Section: Module 7

Pertial Identification / DAG

Kentaro Nakamura

GOV 2002

November 7th, 2025

Logistics

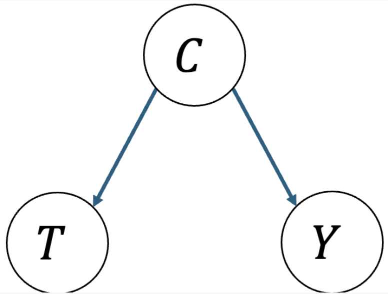
- Problem Set 7 is due on Next Monday
    - From PS7, we release problem set every monday and due on next monday
    - You need to submit 3 out of 5 problem set
    - If you submit revenge midterm, it is counted as 1 submission

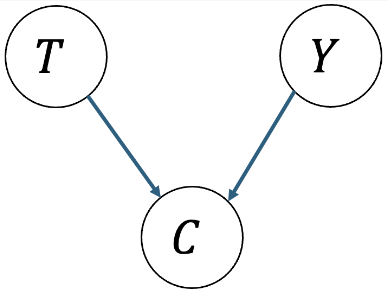
- Today's agenda
    - Directed Acyclic Graph / Structural Causal Model
    - Partial Identification (Problem Set 7)

Directed Acyclic Graph

- DAG: Directed Acyclic Graph
  - Tool to know which variable you need to condition on to achieve conditional ignorability
  - It is also an friendly and useful tool to understand the causal inference
    - Let's use DAG to explore some important results for applied research!

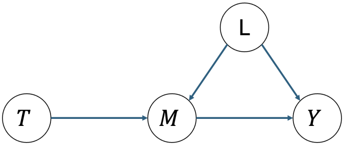
- **Rule of DAG**
  - Confounder
  - Collider





Warmup Question 1: Mediator

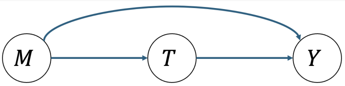
- **Question**: Can we condition on M_i?



- **Implication**: You are not allowed to condition on post-treatment variable
  - This tells us the causal mechanism only under stringent assumption (Blackwell et al. WP)
  - We will learn how to deal with this in Module 9

Warmup Question 2: Moderator

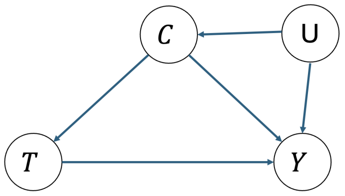
- **Question**: Can we condition on M_i?



- **Implication**: You can condition on pre-treatment variable to explore heterogeneity
  - **CATE**: $\mathbb{E}[Y_i(1) - Y_i(0) \mid M_i]$
  - You cannot condition on post-treatment variable for this purpose

Problem 1: Table 2 Fallacy

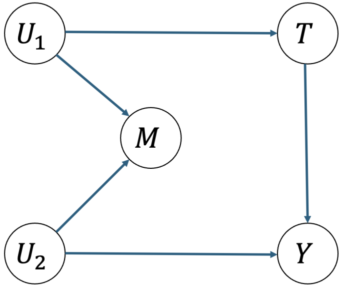
- **Question**: Suppose you are estimating the effect of T_i on Y_i by controlling C_i by regression. You know the sign of C_i in regresssion. Can you interpret the coefficient of C_i?



- **Implication**: You should not interpret the coefficient of control variable!
  - In order to interpret C_i causally, you need to control the confounder of C_i.
  - This mistake is often called as **Table 2 Fallacy** (Westreich and Greenland 2013)

Problem 2: M-Bias

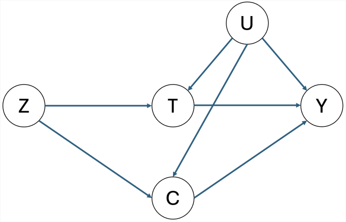
- **Question**: Suppose U is unobserved. Can we achieve ignorability by conditioning observed M_i?



- **Implication**: Controlling the variable that time-wise precedes the treatment can induce the bias
  - Without drawing DAG, you can suffer from collider bias even though you only control pre-treatment variable
  - This is known as **M-Bias**

Problem 3: Post-Instrument Variable

- **Question**: Consider the IV scenario. We know instrument Z_i influences outcome Y_i through C_i. Can we condition on C_i to achieve exclusion restriction?



- **Implication**: You cannot condition on the variable that is influenced by instrument to achieve exclusion restriction
  - Controlling C_i makes Z_i no longer exogeneous
  - This is known as **Post-Instrument Bias** (Schuessler et al. 2025)

Partial Identification

- **Partial Identification**: How much can we know with the minimal amount of assumptions we are willing to make?
  - The more we make assumptions, the more we gains
  - The credibility of inference decreases with the strength of the assumptions maintained

- **Tips**
  - Write down all the constraints (sum of probability, probability is non-negative, etc)
  - Combine every constraints to create the bound

Partial Identification: IV and ATE (1)

- Let's review the case of instrumental variable in the lecture.
  - Setup: Binary instrument Z_i, treatment T_i, outcome Y_i

$$
\begin{aligned}
&\mathbb{P}(Y_i(T_i = 1) = 1) \\
&= \mathbb{P}(Y_i(T_i = 1) = 1 \mid Z_i = 1) \quad (\because \text{Randomization}) \\
&= \mathbb{P}(Y_i(T_i = 1) = 1 \mid T_i = 1, Z_i = 1) \times \mathbb{P}(T_i = 1 \mid Z_i = z) \\
&\quad + \mathbb{P}(Y_i(T_i = 1) = 1 \mid T_i = 0, Z_i = 1) \times \mathbb{P}(T_i = 0 \mid Z_i = z)
\end{aligned}
$$

- By exclusion restriction and consistency

$$
\mathbb{P}(Y_i(T_i = 1) = 1 \mid T_i = 1, Z_i = 1) = \mathbb{P}(Y_i = 1 \mid T_i = 1, Z_i = 1)
$$

- The similar derivation holds for $T_i = 0$; i.e.,

$$
\begin{aligned}
&\mathbb{P}(Y_i(T_i = 0) = 1) \\
&= \underbrace{\mathbb{P}(Y_i(T_i = 0) = 1 \mid T_i = 1, Z_i = 0)}_{\text{Not identifiable}} \times \mathbb{P}(T_i = 1 \mid Z_i = 0) \\
&\quad + \mathbb{P}(Y_i = 1 \mid T_i = 0, Z_i = 0) \times (\underbrace{1 - \mathbb{P}(T_i = 1 \mid Z_i = 0)}_{= \mathbb{P}(T_i = 0 \mid Z_i = 0)})
\end{aligned}
$$

Partial Identification: IV and ATE (2)

- As a result, ATE is obtained as

$$\begin{aligned}
\tau &= \mathbb{P}(Y_i(T_i = 1) = 1) - \mathbb{P}(Y_i(T_i = 0) = 1) \\
&= \mathbb{P}(Y_i = 1 \mid T_i = 1, Z_i = 1) \times \mathbb{P}(T_i = 1 \mid Z_i = z) \\
&+ \underbrace{\mathbb{P}(Y_i(T_i = 1) = 1 \mid T_i = 0, Z_i = 1)}_{\text{Not Identifiable}} \times (1 - \mathbb{P}(T_i = 1 \mid Z_i = z)) \\
&- \underbrace{\mathbb{P}(Y_i(T_i = 0) = 1 \mid T_i = 1, Z_i = 0)}_{\text{Not identifiable}} \times \mathbb{P}(T_i = 1 \mid Z_i = 0) \\
&- \mathbb{P}(Y_i = 1 \mid T_i = 0, Z_i = 0) \times (1 - \mathbb{P}(T_i = 1 \mid Z_i = 0))
\end{aligned}$$

- However, there are two terms we cannot identify.
    - Instead, think about the maximum and minimum values τ can take
    - Recall that probability is bounded between 0 and 1
    - Maximum is when $\mathbb{P}(Y_i(T_i = 1) = 1 \mid T_i = 0, Z_i = 1) = 1$ and $\mathbb{P}(Y_i(T_i = 0) = 1 \mid T_i = 1, Z_i = 0) = 0$
    - Minimum is when when $\mathbb{P}(Y_i(T_i = 1) = 1 \mid T_i = 0, Z_i = 1) = 0$ and $\mathbb{P}(Y_i(T_i = 0) = 1 \mid T_i = 1, Z_i = 0) = 1$

Sharp Bounds (1)

- Manski bound is not sharp!
  - **Sharp Bound**: The bound with smallest width
  - Sharp bounds need to use **all** the possible information in the data
- Let $U_i = (T_i(1), T_i(0), Y_i(1), Y_i(0))$. Then, notice that

$$\tau = \mathbb{P}(Y_i(T_i = 1) = 1) - \mathbb{P}(Y_i(T_i = 0) = 1)$$

$$= \sum_u \Big(\mathbb{P}(Y_i(1) = 1 \mid U_i = u) - \mathbb{P}(Y_i(0) = 1 \mid U_i = u)\Big)\mathbb{P}(U_i = u)$$

$$= \sum_u \Big(\mathbf{1}\{Y_i^{(u)}(1) = 1\} - \mathbf{1}\{Y_i^{(u)}(0) = 1\}\Big)\mathbb{P}(U_i = u)$$

$$= \sum_u \Big(Y_i^{(u)}(1) - Y_i^{(u)}(0)\Big)\mathbb{P}(U_i = u) \quad (\because \; Y_i \text{ is binary})$$

where

  - $Y_i^{(u)}(t)$ is the value $Y_i(t)$ takes under the strata $U_i = u$
- As $U_i = (T_i(1), T_i(0), Y_i(1), Y_i(0))$ tells us $Y_i^{(u)}(1) - Y_i^{(u)}(0)$, once we identify $\mathbb{P}(U_i = u)$, we know ATE

Sharp Bounds (2)

- What condition does U_i need to satisfy?

$$
\begin{aligned}
&\overbrace{\mathbb{P}(Y_i = 1, T_i = t, Z_i = z)}^{\text{Observed}} \\
&= \mathbb{P}(Y_i(T_i(z)) = 1, T_i(z) = t, Z_i = z) \\
&= \mathbb{P}(Y_i(T_i(z)) = 1, T_i(z) = t)\mathbb{P}(Z_i = z) \quad (\because \text{Randomization}) \\
&= \sum_u \mathbb{P}(Y_i(T_i(z)) = 1, T_i(z) = t \mid U_i = u)\mathbb{P}(U_i = u)\mathbb{P}(Z_i = z) \\
&= \sum_u \underbrace{\mathbf{1}\{Y_i^{(u)}(T_i^{(u)}(z)) = 1, T_i^{(u)}(z) = t\}}_{\text{Determined by each strata}} \mathbb{P}(U_i = u) \underbrace{\mathbb{P}(Z_i = z)}_{\text{Observed}}
\end{aligned}
$$

where

- $T_i^{(u)}(z)$ is the value $T_i(z)$ takes under the strata $\mathbb{P}(U_i = u)$ needs to satisfy

- This becomes the constraint of $\mathbb{P}(U_i = u)$

Sharp Bounds (3): Optimization Problem

- Therefore, we can formalize the problem of obtaining upper / lower bound as the following optimization problem
  - Below is the one for obtaining the upper bound. For lower bound, replace max with min.

$$\max_{p_u} \quad \underbrace{\sum_u \left(Y_i^{(u)}(1) - Y_i^{(u)}(0) \right) p_u}_{\text{ATE}}$$

$$\text{such that} \quad 0 \leq p_u \leq 1, \quad \sum_u p_u = 1$$

$$\mathbb{P}(Y_i = 1, T_i = t, Z_i = z)$$

$$= \sum_u \mathbf{1}\{Y_i^{(u)}(T_i^{(u)}(z)) = 1, T_i^{(u)}(z) = t\} \mathbb{P}(Z_i = z) \times p_u$$

where $p_u = \mathbb{P}(U_i = u)$.

- This is the basic idea of automation of bounds (autobounds)

Extra: What is linear programming

- **Optimization problem** contains two components:
  - Objective function: the function to minimize / maximize
  - Constraints that solution need to satisfy

- Standard approach: transform the optimization problem to the specific form so that solver can solve automatically

- **Linear programming**: One form of optimization problem that can be easily solved by solver
  - Both constraint and objective function are linear

$$\begin{aligned} \max_x \quad & c^\top x \\ \text{such that} \quad & x \geq 0, Ax \leq b \end{aligned}$$

Partial Identification: Case of Binary Outcome (1)

- Let's think about the binary outcome and treatment. We have the following principal strata:

$$(Y_i(0), Y_i(1)) \in \{(0,0),(0,1),(1,0),(1,1)\}$$

- Suppose that we want to assign treatment to maximize the effect $Y_i(1) - Y_i(0)$
    - That is, assigning treatment to the strata $(Y_i(0), Y_i(1)) = (0,1)$ and not assigning to the strata $(Y_i(0), Y_i(1)) = (1,0)$
    - The only people whose outcome is 0 is those in strata $(Y_i(0), Y_i(1)) = (0,0)$
- **Question**: How can we maximize the outcome value by optimizing the treatment assignment?
- If we optimally assign the treatment effect, the observed outcome δ will be

$$\begin{aligned}\delta :=& 1 \times \mathbb{P}(Y_i(0) = 1, Y_i(1) = 1) \ + 1 \times \mathbb{P}(Y_i(0) = 0, Y_i(1) = 1) \\ &+ 1 \times \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0) + 0 \times \mathbb{P}(Y_i(0) = 0, Y_i(1) = 0)\end{aligned}$$

Partial Identification: Case of Binary Outcome (2)

- Now,

$$\mathbb{P}(Y_i(1) = 1) = \mathbb{P}(Y_i(0) = 0, Y_i(1) = 1) + \mathbb{P}(Y_i(0) = 1, Y_i(1) = 1)$$
$$\underbrace{\mathbb{P}(Y_i(0) = 1)}_{\text{Identifiable}} = \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0) + \mathbb{P}(Y_i(0) = 1, Y_i(1) = 1)$$

but we do not observe the probability of each principal strata.

- But we know that

$$\delta = \underbrace{\mathbb{P}(Y_i(1) = 1)}_{\text{Identifiable}} + \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)$$

so we need to think about how to maximize
$\mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)$

Partial Identification: Case of Binary Outcome (3)

- Let's write down all the constraints:
  - Firstly, each probability is bounded between 0 and 1
  - Then, we can identify $\mathbb{P}(Y_i(1) = 1)$ and $\mathbb{P}(Y_i(0) = 1)$

- In this case, each strata probability can be written as observed quantity and $\mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)$. I.e.,

$$
\begin{aligned}
&0 \leq \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0) \leq 1 \\
&0 \leq \underbrace{\mathbb{P}(Y_i = 1 \mid T_i = 0) - \mathbb{P}(Y_i(0) = 1, Y_i(1) = 0)}_{=\mathbb{P}(Y_i(0)=1, Y_i(1)=1)} \leq 1
\end{aligned}
$$

- You can also derive $\mathbb{P}(Y_i(0) = 0, Y_i(1) = 1)$ and $\mathbb{P}(Y_i(0) = 0, Y_i(1) = 0)$

- Under these constraints, think about how much you can maximize the quantity of interest.