GRAPHICAL REPRESENTATION OF CAUSAL EFFECTS

Alexander Quispe

November 5, 2021

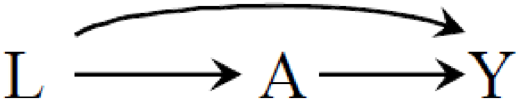
Causal Diagrams

- This lecture introduces a graphical tool to represent our qualitative expert knowledge and a priori assumptions about the causal structure of interest.
- Graphs help clarify conceptual problems and enhance communication among investigators.
- The use of graphs in causal inference problems makes it easier to follow a sensible advice: draw your assumptions before your conclusions.

Causal diagrams

- Figure comprises three nodes representing random variables (L, A, Y) and three edges (the arrows).
- Time flows from left to right, and thus L is temporally prior to A and Y , and A is temporally prior to Y.
- L, A, Y represent disease severity, heart transplant, and death, respectively.

Figure: 1



Directed Acyclic Graphs

- **“Directed”** because the edges imply a direction: because the arrow from L to A is into A, L may cause A, but not the other way around.
- **“Acyclic”** because there are no cycles: a variable can not cause itself, either directly or through another variable
- **Suppose in our study individuals are randomly assigned to heart transplant A with a probability that depends on the severity of their disease L. Then L is a common cause of A and Y**

Causal DAG

- Now all individuals are randomly assigned to heart transplant with the same probability regardless of their disease severity.
- Then L is not a common cause of A and Y and need not be included in the causal diagram.
- represents a conditionally randomized experiment

Figure: 2

$$A \longrightarrow Y$$

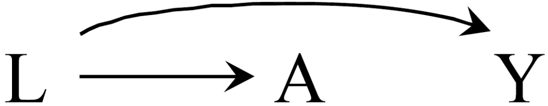
Causal DAG

- Figure 1 represents an observational study in which we are willing to assume that the assignment of heart transplant A has as parent disease severity L and no other causes of Y. This also represents a conditionally randomized experiment.
- Figure 2 represents a marginally randomized experiment.

Causal diagrams and marginal independence

- Lets analyse figure 3, which is an observational study.
- A is carrying a lighter, L cigarette smoking and Y is lung cancer.
- We know that carrying a lighter A has no causal effect on lung cancer Y.
- MAIN QUESTION = carrying a lighter A is associated with lung cancer Y?
- **how do you attack this research question?**

Figure: 3



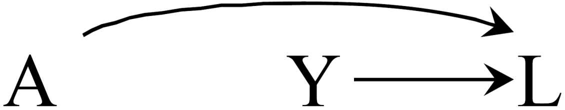
Causal diagrams and marginal independence

- Having information about the treatment A improves our ability to predict the outcome Y, even though A does not have a causal effect on Y.
- The investigator will make a mistake if he concludes that A has a causal effect on Y just because A and Y are associated.

Causal diagrams and marginal independence

- Suppose you know that certain genetic haplotype A has no causal effect on anyone's risk of becoming a cigarette smoker Y
- Both the haplotype A and cigarette smoking Y have a causal effect on the risk of heart disease L.
- The common effect L is referred to as a ***collider*** on the path $A \rightarrow L \leftarrow Y$
- **Question** : **A and Y are associated?**

Figure: 4

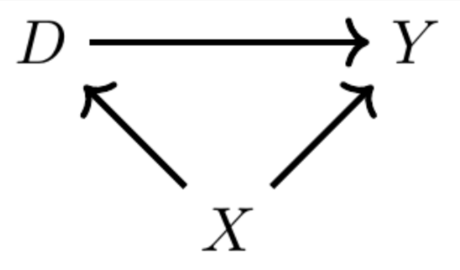


Causal diagrams and marginal independence

- A and Y are independent.
- The knowledge that both A and Y cause heart disease L is irrelevant when considering the association between A and Y.
- Colliders block the flow of association along the path on which they lie. Thus A and Y are independent because the only path between them, $A \rightarrow L \leftarrow Y$, is blocked by the collider L.

Cofounder

Figure: 5



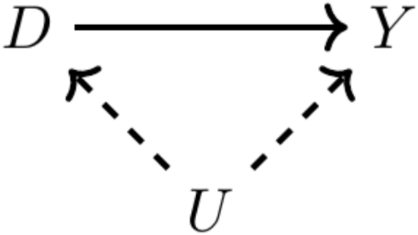
- **Direct Path** = $D \rightarrow Y$, *this is causal*
- **Backdor Path** = $D \leftarrow X \rightarrow Y$, *this is not causal*
- **Backdoor Path** creates spurious correlations between D and Y

Backdoor Path

- The most important things we can learn from the DAG
- Similar to the notion of omitted variable bias in that it represents a variable that determines the outcome and the treatment variable
- Just as not controlling for a variable like that in a regression creates omitted variable bias, **leaving a backdoor open creates bias.**

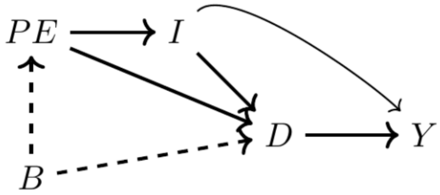
Confounder is unobservable

Figure: 6



Human capital model - DAG

Figure: 4



- D be the treatment (e.g., college education)
- Y be the outcome of interest (e.g., earnings).
- PE be parental education
- I be family income
- B be unobserved background factors, such as genetics, family environment, and mental ability.

Human capital model - DAG

Figure: 8

1. $D \rightarrow Y$ (the causal effect of education on earnings)
2. $D \leftarrow I \rightarrow Y$ (backdoor path 1)
3. $D \leftarrow PE \rightarrow I \rightarrow Y$ (backdoor path 2)
4. $D \leftarrow B \rightarrow PE \rightarrow I \rightarrow Y$ (backdoor path 3)

The problem with open backdoor paths is that they create systematic and independent correlations between D and Y

Backdoor criterion

You, as researcher, have the goal of closing these backdoor paths. And if we can close all of the otherwise open backdoor paths, then we can isolate the causal effect of D on Y.
SOLUTIONS

- you can close that path by **conditioning** on the confounder. **"Controlling for"** the variable in a regression.
- By **not conditioning on a collider**, you will have closed that backdoor path and that takes you closer to your larger ambition to isolate some causal effect.

When **all backdoor paths have been closed**, we say that you have come up with a research design that satisfies the **backdoor criterion**.

Backdoor criterion

Figure: 9

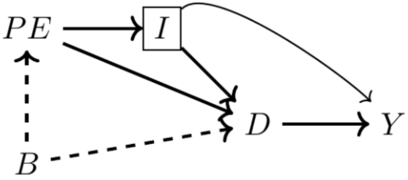


Figure: 10

$$Y_i = \alpha + \delta D_i + \beta I_i + \varepsilon_i$$