5. Exchangeability: Assumptions to block backdoor paths

Ian Lundberg
Cornell Info 6751: Causal Inference in Observational Settings
Fall 2022

6 Sep 2022

Responding to feedback

- Group exercises are good for learning
- Would be better with written instructions
- Would be better with more time allocated to the hard one

Learning goals for today

At the end of class, you will be able to:

1. Encode causal theories in Directed Acyclic Graphs (DAGs)
2. Identify causal effects by blocking backdoor paths
3. Understand collider variables

What is a **Directed Acyclic Graph (DAG)**?

What is a **Directed Acyclic Graph (DAG)**?

A DAG is a formal graph, used for causal assumptions

What is a **Directed Acyclic Graph (DAG)**?

A DAG is a formal graph, used for causal assumptions

- Each **node** is a variable
- Each **edge** is a causal relation

$$\begin{array}{ccc} & B & \\ \nearrow & & \searrow \\ A & \longrightarrow & C \end{array}$$

What is a **Directed Acyclic Graph (DAG)**?

A DAG is a formal graph, used for causal assumptions

- Each **node** is a variable
- Each **edge** is a causal relation

$$\begin{tikzcd} & B \arrow[dr] & \\ A \arrow[ur] \arrow[rr] & & C \end{tikzcd}$$

What makes it a DAG?

What is a **Directed Acyclic Graph (DAG)**?

A DAG is a formal graph, used for causal assumptions

- Each **node** is a variable
- Each **edge** is a causal relation

$$\begin{tikzcd}
 & B \arrow[dr] & \\
A \arrow[ur] \arrow[rr] & & C
\end{tikzcd}$$

What makes it a DAG? It is directed and acyclic.

- Directed: Every edge has an arrow. Causality flows one way.
- Acyclic: There are no cycles

Not DAGs:

$$\begin{tikzcd}
 & B \arrow[dr, dash] & \\
A \arrow[ur, dash] \arrow[rr, dash] & & C
\end{tikzcd}$$

Undirected

$$\begin{tikzcd}
 & B \arrow[dr] & \\
A \arrow[ur] & & C \arrow[ll]
\end{tikzcd}$$

Cyclic

Why draw a DAG? Four reasons

Why draw a DAG? Four reasons

1. DAGs formalize the theory we believe before we see any data

Why draw a DAG? Four reasons

1. DAGs formalize the theory we believe before we see any data
2. DAGs often formalize things the data cannot tell us

Why draw a DAG? Four reasons

1. DAGs formalize the theory we believe before we see any data
2. DAGs often formalize things the data cannot tell us
3. DAGs are mathematically precise, with formal properties

Why draw a DAG? Four reasons

1. DAGs formalize the theory we believe before we see any data
2. DAGs often formalize things the data cannot tell us
3. DAGs are mathematically precise, with formal properties
4. DAGs are intuitive

1) Completely randomized experiment

1) Completely randomized experiment

Flip a coin. Assign job training. Observe employment.

1) Completely randomized experiment

Flip a coin. Assign job training. Observe employment.

Coin Flip

Z

1) Completely randomized experiment

Flip a coin. Assign job training. Observe employment.

Coin Flip

$$\begin{tikzcd} Z \arrow[d] \\ A \end{tikzcd}$$

Job Training

1) Completely randomized experiment

Flip a coin. Assign job training. Observe employment.

$$\begin{tikzcd}
\text{Coin Flip} \\
Z \arrow[d] \\
A \arrow[r] & Y \\
\text{Job Training} & \text{Employment}
\end{tikzcd}$$

1) Completely randomized experiment

Flip a coin. Assign job training. Observe employment.

$$
\begin{array}{ccc}
\text{Coin Flip} & & \\
Z & & \\
\downarrow & & \\
A & \longrightarrow & Y \\
\text{Job Training} & & \text{Employment}
\end{array}
$$

Recall: Heads and tails are **exchangeable**.

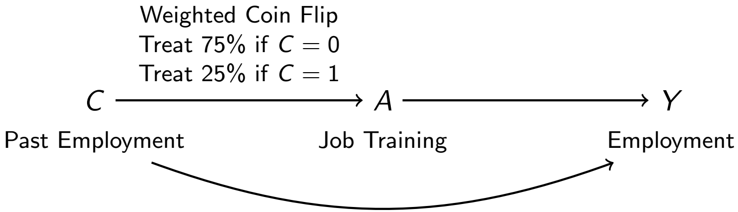
2) Randomized experiment: Unequal probabilities

2) Randomized experiment: Unequal probabilities

Assign job training with higher probability
to those who were not employed last year.

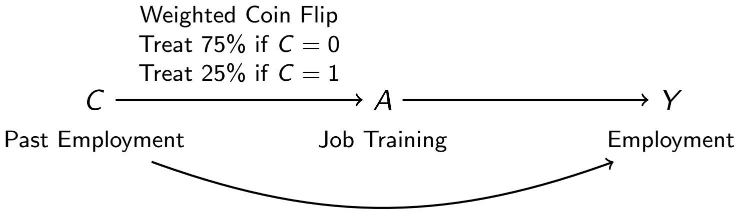
2) Randomized experiment: Unequal probabilities

Assign job training with higher probability
to those who were not employed last year.



2) Randomized experiment: Unequal probabilities

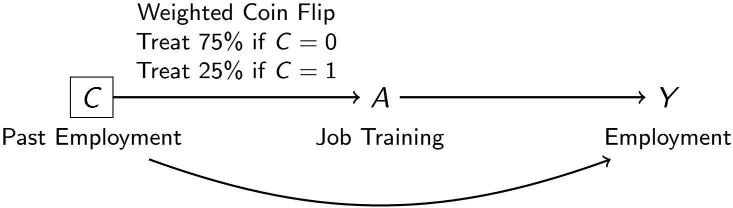
Assign job training with higher probability
to those who were not employed last year.



Question: How can we analyze data from this experiment?

2) Randomized experiment: Unequal probabilities

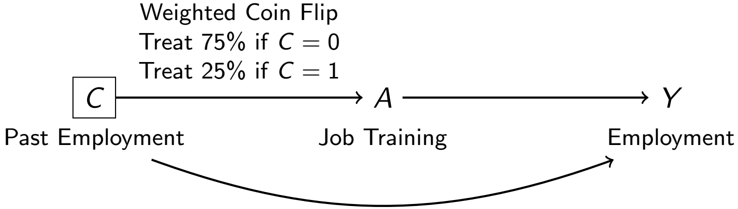
Assign job training with higher probability
to those who were not employed last year.



Question: How can we analyze data from this experiment?
Answer: We should analyze within subgroups of C

2) Randomized experiment: Unequal probabilities

Assign job training with higher probability
to those who were not employed last year.



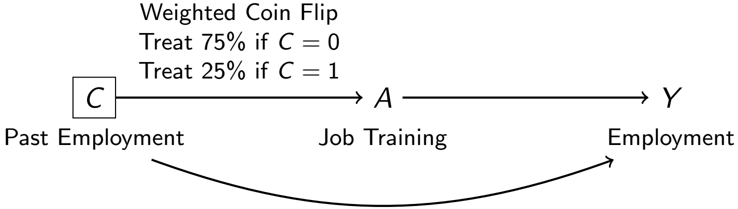
Question: How can we analyze data from this experiment?

Answer: We should analyze within subgroups of C

Heads and tails are **conditionally exchangeable**

2) Randomized experiment: Unequal probabilities

Assign job training with higher probability
to those who were not employed last year.



Question: How can we analyze data from this experiment?
Answer: We should analyze within subgroups of C
Heads and tails are **conditionally exchangeable**
Next up: Formalize this intuition

Key Concept: Backdoor Adjustment

Key Concept: Backdoor Adjustment

There are two reasons A and Y are associated

Key Concept: Backdoor Adjustment

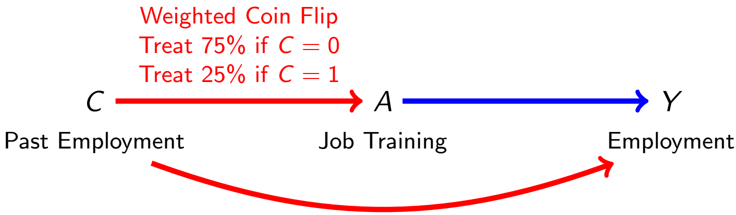
There are two reasons A and Y are associated

- A **causal path** $A \to Y$

Key Concept: Backdoor Adjustment

There are two reasons A and Y are associated

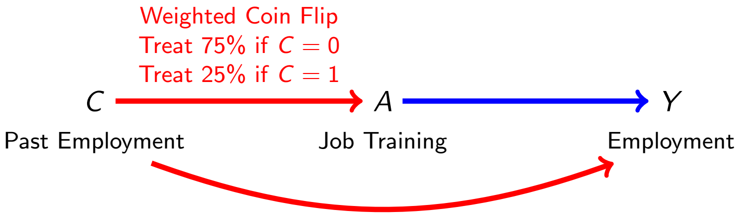
- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$



Key Concept: Backdoor Adjustment

There are two reasons A and Y are associated

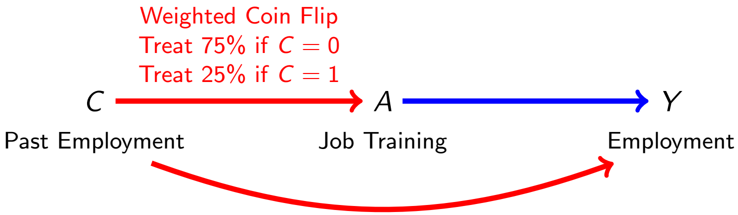
- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$
  - Think of a house at A



Key Concept: Backdoor Adjustment

There are two reasons A and Y are associated

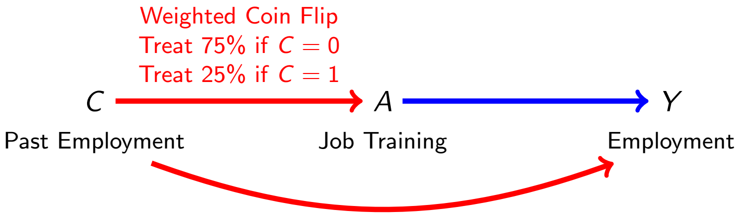
- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$
  - Think of a house at A
  - Walk out the back door to C, then around the house to Y



Key Concept: Backdoor Adjustment

There are two reasons A and Y are associated

- A **causal path** $A \rightarrow Y$
- A **backdoor path** $A \leftarrow C \rightarrow Y$
  - Think of a house at A
  - Walk out the back door to C, then around the house to Y

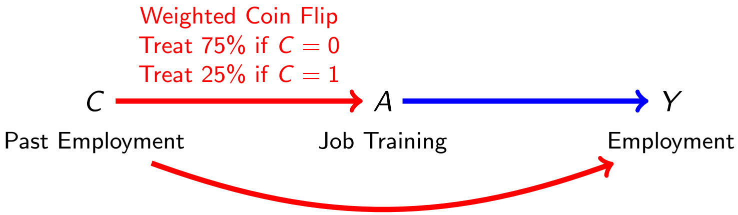


To block the backdoor path, condition on C

Key Concept: Backdoor Adjustment

There are two reasons A and Y are associated

- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$
  - Think of a house at A
  - Walk out the back door to C, then around the house to Y



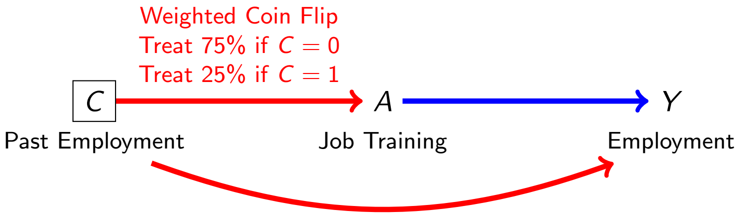
To block the backdoor path, condition on C

- Analyze within strata of C

Key Concept: Backdoor Adjustment

There are two reasons A and Y are associated

- A **causal path** $A \to Y$
- A **backdoor path** $A \leftarrow C \to Y$
  - Think of a house at A
  - Walk out the back door to C, then around the house to Y



To block the backdoor path, condition on C

- Analyze within strata of C
- Denoted by the box

Check: The power of backdoor adjustment

Check: The power of backdoor adjustment

- If my DAG is true

Check: The power of backdoor adjustment

- If my DAG is true
- and I adjust for the required variables

Check: The power of backdoor adjustment

- If my DAG is true
- and I adjust for the required variables
- then my observational data support a causal claim (just like an experiment would)

Check: The power of backdoor adjustment

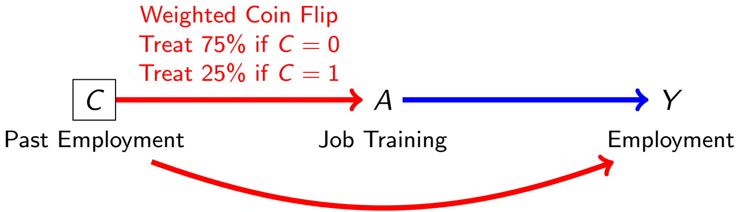
- If my DAG is true
- and I adjust for the required variables
- then my observational data support a causal claim (just like an experiment would)

Check: The power of backdoor adjustment

- If my DAG is true
- and I adjust for the required variables
- then my observational data support a causal claim (just like an experiment would)

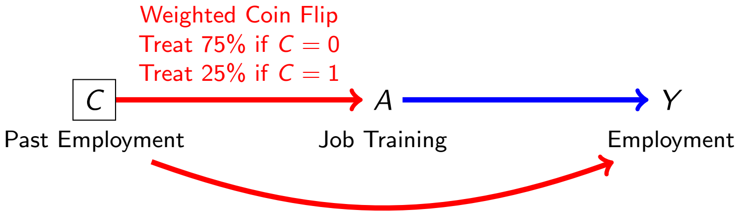
Only difference: In an experiment we know the DAG is true.

Key Concept: Causal Identification



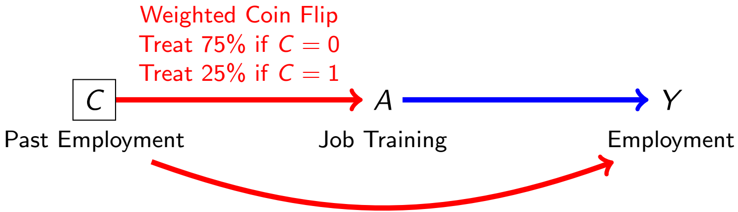
Key Concept: Causal Identification

Causal identification links
causal quantities (involving potential outcomes, e.g. Y^a)
to statistical quantities (involving observable variables, e.g. Y)



Key Concept: Causal Identification

Causal identification links
causal quantities (involving potential outcomes, e.g. Y^a)
to statistical quantities (involving observable variables, e.g. Y)

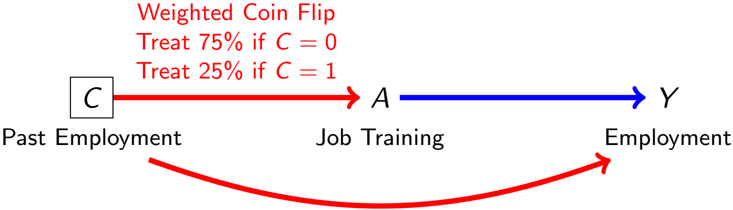


Once we block the backdoor path,

$$\underbrace{\mathrm{E}(Y^a \mid C = c)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid C = c, A = a)}_{\text{Statistical Quantity}}$$

Key Concept: Causal Identification

Causal identification links
causal quantities (involving potential outcomes, e.g. Y^a)
to statistical quantities (involving observable variables, e.g. Y)



Once we block the backdoor path,

$$\underbrace{\mathrm{E}(Y^a \mid C = c)}_{\text{Causal Quantity}} = \underbrace{\mathrm{E}(Y \mid C = c, A = a)}_{\text{Statistical Quantity}}$$

Also known as **conditional exchangeability** (Hernán & Robins 3.2)

Key Concept: Sufficient Adjustment Set

Key Concept: Sufficient Adjustment Set

A **sufficient adjustment set** is any set of variables that blocks all backdoor paths between the treatment and outcome

Key Concept: Sufficient Adjustment Set

A **sufficient adjustment set** is any set of variables that blocks all backdoor paths between the treatment and outcome

$$\begin{tikzcd}
X_1 \arrow[r] \arrow[d] & A \arrow[r] & Y \\
X_2 \arrow[r] & X_3 \arrow[ur] &
\end{tikzcd}$$

Key Concept: Sufficient Adjustment Set

A **sufficient adjustment set** is any set of variables that blocks all backdoor paths between the treatment and outcome

$$\begin{tikzcd}
X_1 \arrow[r] \arrow[d] & A \arrow[r] & Y \\
X_2 \arrow[r] & X_3 \arrow[ur] &
\end{tikzcd}$$

We can block the backdoor path in any of 3 ways:

Key Concept: Sufficient Adjustment Set

A **sufficient adjustment set** is any set of variables that blocks all backdoor paths between the treatment and outcome

$$\begin{tikzcd}
X_1 \arrow[r] \arrow[d] & A \arrow[r] & Y \\
X_2 \arrow[r] & X_3 \arrow[ur] &
\end{tikzcd}$$

We can block the backdoor path in any of 3 ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$

Key Concept: Sufficient Adjustment Set

A **sufficient adjustment set** is any set of variables that blocks all backdoor paths between the treatment and outcome

$$\begin{tikzcd} X_1 \arrow[r] \arrow[d] & A \arrow[r] & Y \\ X_2 \arrow[r] & X_3 \arrow[ur] & \end{tikzcd}$$

We can block the backdoor path in any of 3 ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$
- Condition on X_2: $A \leftarrow X_1 \rightarrow \boxed{X_2} \rightarrow X_3 \rightarrow Y$

Key Concept: Sufficient Adjustment Set

A **sufficient adjustment set** is any set of variables that blocks all backdoor paths between the treatment and outcome

$$\begin{tikzcd}
X_1 \arrow[r] \arrow[d] & A \arrow[r] & Y \\
X_2 \arrow[r] & X_3 \arrow[ur] &
\end{tikzcd}$$

We can block the backdoor path in any of 3 ways:

- Condition on X_1: $A \leftarrow \boxed{X_1} \rightarrow X_2 \rightarrow X_3 \rightarrow Y$
- Condition on X_2: $A \leftarrow X_1 \rightarrow \boxed{X_2} \rightarrow X_3 \rightarrow Y$
- Condition on X_3: $A \leftarrow X_1 \rightarrow X_2 \rightarrow \boxed{X_3} \rightarrow Y$

Key Concept: Colliders[1]

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

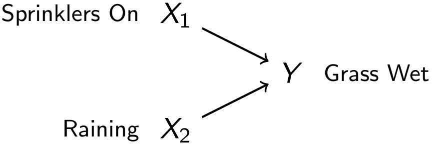
Key Concept: Colliders[1]

Suppose I have sprinklers on a timer.

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Key Concept: Colliders[1]

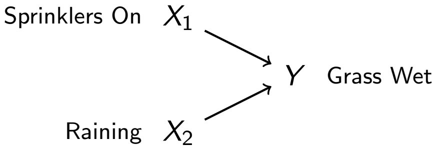
Suppose I have sprinklers on a timer.



[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Key Concept: Colliders[1]

Suppose I have sprinklers on a timer.

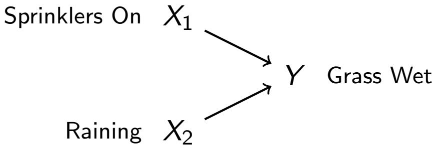


We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Key Concept: Colliders[1]

Suppose I have sprinklers on a timer.



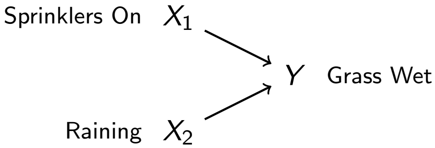
We say Y is a **collider** along the path $X_1 \to Y \leftarrow X_2$

- The collider blocks the path

[1]Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Key Concept: Colliders[1]

Suppose I have sprinklers on a timer.



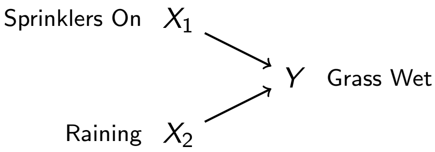
We say Y is a **collider** along the path $X_1 \rightarrow Y \leftarrow X_2$

- The collider blocks the path
- X_1 is independent of X_2
  - (Sprinklers On) is uninformative about (Raining)

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Key Concept: Colliders[1]

Suppose I have sprinklers on a timer.



We say Y is a **collider** along the path $X_1 \to Y \leftarrow X_2$

- The collider blocks the path
- X_1 is independent of X_2
  - (Sprinklers On) is uninformative about (Raining)
- Conditioning on Y opens the path
  - If the grass is wet (conditional on $Y = 1$), then either (Sprinklers On) or (Raining)

---

[1] Example from Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference.

Ancestors vs. Colliders

Conditioning on an ancestor
closes an open path

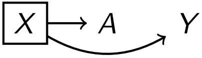
$$\begin{tikzcd}
\boxed{X} \arrow[r] \arrow[rr, bend right] & A & Y
\end{tikzcd}$$

Conditioning on an collider
opens a closed path

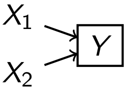
$$\begin{tikzcd}
X_1 \arrow[dr] & \\
& \boxed{Y} \\
X_2 \arrow[ur] &
\end{tikzcd}$$

Ancestors vs. Colliders

Conditioning on an ancestor
closes an open path



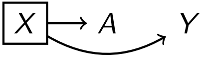
Conditioning on an collider
opens a closed path



In the population,
A and Y are **related**

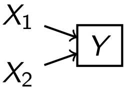
Ancestors vs. Colliders

Conditioning on an ancestor
closes an open path



In the population,
A and Y are **related**

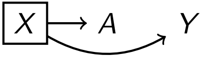
Conditioning on an collider
opens a closed path



In the population,
X_1 and X_2 are **independent**

Ancestors vs. Colliders

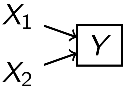
Conditioning on an ancestor **closes** an open path



In the population,
A and Y are **related**

Within strata of X,
A and Y are **independent**

Conditioning on an collider **opens** a closed path



In the population,
X_1 and X_2 are **independent**

Ancestors vs. Colliders

Conditioning on an ancestor **closes** an open path

$$\begin{tikzcd} \boxed{X} \arrow[r] \arrow[rr, bend right] & A & Y \end{tikzcd}$$

In the population,
A and Y are **related**

Within strata of X,
A and Y are **independent**

Conditioning on an collider **opens** a closed path

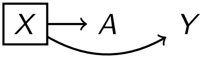
$$\begin{tikzcd} X_1 \arrow[dr] & \\ & \boxed{Y} \\ X_2 \arrow[ur] & \end{tikzcd}$$

In the population,
X_1 and X_2 are **independent**

Within strata of Y,
X_1 and X_2 are **related**

Ancestors vs. Colliders

Conditioning on an ancestor **closes** an open path

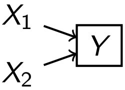


In the population,
A and Y are **related**

Within strata of X,
A and Y are **independent**

Example
— X is past unemployment
— A is ineffective job training
— Y is future employment

Conditioning on an collider **opens** a closed path

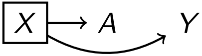


In the population,
X_1 and X_2 are **independent**

Within strata of Y,
X_1 and X_2 are **related**

Ancestors vs. Colliders

Conditioning on an ancestor **closes** an open path

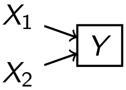


In the population,
A and Y are **related**

Within strata of X,
A and Y are **independent**

Example
— X is past unemployment
— A is ineffective job training
— Y is future employment

Conditioning on an collider **opens** a closed path



In the population,
X_1 and X_2 are **independent**

Within strata of Y,
X_1 and X_2 are **related**

Example
— X_1 is sprinklers on
— X_2 is rain
— Y is wet grass

Ancestors vs. Colliders[2]

The consequences of conditioning on a descendant differ

[2] A great paper on colliders: Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. Annual Review of Sociology, 40.

Ancestors vs. Colliders[2]

The consequences of conditioning on a descendant differ

Conditioning on
the descendant of an ancestor
does not block a path

$$
\begin{tikzcd}
\boxed{\tilde{X}} & & \\
X \arrow[u] \arrow[r] \arrow[rr, bend right] & A & Y
\end{tikzcd}
$$

Ancestors vs. Colliders[2]

The consequences of conditioning on a descendant differ

Conditioning on the descendant of an ancestor does not block a path

$$
\begin{tikzcd}
\boxed{\tilde{X}} & & \\
X \arrow[u] \arrow[r] \arrow[rr, bend right] & A & Y
\end{tikzcd}
$$

But conditioning on the descendant of a collider still opens one

$$
\begin{tikzcd}
X_1 \arrow[dr] & & \\
& Y \arrow[r] & \boxed{Z} \\
X_2 \arrow[ur] & &
\end{tikzcd}
$$

[2] A great paper on colliders: Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. Annual Review of Sociology, 40.

d-separation: Formal definition of blocking backdoor paths

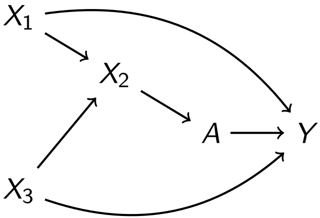
From Greenland, Pearl, & Robins 1999, p. 45:

> *...we say that a set of variables S separates two other sets R and T, or S blocks every path between R and T, if the following criteria are met:*
>
> 1. *Every unblocked path from R to T is intercepted by a variable in S, and*
> 2. *Every unblocked path from R to T generated by adjustment for the variables in S is intercepted by a variable in S*
>
> *(This concept is usually called "d-separation of R and T by S" in the graphical literature, where d stands for "directional."*

Intuition: $R \leftarrow \boxed{S} \rightarrow T$

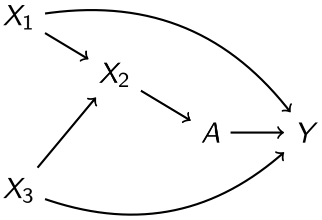
Exercise

Find 3 sufficient adjustment sets to identify $A \to Y$



Exercise

Find 3 sufficient adjustment sets to identify $A \to Y$



Answer: $\{X_2\}, \{X_1, X_3\}, \{X_1, X_2, X_3\}$

Exercise

What is the smallest adjustment set that identifies $A \to Y$?

$$\begin{tikzcd}
X_1 \arrow[dr] \arrow[ddrrr, bend left] & & & \\
& X_2 & & \\
X_3 \arrow[ur] \arrow[rr] & & A \arrow[r] & Y
\end{tikzcd}$$

Exercise

What is the smallest adjustment set that identifies $A \to Y$?

$$
\begin{tikzcd}
X_1 \arrow[dr] \arrow[ddrrr, bend left] & & & \\
& X_2 & & \\
X_3 \arrow[ur] \arrow[rr] & & A \arrow[r] & Y
\end{tikzcd}
$$

Answer: The empty set! Don't condition on anything.
The collider X_2 already blocks the path.

Key Concept: DAGs are better than rules of thumb

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

Examples where that rule fails:

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

Examples where that rule fails:

$$\begin{tikzcd} & M \arrow[dr] & \\ A \arrow[ur] \arrow[rr] & & Y \end{tikzcd}$$

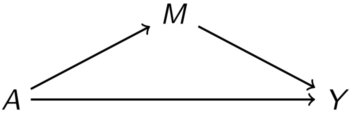
— M is related to A
— M is related to Y

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

Examples where that rule fails:



— M is related to A
— M is related to Y
— But conditioning on M blocks a causal path!

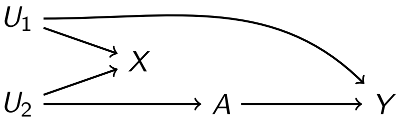
Imai, K., Keele, L., Tingley, D., & Yamamoto, T. (2011). Unpacking the black box of causality: Learning about causal mechanisms from experimental and observational studies. American Political Science Review, 105(4), 765-789.

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

Examples where that rule fails:



— X is related to A
— X is related to Y

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

Examples where that rule fails:

$$
\begin{tikzcd}
U_1 \arrow[dr] \arrow[drrr, bend left] & & & \\
& X & & \\
U_2 \arrow[ur] \arrow[rr] & & A \arrow[r] & Y
\end{tikzcd}
$$

— X is related to A

— X is related to Y

— But conditioning on X opens the backdoor path!

Greenland, S. (2003). Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology, 14(3), 300-306.

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

Examples where that rule fails:

$$\begin{tikzcd} & & & U \arrow[dl] \arrow[dr] & \\ Z \arrow[rr] & & A \arrow[rr] & & Y \end{tikzcd}$$

— Z is related to A
— Z is related to Y, given A

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

Examples where that rule fails:

$$\begin{tikzcd} & & U \arrow[dl] \arrow[dr] & \\ Z \arrow[r] & A \arrow[rr] & & Y \end{tikzcd}$$

— Z is related to A
— Z is related to Y, given A
— But conditioning on Z amplifies bias from U!

Pearl, J. (2011). Invited commentary: Understanding bias amplification. American Journal of Epidemiology, 174(11), 1223-1227.

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

The rule of thumb commits a **fundamental error**.

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

The rule of thumb commits a **fundamental error**.

It presents a criterion that is

statistical

(involving observed relationships among variables)

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

The rule of thumb commits a **fundamental error**.

It presents a criterion that is

statistical

(involving observed relationships among variables)

when we need a criterion that is

causal

(involving counterfactual values)

Key Concept: DAGs are better than rules of thumb

Incorrect Rule of Thumb:

> Adjust for all variables that are related to the treatment and related to the outcome.

The rule of thumb commits a **fundamental error**.

It presents a criterion that is

statistical

(involving observed relationships among variables)

when we need a criterion that is

causal

(involving counterfactual values)

This is why we need a DAG.

Learning goals for today

At the end of class, you will be able to:

1. Encode causal theories in Directed Acyclic Graphs (DAGs)
2. Identify causal effects by blocking backdoor paths
3. Understand collider variables

Let me know what you are thinking

tinyurl.com/CausalQuestions

Office hours TTh 11am-12pm and at
calendly.com/ianlundberg/office-hours
Come say hi!