# Module 2

Friday, October 4, 2024 2:21 PM

## Confounding

Recall that we are interested in the means of different potential outcomes, like Average Causal Effect: $E(Y^1 - Y^0)$. To get this from observational data, we have to make several assumptions, including ignorability: $Y^0, Y^1 \perp A|X$.

Suppose that treatment assignment depends on the potential outcomes, such as sicker patients being more likely to be treated. Treated patients have a higher risk of a bad outcome and we need to account for many potential pre-treatment differences in health.

Suppose that the covariates X are different measures of health like history of various diseases, age, weight, smoking, alcohol intake, etc. It is possible that within levels of X, it may be the case that "sicker" patients are not more likely to get treatment. Perhaps people of the same weight are equally likely to get treatment. This is ignorability.

Confounders are often defined as variables that affect both the treatment and the outcome.

- If treatment is assigned via a coin flip, then that affects treatment but not outcome and the coin flip isn't a confounder
- If people with a family history of cancer are more likely to develop cancer (the outcome), but family history was not a factor in treatment decision, then family history is not a confounder
- If older people are at higher risk of cardiovascular disease (the outcome) and are more likely to receive statins (the treatment), then age is a confounder

Note: a variable that only effects the outcome is sometimes called a Risk Factor.

So if we are interested in controlling for confounders, then:

1. We should identify a set of variables X that will make the ignorability assumption hold. If we are able to find these variables, then the set X is sufficient to control for confounders.
2. Using statistical methods to control for the confounding variables and estimate causal effects.

Discovering which variables to control for isn't simple, but we can use Causal Graphs to help answer this question and formalize key ideas.

## Causal Graphs

Causal Graphs are considered useful for causal inference. These are also known as Directed Acyclic Graphs (DAGs). They are helpful in several ways:

- Identifying which variables to control for
- Making assumptions explicit

Graphical models:

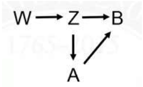
- Encode assumptions about relationships among variables, like which variables are independent, dependent, conditionally dependent, etc.
- Can be used to derive nonparametric causal effect estimators

$A \rightarrow Y$

This is a directed graph, which shows that A affects Y. A and Y are known as nodes / vertices, but can be thought of as variables or set of variables. The link (edge) between A and Y is an arrow, which means there is a direction (directed path). Variables connected by an edge are adjacent. For a graph to be a directed graph, all edges between nodes must be directed.

$A - Y$

This is an undirected graph, which shows that A and Y are associated with each other.



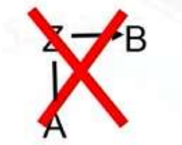
A path is a way to get from one vertex to another, travelling along edges. There are two paths from W to B:

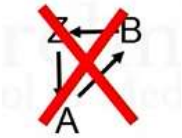
- $W \rightarrow Z \rightarrow B$
- $W \rightarrow Z \rightarrow A \rightarrow B$

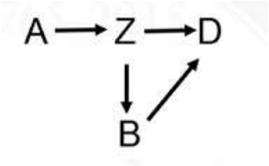
There is one path from Z to W:

- $Z \leftarrow W$

Directed Acyclic Graphs (DAGs) have no undirected paths and no cycles. DAGs will help determine the set of variables that we need to control for in order to achieve ignorability.







In the above DAG:

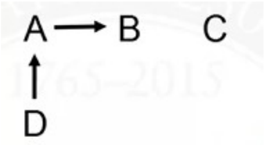
- A is Z's parent
- B is a child of Z
- D is a descendant of A
- Z is an ancestor of D
- D has two parents, Z and B

## Relationships between DAGs and Probability Distributions

Recall that graphical models encode assumptions about relationships among variables.

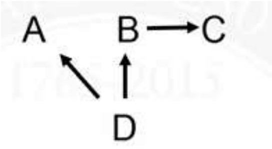
We can use the DAG to decompose the joint distribution by sequential conditioning only on sets of parents.

1. Start with roots, the nodes that do not have any parents.
2. Proceed down the descendant line, always conditioning on parents.



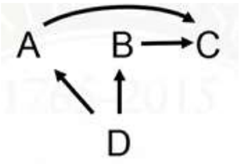
- C is independent from D, A, and B: $P(C|A,B,D) = P(C)$
- B is independent from C and D conditioned on A: $P(B|A,C,D) = P(B|A)$
- B and D are marginally dependent: $P(B|D) \neq P(B)$
- D is independent from C and B conditioned on A: $P(D|A,B,C) = P(D|A)$

Thus the decomposition is: $P(A,B,C,D) = P(C)P(D)P(A|D)P(B|A)$



- A is independent from B and C conditioned on D: $P(A|B,C,D) = P(A|D)$
- D is independent of C conditioned on A and B: $P(D|A,B,C) = P(D|A,B)$

Thus the decomposition is: $P(A,B,C,D) = P(D)P(A|D)P(B|D)P(C|B)$



- A is independent of B conditioned on C and D: $P(A|B,C,D) = P(A|C,D)$
- D is independent of C conditioned on A and B: $P(D|A,B,C) = P(D|A,B)$

Thus the decomposition is: $P(A,B,C,D) = P(D)P(A|D)P(B|D)P(C|A,B)$

When the decomposition matches the DAG, that is called compatible. DAGs that are compatible with a particular probability function aren't necessarily unique.

- If you start with the probability function, that doesn't imply a unique DAG.
- If you start with a particular DAG, that does imply something about a probability function

## Paths and Associations

Types of paths:

Fork: $D \leftarrow E \rightarrow F$

Chain: $D \rightarrow E \rightarrow F$

Inverted Fork: $D \rightarrow E \leftarrow F$

If nodes A and B are on the ends of a path, then they are associated via this path if:

- Some information flows to both of them
- Information from one makes it to the other

$D \leftarrow E \rightarrow F$

- D and F are not independent since they both have information flowing from E.

$A \leftarrow C \leftarrow D \leftarrow E \rightarrow G \rightarrow B$

- A and B get information from E. A and B are associated with each other via this path.

$A \rightarrow G \rightarrow B$

- Information flows from A to B. A and B are associated with each other via this path.

$A \rightarrow G \leftarrow B$

- Information from A and B collide at G
- G is known as a collider
- A and B do not have an association via this path because no information goes from A to B or B to A
- A and B are independent of each other

$A \rightarrow G \leftarrow D \leftarrow B$

- If there is a collider anywhere on the path from A to B, then no association between A and B comes from this path
- G is a collider

## Conditional Independence (d-separation)

Paths can be blocked by conditioning on nodes in the path.

Consider the path: $A \rightarrow G \rightarrow B$

- A and B are dependent through G.
- A affects G which affects B, so it's really G that causes the association between A and B.
- If we condition on G, a node in the middle of a chain, then we block the path from A to B.
- "Condition on G" = "controlling for G"

As an example, suppose that Temperature (A) affects whether or not Sidewalks are Icy (G), which affects whether or not Someone Falls (B).  If we restrict to situations where sidewalks are icy (condition on G), then temperature and falling are not associated via this path.  Recall that conditioning on something means you are limiting to a subpopulation, which in this case is a world where all sidewalks are always icy.

Inverted forks can also be blocked. Consider the path: $A \leftarrow G \rightarrow B$

- If we condition on G, the path from A to B is blocked.  And A and B are independent of one another.

However, the opposite situation occurs if a collider is conditioned on.  Consider the path: $A \rightarrow G \leftarrow B$

- A and B are not associated via this path because information collides at G.  A and B are independent from one another.
- Conditioning on the collider G induces an association between A and B.

As an example, consider the state of an on/off switch (A), the state of a second on/off switch (B), and whether a lightbulb is lit up (G). In this case, A is determined by a coin flip, B is determined by a separate coin flip, and G is lit up only if both A and B are "on".  A and B both affect G and G is a collider: $A \rightarrow G \leftarrow B$

- A and B are independent. Knowing that B is on does not tell you anything about A and vice versa.
- However, A and B are dependent given G. If we know that the light if off (G), then A must be off if B is on or vice versa. Thus conditioning on G opens up a path between A and B.

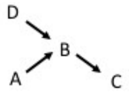
A path is d-separated (dependence-separated) by a set of nodes C if:

- It contains a chain ($D \rightarrow E \rightarrow F$) and the middle portion of the chain is in C.  $E \in C$
  OR
- it contains a fork ($D \leftarrow E \rightarrow F$) and the middle part is in C.  $E \in C$
  OR
- It contains an inverted fork ($D \rightarrow E \leftarrow F$) and the middle part is not in C, nor are any descendants of it. $E \notin C$

Two nodes, A and B, are d-separated by a set of nodes C if C blocks every path from A to B.

- Then A is independent of B conditioned on C: $A \perp B|C$
- Keep in mind the ignorability assumption: $Y^0, Y^1 \perp A|X$

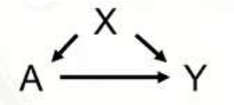
## Practice Quiz

1. What type of path is this? $A \leftarrow X \rightarrow Y$
   a. Fork
2. What type of path is this? $A \rightarrow B \rightarrow C$
   a. Chain
3. This DAG is consistent with which factorization of the joint distribution?
   a. 

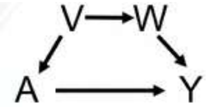
   b. $P(A, B, C, D) = P(D)P(A)P(B|D \cap A)P(C|B)$
4. Is there a collider on the path from A to C? $A \rightarrow B \rightarrow C$
   a. No

## Confounding Revisited

Recall that confounders are defined as variables that affect both the treatment and the outcome.



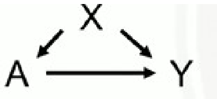
V affects both treatment A (directly) and outcome Y (indirectly through W), so it can be argued that V is a confounder.



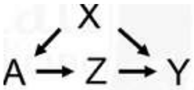
V affects both treatment A (directly) and outcome Y (indirectly through W), so it can be argued that V is a confounder.

We handle this by identifying a set of variables that are sufficient to control for confounding, blocking backdoor paths from treatment to outcome.

A frontdoor path from A to Y is one that begins with an arrow emanating out of A. Frontdoor paths capture the effects of treatment.
A backdoor path from A to outcome Y is one that travels through arrows going into A. Backdoor paths confound the relationship between the treatment and outcome and need to be blocked.



$A \to Y$ is a frontdoor path from A to Y
$A \leftarrow X \to Y$ is a backdoor path from A to Y



$A \to Z \to Y$ is a frontdoor path from A to Y; we do not need to control for Z as it captures some information about the effects of treatment. Causal mediation analysis involves understanding frontdoor paths from A to Y, i.e. how much of the treatment effect are through A's impact on Z?

Let X be the set of variables that block all backdoor paths from treatment to outcome. Then we have the ignorability of treatment mechanism given X:

$$Y^0, Y^1 \perp A | X.$$

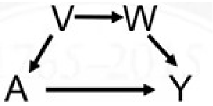
Two ways of identifying backdoor paths are the Backdoor Path Criterion and Disjunctive Cause Criterion.

## Backdoor Path Criterion

A set of variables X is sufficient to control for confounding if:

1. It blocks all backdoor paths from treatment to the outcome
2. It does not include any descendants of treatment

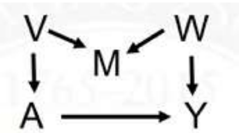
This is the backdoor path criterion. It isn't necessarily unique; there may be several sets X that successfully fit the criterion.



One backdoor path that isn't blocked by a collider: $A \leftarrow V \to W \to Y$

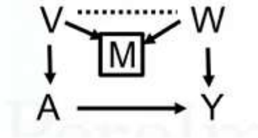
Sets of variables that are sufficient to control for confounding:

- {V}
- {W}
- {V,W}



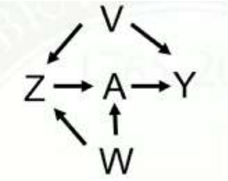
One backdoor path that is blocked by collider M - thus no confounding here: $A \leftarrow V \to M \leftarrow W \to Y$

However, if M is controlled for, it opens a path between V and W and introduces confounding:



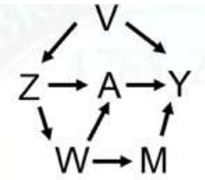
Sets of variables that are sufficient to control for confounding:

- {}
- {V}
- {W}
- {M,W}
- {M, V}
- {M, V, W}



Two backdoor paths:

- $A \leftarrow Z \leftarrow V \rightarrow Y$
  - No colliders
  - Sets of variables that are sufficient for controlling:
    - {Z}
    - {V}
    - {Z,V}
- $A \leftarrow W \rightarrow Z \leftarrow V \rightarrow Y$
  - One collider Z, so controlling for Z opens up backdoor path from W to V.
  - Sets of variables that are sufficient for controlling:
    - {}
    - {V}
    - {W}
    - {Z,V}
    - {Z,W}
- Thus the set of variables sufficient to control for confounding in this DAG:
  - {}
  - {V}
  - {V,Z}
  - {Z,W}
  - {V,Z,W}
  - (but not {Z} or {W})



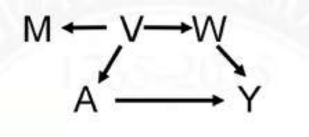
Three backdoor Paths:

- $A \leftarrow Z \leftarrow V \rightarrow Y$
  - No colliders
  - Sets of variables sufficient for controlling:
    - {Z}
    - {V}
    - {Z, V}
- $A \leftarrow W \leftarrow Z \leftarrow V \rightarrow Y$
  - No colliders
  - Sets of variables sufficient for controlling:
    - {W}
    - {Z}
    - {V}
    - {W, Z}
    - {W, V}
    - {V, Z}
    - {W, Z, V}
- $A \leftarrow W \rightarrow M \rightarrow Y$
  - No colliders
  - Sets of variables sufficient for controlling:
    - {W}
    - {M}
    - {W, M}
- Thus the set of variables that are sufficient for controlling this DAG:
  - {W, Z}
  - {W, V}
  - {M, Z}
  - {M, V}
  - {W, Z, V}
  - {M, Z, V}
  - {W, M, Z}

    - {W, M, V}
    - {W, M, Z, V}

## Disjunctive Cause Criterion

Disjunctive Cause Criterion: control for all (observed) causes of the exposure, the outcome, or both. Investigators do not need to know the whole graph, but rather the list of variables that affect the outcome or exposure. If there is a set of observed variables that satisfy the backdoor path criterion, then the variables selected based on the disjunctive cause criterion will be sufficient to control for confounding.



Observed pre-treatment variables: {M, W, V}
Unobserved pre-treatment variables: $\{U_1, U_2\}$

Suppose we know that W & V are causes of A, Y, or both. Suppose M is not a cause of either A or Y.

Then two methods for selecting variables:

- Use all pre-treatment covariates: {M, W, V}
- Use variables based on disjunctive cause criterion: {W, V}