# Causal Graphs

Hans Jarett Ong

September 8, 2022

# Contents

# Why use Causal Graphs?

# Causal Graphs vs. Potential Outcome Model

- Causal graphs provide a direct and powerful way of thinking and reasoning about full (and thus more complex) causal systems.
- Pearl 2009 has shown that the causal graph approach and the potential outcome model are *equivalent*.
- Causal Graphs and Potential Outcomes are **complementary** – it is important to learn both approaches.

# Features of Causal Graphs

3 important features of Causal Graphs:

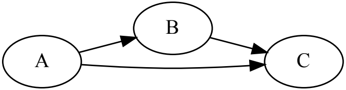
1. Causal graphs are completely non-parametric. This means that we don't need to assume the functional dependence of the variables. For example, if we say that $X \to Y$, all we know is that changing $X$ will change $Y$, but we don't need to specify if the relationship is linear, quadratic, nonlinear, etc.
2. Causal graphs clearly show **collider variable bias** which is hard to catch with just the potential outcomes model. Collider variable bias occurs when conditioning on a variable creates non-causal associations between treatment and outcome.
3. Causal graphs are used to develop transparent and clear justifications for estimating causal effects. It allows us to clearly state our assumptions and to justify the choice of estimation method to be used.

# Identification Analysis

- Before attempting to estimate causal effects, we must first see if this causal effect can be computed.
- Identification analysis assesses the feasibility of estimating a causal effect in an ideal scenario where the sample size is infinite.
- Identification analysis is first since it doesn't make sense to attempt to use a finite sample if it won't even work for an infinite sample.
- It relies on **domain knowledge** and the **joint distribution** of all variables involved – these are best represented with causal graphs.

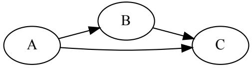
# Basic Elements of Causal Graphs

# Nodes, Edges, Paths, and Cycles

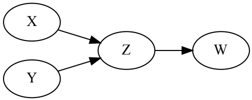



- **Nodes** such as $A$, $B$, and $C$ represent random variables.
- **Directed edges** ($\rightarrow$) represent the direction of the causal effect. e.g. if $A \rightarrow B$, then changing A will affect B but not the other way around; $A$ is called the *parent* while $B$ is called the *child*.
- A **path** is a sequence of edges that connect 2 nodes. For directed graphs, the path should follow the direction of the edges. In this example there are 2 paths connecting $A$ and $C$:
  - $A \rightarrow B \rightarrow C$
  - $A \rightarrow C$

# Directed Acyclic Graphs (DAGs)



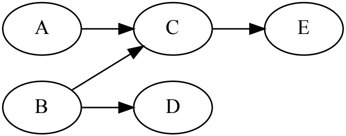

Acyclic Graph




Cyclic Graph

- Directed Acyclic graphs are graphs with **directed edges** and **no cycles**.
- Causal graphs are DAGs, hence they are sometimes called **Causal DAGs**.

# Joint Distribution as Factor Product




- Causal DAGs represent the joint probability distribution as a factor product:

$$P(x_1, x_2, \ldots, x_n) = \prod_i P(x_i | \mathsf{pa}_i)$$

  where $\mathsf{pa}_i$ are the parent nodes of $x_i$.

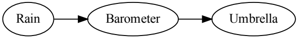
- For example, for the graph above:

$$P(A, B, C, D, E) = P(A)P(B)P(C|A, B)P(D|B)P(E|C)$$

- We'll use this factor product to prove some of the properties of causal DAGs.

# How do we come up with Causal DAGs?

- Causal DAGs require prior **Domain Knowledge**.
- For example, both graphs below will have the same joint distribution. Both ways of factoring are technically valid, but we'll have to rely on common sense (prior knowledge) to know which one is the correct causal graph.





- Causal Discovery aims to discover these graphs from data alone (no prior knowledge), but this is still an active area of research.

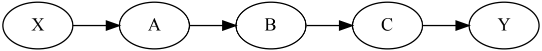
# D-separation Rules

# D-separation Rules

- A set of rules that tells us the conditional dependencies between the variables.
- d-separated variables are **independent**
- There are 3 rules that cover all possible patterns:
  1. Conditional Independence in Chains
  2. Conditional Independence in Forks
  3. Conditional Independence in Colliders

# Rule 1. Conditional Independence in Chains.

- *Two variables, $X$ and $Y$, are conditionally independent given $Z$, if there is only one unidirectional path between $X$ and $Y$ and $Z$ is any set of variables that intercepts that path.*
- In other words, conditioning on a chain **blocks** the path.

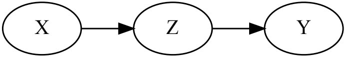



- For example, in the graph above, conditioning on $A$, $B$, or $C$ (or any combination of the three) will make $X$ and $Y$ d-separated. That is,

$$Z \subset \{A, B, C\} \implies X \perp\!\!\!\perp Y | Z$$

# Rule 1. Conditional Independence in Chains.

Proof:



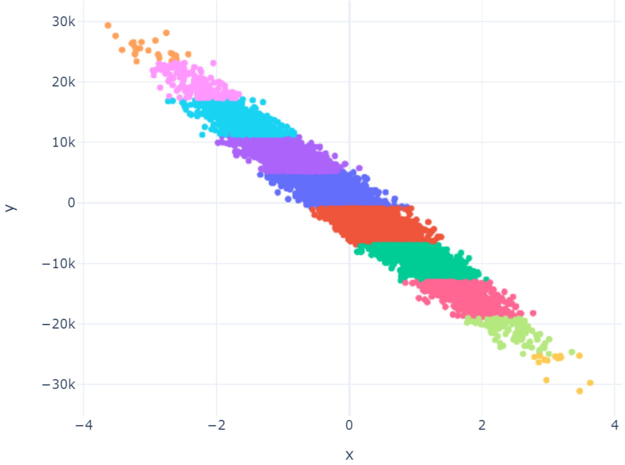

$$P(X,Y,Z) = P(X)P(Z|X)P(Y|Z) \quad \text{(Factor Product)}$$

$$P(X,Y|Z) = \frac{P(X,Y,Z)}{P(Z)} = \frac{P(X)P(Z|X)P(Y|Z)}{P(Z)}$$

$$P(X,Y|Z) = \frac{P(X)(P(X|Z)P(Z)/P(X))P(Y|Z)}{P(Z)} \quad \text{(Bayes' rule)}$$

$$\therefore P(X,Y|Z) = P(X|Z)P(Y|Z) \implies X \perp\!\!\!\perp Y|Z$$

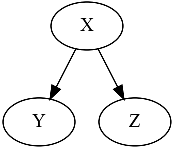
# Rule 1. Conditional Independence in Chains.




Simulated Example for $X \rightarrow Z \rightarrow Y$. Color represents Z.

# Rule 2. Conditional Independence in Forks.

- If a variable $X$ is a common cause of variables $Y$ and $Z$, and there is only one path between $Y$ and $Z$, then $Y$ and $Z$ are independent conditional on $X$.
- In other words, conditioning on a confounder$X$ blocks the non-causal association between $Y$ and $Z$, making them d-separated.

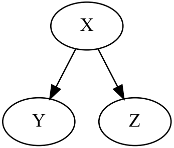



- For this graph,

$$Y \perp\!\!\!\perp Z | X$$

# Rule 2. Conditional Independence in Forks.

Proof:

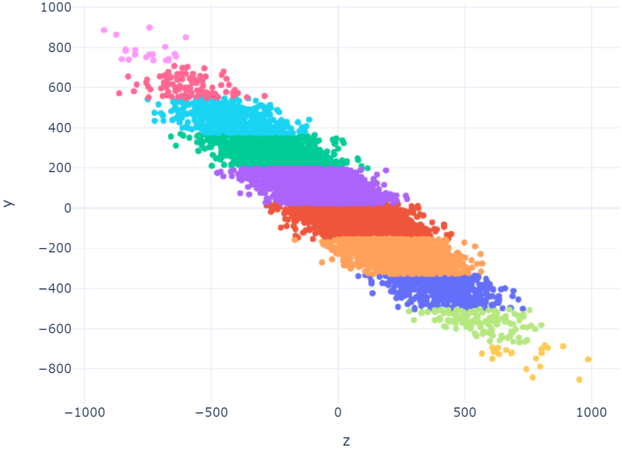



$$P(X,Y,Z) = P(X)P(Y|X)P(Z|X) \qquad \text{(Factor Product)}$$

$$P(Y,Z|X) = \frac{P(X,Y,Z)}{P(X)} = \frac{P(X)P(Y|X)P(Z|X)}{P(X)}$$

$$\therefore P(Y,Z|X) = P(Y|X)P(Z|X) \implies Y \perp\!\!\!\perp Z|X$$

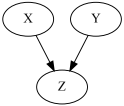
# Rule 2. Conditional Independence in Forks.

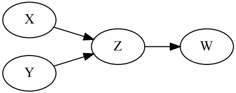



Simulated Example for $Y \leftarrow X \rightarrow Z$. Color represents X.

# Rule 3. Conditional Independence in Colliders.

- *If a variable $Z$ is the collision node between $X$ and $Y$, and there is only one path between $X$ and $Y$, then $X$ and $Y$ are unconditionally independent but are dependent conditional on $Z$ and any descendants of $Z$.*
- This is the source of the **collider variable bias** we mentioned earlier. This is why blindly conditioning or controlling all variables is not a good strategy.

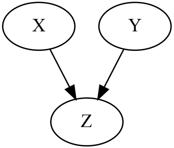



- For these 2 graphs, $X \perp\!\!\!\perp Y$ but $X \not\!\perp\!\!\!\perp Y|Z$ and $X \not\!\perp\!\!\!\perp Y|W$

# Rule 3. Conditional Independence in Colliders.

Proof:



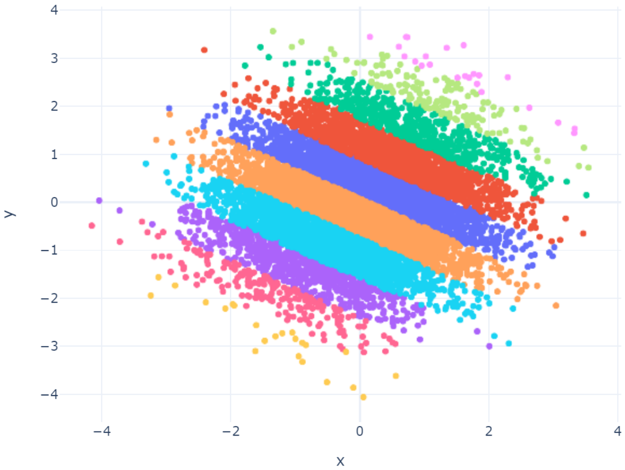

$$P(X, Y, Z) = P(X)P(Y)P(Z|X, Y) \qquad \text{(Factor Product)}$$

$$P(X, Y) = \sum_{z \in Z} P(X, Y, Z) \qquad \text{(Marginalize } Z\text{)}$$

$$P(X, Y) = P(X)P(Y) \sum_{z \in Z} P(Z|X, Y)$$

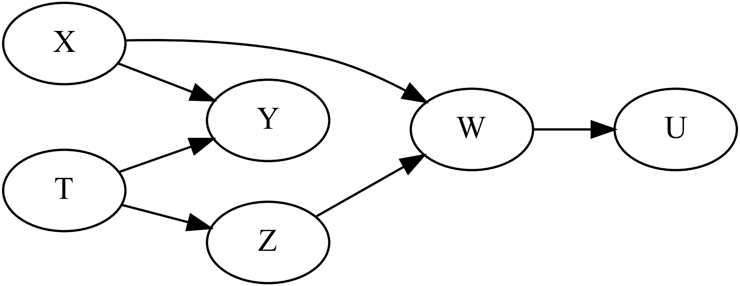
$$P(X, Y) = P(X)P(Y) \implies X \perp\!\!\!\perp Y$$

# Rule 3. Conditional Independence in Colliders.



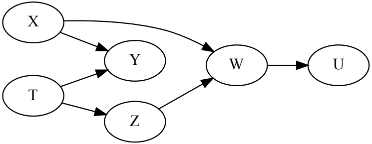

Simulated Example for $X \rightarrow Z \leftarrow Y$. Color represents Z.

# D-separation Exercise




That variables can I condition on to make $Z$ and $Y$ d-separated?

# D-separation Exercise




Possible Answers:

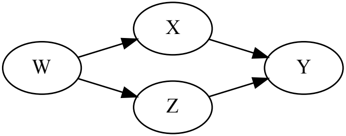
- {T}
- {X, T}
- {W, X, T}
- {U, X, T}

# Backdoor Criterion

# Backdoor Path Criterion

- Intuition: we can estimate the effect of $X$ on $Y$ by **blocking** (using d-separation rules) all non-causal pathways.
- More formally: A set of variables $Z$ is *sufficient* to control for confounding if:
  1. it blocks all backdoor paths from treatment to outcome
  2. it does not include any descendants of treatment
- The set of sufficient variables that satisfy the backdoor path criterion is not necessarily unique.

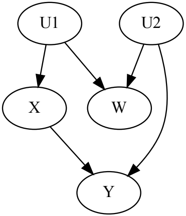
# Backdoor Path Criterion: Example




To estimate the causal effect of $X \to Y$, we can condition on:

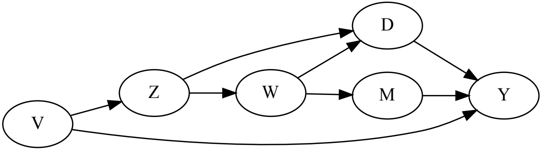
- {W}
- {Z}
- {W,Z}

# Backdoor Path Criterion: Example



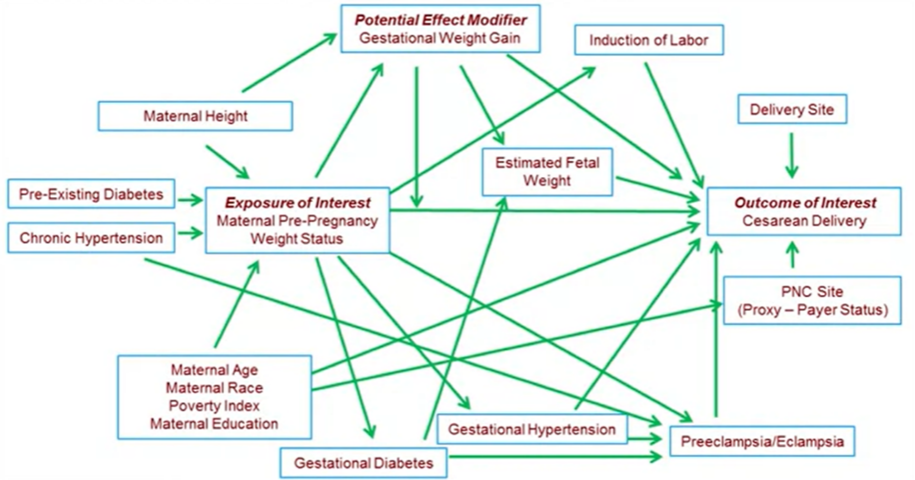

- Let $U1$ and $U2$ be unmeasured variables.
- In this case, we don't have to condition on anything. In fact, conditioning on $W$ will introduce bias.

# Backdoor Path Criterion: Example




- Backdoor paths:
  - block $D \leftarrow Z \leftarrow V \rightarrow Y$ with $Z$, $V$, or both
  - block $D \leftarrow W \leftarrow Z \leftarrow V \rightarrow Y$ with $W$, $Z$, $V$, or any combination of these
  - block $D \leftarrow W \rightarrow M \rightarrow Y$ with $W$, $M$, or both
- Thus, the sufficient sets to control for are: {W, Z}, {W, V}, {M, Z}, {M, V}, {W, Z, V}, {M, Z, V}, {W, M, Z}, {W, M, V}, {W, M, Z, V}

# Backdoor Path Criterion: Real-Life Example




Causal DAG to assess how pre-pregnancy weight affect the risk of cesarean delivery. (*Vahratian et al. 2005*)

# Relation to the Potential Outcome Model

# Complementary Value

- Use Causal DAGs for **identification analysis** – to determine which variables to control for to block the *backdoor paths*.
- Then use the estimation techniques from Potential Outcomes (matching, re-weighting, regression) to estimate ATE.

# There's more to Causal Graphs than this!

- We haven't discussed **do-notation** and **Structural Causal Models**.
- We'll cover these in future lectures. The d-separation rules are sufficient for the following lectures.

# References

# References


1. Hernan, Miquel A., and James M. Robins. Causal Inference. CRC Press, 2019.
2. Roy, Jason. "A Crash Course in Causality: Inferring Causal Effects from Observational Data — Coursera." Coursera, https://www.coursera.org/learn/crash-course-in-causality. Accessed 15 Aug. 2022.
3. Matheus, Facure. "Causal Inference for The Brave and True." Matheus Facure, https://matheusfacure.github.io/python-causality-handbook/landing-page.html. Accessed 15 Aug. 2022.
4. Morgan, Stephen L., and Christopher Winship. Counterfactuals and Causal Inference. Cambridge University Press, 2014.
5. Pearl, Judea, et al. Causal Inference in Statistics. John Wiley & Sons, 2016.
6. Yao, Liuyi, et al. "A Survey on Causal Inference." ACM Transactions on Knowledge Discovery from Data, no. 5, Association for Computing Machinery (ACM), Oct. 2021, pp. 1–46. Crossref, doi:10.1145/3444944.