Causal inference, week 3
Graphical models

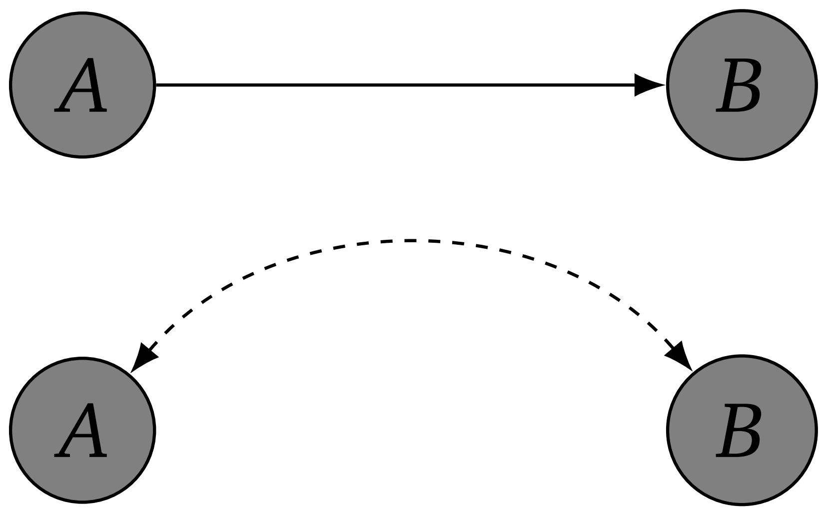
Daniel Stegmueller

Duke University

Graphs

- It can be difficult to think about mechanisms and interventions in terms of counterfactual variables and independence structures
- Sometimes what we need are pictures
- Graphical causal model (Pearl 2000):
  - Edges and directed vertices (with no cycles, DAG)
  - An associated probability distribution that factors according to the graph structure

Graphical models



- A and B are random variables
- Solid arrows indicate causal relations and their conditional probability distributions
- Dotted arrows indicate correlations due to other unobserved variables

Graphical models

- Arrows:
  - are (specified or unspecified) conditional probability relationships
  - may also represent stable mechanisms which give rise to causal effects
- We build and defend causal theories by connecting these basic elements
- Most important aspects of such a theory are where there are no arrows .... and if there are arrows which way they point
- Most importantly: many graph structures are not identifiable from the data alone!

DAGs and potential outcomes

- Concepts such as 'Confounding', 'selection on observables', 'mediation' are particular **graph structures**
- 'Experimental intervention' represents a **graph surgery**
- 'Controlling for' represents statistical **conditioning** (separate analyses by a third variable)
- Graphs allow us to explore more complicated variable relationships
- Graphical and the potential outcomes approach to causation are provably equivalent (Pearl 2000, ch.7)

Causal relationships in DAGs

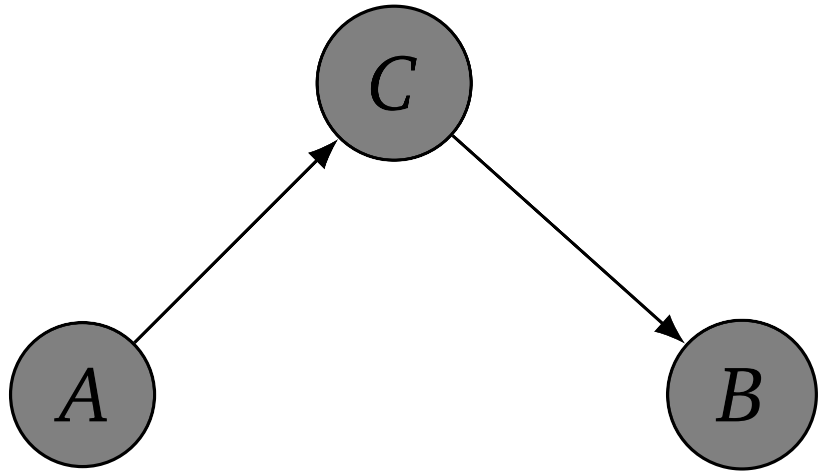
Consider three minimal stylized casual structures involving A, B and C

In the following assume that all variables are observable

Say we are (mostly) interested in the causal influence of A on B

Crucial question: when should we condition on (or experimentally manipulate) C in an analysis?

Causal relationships I: mediation



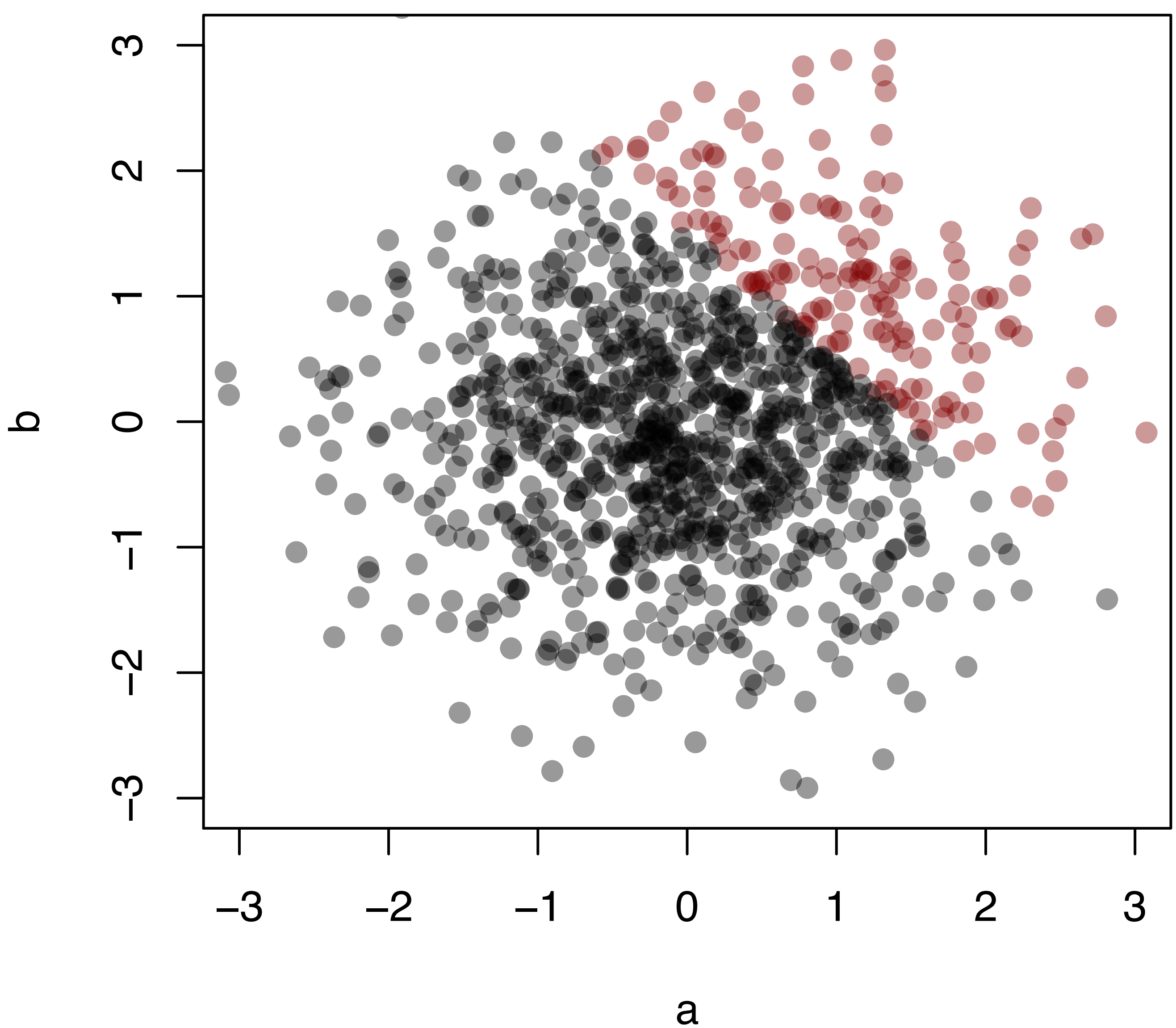
If interested in the causal effect of A on B, should we condition on C?

Example

Example: ability, civil service, money

Assume (optimistically or pessimistically?):

- ▷ A is ability, C is getting into the civil service, B is money
- ▷ A is normal with mean 0
- ▷ $C = 1$ if $A > 1$ and $C = 0$ otherwise
- ▷ B is normal with mean 0 if $C = 0$ and mean 2 if $C = 1$

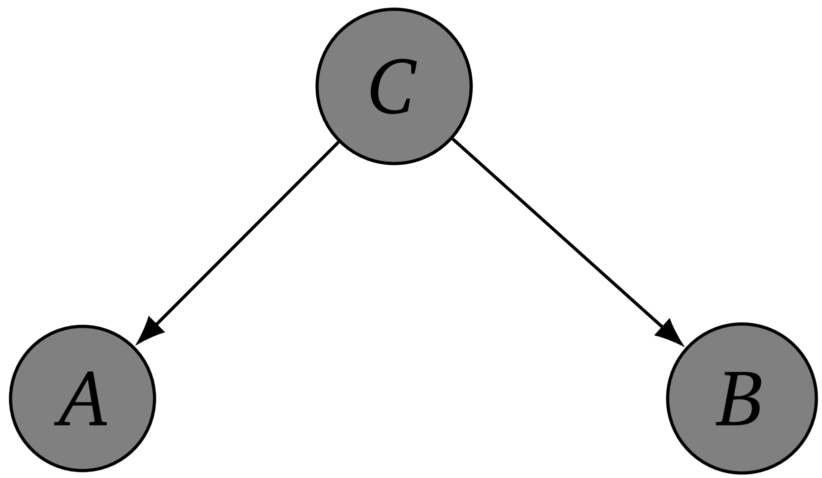


b
a
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3

Example

- In simulated data:
  - Correlation of A and B: 0.38
  - Correlation for A and B when C=1 (civil): -0.08
  - Correlation for A and B when C=0 (no civil): 0.03
- A and B are *marginally dependent* but *conditionally independent* given C
- C ‘screens off’ A from B

Causal relationships II: common cause / mutual dependence



Does A cause B?

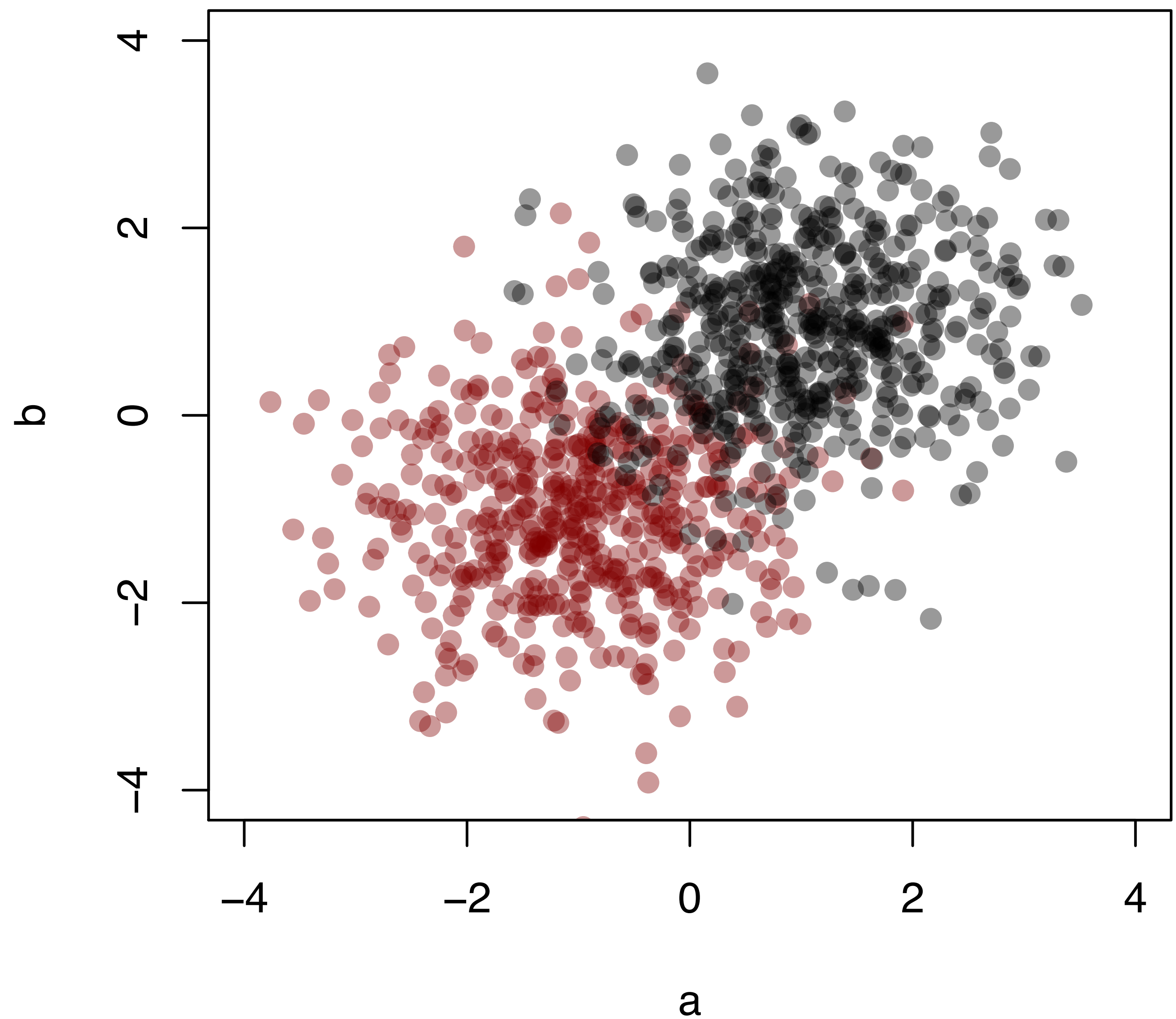
If we wanted to estimate the causal effect of A on B, should we condition on/manipulate C?

Example

Example: work, results, and ability

Assume:

- A is hard work, B is results and C is ability (0=low, 1=high)
- If $C = 0$ then $E[A] = E[B] = -1$
- If $C = 1$ then $E[A] = E[B] = 1$
- Thus, only ability 'matters' $\rightarrow$ A and B are uncorrelated at each ability level

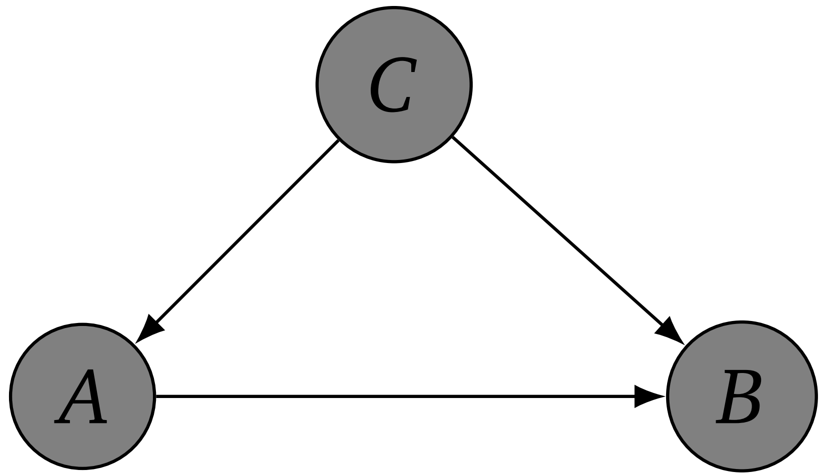

4
2
0
−2
−4
b
−4
−2
0
2
4
a

Example

- In simulated data:
  - Correlation of A and B: 0.54
  - Correlation for A and B when C=0 (low): 0.07
  - Correlation for A and B when C=1 (high): 0.006
- A and B are *marginally dependent* but *conditionally independent* given C

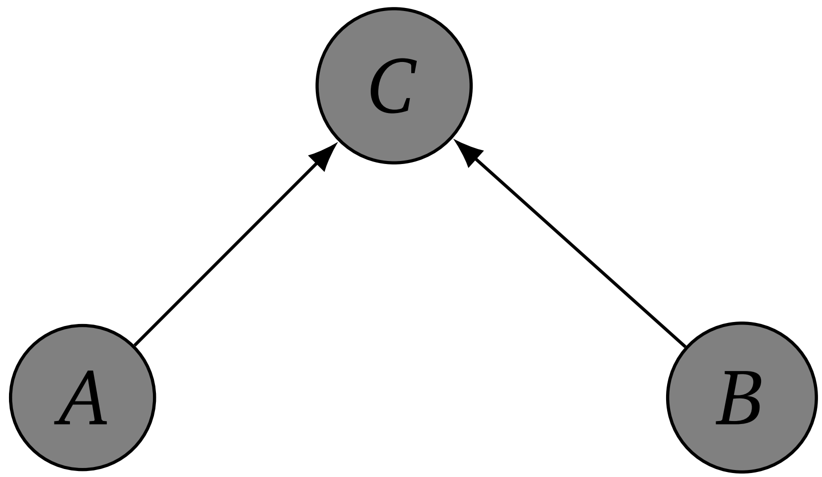
e.g., Fisher: Smoking $\leftarrow$ Genetics $\rightarrow$ Cancer

Of course the natural application of this observation is to the following situation:



where the issue is disentangling the real causal effect (A $\rightarrow$ B) from the the common cause

Causal relationships III: mutual causation



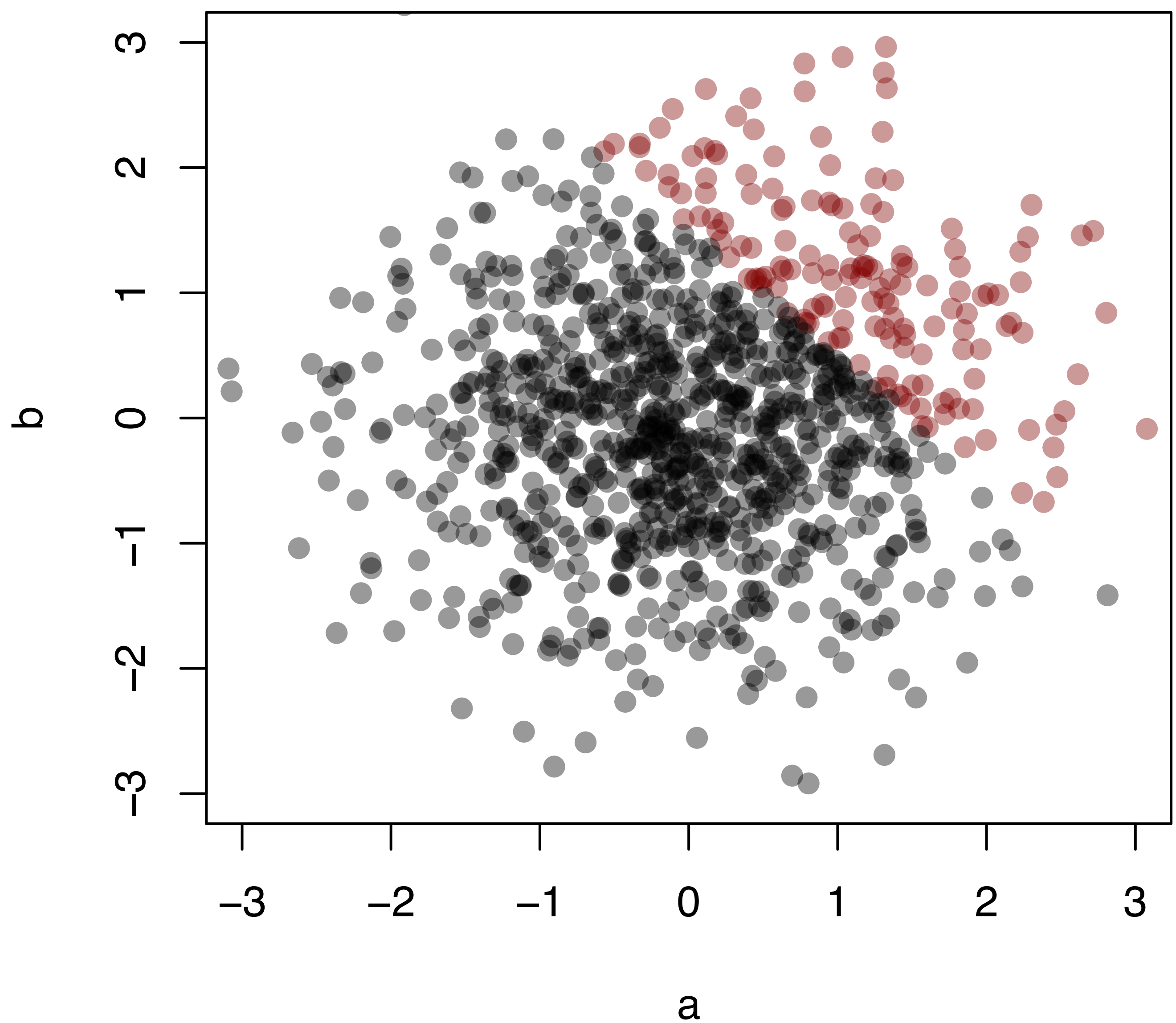
If we were interested in the causal effect of A on B, should we condition on/manipulate C?

Example

Example: test performance

Assume:

- ▹ A and B are normal with mean 0, e.g., speaking and writing abilities
- ▹ C is a test result: $C = 1$ (pass) if $A + B > 1.5$ else $C = 0$ (fail)



b
a
3
2
1
0
−1
−2
−3

Example

- In simulated data:
  - Correlation of A and B: approx. 0
  - Correlation of A and B for $C = 1$ (pass): -0.65
  - Correlation of A and B for $C = 0$ (fail): -0.19
- A and B are *marginally independent* but ***made dependent by conditioning on C***
- Conditioning on C **creates** an association

In the structure $A \rightarrow C \leftarrow B$, C is called a *collider variable* by Pearl

Note: any endogenous variable with more than one cause is a collider

Summa summarum

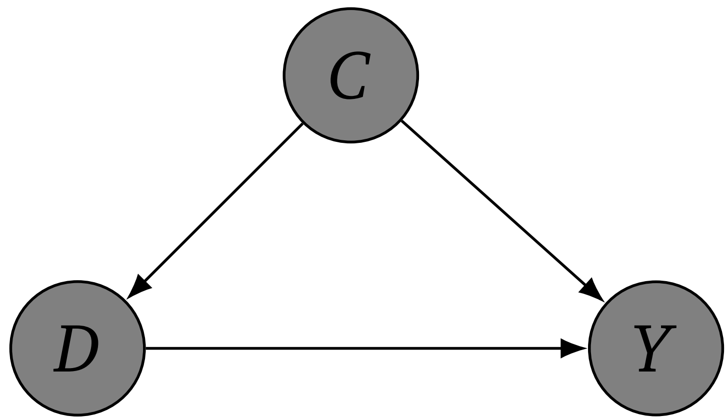
- Nothing said above
  - depends on any parametric assumptions (e.g., linear model)
  - would be discerned by inspecting three variables
- Conditioning on C can remove or create (or leave alone) associations between A and B, depending on the causal graph structure.

Summa summarum

- ‘Confounding’ variables generate spurious associations – we only want the association corresponding to the causal effect
- If we are interested in identifying A $\rightarrow$ B and we have measured C:
  - When there is A $\rightarrow$ C $\rightarrow$ B, condition on C (unless you want the total effect of A on B, a different quantity)
  - When there is A $\leftarrow$ C $\rightarrow$ B, condition on C (ignoring C adds extra spurious association a.k.a confounding)
  - When there is A $\rightarrow$ C $\leftarrow$ B, do **not** condition on C, or its children (conditioning on C creates extra spurious association)

What about counterfactuals?

- So far, we have said nothing explicit about counterfactual states
- What is the equivalent of potential outcomes in the graphical language
- Consider this simple example



- Assume (for simplicity) that D takes on two states, 0, 1

- Pearl distinguishes two regimes changing D
  - *Before-intervention regime*: the value of D is determined by

$$D = f_D(C, e_D)$$

  - *Under-intervention regime:* value of D is determined by manipulation (or "atomic intervention"),

$$do(D = 1), \text{ or } do(D = 0)$$

- Two probability distributions of causal effects are thus

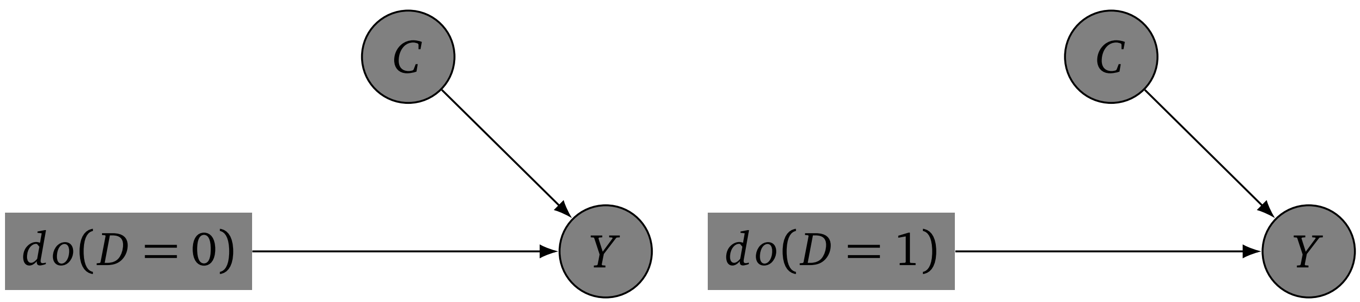
$$Pr(Y|do(D = 1)), \; Pr(Y|do(D = 0))$$

- Note that in general these quantities can be quite different from empirical quantities such as $Pr(Y|D = 1)$

- The average causal effect is

$$E(Y|do(D=1)) - E(Y|do(D=0))$$

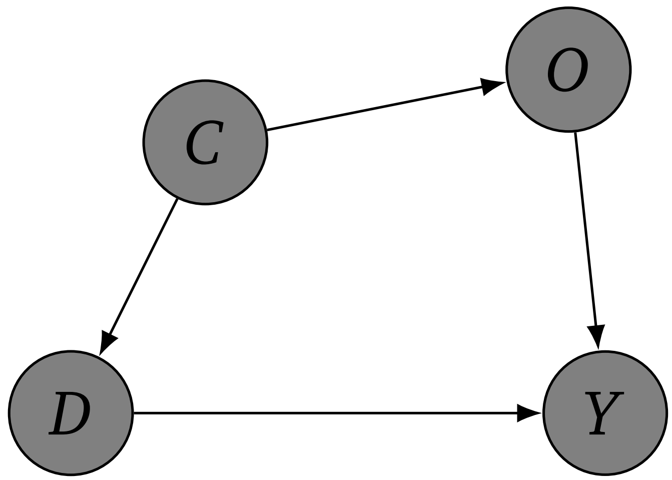
- Thus, the $do(\cdot)$ operator is the equivalent to the potential state notation (subscripts, superscripts etc) that we have introduced before
- It designates the relevant causal states and makes clear what is being varies and what is being held constant



- Again, this causal quantity is different from the empirical difference $E(Y|D=1) - E(Y|D=0)$ (which is not the average causal effect due to dependence on C)

Back-door paths

- Consider the following causal structure



- Here $D \to Y$ is confounded by the "back-door" path $D \leftarrow C \to O \to Y$
- What should you condition on?
- A back-door path is a path between a causally ordered sequence of two variables (say, D and Y) that begins with a directed edge that points to the first variable

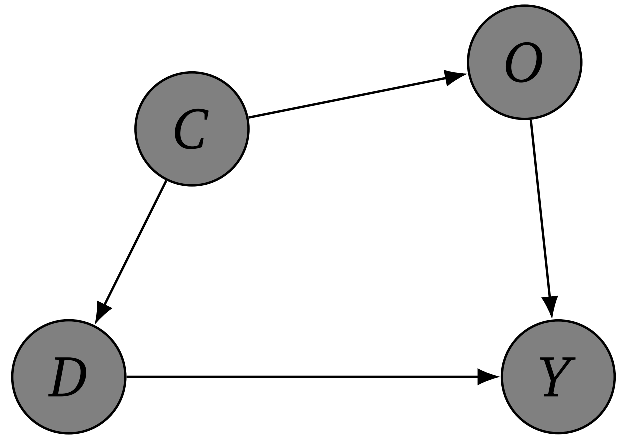
The back-door criterion

- Mechanical rules for determining if a (set of) variable(s) should be conditioned on
- Goal: *block* all paths that generate noncausal associations

Criterion conditions:

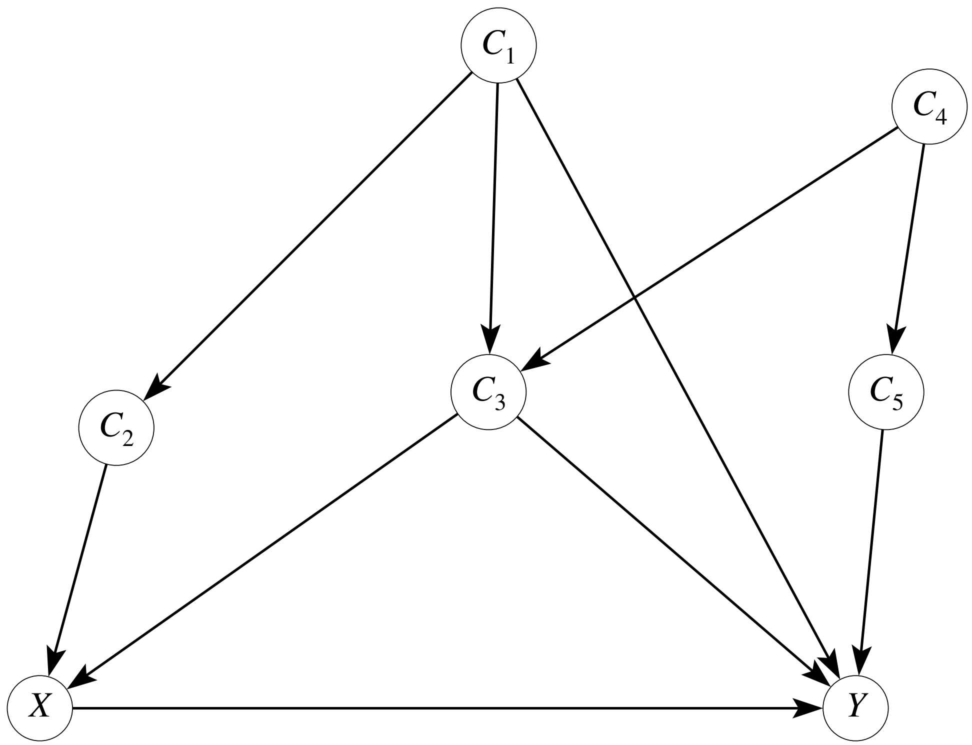
1. All back-door paths between causal variable and outcome are blocked after conditioning on set Z, which will ve true if each back-door path
   (a) contains a mediation chain, $A \rightarrow C \rightarrow B$ with C in Z **OR**
   (b) contains mutual dependence/common cause chain $A \leftarrow C \rightarrow B$ with C in Z **OR**
   (c) contains a mutual causation chain $A \rightarrow C \leftarrow B$ with C and all of its descendants *not* in Z
2. No variables in Z are descendants of the causal variable that lie on any of the directed paths beginning at the causal variable and reach the outcome

In our example



- Path $D \leftarrow C \rightarrow O \rightarrow Y$ fits both 1a and 1b.
- Thus, condition on C or O satisfies part 1 of the criterion (part 2 is trivially satisfied in a simple model like this)
- Criterion 1c (a.k.a."no included colliders!!") is more sophisticated. We will return to a prominent example in our discussion of panel data

Applied example 1



$$P_X(Y) = \sum_{C_1, C_3} P(Y \mid X, C_1, C_3) P(C_1, C_3)$$

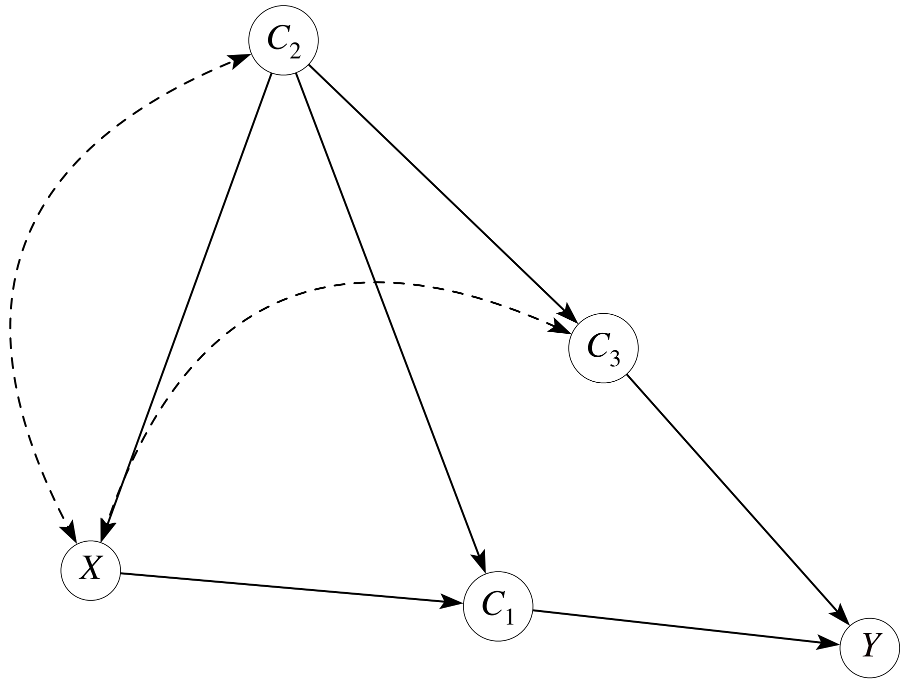
Other possible conditioning set: $\{C2, C3\}$

Why?

$X \leftarrow C_2 \leftarrow C_1 \rightarrow Y$	Cond. on C_1
$X \leftarrow C_2 \leftarrow C_1 \rightarrow C_3 \rightarrow Y$	Cond. on C_1 or C_3
$X \leftarrow C_3 \leftarrow C_1 \rightarrow Y$	Cond. on C_1 or C_3
$X \leftarrow C_3 \rightarrow Y$	Cond on C_3
$X \leftarrow C_3 \leftarrow C_4 \rightarrow C_5 \rightarrow Y$	Cond. on C_3

Can't drop $C1$ or $C3$ without opening backdoor paths.

Applied example 2



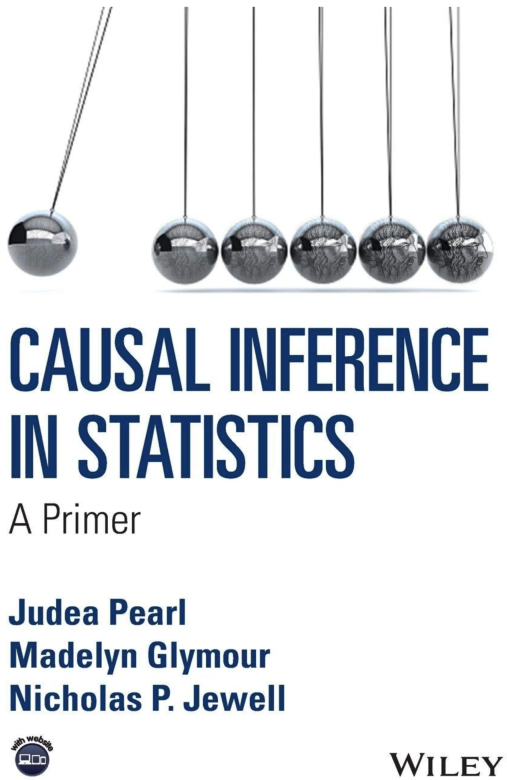
Minimal conditioning set is $\{C2, C3\}$.

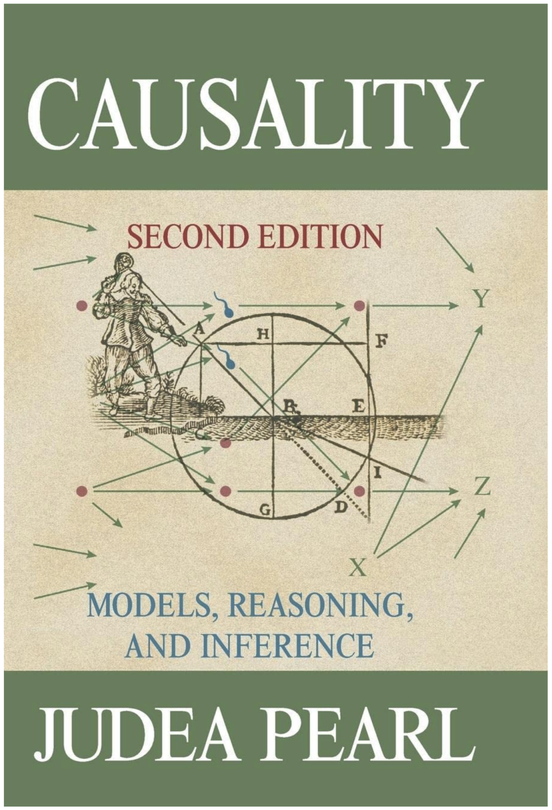
BTW, ChatGPT5, Sept 9, 2025:

✅ **Answer:**

For this DAG, the **minimal admissible adjustment set is** $\{C_2\}$.

Read





Talk to me for more literature...