Causal Directed Acyclic Graphs

Causal Inference

Spring 2026

Why DAGs?

- To visualize the causal structure of the world.
- To identify sources of bias in our estimates.
- To formally determine which variables we must condition on

Causal Directed Acyclic Graphs (DAGs)

- A DAG is a visual representation of hypothesized causal relationships.
- **Directed:** Each arrow implies a direction of causation (e.g., $X \to Y$: X causes Y).
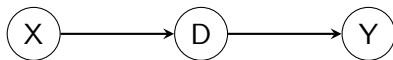- **Acyclic:** No cycles are allowed (a variable cannot cause itself).

Figure: A simple Causal Chain: $X \to D \to Y$

- A **Path** is any sequence of edges connecting two variables, regardless of arrow direction.
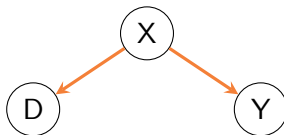
The Causal Chain

- **Structure:** $D \rightarrow X \rightarrow Y$
- Causal association flows freely along this path.

- **Conditioning on X:**
  - X is an **mediator**
  - Conditioning on X blocks association between D and Y.

The Confounder (The Backdoor Path)

- **Structure:** $D \leftarrow X \rightarrow Y$
- X is a **Confounder** (or common cause) of D and Y.
- **Example:** D=Smoking, Y=Lung Cancer, X=Genetics.

- **The Problem:** The path $D \leftarrow X \rightarrow Y$ is a **Backdoor Path**.
- Spurious relationship between D and Y.
- $Pr[Y|D]$ is biased because it includes the effect of X.

Conditioning on a Confounder
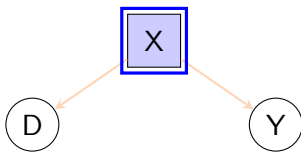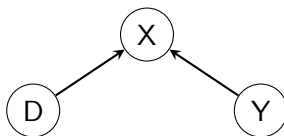
- Condition on the Confounder X Block the Backdoor Path

Figure: Conditioning on X blocks the flow.

The Collider

- **Structure:** $D \to X \leftarrow Y$
- **Collider** a node that two arrows point into / shared outcome.
- **Example:** D=Student, Y=TA, X=In This Room.

- The flow of association is naturally blocked by a collider.
- Marginally, D and Y are independent (e.g., being a student doesn't make you more or less likely to be a TA).
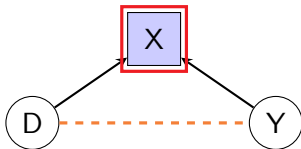
Conditioning on a Collider

Figure: Conditioning on X opens the flow.

- Conditioning on the Collider opens the path, creating a spurious association between D and Y.
- This is called *Collider Bias*.
- **Example:** If we only look at people *in this room* ($X = 1$), knowing someone is a student ($D = 1$) makes it less likely they are a TA ($Y = 0$), because the room is a mix of students and TAs.

Practice

- The key is:
  1. **close all Backdoor Paths** (Confounders) and
  2. **not open any Collider Paths**.

- Which variables must we condition on?
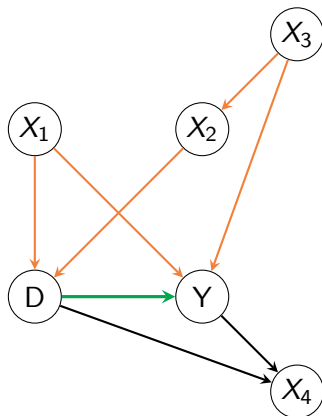
Summary

- **DAGs** are a powerful tool for visualizing and formalizing causal assumptions.

- Backdoor Criterion Revisited:

  A set S is sufficient for adjustment to identify the causal effect of X on Y if:
  1. No element of S is a descendant of X and
  2. The elements of S block all back-door paths from X to Y

- Condition 1: **Colliders** must **not** be opened by conditioning.
- Condition 2: **Confounders** (Backdoor Paths) must be closed by conditioning.