Lecture 4

Alex Stephenson

9-1-2021

What did we do last time?

- Defined Causal Inference
- Introduced Potential Outcomes
- Introduced our first estimand

What are we doing today?

- Continue our discussion of SUTVA
- Discuss DAGs
- Begin discussing what makes a good estimator

Stability Assumptions

In order to get a situation where we have two potential outcomes, it must be the case that potential outcomes depend **solely** on whether a unit **itself** receives a singular **treatment**.

Solely is an *exclusion* restriction assumption

Itself is a *non-interference* assumption

Treatment here means that there are not different levels or different kind of treatments.

Stability Assumptions: Excludability

That means that the treatment received d_i must be separated from the treatment group assigned. We call the latter z_i where $z_i \in 0, 1$ is the observed group.

Formally:

$$Y_i(z_i = 1, d_i) = Y_i(z_i = 0, d_i)$$

In words: The value of z is irrelevant to the potential outcomes. Potential outcomes only respond to the treatment.

No interference

No interference means that a unit's potential outcomes are not affected by the treatment applied to a different unit. It also means that the treatment is the same for all units that receive treatment.

Consider our example about the effect of the coal industry on local government. If the coal industry in one county affects the decisions made by a local government in a different county $\rightarrow$ interference is present.

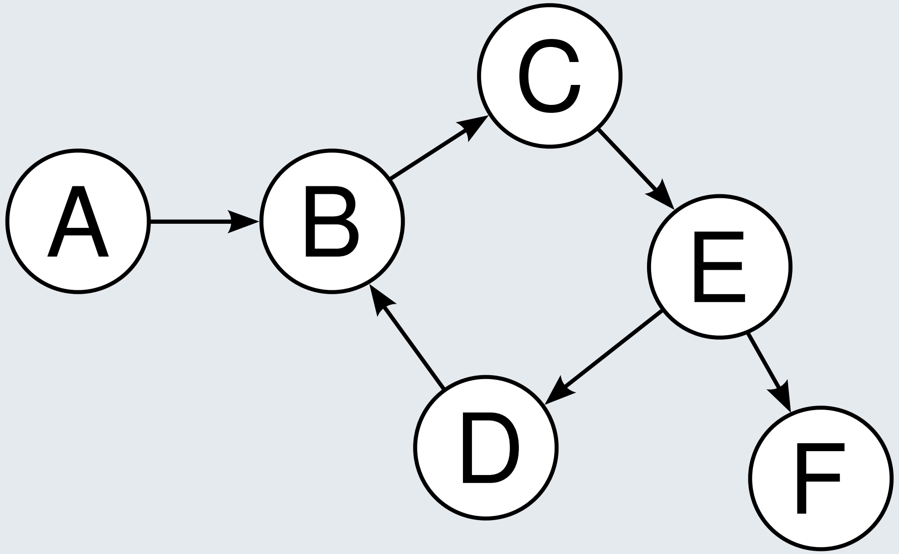
Formally $Y_i(\mathbf{d}) = Y_i(d)$ where $\mathbf{d}$ describes all treatments administered to every unit

DAG

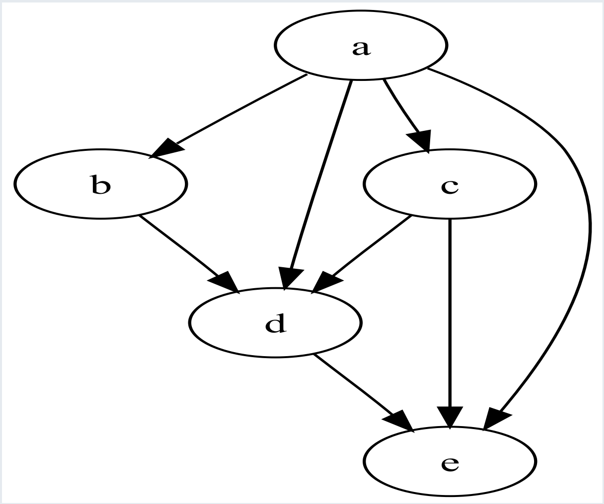
A Directed Acyclic Graph (DAG) is a graphical representation of a chain of causal effects

Use nodes to represent random variables and arrows to represent the causal effect between variables

DAG



Directed Cyclic Graph



Directed Acyclic Graph

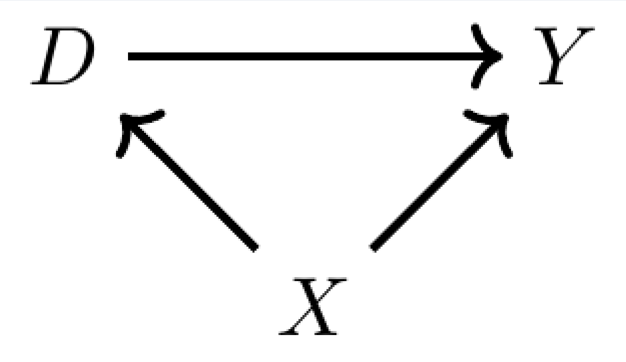
"Reading" DAGs

There are three variables.

There is a direct path from D $\rightarrow$ Y

There is a *backdoor* path from

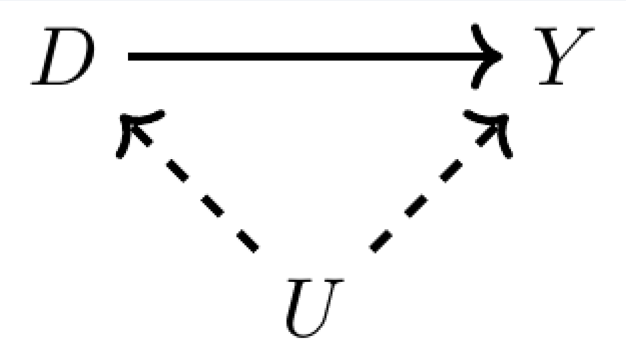
D $\leftarrow$ X $\rightarrow$ Y



Reading DAGs

This is an example of a DAG with an unobserved variable U

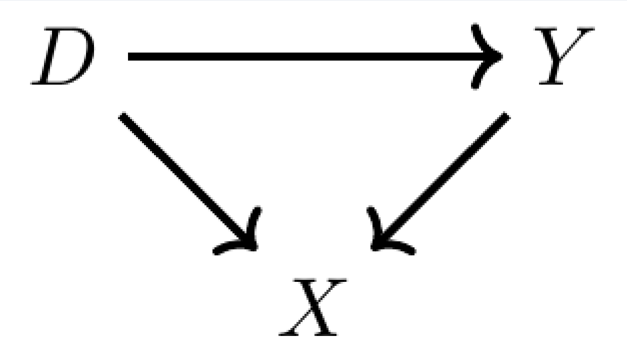
This is a confounder we don't know about and do not observe.



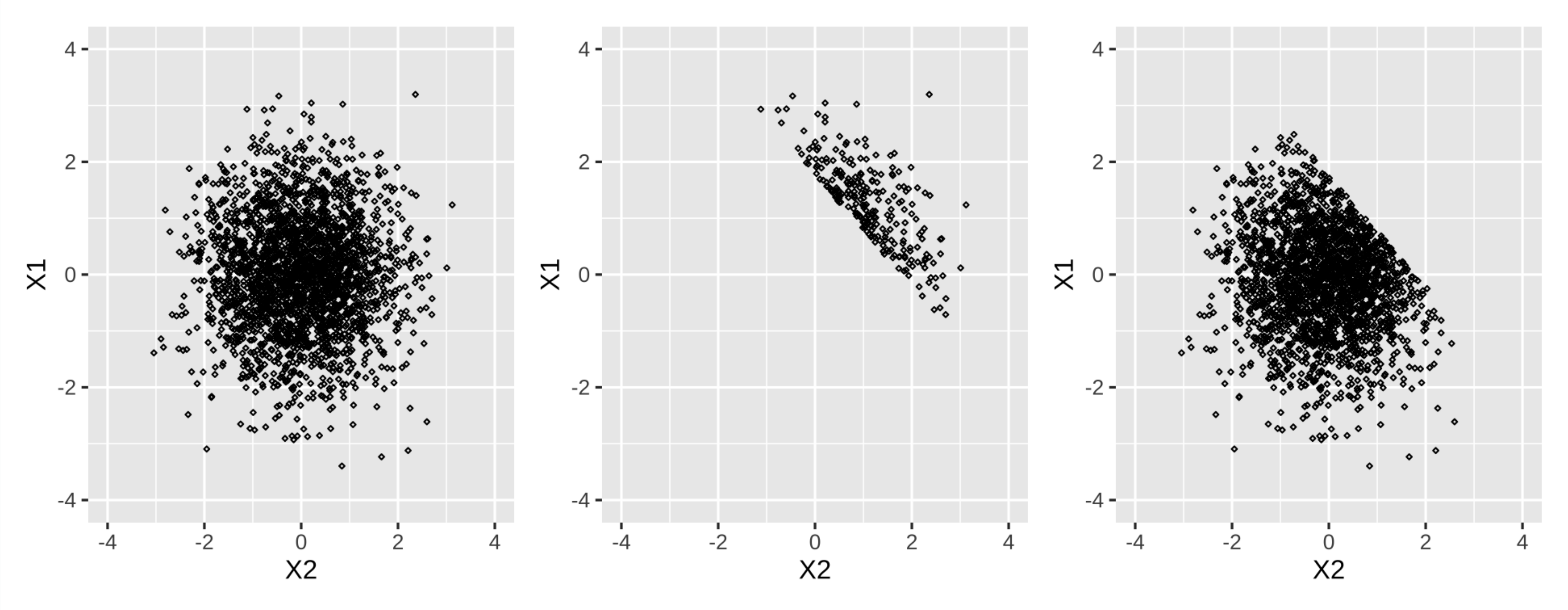
Reading DAGs

Here is an example of a collider

In this case, the causal path is the same but the backdoor path arrows both point to X



Collider Bias



Estimators

An estimator is the procedure that generates a guess about the estimand

Example: Our estimand is the

$$ATE = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

With random sampling the plug-in estimator is the difference in means

$$\hat{E}_{DM} = \hat{E}[Y_i(1)] - \hat{E}[Y_i(0)]$$

Estimators

There are three properties of an estimator that we care about when considering whether it is good.

Consistency

Unbiasedness

Precision

Estimators

There are three properties of an estimator that we care about when considering whether it is good.

Consistency: If we had enough data, the probability of our estimate would be far from the estimand goes to 0.

Unbiasedness

Precision

Estimators

There are three properties of an estimator that we care about when considering whether it is good.

Consistency: If we had enough data, the probability of our estimate would be far from the estimand goes to 0.

Unbiasedness: The difference between the expected value of an estimator and the true value of the estimand in 0

Precision

Estimators

There are three properties of an estimator that we care about when considering whether it is good.

Consistency: If we had enough data, the probability of our estimate would be far from the estimand goes to 0.

Unbiasedness: The difference between the expected value of an estimator and the true value of the estimand in 0

Precision An estimator is more precise than another estimator if it produces estimates closer to the estimand on average

Estimators

The most common metric for an estimator's precision is Mean Square Error

Formally for an estimator $\hat{\theta}$

$$MSE = E[(\hat{\theta} - \theta)^2] = V[\hat{\theta}] + (E[\hat{\theta}] - \theta)^2$$

An estimator with a lower MSE than another estimator is more efficient

What are we doing on Friday

Fill in the reason why randomization is great

Give an overview of common estimators