Structural Causal Models (SCMs)

EC 421, Set 11

Connor Lennon
Spring 2021

Prologue

Schedule

Last time

Rubin-Neyman Causal Model

Today

We're covering more complex causal relationships than we did last time - these are represented by what are known as Directed Acyclic Graphs (DAGs)

Upcoming

Instrumental Variables

SCMs

Causality is Complicated

Though we'd ideally be able to run experiments to identify causal effects, there are a number of ways experiments can fail. Just to name a few -

Pre-administration

- Where can we 'intervene'?
- Do we have access to a truly representative sample?
- Is anyone willing to undertake the treatment/control?
- How close are we to a laboratory sample?

Post-administration

- Placebo Effect
- Reverse causality
- Defiers and Always-Takers
- Mediators (ie. Is our experiment causing this or causing something else)?

Causality is Complicated

This means we have to think about

- experimental setting AND
- what causal system we are acting on

Causality is Complicated

This means we have to think about

- experimental setting AND
- what causal system we are acting on

Therefore it's useful to have a nice way of visualizing and understanding causal systems. Like **DAGs**.

- It's useful, however, to understand what a 'graph' actually is

Causality is Complicated

We can use the tools of formal graphs to organize our assumptions about the problem space into:

- Variables of interest
- Direct Causal Effects between them
- Potential exogeneity concerns

We can then find out...

Causality is Complicated

We can use the tools of formal graphs to organize our assumptions about the problem space into:

- Variables of interest
- Direct Causal Effects between them
- Potential exogeneity concerns

We can then find out...

- whether or not the problem we are interested in is tractable

Causality is Complicated

We can use the tools of formal graphs to organize our assumptions about the problem space into:

- Variables of interest
- Direct Causal Effects between them
- Potential exogeneity concerns

We can then find out…

- whether or not the problem we are interested in is tractable
- If it is tractable, what variables do we need to measure and control for to estimate causal effects

Causality is Complicated

We can use the tools of formal graphs to organize our assumptions about the problem space into:

- Variables of interest
- Direct Causal Effects between them
- Potential exogeneity concerns

We can then find out...

- whether or not the problem we are interested in is tractable
- If it is tractable, what variables do we need to measure and control for to estimate causal effects
- How existing experiments might get polluted by poor specification

Causality is Complicated

The graphs we are interested in for causal effects are

Causality is Complicated

The graphs we are interested in for causal effects are

- directed
- acyclic

We restrict our problem space to these mostly because it reduces the complexity of the math.

So let's learn about graphs 📊

Graphs

Graphs

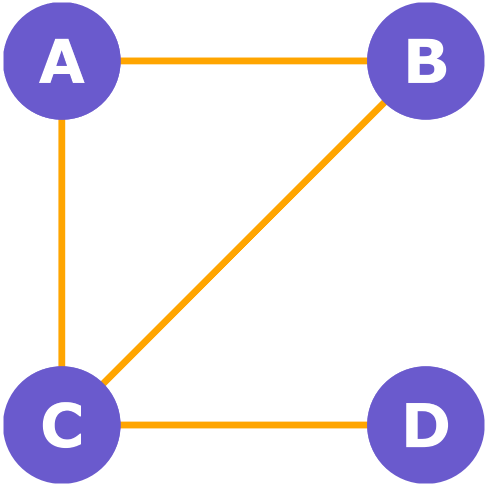
More formally

In graph theory, a graph is a collection of nodes connected by edges.

Graphs

More formally

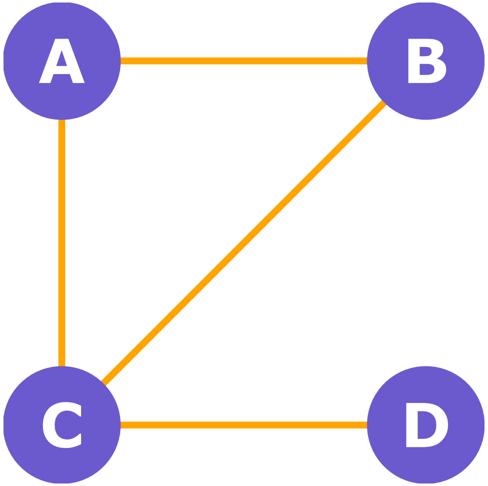
In graph theory, a graph is a collection of nodes connected by edges.



Graphs

More formally

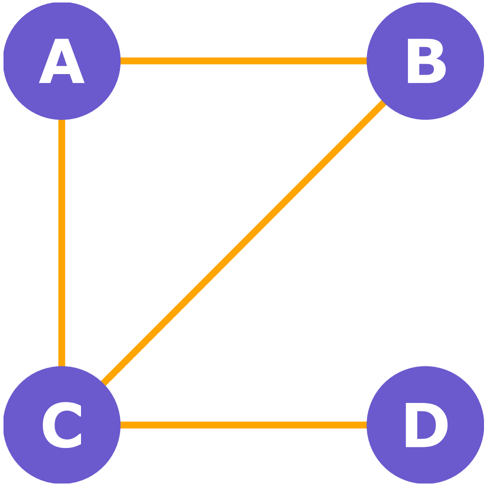
In graph theory, a graph is a collection of nodes connected by edges.



- Nodes connected by an edge are called adjacent.

More formally

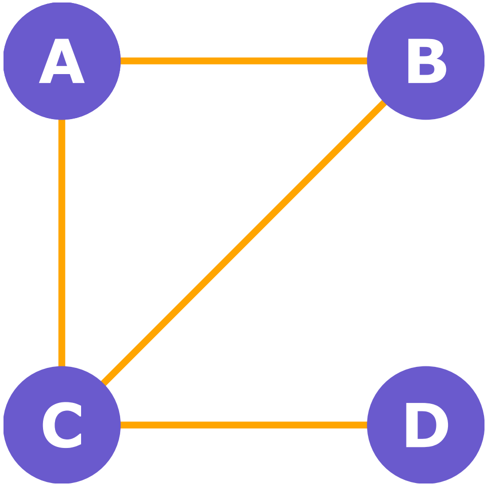
In graph theory, a graph is a collection of nodes connected by edges.



- Nodes connected by an edge are called adjacent.
- Paths run along adjacent nodes, e.g., $\mathbf{A} - \mathbf{B} - \mathbf{C}$.

More formally

In graph theory, a graph is a collection of nodes connected by edges.

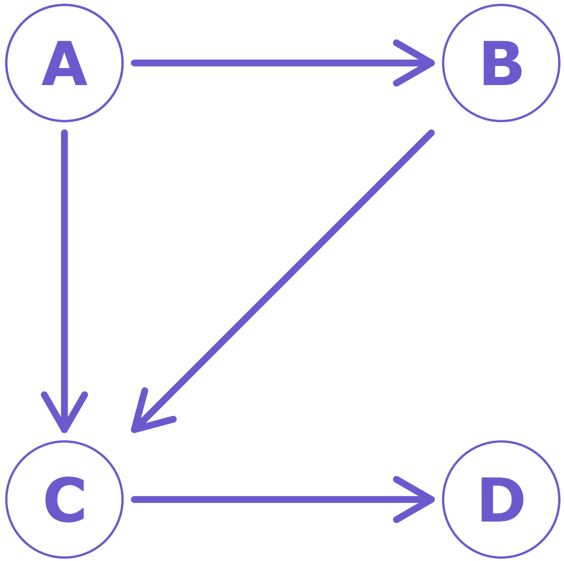


- Nodes connected by an edge are called adjacent.
- Paths run along adjacent nodes, e.g., $\mathbf{A} - \mathbf{B} - \mathbf{C}$.
- The graph above is undirected, since the edges don't have direction. We can give our graphs more information by defining direction for the edges

Graphs

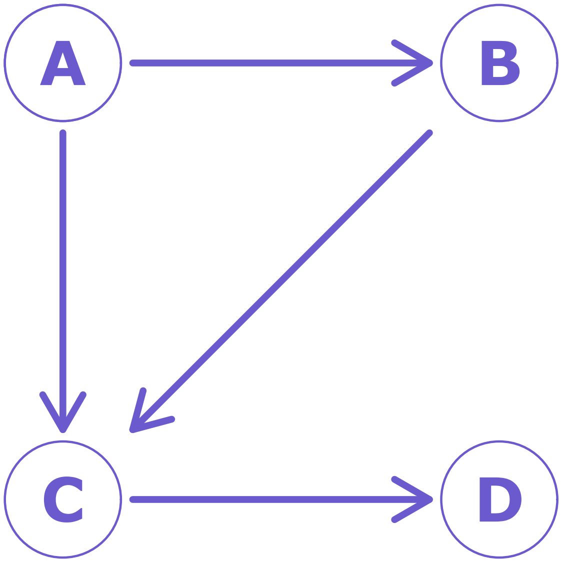
Directed

Directed graphs have edges with direction.



Directed

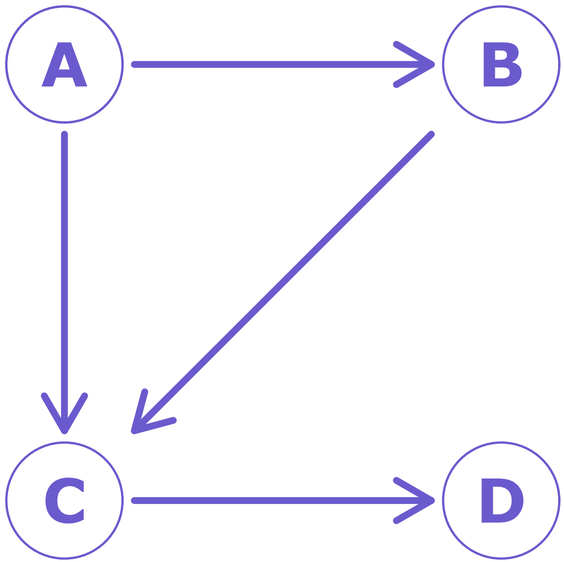
Directed graphs have edges with direction.



- Directed paths follow edges' directions, *e.g.*, $\mathbf{A} \rightarrow \mathbf{B} \rightarrow \mathbf{C}$.

Directed

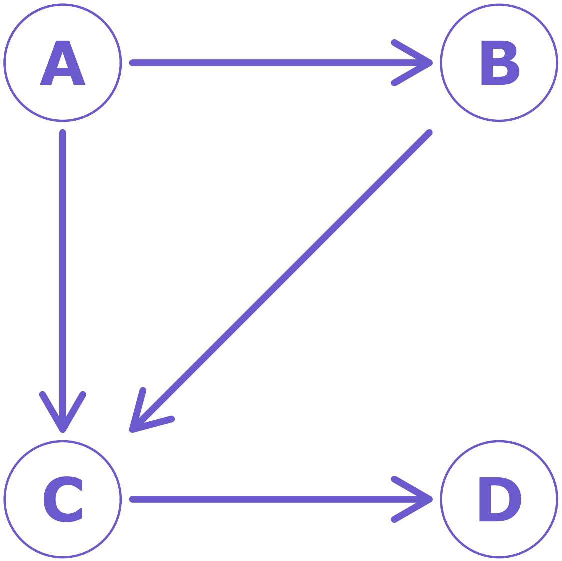
Directed graphs have edges with direction.



- Directed paths follow edges' directions, *e.g.*, $\mathbf{A} \to \mathbf{B} \to \mathbf{C}$.
- Nodes that precede a given node in a directed path are its ancestors.

Directed

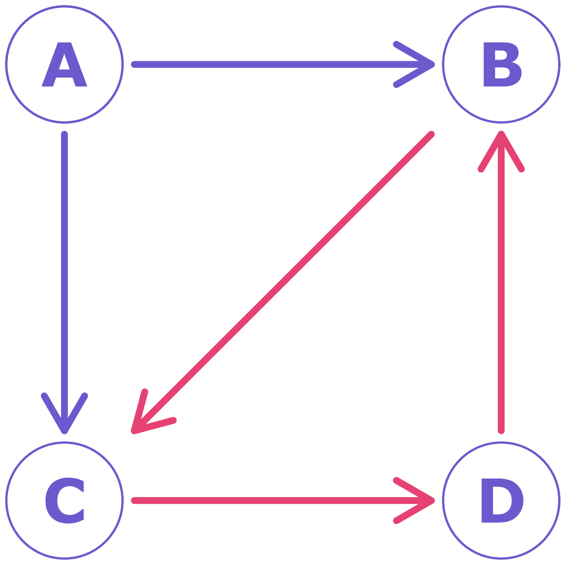
Directed graphs have edges with direction.



- Directed paths follow edges' directions, *e.g.*, $\mathbf{A} \rightarrow \mathbf{B} \rightarrow \mathbf{C}$.
- Nodes that precede a given node in a directed path are its ancestors.
- The opposite: descendants come after the node, *e.g.*, $\mathbf{D} = \mathbf{de}(\mathbf{C})$.

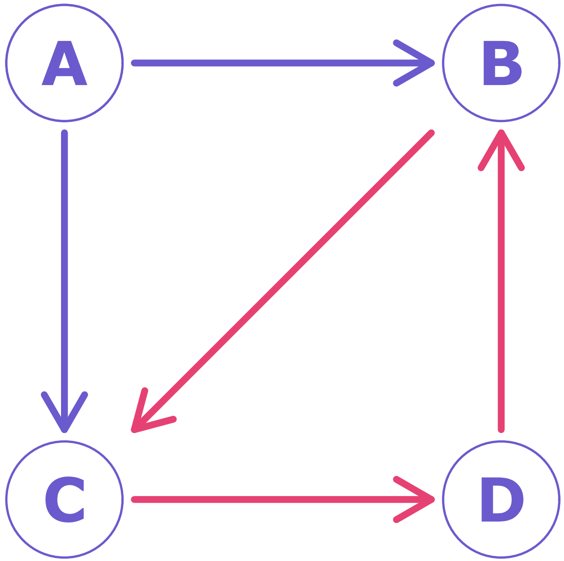
Cycles

If a node is its own descendant (*e.g.*, $\mathbf{de}(\mathrm{D}) = \mathrm{D}$), your graph has a cycle.



Cycles

If a node is its own descendant (*e.g.*, $\mathbf{de}(\mathrm{D}) = \mathrm{D}$), your graph has a cycle.



If your directed graph does not have any cycles, then you have a directed acyclic graph (DAG).

DAGs

What's a DAG?

DAG stands for directed acyclic graph.

DAGs

What's a DAG?

DAG stands for directed acyclic graph.

Formally, any graph is a collection of nodes and edges defined by their connections.

DAGs

What's a DAG?

DAG stands for directed acyclic graph.

Formally, any graph is a collection of nodes and edges defined by their connections.

DAGs are special graphs that help us understand causality in much the same way as microeconomic models help us understand markets

DAGs

What's a DAG?

DAG stands for directed acyclic graph.

More helpful...

A DAG is a graph that features a collection of nodes and **directed** edges

DAGs

What's a DAG?

DAG stands for directed acyclic graph.

More helpful...

A DAG is a graph that features a collection of nodes and **directed** edges

The graph illustrates and differentiates the causal associations and non-causal associations within a network of "random" variables.

Why a DAG?

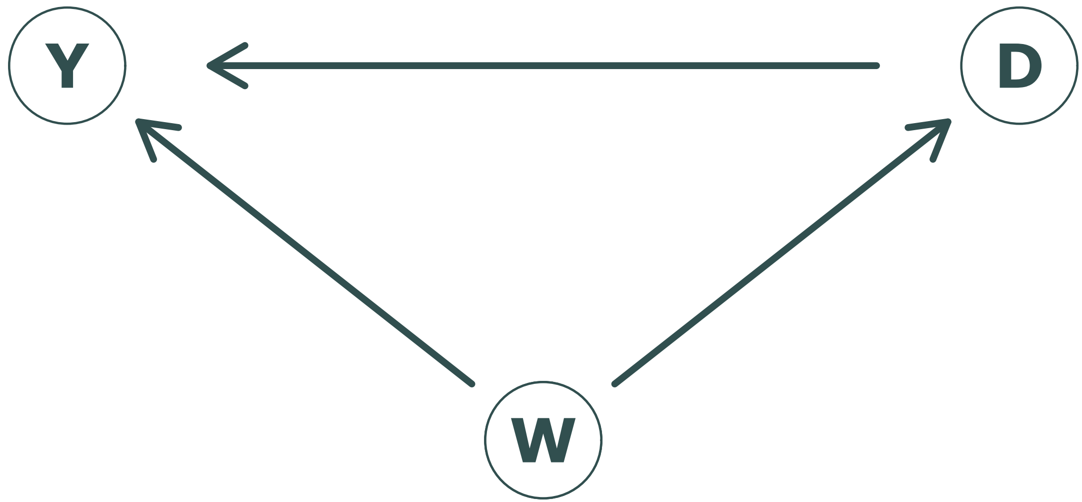
Following a fairly simple set of rules, DAGs allow researchers to visualize complex systems.

It separates -

- causal 'associations'
- noncausal 'associations'

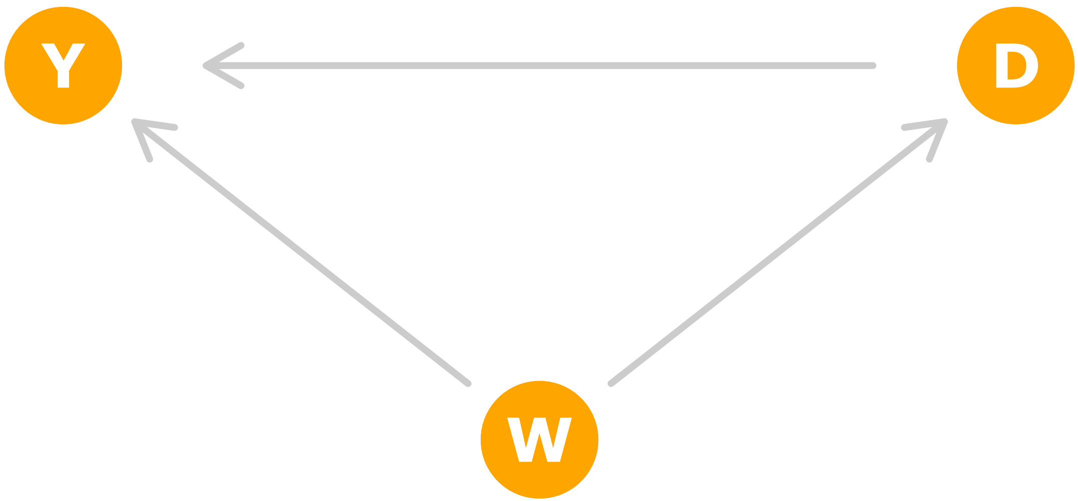
  in any assumption space without cycles

Example Omitted-variable bias in a DAG



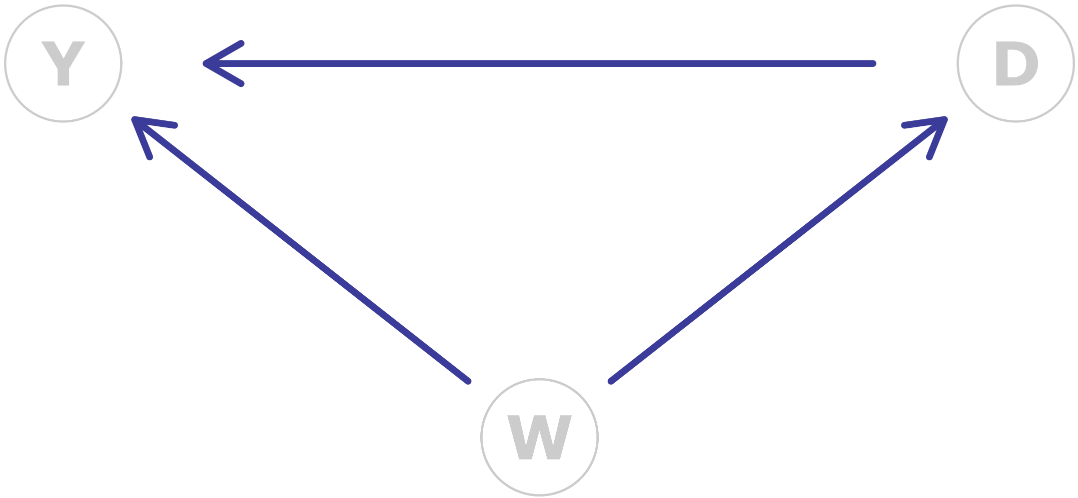
A pretty standard DAG.

Example Omitted-variable bias in a DAG



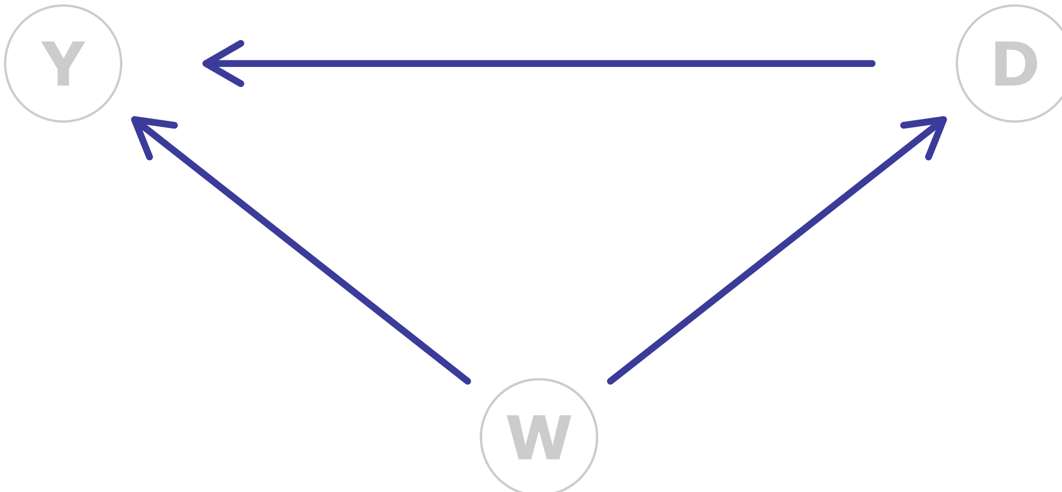
Nodes are random variables.

Example Omitted-variable bias in a DAG



Edges depict causal links. Causality flows in the direction of the arrows.

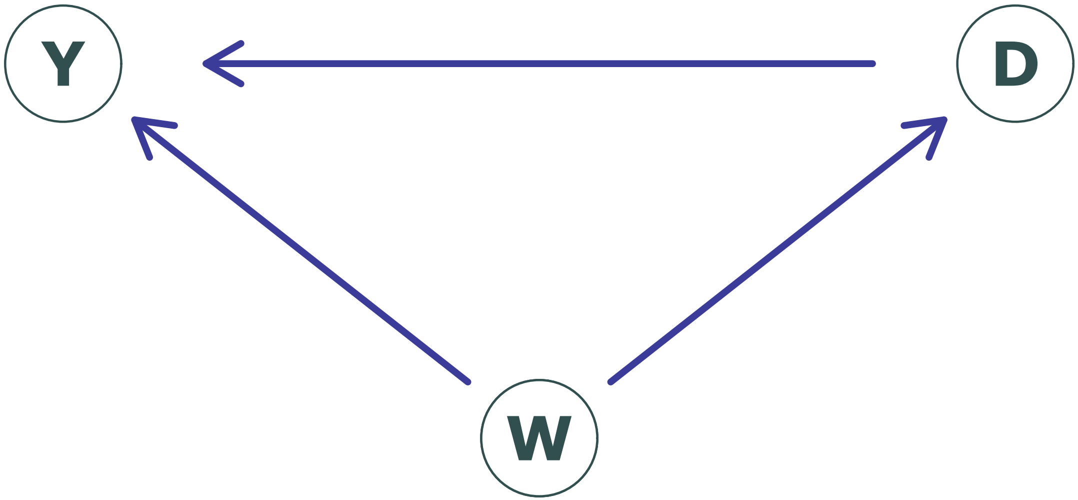
Example Omitted-variable bias in a DAG



Edges depict causal links. Causality flows in the direction of the arrows.

- Connections are drawn between nodes that **directly** cause one another
- Direction matters (for causality).
- Non-connections also (sometimes) matter! (More on this topic soon.)

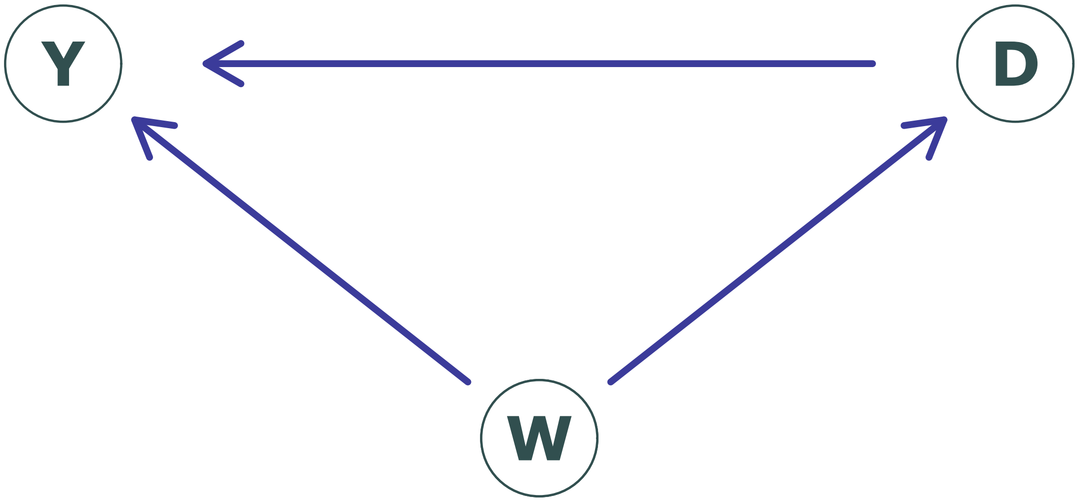
Example Omitted-variable bias in a DAG



Here we can see that Y is affected by both D and W.

W also affects D.

Example Omitted-variable bias in a DAG

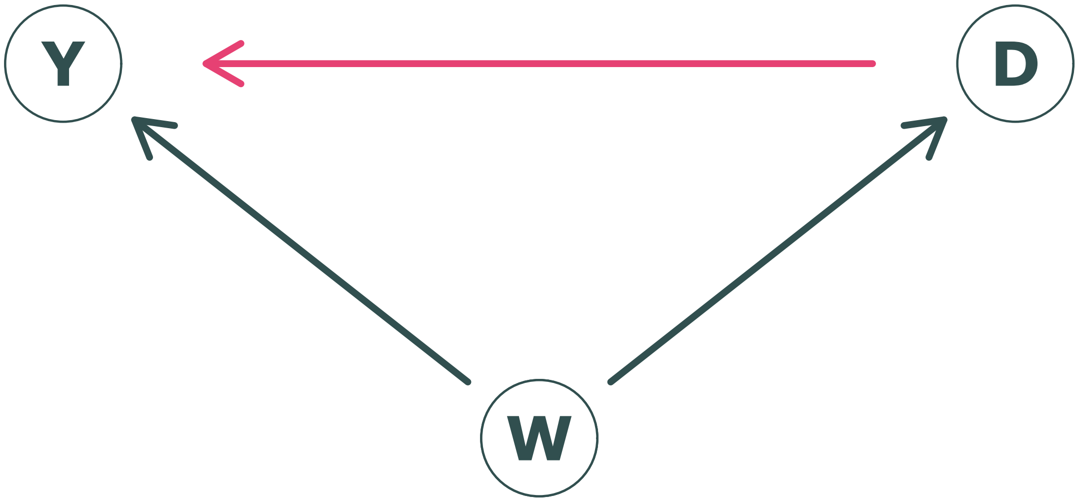


Here we can see that Y is affected by both D and W.

W also affects D.

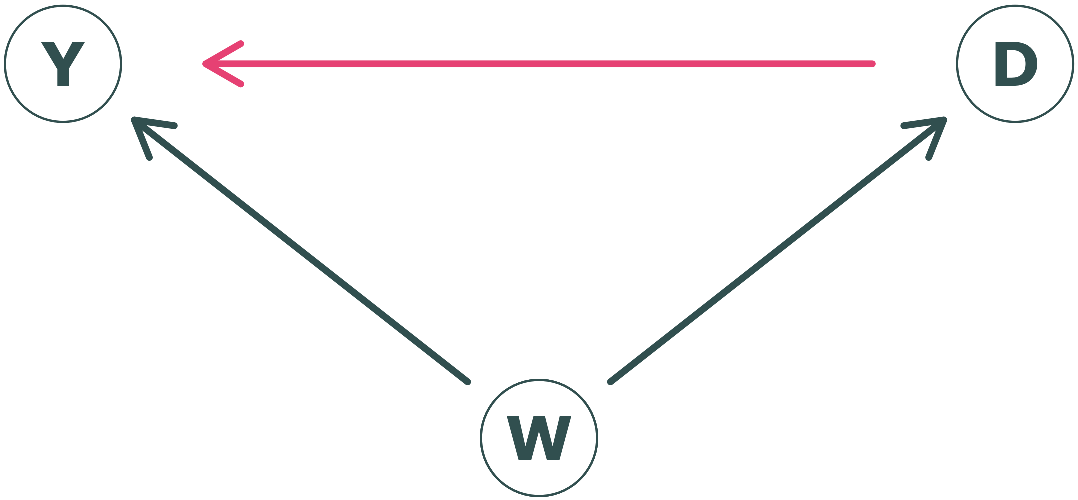
Q How does this graph exhibit OVB?

Example Omitted-variable bias in a DAG



There are two pathways from D to Y.

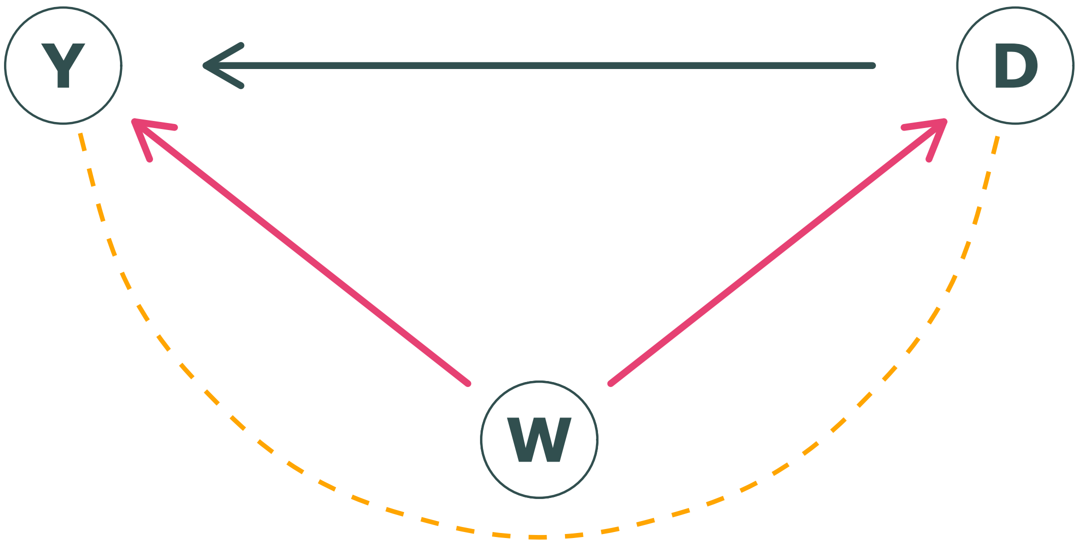
Example Omitted-variable bias in a DAG



There are two pathways from D to Y.

1. The path from D to Y $(\mathbf{D} \rightarrow \mathbf{Y})$ is our casual relationship of interest.

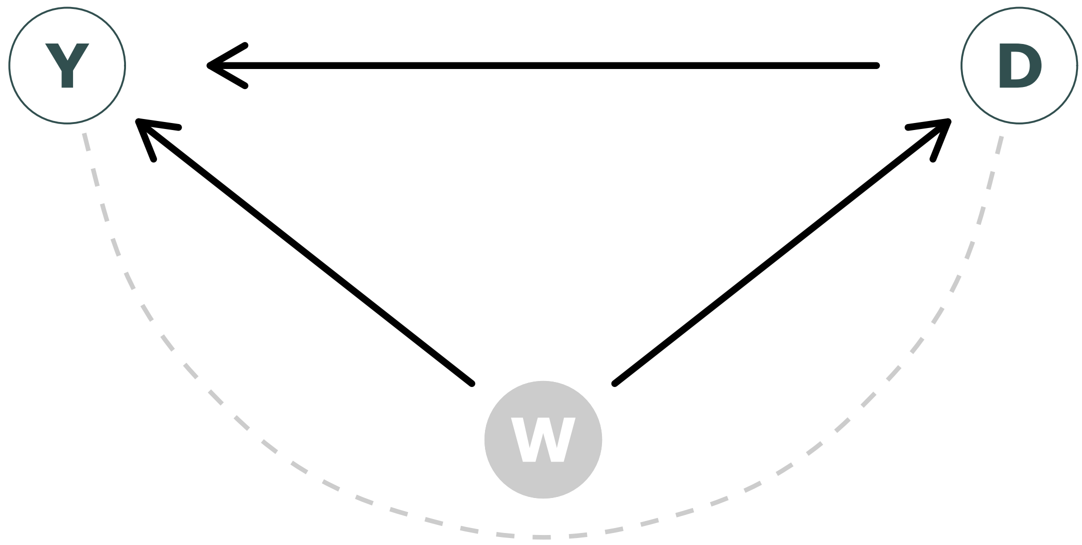
Example Omitted-variable bias in a DAG



There are two pathways from D to Y.

1. The path from D to Y $(\mathbf{D} \rightarrow \mathbf{Y})$ is our casual relationship of interest.
2. The path $(\mathbf{Y} \leftarrow \mathbf{W} \rightarrow \mathbf{D})$ creates a non-causal association btn D and Y.

Example Omitted-variable bias in a DAG



There are two pathways from D to Y.

1. The path from D to Y ($\mathbf{D} \rightarrow \mathbf{Y}$) is our casual relationship of interest.
2. The path ($\mathbf{Y} \leftarrow \mathbf{W} \rightarrow \mathbf{D}$) creates a non-causal association btn D and Y.

To shut down this pathway creating a non-causal association, we must condition on W. Sound familiar?

DAGs

DAGs

The origin story

Many developments in causal graphical models came from work in probabilistic graphical models—especially Bayesian networks.

The origin story

Many developments in causal graphical models came from work in probabilistic graphical models—especially Bayesian networks.

Using bayes rule, we can decompose any joint probability distribution into a product of conditional distributions

The origin story

Many developments in causal graphical models came from work in probabilistic graphical models—especially Bayesian networks.

Using bayes rule, we can decompose any joint probability distribution into a product of conditional distributions

I won't require you guys to learn the formal math behind Dags, though it is totally doable.

If you go to the end of the lecture, there will be an extensive section covering the theory/proofs.

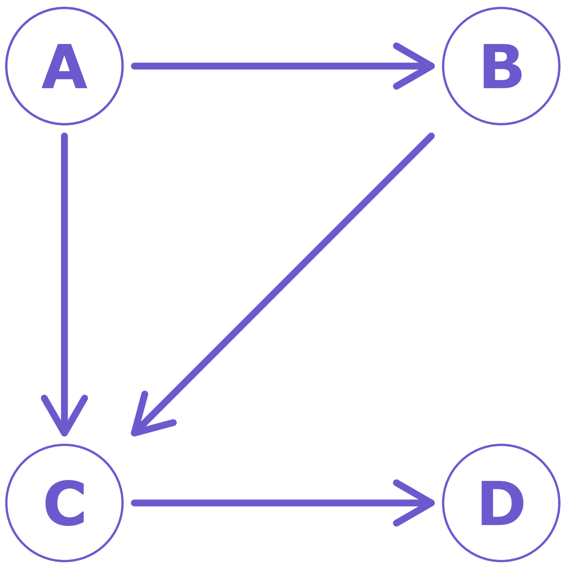
Thinking locally

DAGs help us think through how we can break down complex causal systems of many variables - simplifying $P(x_k|x_{k-1}, x_{k-2}, \ldots, x_1)$.

DAGs

Thinking locally

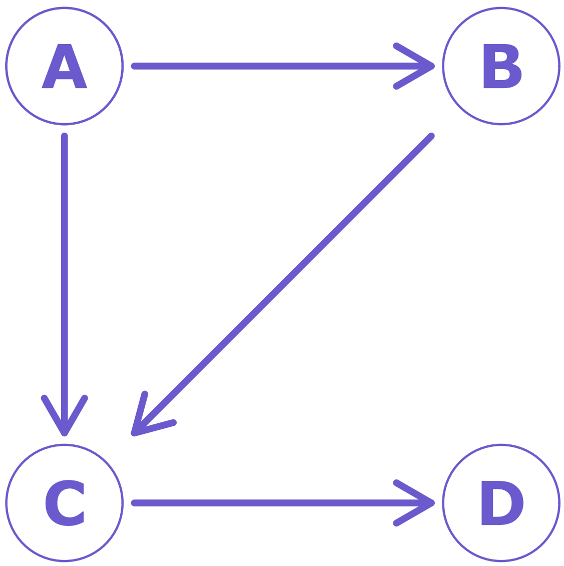
DAGs help us think through how we can break down complex causal systems of many variables - simplifying $P(x_k|x_{k-1}, x_{k-2}, \ldots, x_1)$.



Given a prob. dist. and a DAG, can we assume some independencies?

Thinking locally

DAGs help us think through how we can break down complex causal systems of many variables - simplifying $P(x_k|x_{k-1}, x_{k-2}, \ldots, x_1)$.



Given a prob. dist. and a DAG, can we assume some independencies?

Given C, is it reasonable to assume D is independent of A and B?

DAGs

Local Markov

This intuitive approach *is* the Local Markov Assumption

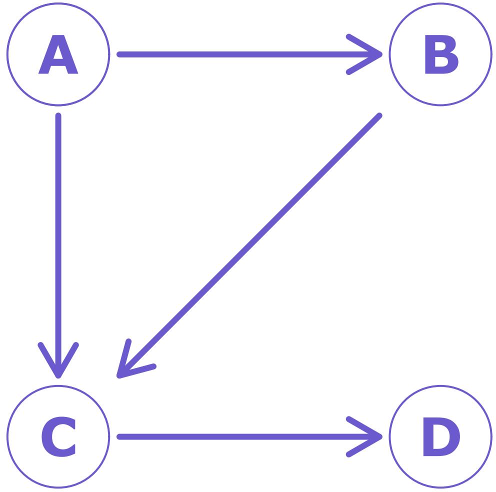
> Given its parents in the DAG, a node X is independent of all of its non-descendants.

Local Markov

This intuitive approach *is* the Local Markov Assumption

> Given its parents in the DAG, a node X is independent of all of its non-descendants.

Ex. Consider the DAG to the right:



With the Local Markov Assumption, $P(\mathrm{D}|\mathrm{A}, \mathrm{B}, \mathrm{C})$ simplifies to $P(\mathrm{D}|\mathrm{C})$.

Conditional on its parent (C), D is independent of A and B.

Independence

What have we learned so far? (Why should you care about this stuff?)

Local Markov tells us abound independencies within a probability distribution implied by the given DAG.

You're now able to say something about which variables are independent.

Independence

What have we learned so far? (Why should you care about this stuff?)

Local Markov tells us abound independencies within a probability distribution implied by the given DAG.

You're now able to say something about which variables are independent.

There's more: Great start, but there's more to life than independence.
We also want to say something about dependence.

DAGs

Dependence

We need to strengthen our Local Markov assumption to be able to interpret adjacent nodes as dependent. (*I.e.*, add it to our small set of assumptions.)

DAGs

Dependence

We need to strengthen our Local Markov assumption to be able to interpret adjacent nodes as dependent. (*I.e.*, add it to our small set of assumptions.)

The Minimality Assumption†

> 1. Local Markov Given its parents in the DAG, a node X is independent of all of its non-descendants.
> 2. (NEW) Adjacent nodes in the DAG are dependent.

† The name minimality refers to the minimal set of independencies for P and G—we cannot remove any more edges from the graph (while staying Markov compatible with G).

Dependence

We need to strengthen our Local Markov assumption to be able to interpret adjacent nodes as dependent. (*I.e.*, add it to our small set of assumptions.)

The Minimality Assumption†

> 1. Local Markov Given its parents in the DAG, a node X is independent of all of its non-descendants.
> 2. (NEW) Adjacent nodes in the DAG are dependent.

With the minimality assumption, we can learn both dependence and independence from connections (or non-connections) in a DAG.

† The name minimality refers to the minimal set of independencies for P and G—we cannot remove any more edges from the graph (while staying Markov compatible with G).

Causality

We need one last assumption move DAGs from statistical to causal models.

Causality

We need one last assumption move DAGs from statistical to causal models.

Strict Causal Edges Assumption

> Every parent is a direct cause of each of its children.

DAGs

Causality

We need one last assumption move DAGs from statistical to causal models.

Strict Causal Edges Assumption

> Every parent is a direct cause of each of its children. For Y, the set of direct causes is the set of variables to which Y responds.

DAGs

Causality

We need one last assumption move DAGs from statistical to causal models.

Strict Causal Edges Assumption

> Every parent is a direct cause of each of its children. For Y, the set of direct causes is the set of variables to which Y responds.

This assumption actually strengthens the second part of Minimality:

> 2. Adjacent nodes in the DAG are dependent.

Assumptions

Thus, we only need two assumptions to turn DAGs into causal models:

1. Local Markov Given its parents in the DAG, a node X is independent of all of its non-descendants.

2. Strict Causal Edges Every parent is a direct cause of each of its children.

Assumptions

Thus, we only need two assumptions to turn DAGs into causal models:

1. Local Markov Given its parents in the DAG, a node X is independent of all of its non-descendants.

2. Strict Causal Edges Every parent is a direct cause of each of its children.

Not bad, right?

DAGs

Flows

Brady Neal emphasizes the flow(s) of association and causation in DAGs, and I find it to be a super helpful way to think about these models.

Flow of association refers to whether two nodes are associated (statistically dependent) or not (statistically independent).

We will be interested in unconditional and conditional associations.

Building blocks

We will run through a few simple building blocks (DAGs) that make up more complex DAGs.

For each simple DAG, we want to ask a few questions:

1. Which nodes are unconditionally or conditionally independent?†
2. Which nodes are dependent?
3. What is the intuition?

† To prove A and B are conditionally independent, we can show $P(\mathrm{A}, \mathrm{B}|\mathrm{C})$ factorizes as $P(\mathrm{A}|\mathrm{C})P(\mathrm{B}|\mathrm{C})$.

Building block 1: Two unconnected nodes

Building block 1: Two unconnected nodes

Intuition:

Building block 1: Two unconnected nodes

Intuition: $\mathbf{A}$ and $\mathbf{B}$ appear independent—no link between the nodes.

Building block 1: Two unconnected nodes

Intuition: $\mathbf{A}$ and $\mathbf{B}$ appear independent—no link between the nodes.

Proof:

Building block 1: Two unconnected nodes

Intuition: A and B appear independent—no link between the nodes.

Proof: By Bayesian network factorization,

$$P(\mathrm{A}, \mathrm{B}) = P(\mathrm{A})P(\mathrm{B})$$

(since neither node has parents). ✓

Building block 2: Two connected nodes

Building block 2: Two connected nodes

Intuition:

Building block 2: Two connected nodes

Intuition: $\mathbf{A}$ "is a cause of" $\mathbf{B}$: there is clear (causal) dependence.

Building block 2: Two connected nodes

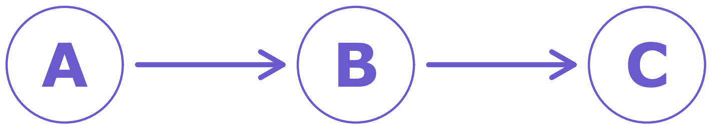
Intuition: $\mathbf{A}$ "is a cause of" $\mathbf{B}$: there is clear (causal) dependence. Proof:

Building block 2: Two connected nodes

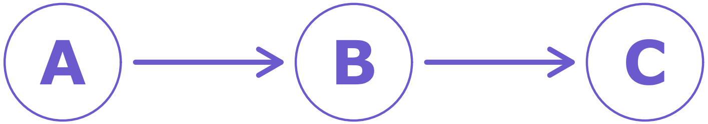
$$\begin{tikzcd} \mathbf{A} \arrow[r] & \mathbf{B} \end{tikzcd}$$

Intuition: $\mathbf{A}$ "is a cause of" $\mathbf{B}$: there is clear (causal) dependence.Proof: By the Strict Causal Edges Assumption, every parent (here, $\mathbf{A}$) is a direct cause of each of its children $(\mathbf{B})$. ✓

Building block 3: Chains



Building block 3: Chains

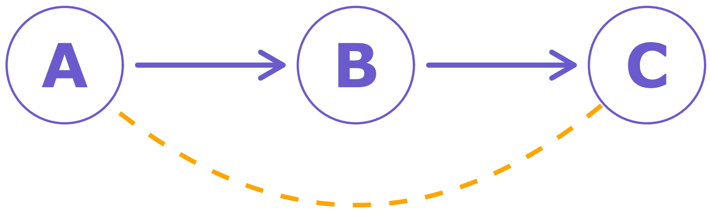


Intuition: We already showed two connected nodes are dependent:

- $\mathbf{A}$ and $\mathbf{B}$ are dependent.
- $\mathbf{B}$ and $\mathbf{C}$ are dependent.

The question is whether $\mathbf{A}$ and $\mathbf{C}$ are dependent:
Does association flow from $\mathbf{A}$ to $\mathbf{C}$ through $\mathbf{B}$?

Building block 3: Chains



Intuition: We already showed two connected nodes are dependent:

- A and B are dependent.
- B and C are dependent.

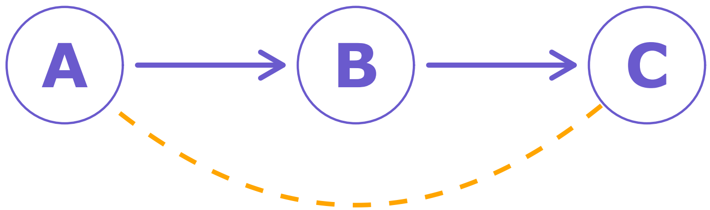
The question is whether A and C are dependent:

Does association flow from A to C through B?

The answer generally[†] is "yes": changes in A typically cause changes in C.

† Section 2.2 of Pearl, Glymour, and Jewell provides a "pathological" example of "intransitive dependence". It's basically when A induces variation in B that is not relevant to C' outcome.

Building block 3: Chains

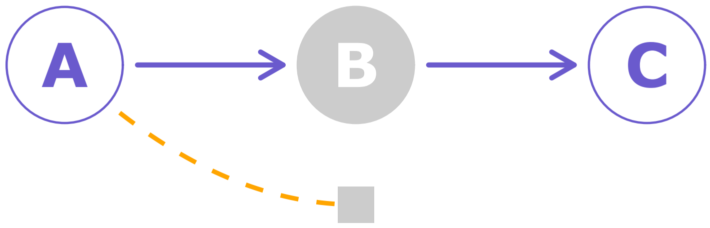


Proof: Here's the unsatisfying part.

Without more assumptions, we can't *prove* this association of $\mathbf{A}$ and $\mathbf{C}$.

We'll think of this as a potential (even likely) association.

Building block 3: Chains with conditions



Q How does conditioning on $\mathbf{B}$ affect the association between $\mathbf{A}$ and $\mathbf{C}$?

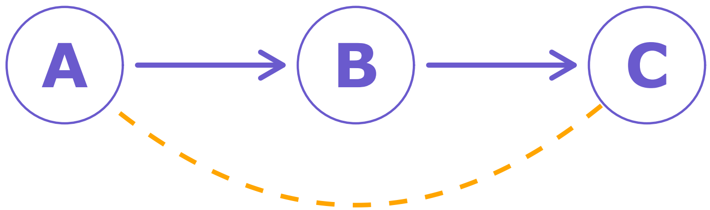
Intuition:

1. $\mathbf{A}$ affects $\mathbf{C}$ by changing $\mathbf{B}$.
2. When we hold $\mathbf{B}$ constant, $\mathbf{A}$ cannot "reach" $\mathbf{C}$.

We've blocked the path of association between $\mathbf{A}$ and $\mathbf{C}$.

Conditioning blocks the flow of association in chains. ("Good" control!)

Building block 3: Chains

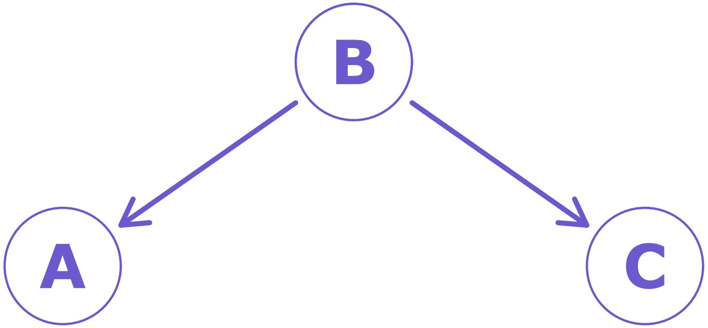


Note This association of $\mathbf{A}$ and $\mathbf{C}$ is not directional. (It is symmetric.)

On the other hand, causation is directional (and asymmetric).

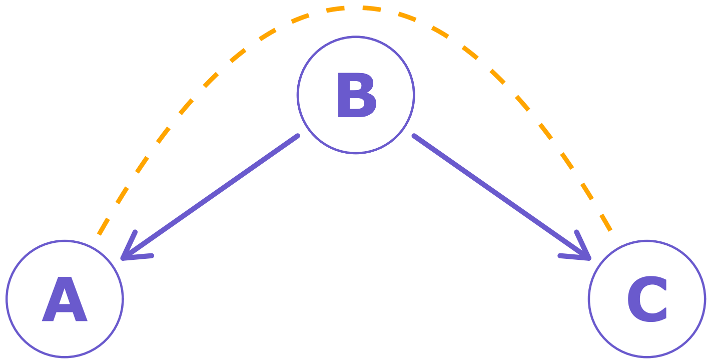
As you've been warned for years: Associations are not necessarily causal.

Building block 4: Forks



Forks are another very common structure in DAGs: $\mathbf{A} \leftarrow \mathbf{B} \rightarrow \mathbf{C}$.

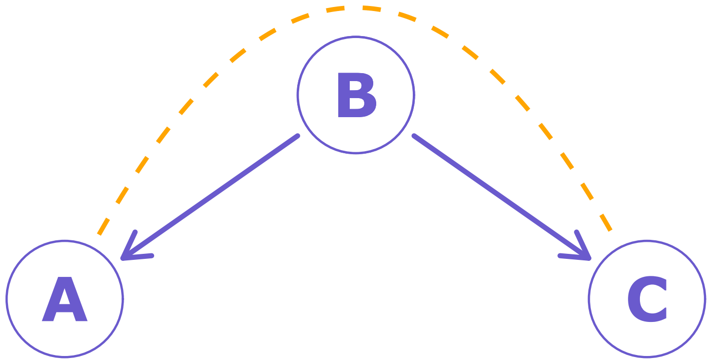
Building block 4: Forks



$\mathbf{A}$ and $\mathbf{C}$ are *usually* associated in forks. (As with chains.)

This chain of association follows the path $\mathbf{A} \leftarrow \mathbf{B} \rightarrow \mathbf{C}$.

Building block 4: Forks

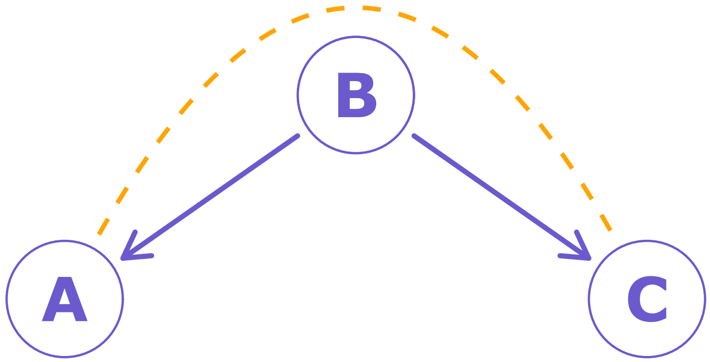


$\mathbf{A}$ and $\mathbf{C}$ are *usually* associated in forks. (As with chains.)

This chain of association follows the path $\mathbf{A} \leftarrow \mathbf{B} \rightarrow \mathbf{C}$.

Intuition:

Building block 4: Forks

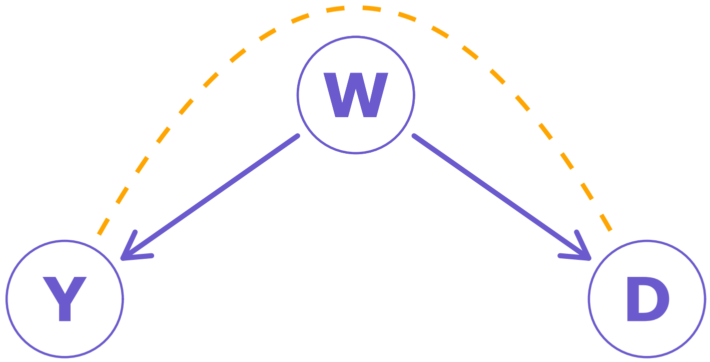


$\mathbf{A}$ and $\mathbf{C}$ are *usually* associated in forks. (As with chains.)

This chain of association follows the path $\mathbf{A} \leftarrow \mathbf{B} \rightarrow \mathbf{C}$.

Intuition: $\mathbf{B}$ induces changes in $\mathbf{A}$ and $\mathbf{B}$. An observer will see $\mathbf{A}$ change when $\mathbf{C}$ also changes—they are associated due to their common cause.

Building block 4: Forks

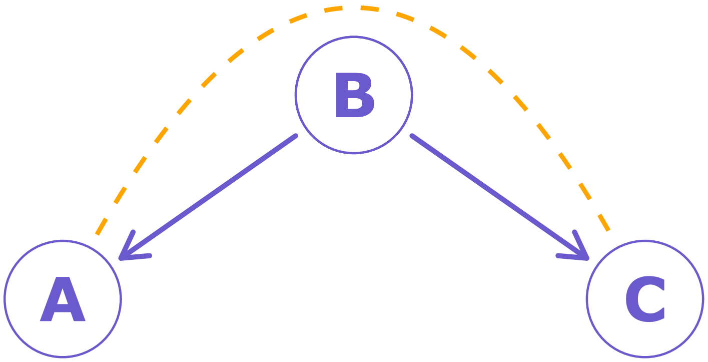


Another way to think about forks:

OVB when a treatment $\mathbf{D}$ does not affect the outcome $\mathbf{Y}$.

Without controlling for $\mathbf{W}$, $\mathbf{Y}$ and $\mathbf{D}$ are (usually) non-causally associated.

Building block 4: Forks

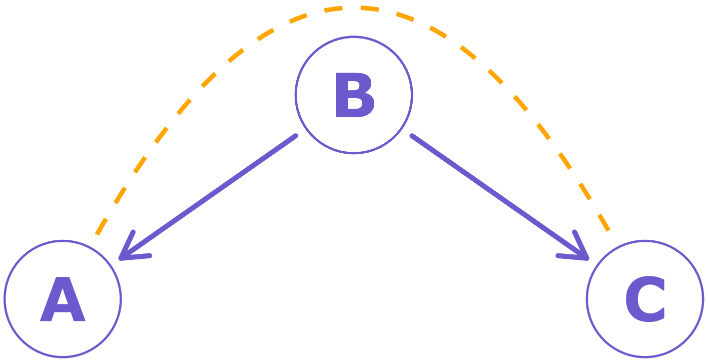


$\mathbf{A}$ and $\mathbf{C}$ are *usually* associated in forks. (As with chains.)

This chain of association follows the path $\mathbf{A} \leftarrow \mathbf{B} \rightarrow \mathbf{C}$.

Proof:

Building block 4: Forks

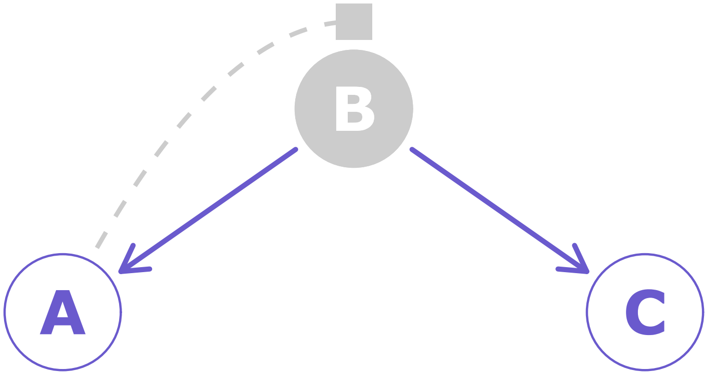


$\mathbf{A}$ and $\mathbf{C}$ are *usually* associated in forks. (As with chains.)

This chain of association follows the path $\mathbf{A} \leftarrow \mathbf{B} \rightarrow \mathbf{C}$.

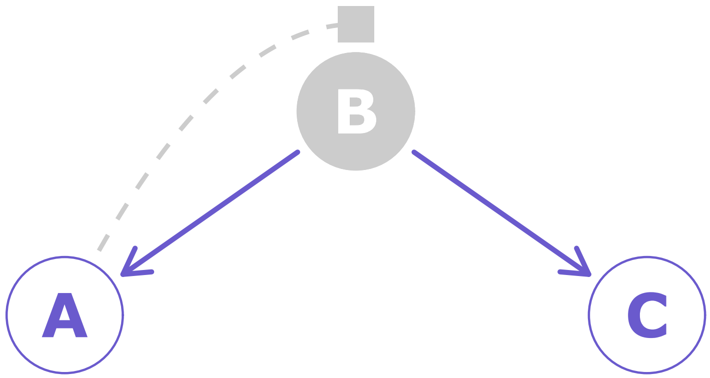
Proof: Same problem as chains: We can't show $\mathbf{A}$ and $\mathbf{C}$ are independent, so we assume they're likely (potentially?) dependent.

Building block 4: Blocked forks



Conditioning on $\mathbf{B}$ makes $\mathbf{A}$ and $\mathbf{C}$

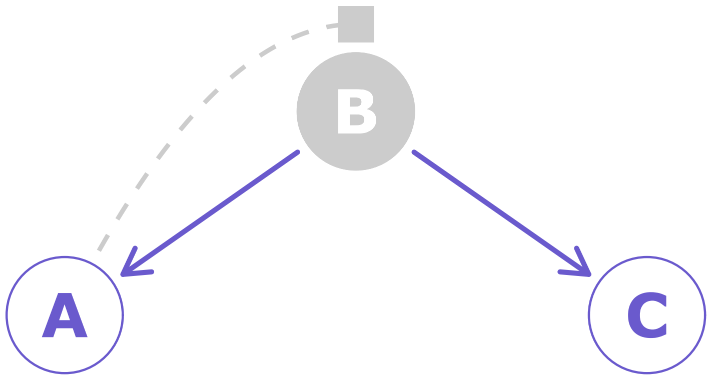
Building block 4: Blocked forks



Conditioning on $\mathbf{B}$ makes $\mathbf{A}$ and $\mathbf{C}$ independent. (As with chains.)

Intuition:

Building block 4: Blocked forks

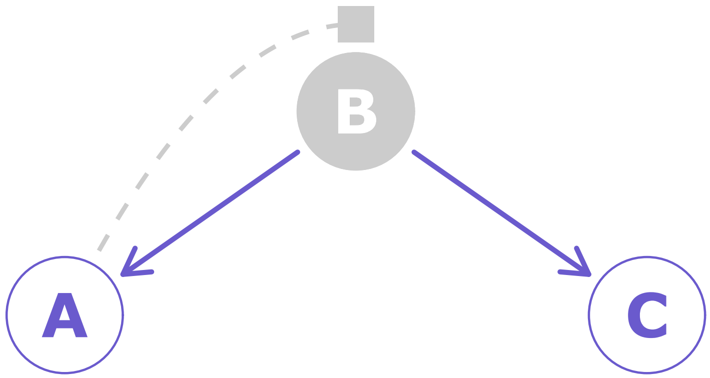


Conditioning on $\mathbf{B}$ makes $\mathbf{A}$ and $\mathbf{C}$ independent. (As with chains.)

Intuition: $\mathbf{A}$ and $\mathbf{C}$ are only associated due to their common cause $\mathbf{B}$.

When we shutdown (hold constant) this common cause $(\mathbf{B})$,
there is way for $\mathbf{A}$ and $\mathbf{C}$ to associate.

Building block 4: Blocked forks



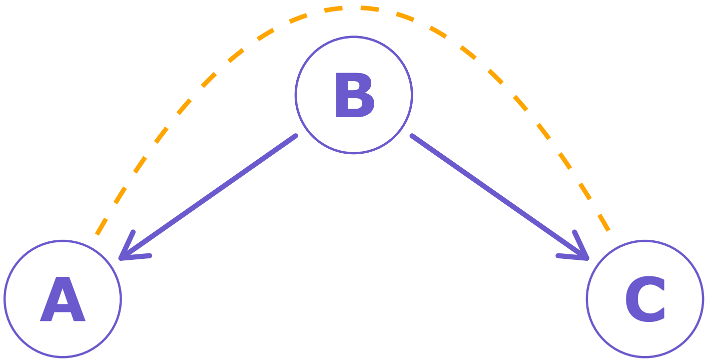
Conditioning on $\mathbf{B}$ makes $\mathbf{A}$ and $\mathbf{C}$ independent. (As with chains.)

Intuition: $\mathbf{A}$ and $\mathbf{C}$ are only associated due to their common cause $\mathbf{B}$.

When we shutdown (hold constant) this common cause $(\mathbf{B})$,
there is way for $\mathbf{A}$ and $\mathbf{C}$ to associate.

Also: Think about Local Markov. Or think about OVB.

Building block 4: Forks

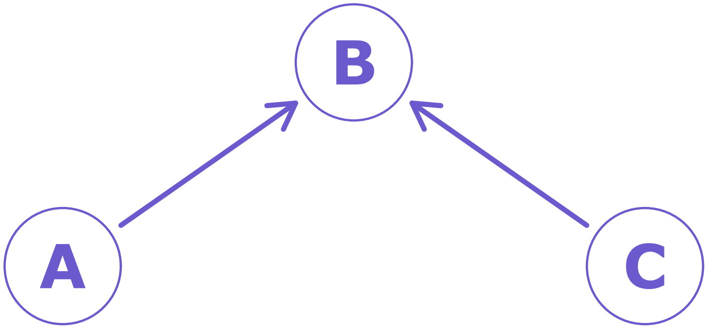


Two more items to emphasize:

1. Association need not follow paths' directions, *e.g.*, $\mathbf{A} \leftarrow \mathbf{B} \rightarrow \mathbf{C}$.

2. Causation follows directed paths.

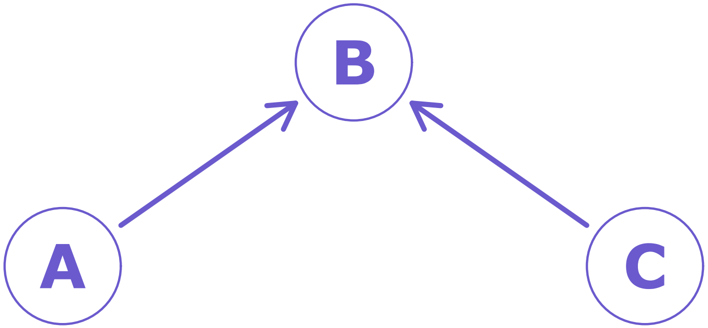
Building block 5: Immoralities



An immorality occurs when two nodes share a child without being otherwise connected.[†] $\mathbf{A} \rightarrow \mathbf{B} \leftarrow \mathbf{C}$

† Yes, this is a stupid term, and I hate it. I way prefer just using 'collider'

Building block 5: Immoralities

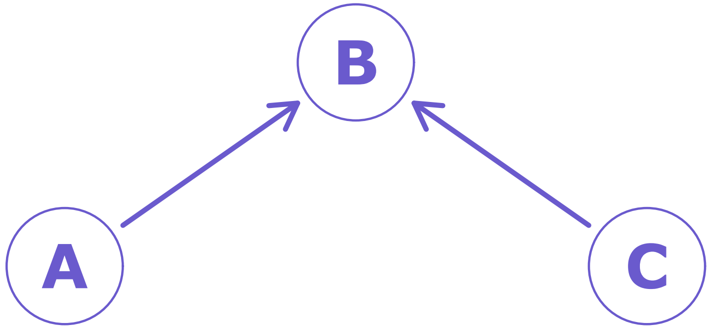


An immorality occurs when two nodes share a child without being otherwise connected.[†] $\mathbf{A} \rightarrow \mathbf{B} \leftarrow \mathbf{C}$

The child (here: $\mathbf{B}$) at the center of this immorality is called a collider.

† Yes, this is a stupid term, and I hate it. I way prefer just using 'collider'

Building block 5: Immoralities



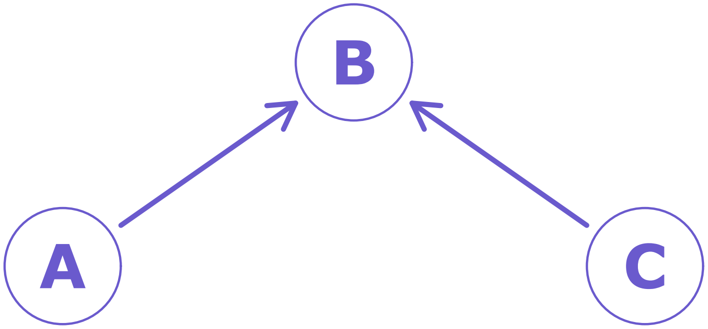
An immorality occurs when two nodes share a child without being otherwise connected.[†] $\mathbf{A} \rightarrow \mathbf{B} \leftarrow \mathbf{C}$

The child (here: $\mathbf{B}$) at the center of this immorality is called a collider.

Notice: An immorality is a fork with reversed directions of the edges.

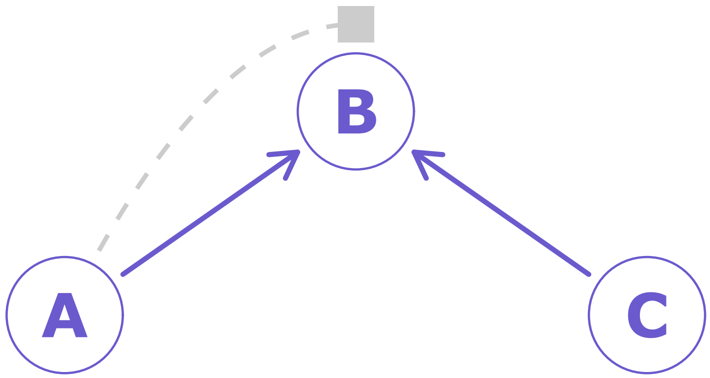
† Yes, this is a stupid term, and I hate it. I way prefer just using 'collider'

Building block 5: Immoralities



Q Are $\mathbf{A}$ and $\mathbf{C}$ independent?

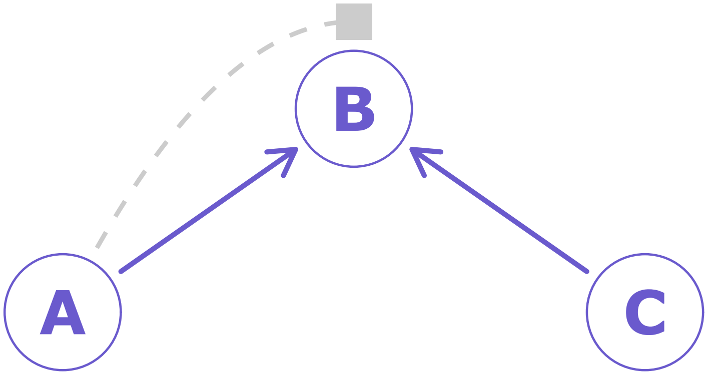
Building block 5: Immoralities



Q Are $\mathbf{A}$ and $\mathbf{C}$ independent?

A Yes. $\mathbf{A} \perp\!\!\!\perp \mathbf{C}$.

Building block 5: Immoralities



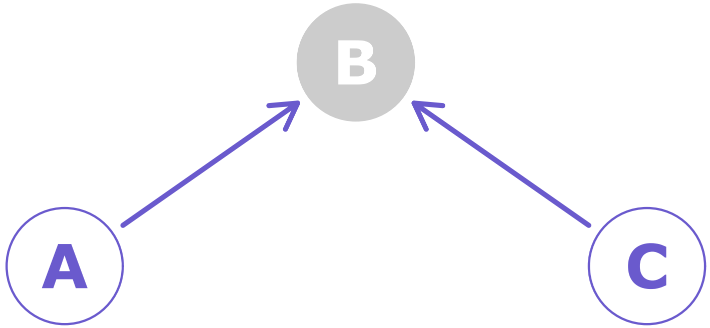
Q Are $\mathbf{A}$ and $\mathbf{C}$ independent?

A Yes. $\mathbf{A} \perp\!\!\!\perp \mathbf{C}$.

Intuition: Causal effects flow from $\mathbf{A}$ and $\mathbf{C}$ and stop there.

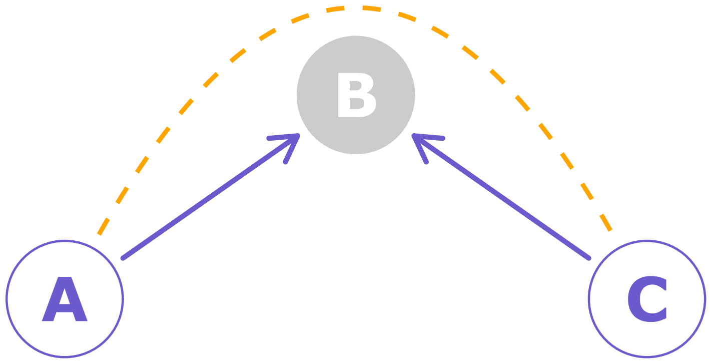
- Neither $\mathbf{A}$ nor $\mathbf{C}$ is a descendant of the other.
- $\mathbf{A}$ and $\mathbf{C}$ do not share any common causes.

Building block 5: Immoralities with conditions



Q What happens when we condition on $\mathbf{B}$?

Building block 5: Immoralities with conditions

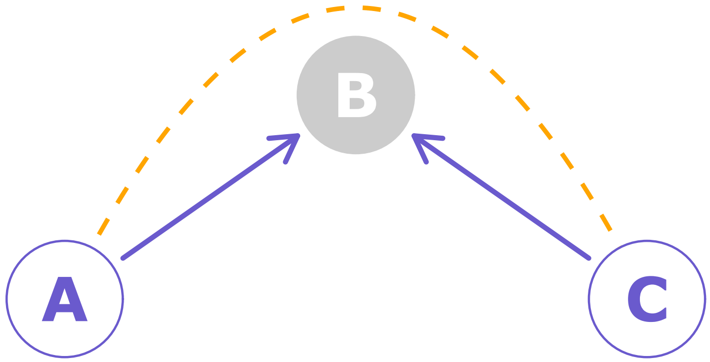


Q What happens when we condition on $\mathbf{B}$?

A We unblock (or open) the previously blocked (closed) path.

While $\mathbf{A}$ and $\mathbf{C}$ are independent, they are conditionally dependent.

Building block 5: Immoralities with conditions



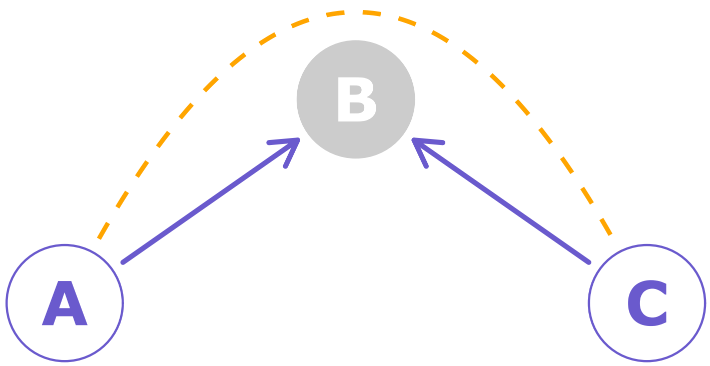
Q What happens when we condition on $\mathbf{B}$?

A We unblock (or open) the previously blocked (closed) path.

While $\mathbf{A}$ and $\mathbf{C}$ are independent, they are conditionally dependent.

Important: When you condition on a collider, you open up the path.

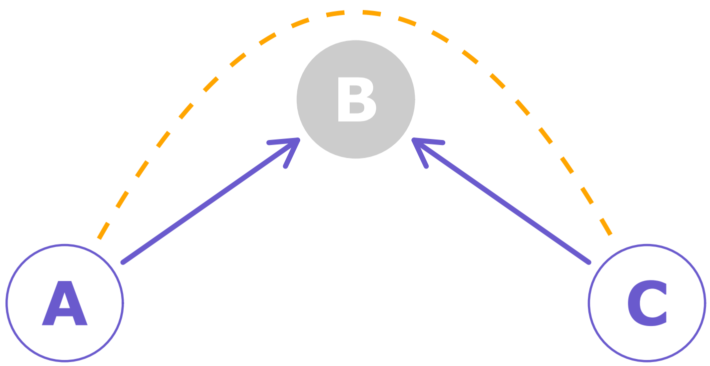
Building block 5: Immoralities with conditions



Intuition: $\mathbf{B}$ is a combination of $\mathbf{A}$ and $\mathbf{C}$.

Conditioning on a value of $\mathbf{B}$ jointly constrains $\mathbf{A}$ and $\mathbf{C}$—they can no longer move independently.

Building block 5: Immoralities with conditions



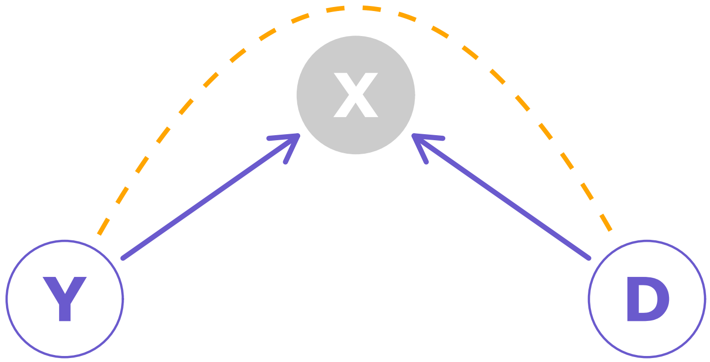
Intuition: $\mathbf{B}$ is a combination of $\mathbf{A}$ and $\mathbf{C}$.

Conditioning on a value of $\mathbf{B}$ jointly constrains $\mathbf{A}$ and $\mathbf{C}$—they can no longer move independently.

Example: Let $\mathbf{A}$ take on $\{0, 1\}$ and $\mathbf{C}$ take on $\{0, 1\}$ (independently).

Conditional on $\mathbf{B} = 1$, $\mathbf{A}$ and $\mathbf{C}$ are perfectly negatively correlated.

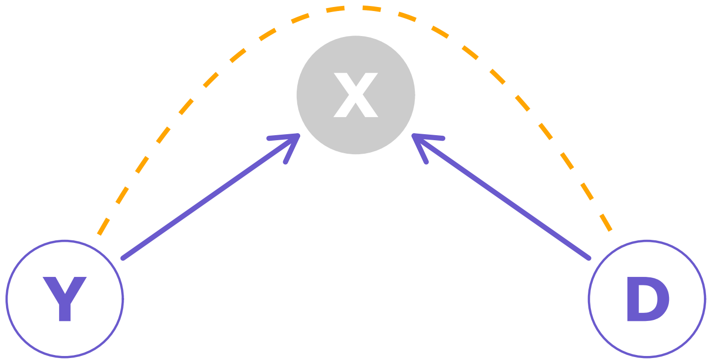
Building block 5: Immoralities with conditions



In *MHE* vocabulary: The collider $\mathbf{X}$ is a *bad control*.

$\mathbf{X}$ is affected by both your treatment $\mathbf{D}$ and outcome $\mathbf{Y}$.

Building block 5: Immoralities with conditions

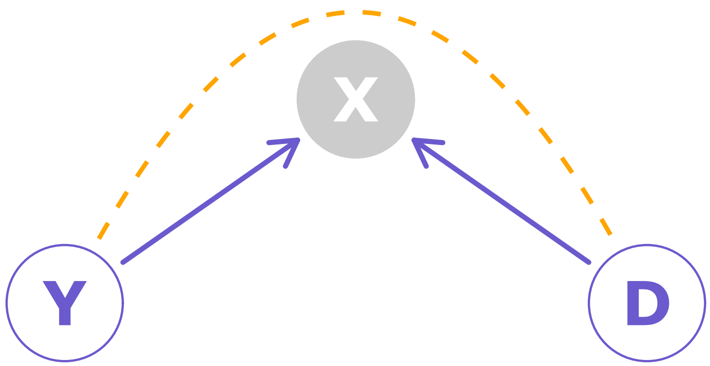


In *MHE* vocabulary: The collider $\mathbf{X}$ is a *bad control*.

$\mathbf{X}$ is affected by both your treatment $\mathbf{D}$ and outcome $\mathbf{Y}$.

The result: A spurious relationship between $\mathbf{Y}$ and $\mathbf{D}$
Remember: they're actually (unconditionally) independent.

Building block 5: Immoralities with conditions



In *MHE* vocabulary: The collider $\mathbf{X}$ is a *bad control*.

$\mathbf{X}$ is affected by both your treatment $\mathbf{D}$ and outcome $\mathbf{Y}$.

The result: A spurious relationship between $\mathbf{Y}$ and $\mathbf{D}$
Remember: they're actually (unconditionally) independent.

This spurious relationship is often called collider bias.

Example Obesity, Mortality Factors and Cardiovascular Disease.

Define O as obesity,

Example Obesity, Mortality Factors and Cardiovascular Disease.

Define $\mathbf{O}$ as obesity, $\mathbf{A}$ as age of patient,

Example Obesity, Mortality Factors and Cardiovascular Disease.

Define O as obesity, A as age of patient, and M as mortality.

Example Obesity, Mortality Factors and Cardiovascular Disease.

Define $\mathbf{O}$ as obesity, $\mathbf{A}$ as age of patient, and $\mathbf{M}$ as mortality.

Suppose for the moment obesity and age

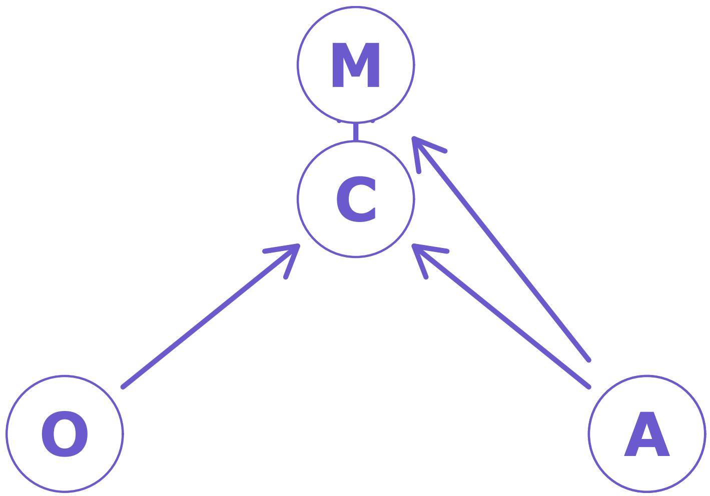
1. are independent of each other
2. each cause cardiovascular disease and cvd $\uparrow \mathbf{M}$

Example Obesity, Mortality Factors and Cardiovascular Disease.

Define $\mathbf{O}$ as obesity, $\mathbf{A}$ as age of patient, and $\mathbf{M}$ as mortality.

Suppose for the moment obesity and age

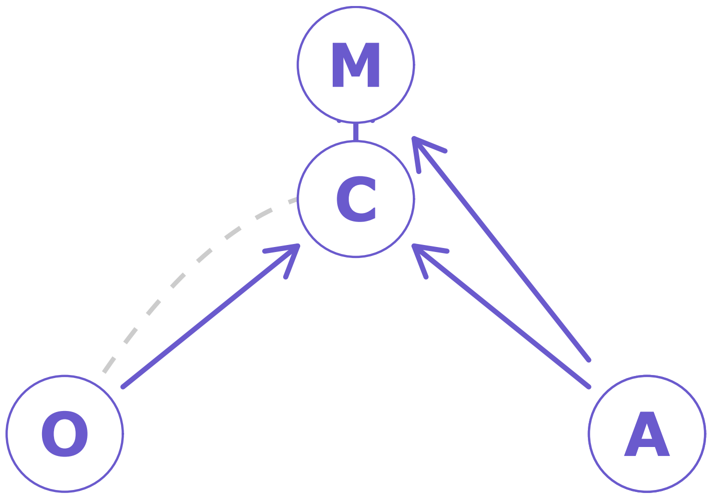
1. are independent of each other
2. each cause cardiovascular disease and cvd $\uparrow \mathbf{M}$



The implied DAG.

Example Obesity, Cardiovascular Disease and Mortality.

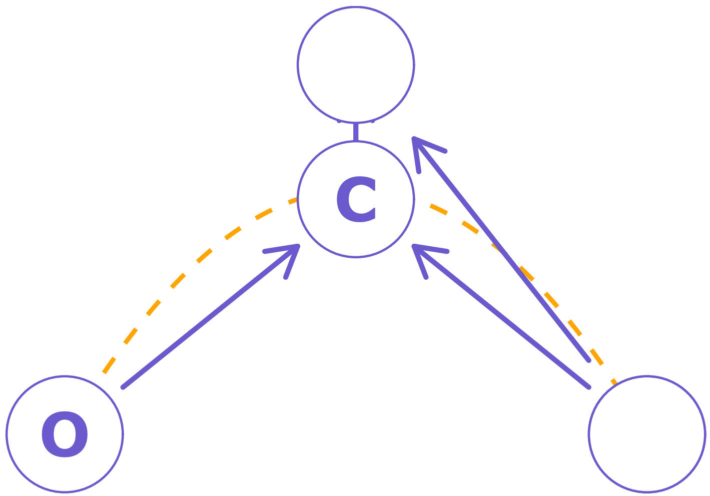
Define $\mathbf{O}$ as obesity, $\mathbf{A}$ as Age, $\mathbf{C}$ as Cardiovascular Disease, and $\mathbf{M}$ as Mortality.



If we do not condition on Cardiovascular Disease, $\mathbf{S} \rightarrow \mathbf{H} \leftarrow \mathbf{A}$ is blocked, but $\mathbf{O} \rightarrow \mathbf{C} \rightarrow \mathbf{M}$ is not

Example Obesity, Cardiovascular Disease and Mortality.

Define $\mathbf{O}$ as obesity, $\mathbf{A}$ as Age, $\mathbf{C}$ as Cardiovascular Disease, and $\mathbf{M}$ as Mortality.

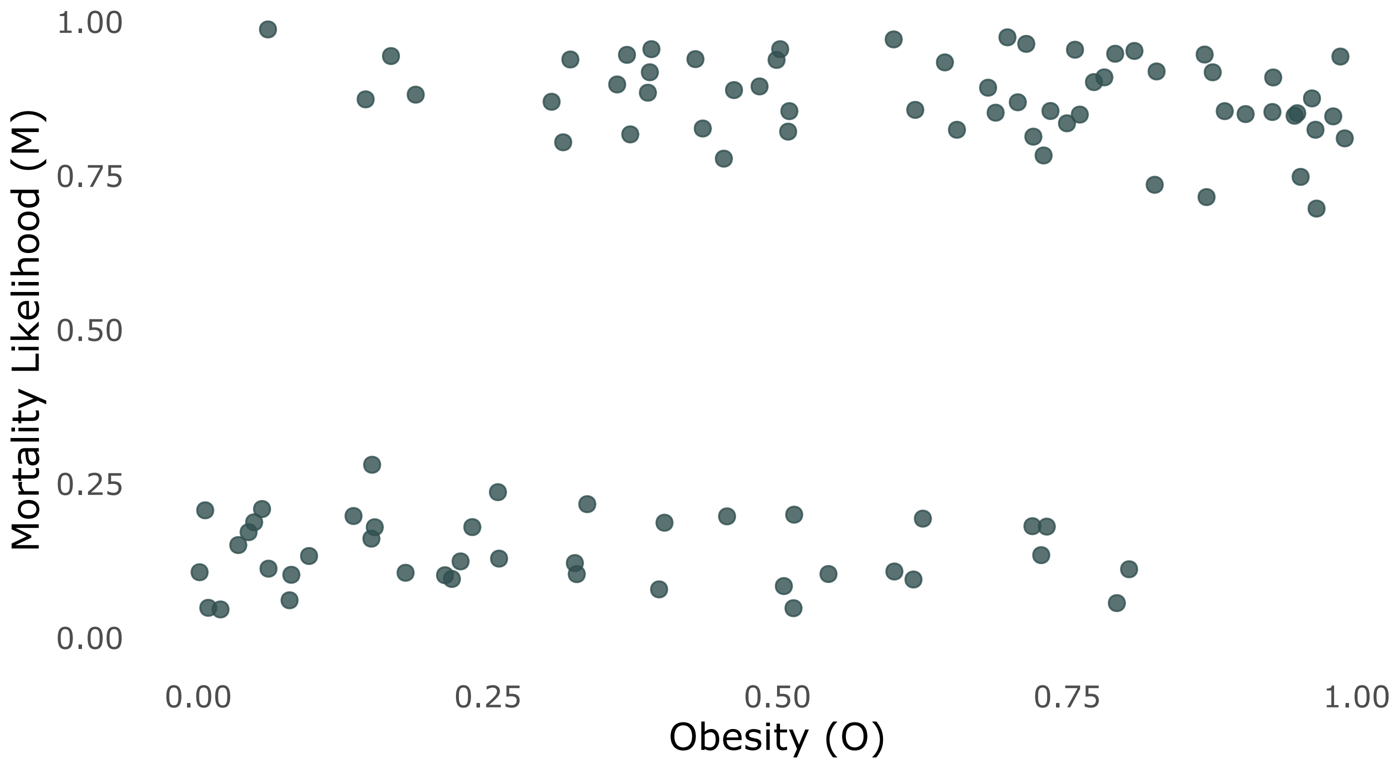


Our data conditions on cardiovascular disease, which opens $\mathbf{O} \rightarrow \mathbf{C} \leftarrow \mathbf{A}$.

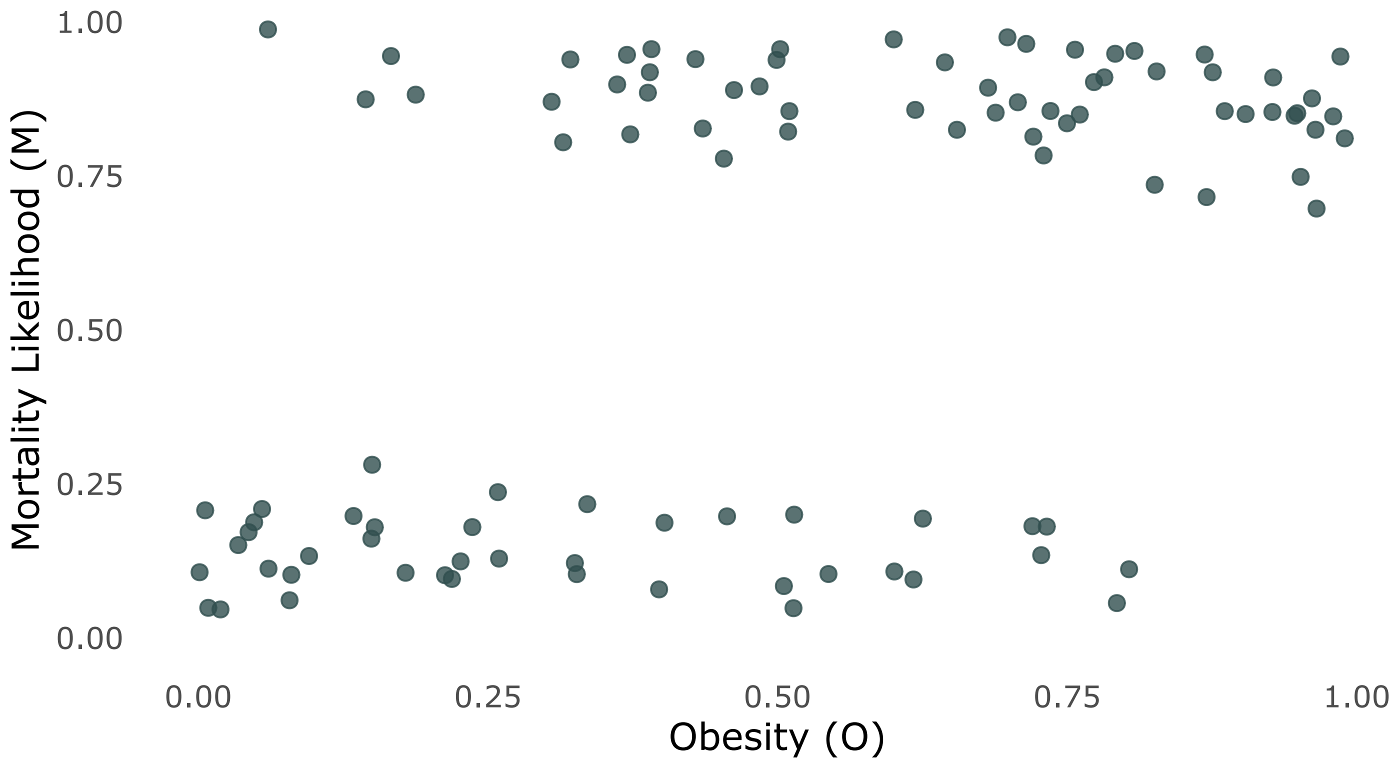
You can also see this example graphically...

Example Obesity, Cardiovascular Disease and Mortality.

Example Obesity, Cardiovascular Disease and Mortality.

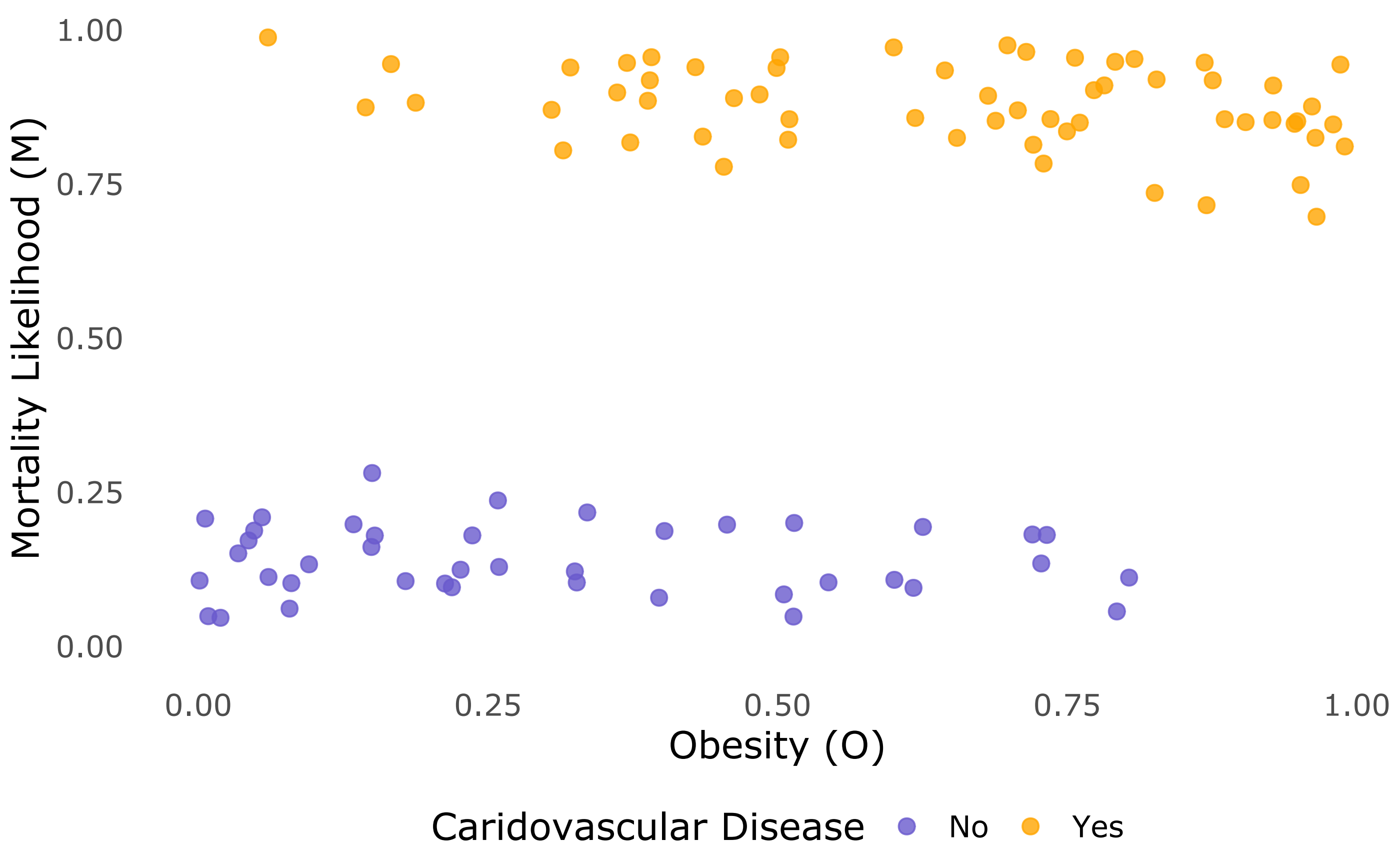


Example Obesity, Cardiovascular Disease and Mortality.



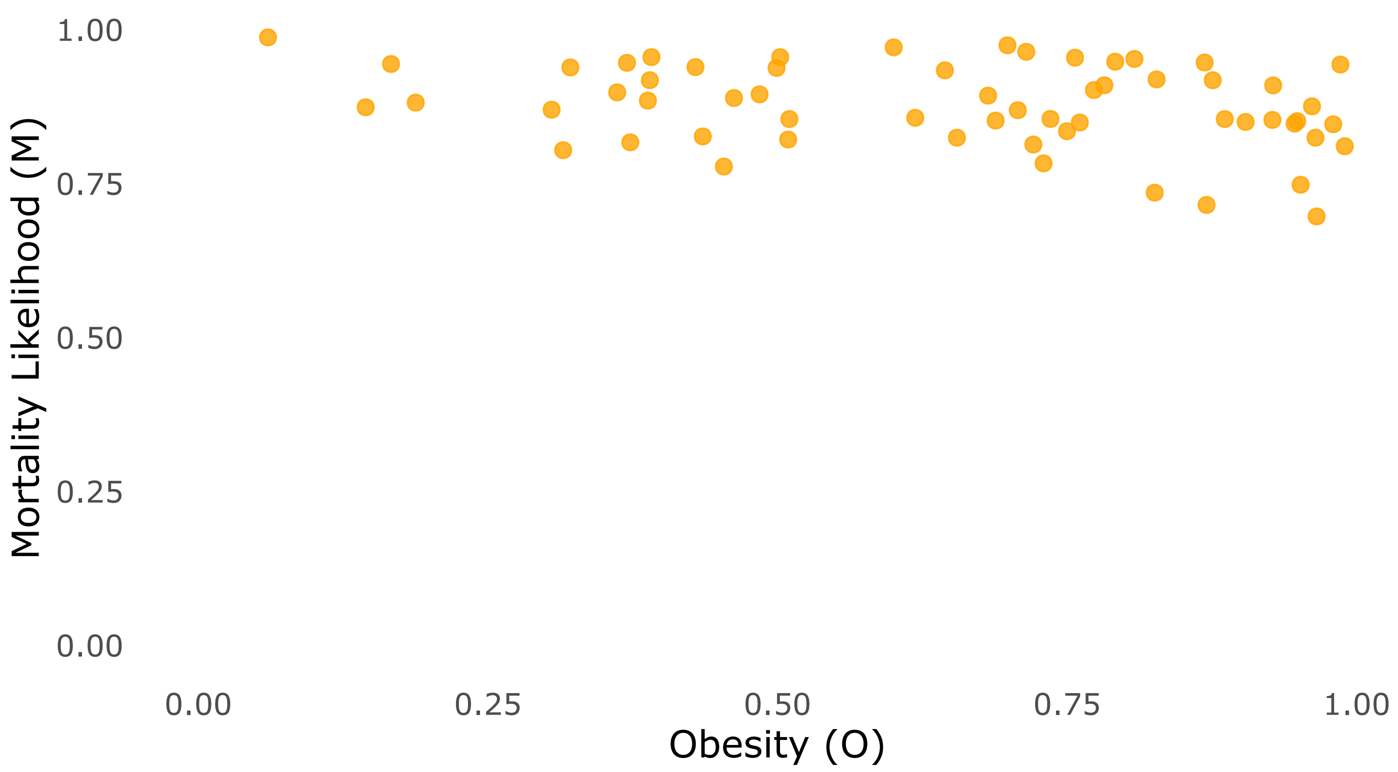
Without conditioning: Positive relationship between obesity and mortality.

Example Obesity, Cardiovascular Disease and Mortality.



Recall: Our sample excludes individuals without cv-disease.

Example Obesity, Cardiovascular Disease and Mortality.



Conditioning on $\mathbf{C}$: Obesity now only increases mortality through Age and there are disproportionately large numbers of young obese patients with cvd.

Across the general population...

Across the general population...

```
lm(data = cb_dt, M ~ O) %>% tidy() %>% filter(term == 'O')
```

term	estimate	std.error	statistic	p.value
O	0.64	0.106	6.04	2.74e-08

Across the general population...

```
lm(data = cb_dt, M ~ O) %>% tidy() %>% filter(term == 'O')
```

term	estimate	std.error	statistic	p.value
O	0.64	0.106	6.04	2.74e-08

Only on patients with CVD...

```
lm(data = cb_dt %>% filter(C == 1), M ~ O) %>% tidy()%>% filter(term == 'O')
```

term	estimate	std.error	statistic	p.value
O	-0.0792	0.0339	-2.34	0.0229

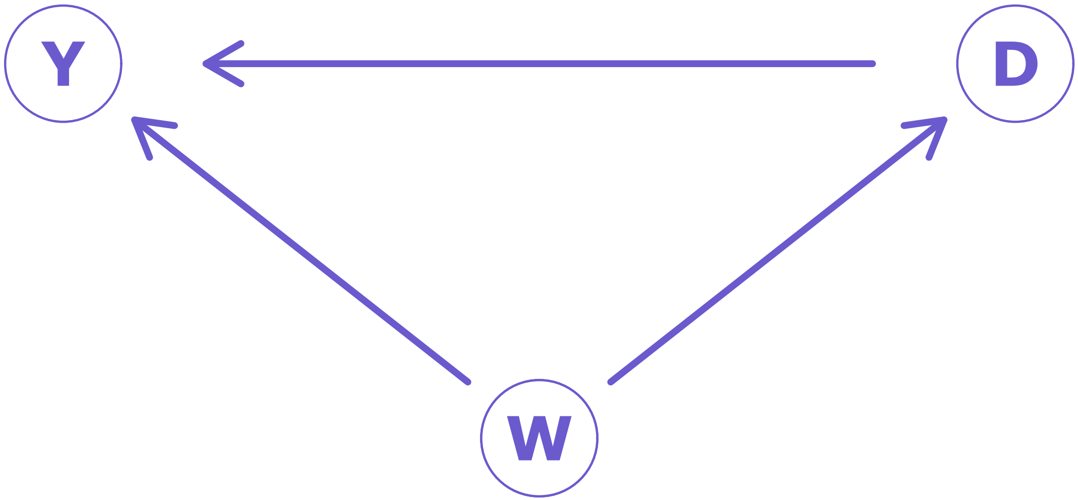
I like this example because it reminds us that conditioning occurs both explicitly (*e.g.*, "controlling for") and implicitly (*e.g.*, sample inclusion/exclusion).

This example of collider bias in hospitalization data comes from V. Viallon & M. Dufournet's 2016 paper Can collider bias fully explain the obesity paradox?.

More generally: You'll hear this called selection bias or Berkson's paradox.

Examples

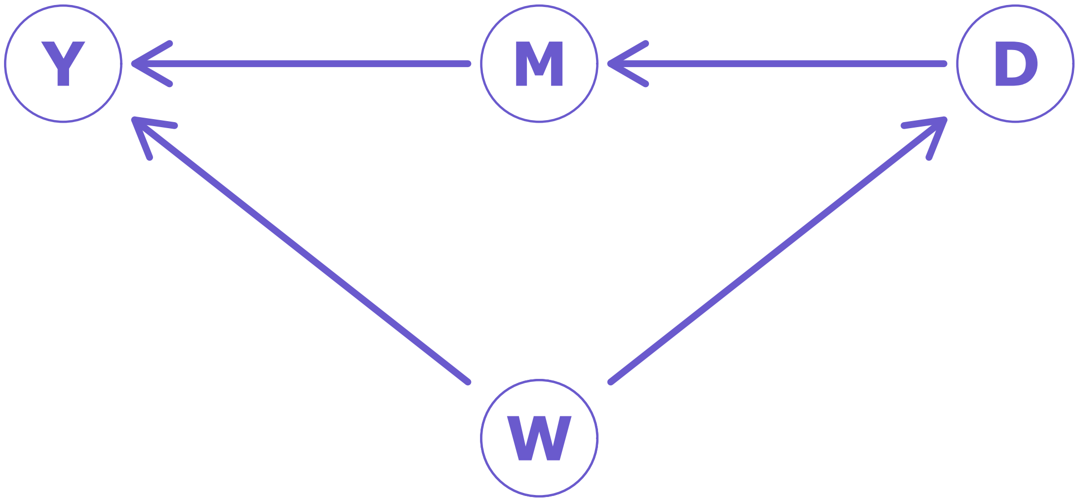
Example 1: OVB



Q OVB using DAG fundamentals: When can we isolate causal effects?

Example 2: Mediation

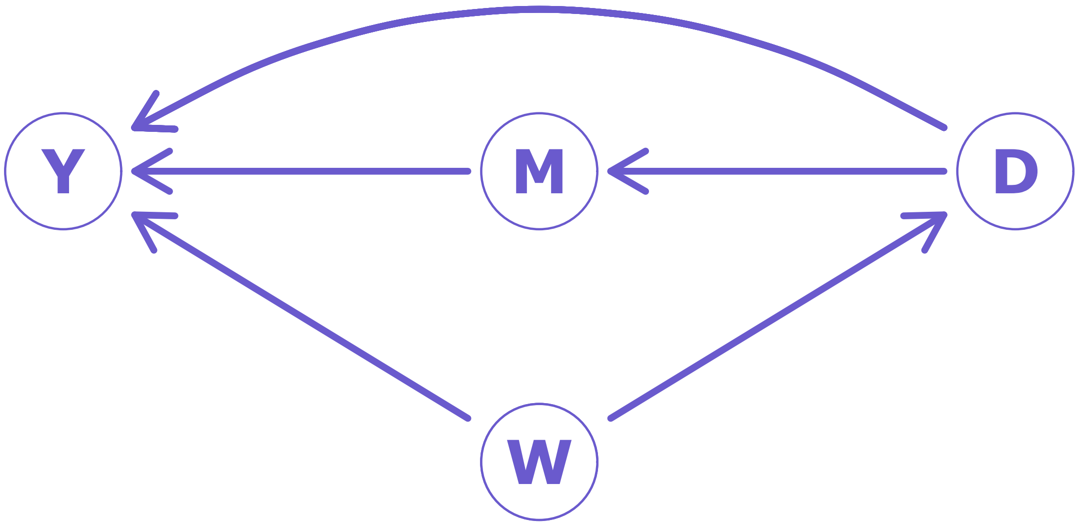
Here $\mathbf{M}$ is a mediator: it mediates the effect of $\mathbf{D}$ on $\mathbf{Y}$.



Q_1 What do we need to condition on to get the effect of $\mathbf{D}$ on $\mathbf{Y}$?

Q_2 What happens if we condition on $\mathbf{W}$ and $\mathbf{M}$?

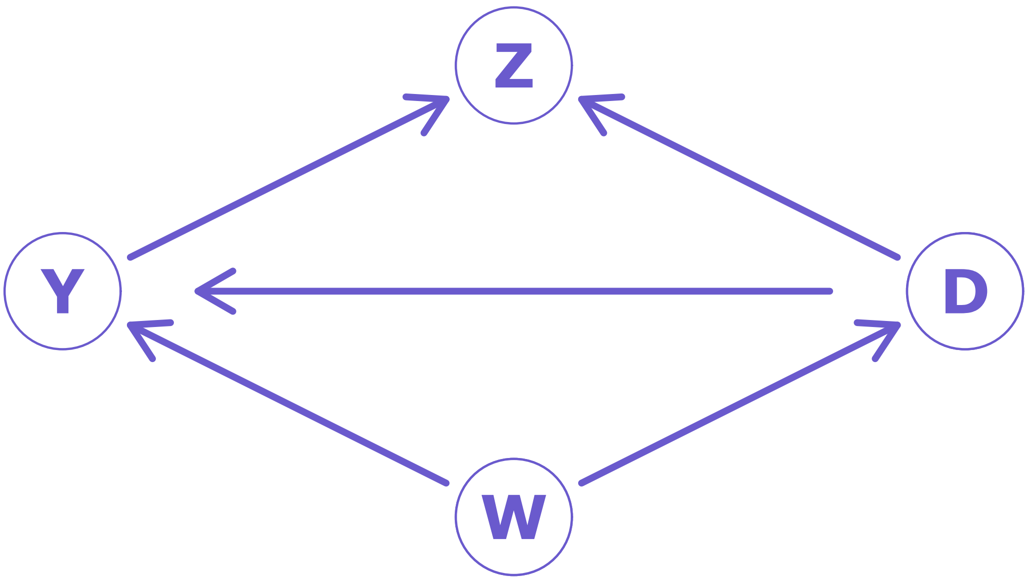
Example 3: Partial mediation



Q_1 What do we need to condition on to get the effect of $\mathbf{D}$ on $\mathbf{Y}$?

Q_2 What happens if we condition on $\mathbf{W}$ and $\mathbf{M}$?

Example 4: Non-mediator descendants

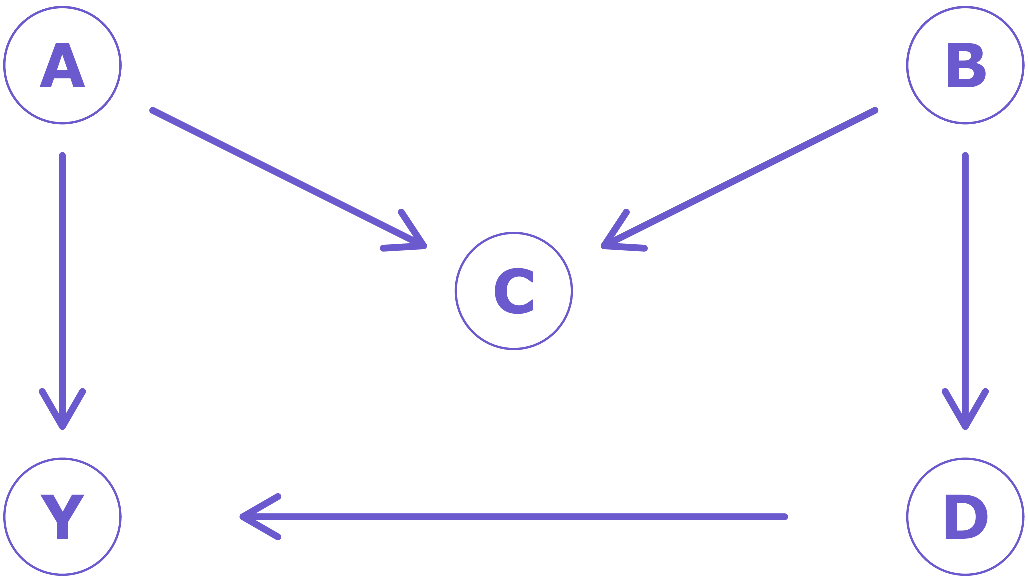


Q_1 What do we need to condition on to get the effect of $\mathbf{D}$ on $\mathbf{Y}$?

Q_2 What happens if we condition on $\mathbf{W}$ and/or $\mathbf{Z}$?

Example 5 M-Bias

Notice that $\mathbf{C}$ here is not a result of treatment (could be "pre-treatment").



Q_1 What do we need to condition on to get the effect of $\mathbf{D}$ on $\mathbf{Y}$?

Q_2 What happens if we condition on $\mathbf{W}$ and/or $\mathbf{Z}$?

One more note:

DAGs are often drawn without "noise variables" (disturbances).

But they still exist—they're just "outside of the model."

Experiments in SCM

Experiments

So - how do we think about a randomized experiment with a SCM?

- Recall, an experiment induces random noise in a variable that is unrelated to other causal factors

We can think about an experiment deleting the edges out of the experiment. Let's return to our fertilizer example.

Experiments

So - how do we think about a randomized experiment with a SCM?

- Recall, an experiment induces random noise in a variable that is unrelated to other causal factors

We can think about an experiment deleting the edges out of the experiment. Let's return to our fertilizer example.

If we induce random noise in fertilizer - we know that no effect that causes fertilizer placement is related to the random placement

Experiments

So - how do we think about a randomized experiment with a SCM?

- Recall, an experiment induces random noise in a variable that is unrelated to other causal factors

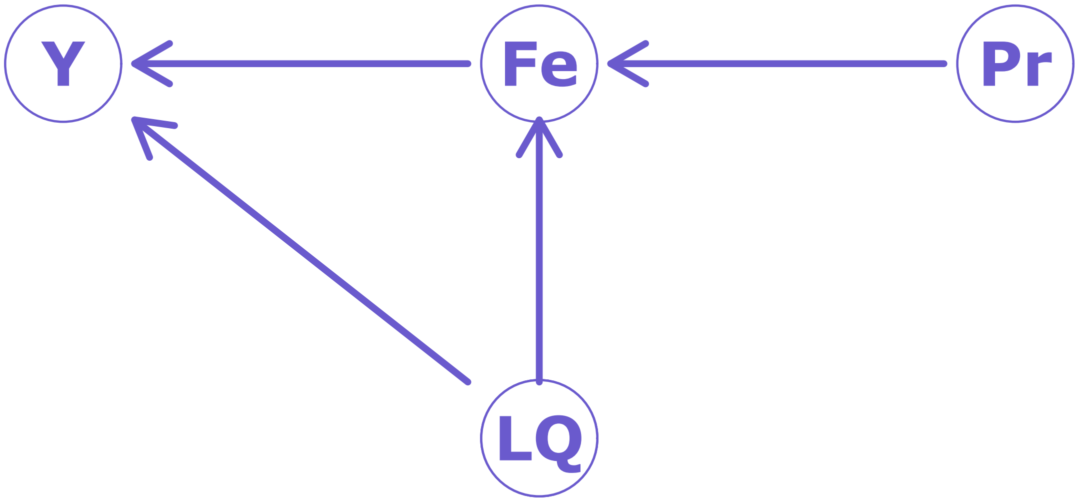
We can think about an experiment deleting the edges out of the experiment. Let's return to our fertilizer example.

If we induce random noise in fertilizer - we know that no effect that causes fertilizer placement is related to the random placement

But if our experiment works, then things caused by fertilizer should still be impacted

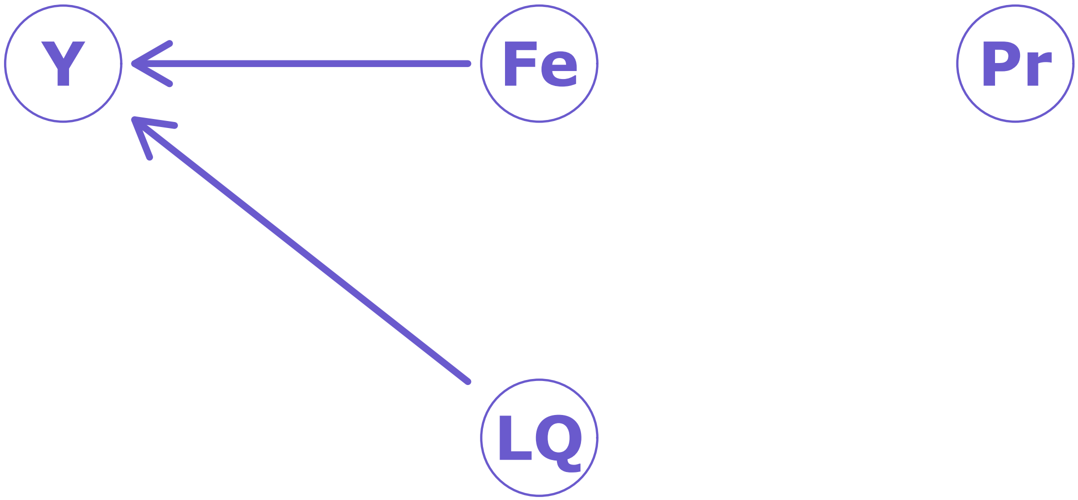
Experiments

Fertilizer on Yield: pre-experiment



Experiments

Fertilizer on Yield: post-experiment



DAGs

Limitations

So what can't DAGs do (well)?

Limitations

So what can't DAGs do (well)?

- Simultaneity: Defined causality as unidirectional and prohibited cycles.

Limitations

So what can't DAGs do (well)?

- Simultaneity: Defined causality as unidirectional and prohibited cycles.
- Dynamics: You can sort of allow a variable to affect itself... $\mathbf{Y}_{t=1} \rightarrow \mathbf{Y}_{t=2}$.

Limitations

So what can't DAGs do (well)?

- Simultaneity: Defined causality as unidirectional and prohibited cycles.
- Dynamics: You can sort of allow a variable to affect itself... $\mathbf{Y}_{t=1} \rightarrow \mathbf{Y}_{t=2}$.
- Uncertainty: DAGs are most useful when you can correctly draw them.

DAGs

Limitations

So what can't DAGs do (well)?

- Simultaneity: Defined causality as unidirectional and prohibited cycles.
- Dynamics: You can sort of allow a variable to affect itself... $\mathbf{Y}_{t=1} \rightarrow \mathbf{Y}_{t=2}$.
- Uncertainty: DAGs are most useful when you can correctly draw them.
- Make friends: There's *a lot* of (angry/uncharitable) fighting about DAGs:

$$\textbf{Philosophy} \rightarrow \textbf{DAGs/Epidemiology} \leftarrow \textbf{Economics}$$

DAGs

Limitations

So what can't DAGs do (well)?

- Simultaneity: Defined causality as unidirectional and prohibited cycles.
- Dynamics: You can sort of allow a variable to affect itself... $\mathbf{Y}_{t=1} \rightarrow \mathbf{Y}_{t=2}$.
- Uncertainty: DAGs are most useful when you can correctly draw them.
- Make friends: There's *a lot* of (angry/uncharitable) fighting about DAGs:

$$\textbf{Philosophy} \rightarrow \textbf{DAGs/Epidemiology} \leftarrow \textbf{Economics}$$

- Functional Form: All of the existing tools are functionally indifferent - but functional forms are useful (as we've seen.)

Sources

Thanks

These notes rely heavily upon Brady Neal's *Introduction to Causal Inference*.

I also borrow from Scott Cunningham's *Causal Inference: The Mixtape*.

I found the Sackett (1978) example on the "Catalog of Bias" website.

Table of contents

Math and Probability

- Let's quickly review some probability concepts (we'll need them for this section)

Brief Probability Review

We can decompose any joint probability into a conditional probability and it's product

a.) $P(A \cap B) = P(A|B)\,P(B)$

Brief Probability Review

We can decompose any joint probability into a conditional probability and it's product

a.) $P(A \cap B) = P(A|B)\,P(B)$

We can decompose any conditional probability into the ratio of the joint probability and the probability of it's ratio like so

b.) $P(A|B) = \frac{P(A \cap B)}{P(B)}$

Theory

Brief Probability Review

We can decompose any joint probability into a conditional probability and it's product

a.) $P(A \cap B) = P(A|B)\,P(B)$

We can decompose any conditional probability into the ratio of the joint probability and the probability of it's ratio like so

b.) $P(A|B) = \frac{P(A \cap B)}{P(B)}$

We can substitute a into b and get bayes rule - ditching the intersection notation

$$P(B|A) = \frac{P(A|B)\,P(B)}{P(A)}$$

Product Rule

The important piece of this however is the product rule.

By using our definitions from before, we can decompose larger and larger joint probability distributions, like so -

$$
\begin{aligned}
&2 \qquad P(x_1, x_2) = P(x_1)P(x_2|x_1) \\
&3 \qquad P(x_1, x_2, x_3) = P(x_1)P(x_2, x_3|x_1) = P(x_1)P(x_2|x_1)P(x_3|x_2, x_1) \\
&\vdots \\
&n \qquad P(x_1, x_2, \ldots, x_n) = P(x_1)\prod_{i=2}^{n} P(x_i|x_{i-1}, \ldots, x_1)
\end{aligned}
$$

Product Rule

The important piece of this however is the product rule.

By using our definitions from before, we can decompose larger and larger joint probability distributions, like so -

$$
\begin{array}{ll}
2 & P(x_1, x_2) = P(x_1)P(x_2|x_1) \\
3 & P(x_1, x_2, x_3) = P(x_1)P(x_2, x_3|x_1) = P(x_1)P(x_2|x_1)P(x_3|x_2, x_1) \\
\vdots & \\
n & P(x_1, x_2, \ldots, x_n) = P(x_1)\prod_{i=2}^{n} P(x_i|x_{i-1}, \ldots, x_1)
\end{array}
$$

This final product can include *a lot* of terms.
E.g., even when x_i are binary, $P(x_4|x_3, x_2, x_1)$ requires $2^3 = 8$ parameters.

Local Markov

This intuitive approach *is* the Local Markov Assumption

> Given its parents in the DAG, a node X is independent of all of its non-descendants.

Local Markov

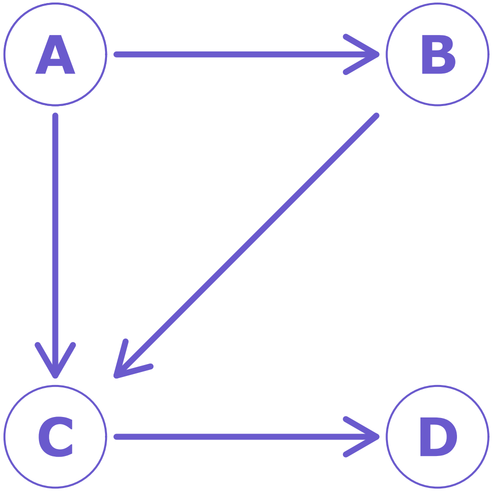
This intuitive approach *is* the Local Markov Assumption

> Given its parents in the DAG, a node X is independent of all of its non-descendants.

Ex. Consider the DAG to the right:

With the Local Markov Assumption, $P(\mathrm{D}|\mathrm{A},\mathrm{B},\mathrm{C})$ simplifies to $P(\mathrm{D}|\mathrm{C})$.

Conditional on its parent (C), D is independent of A and B.



Local Markov and factorization

The Local Markov Assumption is equiv. to Bayesian Network Factorization

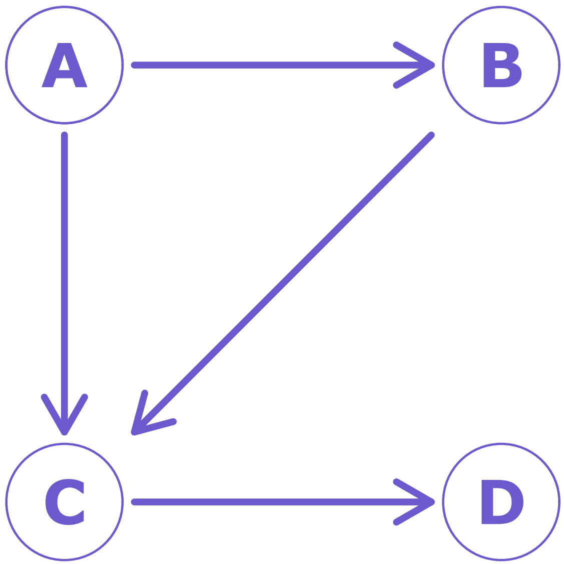
> For prob. dist. P and DAG G, P factorizes according to G if
>
> $$P(x_1, \ldots, x_n) = \prod_i P(x_i | \mathbf{pa}_i)$$
>
> where $\mathbf{pa}_i$ refers to x_i's parents in G. Bayesian network factorization is also called *the chain rule for Bayesian networks* and *Markov compatibility*.

Theory

Factorize!

You can now (more easily) factorize a causal system!

- Let's think about this graph...



Factorize!

You can now (more easily) factorize a causal system!

- Let's think about this graph...

Factorization via B.N. chain rule

$$
\begin{aligned}
&P(\mathrm{A}, \mathrm{B}, \mathrm{C}, \mathrm{D}) \\
&\quad = \prod_i P(x_i | \mathbf{pa}_i) \\
&\quad = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})P(\mathrm{C}|\mathrm{A}, \mathrm{B})P(\mathrm{D}|\mathrm{C})
\end{aligned}
$$

Blocked paths

Let's formally define a blocked path (blocking is important).

Blocked paths

Let's formally define a blocked path (blocking is important).

A path between $\mathbf{X}$ and $\mathbf{Y}$ is blocked by conditioning on variables $\mathbf{Z}$ (potentially no variables) if either of the following statements is true:

1. On the path $\mathbf{X} \rightarrow \ldots \mathbf{Y}$, there is either a chain or a fork, and we condition on some element $\mathbf{w}$ in that chain, ie $\mathbf{w}$ is in $\mathbf{Z}$.

2. On the path, there is a collider $(\cdots \rightarrow \mathbf{W} \leftarrow \ldots)$, and we do not condition on it or any of its descendants $(\mathbf{de}(W))$.

Theory

Blocked paths

Let's formally define a blocked path (blocking is important).

A path between $\mathbf{X}$ and $\mathbf{Y}$ is blocked by conditioning on variables $\mathbf{Z}$ (potentially no variables) if either of the following statements is true:

1. On the path $\mathbf{X} \rightarrow \ldots \mathbf{Y}$, there is either a chain or a fork, and we condition on some element $\mathbf{w}$ in that chain, ie $\mathbf{w}$ is in $\mathbf{Z}$.

2. On the path, there is a collider $(\cdots \rightarrow \mathbf{W} \leftarrow \ldots)$, and we do not condition on it or any of its descendants $(\mathbf{de}(W))$.

Association flows along unblocked paths.

d-separation and d-connected(-ness)

Finally, we'll define whether nodes are separated or connected in DAGs.

d-separation and d-connected(-ness)

Finally, we'll define whether nodes are separated or connected in DAGs.

Separation: Nodes $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by a set of nodes $\mathbf{Z}$ if all paths between $\mathbf{X}$ and $\mathbf{Y}$ are blocked by $\mathbf{Z}$.

d-separation and d-connected(-ness)

Finally, we'll define whether nodes are separated or connected in DAGs.

Separation: Nodes $\mathbf{X}$ and $\mathbf{Y}$ are d-separated by a set of nodes $\mathbf{Z}$ if all paths between $\mathbf{X}$ and $\mathbf{Y}$ are blocked by $\mathbf{Z}$.

Connection: If there is at least one path between $\mathbf{X}$ and $\mathbf{Y}$ that is unblocked, then $\mathbf{X}$ and $\mathbf{Y}$ are d-connected.

d-separation and causality

d-separation tells us that two nodes are not associated.

d-separation and causality

d-separation tells us that two nodes are not associated.

To measure the causal effect of $\mathbf{X}$ on $\mathbf{Y}$:
We must eliminate non-causal association.

d-separation and causality

d-separation tells us that two nodes are not associated.

To measure the causal effect of $\mathbf{X}$ on $\mathbf{Y}$:
We must eliminate non-causal association.

Putting these ideas together, here is our criterion to isolate causal effects:

> If we remove all edges flowing out of $\mathbf{X}$ (its causal effects),
> then $\mathbf{X}$ and $\mathbf{Y}$ should be d-separated.

d-separation and causality

d-separation tells us that two nodes are not associated.

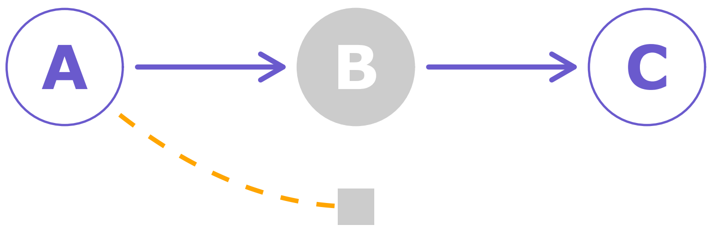
To measure the causal effect of $\mathbf{X}$ on $\mathbf{Y}$:
We must eliminate non-causal association.

Putting these ideas together, here is our criterion to isolate causal effects:

> If we remove all edges flowing out of $\mathbf{X}$ (its causal effects), then $\mathbf{X}$ and $\mathbf{Y}$ should be d-separated. This criterion ensures that we've closed the backdoor paths that generate non-causal associations between $\mathbf{X}$ and $\mathbf{Y}$.

Theory

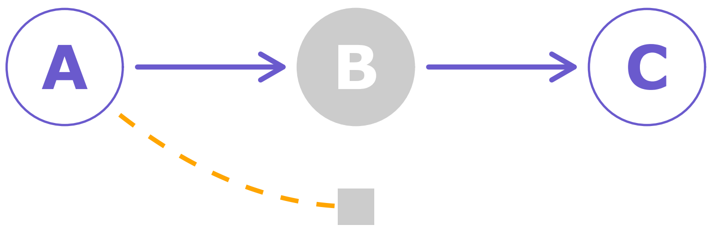
Building block 3: Chains with conditions



Proof: We want to show $\mathbf{A}$ and $\mathbf{C}$ are independent conditional on $\mathbf{B}$, *i.e.*, $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

Theory

Building block 3: Chains with conditions

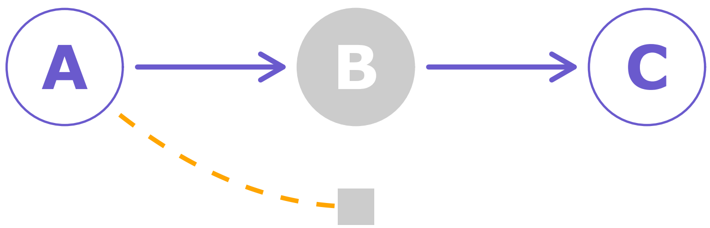


Proof: We want to show $\mathbf{A}$ and $\mathbf{C}$ are independent conditional on $\mathbf{B}$, *i.e.*, $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

Start with BN factorization: $P(\mathrm{A}, \mathrm{B}, \mathrm{C})$

Theory

Building block 3: Chains with conditions

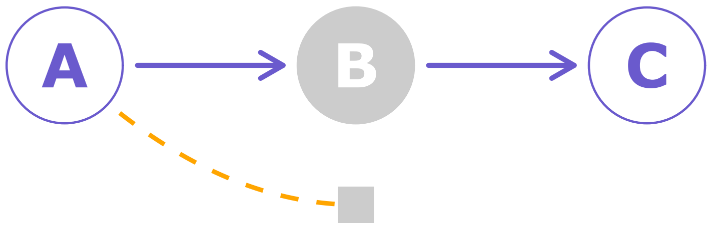


Proof: We want to show $\mathbf{A}$ and $\mathbf{C}$ are independent conditional on $\mathbf{B}$, *i.e.*, $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

Start with BN factorization: $P(\mathrm{A}, \mathrm{B}, \mathrm{C}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})P(\mathrm{C}|\mathrm{B})$.

Theory

Building block 3: Chains with conditions



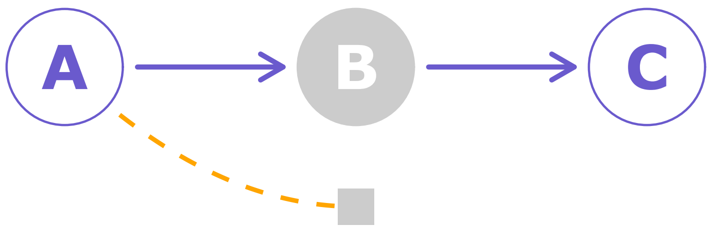
Proof: We want to show A and C are independent conditional on B, *i.e.*, $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

Start with BN factorization: $P(\mathrm{A}, \mathrm{B}, \mathrm{C}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})P(\mathrm{C}|\mathrm{B})$.

Now apply Bayes' rule for the LHS of our goal: $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = \frac{P(\mathrm{A},\mathrm{B},\mathrm{C})}{P(\mathrm{B})}$.

Theory

Building block 3: Chains with conditions



Proof: We want to show $\mathbf{A}$ and $\mathbf{C}$ are independent conditional on $\mathbf{B}$, *i.e.*, $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

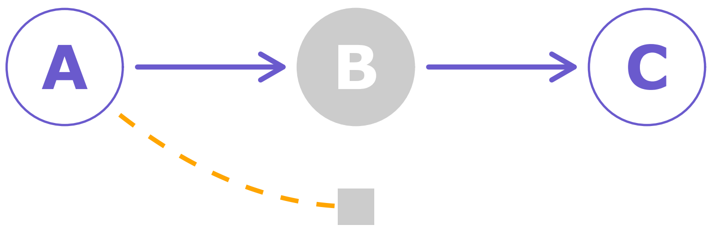
Start with BN factorization: $P(\mathrm{A}, \mathrm{B}, \mathrm{C}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})P(\mathrm{C}|\mathrm{B})$.

Now apply Bayes' rule for the LHS of our goal: $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = \frac{P(\mathrm{A},\mathrm{B},\mathrm{C})}{P(\mathrm{B})}$.

And substitute our factorization into the Bayes' rule expression:

Theory

Building block 3: Chains with conditions



Proof: We want to show $\mathbf{A}$ and $\mathbf{C}$ are independent conditional on $\mathbf{B}$, *i.e.*, $P(\mathrm{A},\mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

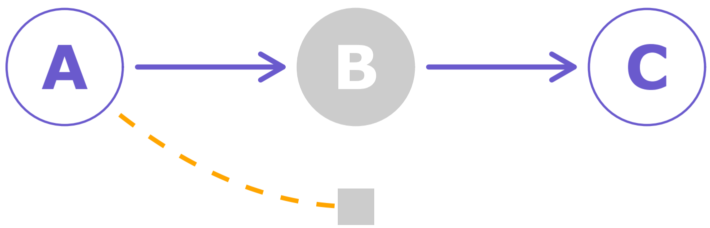
Start with BN factorization: $P(\mathrm{A},\mathrm{B},\mathrm{C}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})P(\mathrm{C}|\mathrm{B})$.

Now apply Bayes' rule for the LHS of our goal: $P(\mathrm{A},\mathrm{C}|\mathrm{B}) = \frac{P(\mathrm{A},\mathrm{B},\mathrm{C})}{P(\mathrm{B})}$.

And substitute our factorization into the Bayes' rule expression:

Theory

Building block 3: Chains with conditions



Proof: We want to show $\mathbf{A}$ and $\mathbf{C}$ are independent conditional on $\mathbf{B}$, *i.e.*, $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

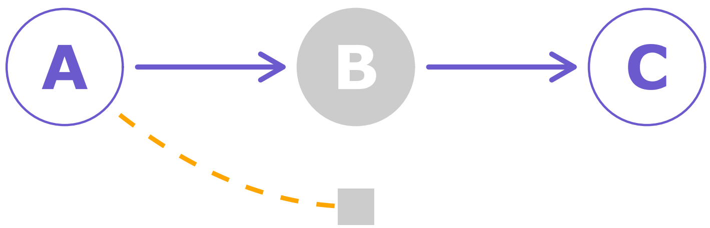
Start with BN factorization: $P(\mathrm{A}, \mathrm{B}, \mathrm{C}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})P(\mathrm{C}|\mathrm{B})$.

Now apply Bayes' rule for the LHS of our goal: $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = \frac{P(\mathrm{A},\mathrm{B},\mathrm{C})}{P(\mathrm{B})}$.

And substitute our factorization into the Bayes' rule expression:

Theory

Building block 3: Chains with conditions



Proof: We want to show $\mathbf{A}$ and $\mathbf{C}$ are independent conditional on B, *i.e.*, $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

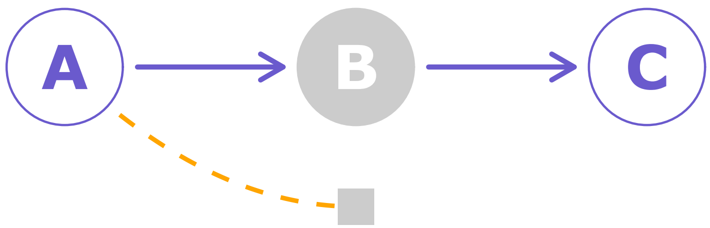
Start with BN factorization: $P(\mathrm{A}, \mathrm{B}, \mathrm{C}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})P(\mathrm{C}|\mathrm{B})$.

Now apply Bayes' rule for the LHS of our goal: $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = \frac{P(\mathrm{A},\mathrm{B},\mathrm{C})}{P(\mathrm{B})}$.

And substitute our factorization into the Bayes' rule expression:

Theory

Building block 3: Chains with conditions



Proof: We want to show $\mathbf{A}$ and $\mathbf{C}$ are independent conditional on $\mathbf{B}$, *i.e.*, $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = P(\mathrm{A}|\mathrm{B})P(\mathrm{C}|\mathrm{B})$.

Start with BN factorization: $P(\mathrm{A}, \mathrm{B}, \mathrm{C}) = P(\mathrm{A})P(\mathrm{B}|\mathrm{A})P(\mathrm{C}|\mathrm{B})$.

Now apply Bayes' rule for the LHS of our goal: $P(\mathrm{A}, \mathrm{C}|\mathrm{B}) = \frac{P(\mathrm{A},\mathrm{B},\mathrm{C})}{P(\mathrm{B})}$.

And substitute our factorization into the Bayes' rule expression: