer colds.

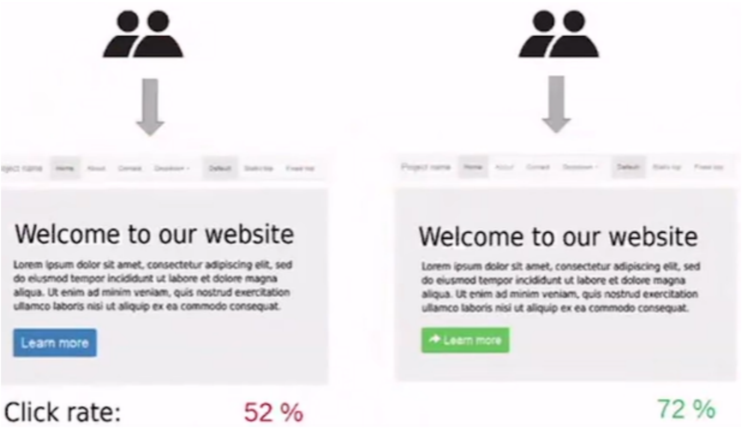
Observational vs Experimental

- Observational Study
  - △ data are observed and collected on each subject
  - △ No manipulation of the subjects' environment occurs.
- Experiment
  - △ Manipulate the subject's environment, then
  - △ measure the outcome

Example: Does exercise prevent cold?

- Observational study
  - △ Randomly select a sample of subjects
  - △ Record data for each subject on amount of exercise and number of colds last year.
- Randomised Experiment
  - △ Obtain a group of study participants (often volunteers)
  - △ Intervention: randomly assign the participants to the treatment (exercise) and control (no exercise)
  - △ After a set amount of time, record amount of exercise and the number of colds for each person.

A/B test



Randomly!!!

Observational Study VS. Experiment

Important

An observational study may reveal correlation between two variables, but only a randomized experiment can directly prove cause - and - effect.

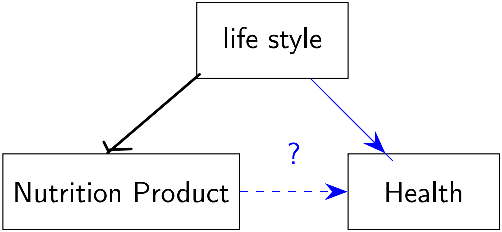
Why???

- Confounding may be present in observational study.
- Randomized assignment to treatment and control groups in an experiment makes all other factors that influence the outcome vary at random, so any change in the outcome is attributable to the treatment.

Observational Study

	(L) Life Style	(N) Nutrition Product	(H) Health
1	Good (1)	More (1)	Good (1)
2	Good	More	Good
3	Good	More	Good
4	Good	More	Good
5	bad (0)	less (0)	bad (0)
6	bad	less	bad
7	bad	less	bad
8	bad	less	bad

Causal Inference Example



Observational Study

	(L) Life Style	(N) Nutrition Product	(H) Health
1	Good (1)	More (1)	(Good 1)
2	Good	More	Good
3	Good	More	Good
4	Good	More	Good
5	bad (0)	More	bad (0)
6	bad	More	bad
7	bad	less	bad
8	bad	less	bad

manipulation $---> P_m(H=1 \mid N=1) = 2/4 = 0.5$

$$P_m(H=1 \mid N=0) = 2/4 = 0.5$$

Placebo

- fake treatment, sugar pill
- Help control for the placebo effect.
  - people who believe they are getting a treatment often get better even if the treatment has no active ingredient.

- Study shows placebo helps 62% of headache sufferers,58% of those with sea-sickness

Control Group and Randomization

- Control group
  - subjects in this group do not receive the treatment but may receive a placebo
  - So we can tell what happens to the outcome without the treatment (baseline)
- Randomization
  - random assignment to treatment and control groups
  - Helps to equalize group with respect to confounders
- Placebo
  - fake treatment,suger pill
  - Helps control for the placebo effect.people who believe they are getting a treatemnt often get better even if th treatment has no active ingradient.

Good Experiments

- Placebo and blinding are not always possible.
  - For example, brain surgery, exercise.
- If observational studies can't prove cause and effect, why don't researchers always do randomized experiments?
  - E.g. left - handedness / right - handedness and mathematical aptitude, height and income.

Example: Pets and Happiness

Study Description

A study is conducted to investigate the relationship between owning pets and happiness. 100 subjects are randomly selected and data on whether or not a pet is owned and a happiness score (1 - 10, 10 being extremely happy) are obtained.

- Treatment and outcome?
- Observational or experimental?
- What possible confounders exist?

Example: Pets and Happiness

Study Description

A study is conducted to investigate the relationship between owning pets and happiness. 100 subjects are randomly selected and data on whether or not a pet is owned and a happiness score (1 - 10, 10 being extremely happy) are obtained.

- Treatment and outcome?(pets;happiness)
- Observational or experimental?(obs)
- What possible confounders exist?(work;income)

Example: Video Games and Aggression

Study Description

210 college students were randomly assigned to play either a violent or nonviolent video game . A short time later, the students who played the violent video game punished an opponent (received a noise blast with varying intensity) for a longer period of time than did students who had played the nonviolent video game.

- Treatment and outcome?
- Observational or experimental?
- Can cause and effect be established?

Example: Video Games and Aggression

Study Description

210 college students were randomly assigned to play either a violent or nonviolent video game. A short time later, the students who played the violent video game punished an opponent (received a noise blast with varying intensity) for a longer period of time than did students who had played the nonviolent video game.

- Treatment and outcome? （game； Aggression）
- Observational or experimental? （ex）
- Can cause and effect be established? （yes）

Pearl's Causal Hierarchy (PCH)

Layer (Symbolic)	Typical Activity	Typical Question
$\mathcal{L}_1$ Associational $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?
$\mathcal{L}_2$ Interventional $P(y\|do(x), c)$	Doing	What if? What if I do X?
$\mathcal{L}_3$ Counterfactual $P(y_x\|x', y')$	Imagining	Why? What if I had acted differently?

Imaginary Intervention

Introduction

Although randomized controlled experiment is considered the golden standard of causal inference, in practice, sometimes it is not possible.

- Weather — Wild fire
- Violent Tv watched by kids — Kid' s behavior
- Smoking — pregnant women' s health

Solution

In these situations, observational data has to be used. For observational data, we introduce "imaginary" intervention.

SCM - Structural Causal Model

$$U = \{u_X, u_Y, u_Z\}$$
$$F = \{f_X, f_Y, f_Z\}$$
$$V = \{Z, X, Y\}$$

where Z : temperature , X : ice cream sales, Y : drowning

Original

$$f_Z : Z = u_Z$$
$$f_X : X = 4Z + u_X = \mathbf{c}$$
$$f_Y : Y = \frac{Z}{10} + u_Y$$

After intervention

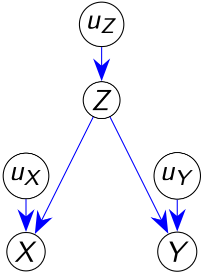
$$f_Z : Z = u_Z$$
$$f_X : \mathbf{do(X = c)}$$
$$f_Y : Y = \frac{Z}{10} + u_Y$$

- We basically remove $X = \frac{Z}{10} + u_X$ and enforce $X = c$, we denote it by $\text{do}(X = c)$.

Intervention in Causal Models

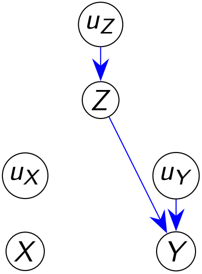
- e.g. If we want to make ice cream sales low we shut down all ice cream shops.
- If we do($X = c$), in graph we remove all edges directed into that variable.

Original



obs($X = c$)

After intervention



do($X = c$)

- After do($X = c$), X no longer associated with Z, thus X becomes independent of Y.

$P(Y=y \mid X=c) \neq P(Y=y \mid do(X=c))$

obs

set

$P(Y=y \mid X=c) \neq P(Y=y)$

$P(Y=y \mid do(X=c)) \triangleq P_m(Y=y \mid X=c) = P_m(Y=y) = P(Y=y)$

notation 上的定义

Y只与U_y有关

- $P(\cdot)$ is based on the model **before** intervention
- $P_m(\cdot)$ is based on the model **after** intervention(x,y独立）

Difference between $P(Y = y \mid X = x)$ and $P(Y = y \mid do(X = x))$

- $P(Y = y \mid X = x)$: distribution of the sub - population of Y among individuals whose $X = x$.
- $P(Y = y \mid do(X = x))$: distribution of the population of Y if everyone in the population had their X value set to x.

Intervening on a variable $do(X = c)$

We change the system, we set its value,

Intervening vs Conditioning on a Variable

Intervening on a variable $\mathrm{do}(X = c)$

We change the system, we set its value, and the value of other variables often change as a result.

- e.g. We shut down ice cream shops to make sales low, regardless the number of people on beach.

Conditioning on a variable observe $X = c$

We change nothing, only narrow our focus to a subset of cases when the variable takes the value we are interested in.

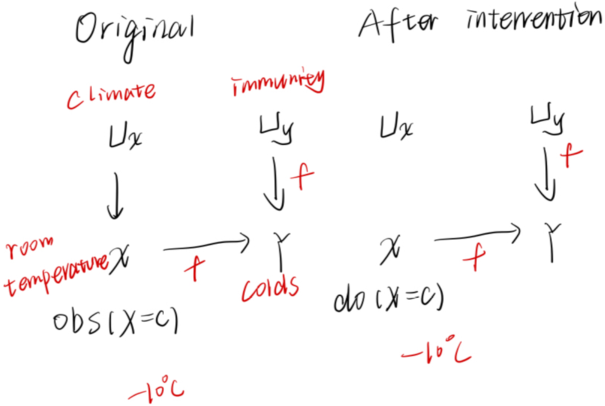
- e.g. We passively observed that sales is low, when number of people on beach is low.

Special Cases in Causal Inference

Equality Condition

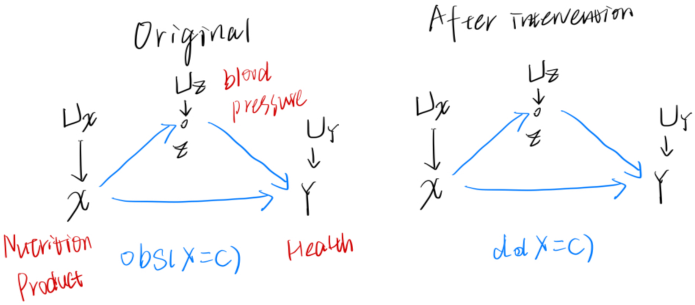
$$P(Y = y \mid X = c) = P(Y = y \mid \text{do}(X = c))$$

Example 1



$$P(Y = y \mid X = c) = P(Y = y \mid \text{do}(X = c))$$

Example 2 in Causal Inference



$$\underbrace{P(Y=y \mid X=c)}_{①} = \underbrace{P(Y=y \mid do(X=c))}_{②} = P_m(Y=y \mid X=c)$$

- ① = ② because the generation process $Y = f(X, u_Z, u_Y)$

Average Causal Effect (ACE)

Definition

$$P(Y = 1 \mid \mathrm{do}(X = 1)) - P(Y = 1 \mid \mathrm{do}(X = 0))$$
$$E(Y \mid \mathrm{do}(X = 1)) - E(Y \mid \mathrm{do}(X = 0))$$

Note

Note that this is different to

$$P(Y = 1 \mid X = 1) - P(Y = 1 \mid X = 0)$$

Adjustment Formula

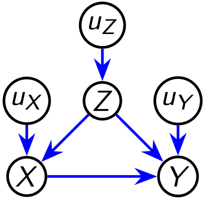
Variable Definitions

X: drug use, Y: recovery, Z: gender

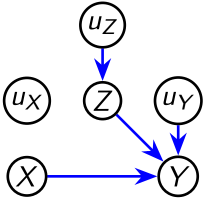
Functional Relationships

$$
\begin{aligned}
f_X &: X = 3Z + u_X \\
f_Y &: Y = 5X + 4Z + u_Y = f_Y(X, Z, u_Y) \\
f_Z &: Z = \frac{u_Z}{10}
\end{aligned}
$$

Original



After intervention



(1)

Distinction Note

Note that this is different to

$$P(Y = 1 \mid X = 1) - P(Y = 1 \mid X = 0)$$

Probability Equalities

$$\begin{aligned} P(Z = z) = P(Z = z \mid \mathrm{do}(X = x)) = \mathbb{E}_m(Z) \\ \triangleq P_m(Z = z \mid X = x) = P_m(Z = z) \end{aligned}$$

Distribution Representations

(2)

Probability Equality

$$
\begin{aligned}
P(Y = y \mid do(X = x), Z = z) &= P_m(Y = y \mid X = x, Z = z) \\
&= P(Y = y \mid X = x, Z = z)
\end{aligned}
$$

Reason

Because the generating process of Y remains the same.

$$
\begin{aligned}
Y &= f(X, Z, u_Y) \\
&= 5X + 4Z + u_Y
\end{aligned}
$$

Derivation of Causal Probability

$$P(Y = y \mid \mathrm{do}(X = x)) \triangleq P_m(Y = y \mid X = x)$$

by Law of Total Probability (LOTP)(全概率公式)

$$= \sum_z P_m(Y = y, Z = z \mid X = x)$$

$$= \sum_z P_m(Y = y \mid X = x, Z = z) \cdot P_m(Z = z \mid X$$

using ①

$$= \sum_z P_m(Y = y \mid X = x, Z = z) \cdot P_m(Z = z)$$

using ① and ②

$$= \sum_z P(Y = y \mid X = x, Z = z) \cdot P(Z = z)$$

Adjustment Formula or G - computation

Adjustment Formula

"adjust / control for Z"

$$P(Y = y \mid \text{do}(X = x)) = \sum_z P(Y = y \mid X = x, Z = z) P(Z = z)$$

$$\mathbb{E}(Y \mid \text{do}(X = x)) = \mathbb{E}_Z[\mathbb{E}_{Y|X,Z}(Y \mid X = x, Z)]$$

[1] J. Pearl and D. Mackenzie, *The Book of Why: The New Science of Cause and Effect*.
Basic Books, May 2018.

[2] J. Pearl, M. Glymour, and N. P. Jewell, *Causal Inference in Statistics: A Primer*.
John Wiley & Sons, reissue ed., 2016.

Thank You