

Stat 256: Causality

Causal Graphs and Causal Structural Models

Chad Hazlett

UCLA

`chazlett@ucla.edu`

Spring 2026

Resources for this material

Today we will just get our feet wet, but for the DAG material in general:

Required reading:

- *What if* (Hernan and Robins), Chapters 6, 7.
- *Primer* (Pearl, Glymour, Jewell), Chapters 1.4, 1.5, 2, 3.

Other suggestions:

- Pearl, J. (2009). *Causality*
- Some online materials:
 - *Causal Diagrams: Draw Your Assumptions Before Your Conclusions* course by Miguel Hernan on EdX
 - *A Crash Course in Causality: Inferring Causal Effects from Observational Data* course by Jason Roy on Coursera.
 - Richard McElreath lecture on “Science before statistics: causal inference”.
<https://www.youtube.com/watch?v=KNPYUVmY3NM&t=8565s>

Goals for this lecture

- Motivation: Why do we need causal structure to make sense of data?
- Introduction to structural causal models (SCMs) and their graphs
- Graph basics: Conditional independence and d -separation

Why we need causal structure

Suppose people self-select into taking a new drug. You find:

	Outcome: Recovery	
	Drug	No Drug
Mild	81 of 87 (93%)	234 of 270 (87%)
Severe	192 of 263 (73%)	55 of 80 (69%)
Combined	273 of 350 (78%)	289 of 350 (83%)

- Among mild cases, those on the drug did better (93% vs. 87%)
- Among severe cases, those on the drug did better (73% vs. 69%)
- But in the combined data, people did *worse* on the drug (78% vs. 83%)

Which comparison answers the causal question?

What's going on? Confounding

	Outcome: Recovery	
	Drug	No Drug
Mild	81 of 87 (93%)	234 of 270 (87%)
Severe	192 of 263 (73%)	55 of 80 (69%)
Combined	273 of 350 (78%)	289 of 350 (83%)

- Severe cases have lower recovery rates *and* are more likely to take drug.
- So the drug group is disproportionately severe cases, dragging down its average — classic confounding.

Knowing to condition on severity here required a causal judgment: that severity is a common cause of drug-taking and recovery.

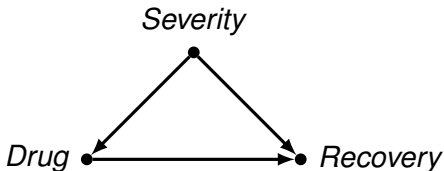
Thus knowing what to do – and the answer you get – depends on the analyst's beliefs about causal structure.

Encoding causal assumptions

In this example, our assumption is:

- 1 Severity influences both recovery and drug-taking, and
- 2 no other factor creates a spurious association between drug-taking and recovery.

We encode this with a DAG:



Preview: this graph shows us that conditioning on *Severity* removes the confounding, isolates the effect of *Drug* on *Recovery*.

Structural Causal Models

Let's now formalize what we mean by “causal structure.”

Structural Causal Models (SCMs) specify a set of causal relationships. They consist of:

- a group of observed variables
- a group of unobserved exogenous variables, $\{U\}$
- functions that relate them, $f_{(\cdot)}$
- an assumed distribution over the exogenous variables, $p(U)$.

E.g.,

$$X = f_x(U_x)$$

$$A = f_a(X, U_a)$$

$$Y = f_y(X, A, U_y)$$

$$U_y \perp\!\!\!\perp U_x, \quad U_y \perp\!\!\!\perp U_a$$

where “exogenous” ancestors, $U = \{U_x, U_a, U_y\}$ with $U \sim p(u)$.

Note that:

- $\{f_x, f_a, f_y\}$ make no parametric commitments.
- Each expression has a causal ("structural") meaning, not just algebraic.
- "X causes Y" means Y's function depends on X

SCMs impose a deterministic, "clockwork" notion of causation:

- Population: The distribution $p(U)$ implies a distribution $p(\text{observables})$.
- Individuals: If you knew all the values of U for a given unit and the details of every f , you would know the exact values of every variable.
- Modularity/invariance: each function f_j does not depend on how its inputs were produced. If you *intervene* and set $A = a$, the downstream functions still apply as usual. We will use this when we discuss interventions.

SCM to Graph

For a model M such as:

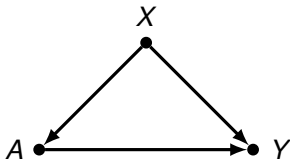
$$X = f_x(U_x)$$

$$A = f_a(X, U_a)$$

$$Y = f_y(X, A, U_y)$$

$$U_y \perp\!\!\!\perp U_x, \quad U_y \perp\!\!\!\perp U_a$$

It is convenient to draw a graph, G :

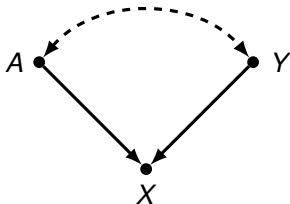


- Arrows go into a node only if the corresponding argument enters the corresponding function in the SCM.
- You could show U_j pointing to each j , but usually omit these.

Missing edges, bidirected arcs

A missing edge means the value of A is not an argument into Y . *It's the missing edges that matter for the arguments we will need to make.*

Some authors show unobserved common causes as a bidirected arc.



Terminology

- ① nodes, vertices: the variables that are written
- ② edges, arrows: the lines, may be “directed” or “bidirected” (arcs)
- ③ parents: for a given node, the causes feeding into it
- ④ root nodes: a node with no (endogenous) parents on the graph
- ⑤ descendants, children: effects of a given cause
- ⑥ paths: any series of connected nodes, regardless of direction
- ⑦ **Directed Acyclic Graph**: a graph where all edges are directed and you can never follow directed paths from a node back to itself.

Factorization

Markov condition: Each node listens only to its parents. For some node X , its parents (PA) and non-parent, non-descendants Q ,

$$p(X|PA, Q) = p(X|PA)$$

We'll often use this to see what should be conditionally independent,

$$X \perp\!\!\!\perp Q \mid PA$$

Further, this allows “Markov factorization” of the joint distribution:

Rule

Product Decomposition (Pearl & Verma 1991): for any model M whose graph G is acyclic and without dependent unobserved causes (i.e. no bidirected arcs; Markovian), the joint distribution of variables in the model is given by

$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i | PA_i)$$

where PA_i denotes the parents of X_i

Example of Product Decomposition

$$A \longrightarrow B \longrightarrow C$$

What is $P(A, B, C)$?

Without the graph, you could always write

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

With the graph we can write:

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

Paths, dependencies, and information

A first use of graphs is to get a handle on what observables will be conditionally or marginally associated with what others.

Three key structures: forks, chains, colliders

Chain:

$$A \longrightarrow B \longrightarrow C$$

Implications:

$A \perp\!\!\!\perp C$? **No** — A influences C through B .

$A \perp\!\!\!\perp C | B$? **Yes** — conditioning on B blocks the chain.

$B \perp\!\!\!\perp C | A$? **No** — B still directly causes C .

Paths, dependencies, and information

Fork (aka common cause):

$$A \longleftarrow B \longrightarrow C$$

Implications:

$A \perp\!\!\!\perp C$? **No** — the common cause B induces association.

$A \perp\!\!\!\perp C|B$? **Yes** — conditioning on the common cause blocks it.

$B \perp\!\!\!\perp C|A$? **No** — B still directly causes C .

Paths, dependencies, and information

Collider:

$$A \longrightarrow B \longleftarrow C$$

$A \perp\!\!\!\perp C$? **Yes** — no common cause, no causal path.

$A \perp\!\!\!\perp C | B$? **No!** — conditioning on the collider *opens* a path.

$B \perp\!\!\!\perp C | A$? **No** — C directly causes B .

Beware colliders!

Colliding is real

Pearl's favorite ($A \rightarrow B \leftarrow C$):

A: It rains

B: The lawn is wet (collider)

C: Sprinkler recently on

Hollywood success ($A \rightarrow B \leftarrow C$):

A: Good looks

B: Fame (collider)

C: Acting skills

Academic tenure ($A \rightarrow B \leftarrow C$):

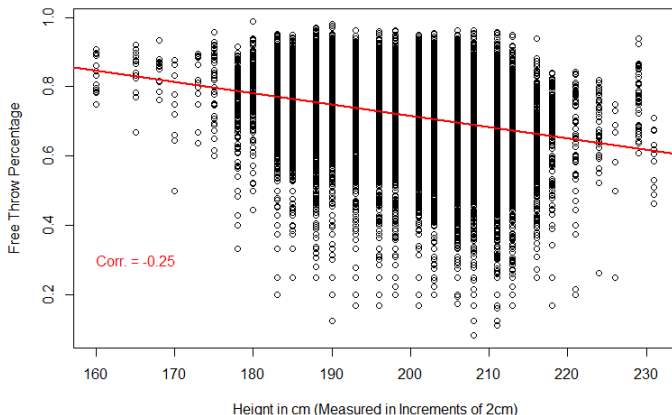
A: Productivity

B: Tenure (collider)

C: Originality¹

¹Thanks to Jason Roy for these examples

Free Throw Percentage vs Height
for All NBA Player-Seasons (1950-2017)



Why do you think that is? Answer with a DAG.

*Source: Adam Rohde, via <https://www.basketball-reference.com/> and <https://www.kaggle.com/datasets/drgilermo/nba-players-stats>.

d-Separation

You now also understand *d* (directional)-Separation, which just follows from these to formalize when we can claim conditional independence

Rule

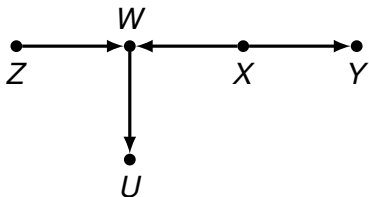
***d*-separation.**

A and C are d-separated by a set of nodes Z if and only if every path between A and C is blocked, where a path is blocked if:

- *it contains a chain ($\rightarrow B \rightarrow$) or fork ($\leftarrow B \rightarrow$) and $B \in Z$ (conditioning blocks it), or*
- *it contains a collider ($\rightarrow B \leftarrow$) and neither B nor any descendant of B is in Z (not conditioning keeps it closed).*

If A and C are *d*-separated by Z, then $A \perp\!\!\!\perp C \mid Z$ in any distribution generated by that graph.

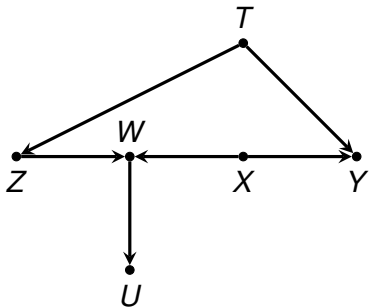
Practice



Looking at Z and Y :

- Are they d -separated conditional on $\{\}$?
- Are they d -separated conditional on W ?
- Are they d -separated conditional on $\{W, X\}$?
- Are they d -separated conditional on U ?

More practice



What sets d -separate Z from Y ?

Review so far

All of this is really to say that for two nodes of interest,

- they come to be associated due to (i) one causing another (directly or through a chain), (ii) a common cause; (iii) conditioning on a common consequence (collider)
- you can render them independent according to the rules of d -separation:
 - if they are independent already
 - if you block the common cause or intermediate causal path by conditioning
 - if you don't condition on collider (or descendant thereof)
- d -separation guarantees conditional independence; non- d -separation risks lack of independence.
- This will lead to more interesting things soon, I promise!