Simpson's Paradox

In a cohort of 1000 students:
500 used a new method for learning a concept (treatment T).
They were judged whether they understood the concept (evaluation E)
for two subpopulations (one with C=*true* and one with C=*false*):

C	T	E=*true*	E=*false*	Rate
true	*true*	90	10	$90/(90+10) = 90\%$
true	*false*	290	110	$290/(290+110) = 72.5\%$
false	*true*	110	290	$110/(110+290) = 27.5\%$
false	*false*	10	90	$10/(10+90) = 10\%$

Does the treatment increase understanding?

Simpson's Paradox

In a cohort of 1000 students:
500 used a new method for learning a concept (treatment T).
They were judged whether they understood the concept (evaluation E)
for two subpopulations (one with C=*true* and one with C=*false*):

C	T	E=*true*	E=*false*	Rate
true	*true*	90	10	$90/(90+10) = 90\%$
true	*false*	290	110	$290/(290+110) = 72.5\%$
false	*true*	110	290	$110/(110+290) = 27.5\%$
false	*false*	10	90	$10/(10+90) = 10\%$

Does the treatment increase understanding?

T	E=*true*	E=*false*	Rate
true	200	300	$200/(200+300) = 40\%$
false	300	200	$300/(300+200) = 60\%$

Causal Model

A causal network is a belief network where

$$P(X \mid parents(X)) = P(X \mid do(parents(X)))$$

for each variable X, intervening on the parents of X has the same effect as observing them.

Inferring Causality

- A confounder, between X and Y is a variable Z such that:
  - $P(Y \mid X, do(Z)) \neq P(Y \mid X)$

Inferring Causality

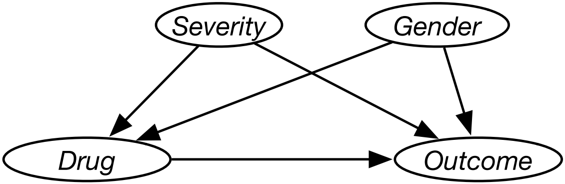
- A confounder, between X and Y is a variable Z such that:
  - $P(Y \mid X, do(Z)) \neq P(Y \mid X)$
  - $P(X \mid do(Z)) \neq P(X)$.

Inferring Causality

- A confounder, between X and Y is a variable Z such that:
  - $P(Y \mid X, do(Z)) \neq P(Y \mid X)$
  - $P(X \mid do(Z)) \neq P(X)$.

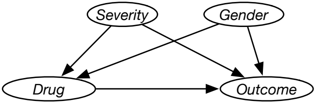
  A confounder can account for the correlation between X and Y by being a common cause of both.

Example



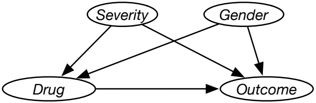
$$P(outcome \mid drug) \neq P(outcome \mid do(drug)).$$

Example



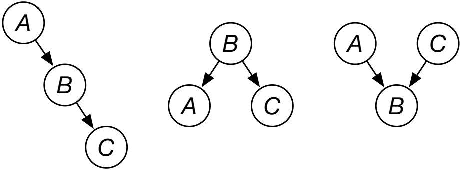
$$
\begin{aligned}
&P(Outcome \mid do(Drug)) \\
&\quad = \sum_{Severity} \sum_{Gender} P(Severity) * P(Gender) \\
&\qquad\qquad * P(Outcome \mid do(Drug), Severity, Gender)
\end{aligned}
$$

Example



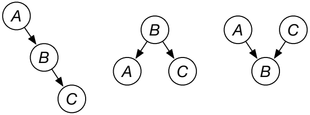
$$
\begin{aligned}
P(Outcome \mid do(Drug)) &= \sum_{Severity} \sum_{Gender} P(Severity) * P(Gender) \\
&\qquad * P(Outcome \mid do(Drug), Severity, Gender) \\
&= \sum_{Severity} \sum_{Gender} P(Severity) * P(Gender) \\
&\qquad * P(Outcome \mid Drug, Severity, Gender)
\end{aligned}
$$

Three types of meetings between arcs



(a) chain (b) fork (c) collider

D-separation



(a) chain (b) fork (c) collider

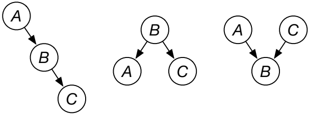
- A path p can follow arrows in either direction.
- Observations Zs block a path p if:

  (a) p contains a chain $A \rightarrow B \rightarrow C$, and $B \in Zs$

  (b) p contains a fork $A \leftarrow B \rightarrow C$, and $B \in Zs$

  (c) p contains a collider $A \rightarrow B \leftarrow C$, and B, and all its descendants, are not in Zs

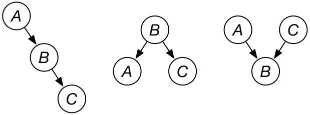
D-separation



(a) chain (b) fork (c) collider

- A path p can follow arrows in either direction.
- Observations Zs block a path p if:
  (a) p contains a chain $A \rightarrow B \rightarrow C$, and $B \in Zs$
  (b) p contains a fork $A \leftarrow B \rightarrow C$, and $B \in Zs$
  (c) p contains a collider $A \rightarrow B \leftarrow C$, and B, and all its descendants, are not in Zs
- X is d-separated from Y given Zs if every path between X and Y is blocked by Zs

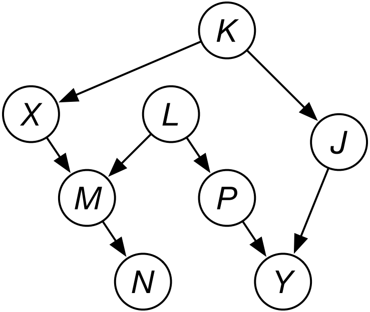
D-separation



(a) chain (b) fork (c) collider

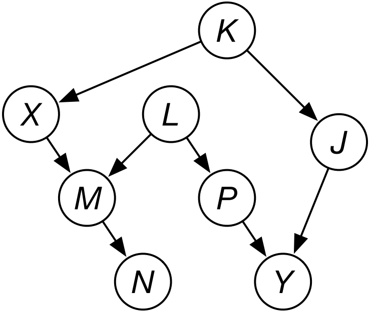
- A path p can follow arrows in either direction.
- Observations Zs block a path p if:
  (a) p contains a chain $A \rightarrow B \rightarrow C$, and $B \in Zs$
  (b) p contains a fork $A \leftarrow B \rightarrow C$, and $B \in Zs$
  (c) p contains a collider $A \rightarrow B \leftarrow C$, and B, and all its descendants, are not in Zs
- X is d-separated from Y given Zs if every path between X and Y is blocked by Zs
- X is independent Y given Zs for all conditional probabilities iff X is d-separated from Y given Zs

Example



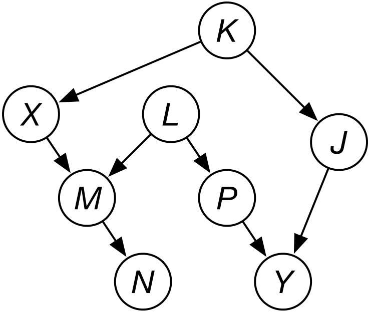
- Are X and Y d-separated by $\{\}$?

Example



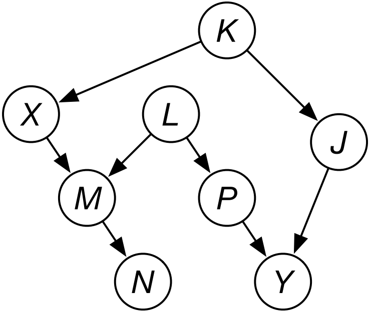
- Are X and Y d-separated by $\{\}$?
- Are X and Y d-separated by $\{K\}$?

Example



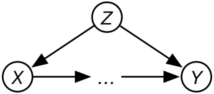
- Are X and Y d-separated by $\{\}$?
- Are X and Y d-separated by $\{K\}$?
- Are X and Y d-separated by $\{K, N\}$?

Example



- Are X and Y d-separated by $\{\}$?
- Are X and Y d-separated by $\{K\}$?
- Are X and Y d-separated by $\{K, N\}$?
- Are X and Y d-separated by $\{K, N, P\}$?

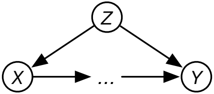
Backdoor criterion



A set of variables Z satisfies the backdoor criterion for X and Y with respect to directed acyclic graph G if

- Z is observed,

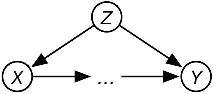
Backdoor criterion



A set of variables Z satisfies the backdoor criterion for X and Y with respect to directed acyclic graph G if

- Z is observed,
- no node in Z is a descendant of X, and

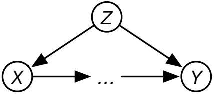
Backdoor criterion



A set of variables Z satisfies the backdoor criterion for X and Y with respect to directed acyclic graph G if

- Z is observed,
- no node in Z is a descendant of X, and
- Z blocks every path between X and Y that contains an arrow into X.

Backdoor criterion

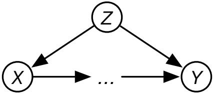


A set of variables Z satisfies the backdoor criterion for X and Y with respect to directed acyclic graph G if

- Z is observed,
- no node in Z is a descendant of X, and
- Z blocks every path between X and Y that contains an arrow into X.

If Z satisfies the backdoor criterion, then

Backdoor criterion



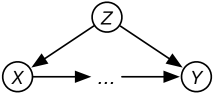
A set of variables Z satisfies the backdoor criterion for X and Y with respect to directed acyclic graph G if

- Z is observed,
- no node in Z is a descendant of X, and
- Z blocks every path between X and Y that contains an arrow into X.

If Z satisfies the backdoor criterion, then

$$P(Y \mid do(X), Z) = P(Y \mid X, Z)$$

Backdoor criterion



A set of variables Z satisfies the backdoor criterion for X and Y with respect to directed acyclic graph G if

- Z is observed,
- no node in Z is a descendant of X, and
- Z blocks every path between X and Y that contains an arrow into X.

If Z satisfies the backdoor criterion, then

$$P(Y \mid do(X), Z) = P(Y \mid X, Z)$$

$$\text{so,} \quad P(Y \mid do(X)) = \sum_{Z} P(Y \mid X, Z) * P(Z)$$

Do-calculus

The do-calculus tells us how probability expressions involving the do-operator can be simplified.

- If Z blocks all of the paths from W to Y in the graph obtained after removing all of the arcs into X, then

$$P(Y \mid do(X), Z, W) = P(Y \mid do(X), Z).$$

  This is d-separation in the manipulated graph.

Do-calculus

The do-calculus tells us how probability expressions involving the do-operator can be simplified.

- If Z blocks all of the paths from W to Y in the graph obtained after removing all of the arcs into X, then

  $$P(Y \mid do(X), Z, W) = P(Y \mid do(X), Z).$$

  This is d-separation in the manipulated graph.

- If Z satisfies the backdoor criterion, for X and Y

  $$P(Y \mid do(X), Z) = P(Y \mid X, Z).$$

  This rule lets us convert an intervention into an observation.

Do-calculus

The do-calculus tells us how probability expressions involving the do-operator can be simplified.

- If Z blocks all of the paths from W to Y in the graph obtained after removing all of the arcs into X, then

  $$P(Y \mid do(X), Z, W) = P(Y \mid do(X), Z).$$

  This is d-separation in the manipulated graph.
- If Z satisfies the backdoor criterion, for X and Y

  $$P(Y \mid do(X), Z) = P(Y \mid X, Z).$$

  This rule lets us convert an intervention into an observation.
- If there are no directed paths from X to Y, or from Y to X:

  $$P(Y \mid do(X)) = P(Y).$$

  This only can be used when there are no observations.

Do-calculus

The do-calculus tells us how probability expressions involving the do-operator can be simplified.

- If Z blocks all of the paths from W to Y in the graph obtained after removing all of the arcs into X, then

$$P(Y \mid do(X), Z, W) = P(Y \mid do(X), Z).$$

  This is d-separation in the manipulated graph.

- If Z satisfies the backdoor criterion, for X and Y

$$P(Y \mid do(X), Z) = P(Y \mid X, Z).$$

  This rule lets us convert an intervention into an observation.

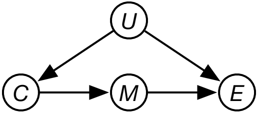
- If there are no directed paths from X to Y, or from Y to X:

$$P(Y \mid do(X)) = P(Y).$$

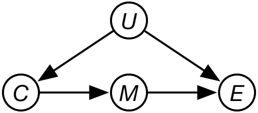
  This only can be used when there are no observations.

These three rules are complete all cases where interventions can be reduced to observations follow from these rules.

Front-door criterion

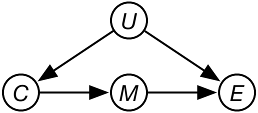


Front-door criterion



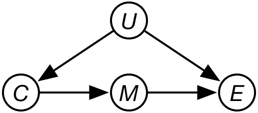
$$P(E \mid do(C)) = \sum_{M} P(E \mid do(C), M) * P(M \mid do(C))$$

Front-door criterion



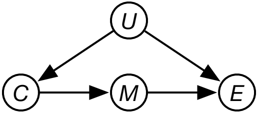
$$
\begin{aligned}
P(E \mid do(C)) &= \sum_M P(E \mid do(C), M) * P(M \mid do(C)) \\
&= \sum_M P(E \mid do(C), do(M)) * P(M \mid do(C))
\end{aligned}
$$

Front-door criterion



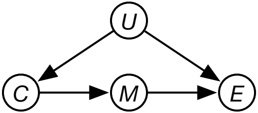
$$
\begin{aligned}
P(E \mid do(C)) &= \sum_{M} P(E \mid do(C), M) * P(M \mid do(C)) \\
&= \sum_{M} P(E \mid do(C), do(M)) * P(M \mid do(C)) \\
&= \sum_{M} P(E \mid do(C), do(M)) * P(M \mid C)
\end{aligned}
$$

Front-door criterion



$$
\begin{aligned}
P(E \mid do(C)) &= \sum_M P(E \mid do(C), M) * P(M \mid do(C)) \\
&= \sum_M P(E \mid do(C), do(M)) * P(M \mid do(C)) \\
&= \sum_M P(E \mid do(C), do(M)) * P(M \mid C) \\
&= \sum_M P(E \mid do(M)) * P(M \mid C)
\end{aligned}
$$

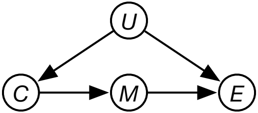
Front-door criterion (Cont.)



From last slide:

$$P(E \mid do(C)) = \sum_M P(E \mid do(M)) * P(M \mid C)$$

Front-door criterion (Cont.)



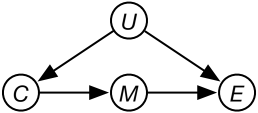
From last slide:

$$P(E \mid do(C)) = \sum_{M} P(E \mid do(M)) * P(M \mid C)$$

C' closes the backdoor between M and E, and there are no backdoors between M and C, so:

$$P(E \mid do(M)) = \sum_{C'} P(E \mid do(M), C') * P(C' \mid do(M))$$

Front-door criterion (Cont.)



From last slide:

$$P(E \mid do(C)) = \sum_{M} P(E \mid do(M)) * P(M \mid C)$$

C' closes the backdoor between M and E, and there are no backdoors between M and C, so:

$$P(E \mid do(M)) = \sum_{C'} P(E \mid do(M), C') * P(C' \mid do(M))$$

So

$$P(E \mid do(C)) = \sum_{M} P(M \mid C) * \sum_{C'} P(E \mid M, C') * P(C').$$

Simpson's Paradox (Revisited)

1000 students, some a particular method for learning a concept (the treatment variable T),
whether they were judged to have understood the concept (evaluation E)
for two subpopulations (one with $C{=}true$ and one with $C{=}false$):

C	T	$E{=}true$	$E{=}false$	Rate
true	*true*	90	10	$90/(90+10) = 90\%$
true	*false*	290	110	$290/(290+110) = 72.5\%$
false	*true*	110	290	$110/(110+290) = 27.5\%$
false	*false*	10	90	$10/(10+90) = 10\%$

Does the treatment increase understanding?

Simpson's Paradox (Revisited)

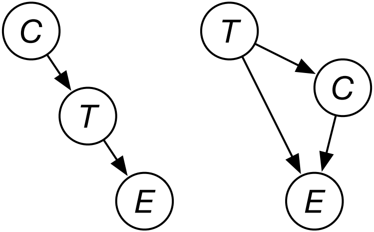
1000 students, some a particular method for learning a concept (the treatment variable T),
whether they were judged to have understood the concept (evaluation E)
for two subpopulations (one with C=*true* and one with C=*false*):

C	T	E=*true*	E=*false*	Rate
true	*true*	90	10	$90/(90+10) = 90\%$
true	*false*	290	110	$290/(290+110) = 72.5\%$
false	*true*	110	290	$110/(110+290) = 27.5\%$
false	*false*	10	90	$10/(10+90) = 10\%$

Does the treatment increase understanding?

T	E=*true*	E=*false*	Rate
true	200	300	$200/(200+300) = 40\%$
false	300	200	$300/(300+200) = 60\%$

Simpson's Paradox



(a) (b)

For each one, should we use subpopulations, or the combined population?

Instrumental Variables

An instrumental variable is a variable that can be used as a surrogate for a variable that is difficult to manipulate.

Instrumental Variables

An instrumental variable is a variable that can be used as a surrogate for a variable that is difficult to manipulate.

Observable or controllable variable Z is an instrumental variable for variable X in predicting Y if:

- Z is independent of the possible confounders between X and Y. One way to ensure independence is to randomize Z.

Instrumental Variables

An instrumental variable is a variable that can be used as a surrogate for a variable that is difficult to manipulate.
Observable or controllable variable Z is an instrumental variable for variable X in predicting Y if:

- Z is independent of the possible confounders between X and Y. One way to ensure independence is to randomize Z.
- Y is independent of Z given X. The only way for Z to affect Y is to affect X.

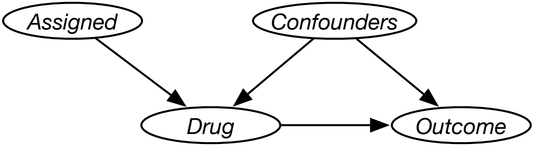
Instrumental Variables

An instrumental variable is a variable that can be used as a surrogate for a variable that is difficult to manipulate.

Observable or controllable variable Z is an instrumental variable for variable X in predicting Y if:

- Z is independent of the possible confounders between X and Y. One way to ensure independence is to randomize Z.
- Y is independent of Z given X. The only way for Z to affect Y is to affect X.
- There is a strong association between Z and X.

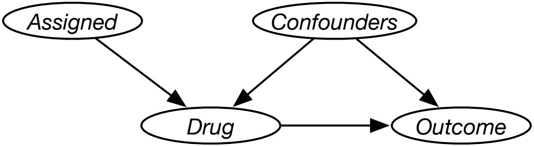
Example



- You want $P(Disease \mid do(Drug))$
- You create a randomized experiment where some people are assigned the drug and some are assigned a placebo.
- However, some people might not take the pill prescribed for them.

The do-calculus does not help here; the propensity to not take the drug might be highly correlated with the outcome.

Example



Assigned	Drug	Outcome	count
true	true	good	300
true	true	bad	50
true	false	good	25
true	false	bad	125
false	true	good	0
false	true	bad	0
false	false	good	100
false	false	bad	400

Example

Assigned	Drug	Outcome	count	
true	true	good	300	
true	true	bad	50	
true	false	good	25	– non-compliers
true	false	bad	125	– non-compliers
false	true	good	0	
false	true	bad	0	
false	false	good	100	
false	false	bad	400	

– If no non-compliers would have good outcome if they took the drug, ___ patients taking the drug would have a good outcome.

Example

Assigned	Drug	Outcome	count	
true	true	good	300	
true	true	bad	50	
true	false	good	25	– non-compliers
true	false	bad	125	– non-compliers
false	true	good	0	
false	true	bad	0	
false	false	good	100	
false	false	bad	400	

– If no non-compliers would have good outcome if they took the drug, 300 patients taking the drug would have a good outcome.

Example

Assigned	Drug	Outcome	count	
true	true	good	300	
true	true	bad	50	
true	false	good	25	– non-compliers
true	false	bad	125	– non-compliers
false	true	good	0	
false	true	bad	0	
false	false	good	100	
false	false	bad	400	

– If no non-compliers would have good outcome if they took the drug, 300 patients taking the drug would have a good outcome.
– If all non-compliers would have good outcome if they took the drug, ___ of the drug-taking patients would have a good outcome.

Example

Assigned	Drug	Outcome	count	
true	true	good	300	
true	true	bad	50	
true	false	good	25	– non-compliers
true	false	bad	125	– non-compliers
false	true	good	0	
false	true	bad	0	
false	false	good	100	
false	false	bad	400	

– If no non-compliers would have good outcome if they took the drug, 300 patients taking the drug would have a good outcome.
– If all non-compliers would have good outcome if they took the drug, 450 of the drug-taking patients would have a good outcome.

Example

Assigned	Drug	Outcome	count	
true	true	good	300	
true	true	bad	50	
true	false	good	25	– non-compliers
true	false	bad	125	– non-compliers
false	true	good	0	
false	true	bad	0	
false	false	good	100	
false	false	bad	400	

– If no non-compliers would have good outcome if they took the drug, 300 patients taking the drug would have a good outcome.
– If all non-compliers would have good outcome if they took the drug, 450 of the drug-taking patients would have a good outcome.

$$0.6 \leq P(Outcome{=}good \mid do(Drug = true)) \leq 0.9$$

$$P(Outcome{=}good \mid do(Drug = false)) = 0.2$$