Directed Acyclic Graphs

A. Colin Cameron
Univ. of California, Davis

These slides are part of the set of slides
A. Colin Cameron, Introduction to Causal Methods
https://cameron.econ.ucdavis.edu/causal/

March 2023

Introduction

- These slides present an introduction to graphs for causal relationships.
- Directed acyclic graphs (DAGs) present all the paths from a causal variable D to outcome Y, including the role of any intermediate variables.
- The introduction here is very brief. For more detail see
  - Cunningham (2021), chapter 3, provides a good introduction
  - or Morgan and Winship (2015), chapters 1.4, 3 and 4.
  - Judea Pearl (2009), *Causality*, Cambridge University Press, is the main reference.

Outline

Directed Acyclic Graphs

- Directed graphs use arrows to show direction of causation.
  - A $\longrightarrow$ B means A causes B
  - A $\longleftrightarrow$ B means A causes B and B causes A (there is feedback).
- We consider directed acylic graphs (DAGs) that rule out cycles such as

$$\begin{tikzcd} A \arrow[rr] & & B \arrow[dl] \\ & C \arrow[ul] & \end{tikzcd}$$

- DAGs do not handle all cases of causality
  - they do not handle reverse causality
  - they do not handle simultaneity, such as the demand-supply model
- But they do handle most applications in modern microeconometrics.

Backdoor Path and Confounder

- We are interested in the causal effect of D on Y: $D \longrightarrow Y$.
- A **backdoor path** is a variable that affects both D and Y

$$\begin{tikzcd} D \arrow[rr] & & Y \\ & X \arrow[ul] \arrow[ur] & \end{tikzcd}$$

- D is associated with Y for two reasons
  - a direct causal path $D \longrightarrow Y$
  - an indirect or backdoor path as different values of X lead to changes in both D and Y.
- X is called a **confounder** as it confounds the causal effect $D \longrightarrow Y$.
- If we appropriately control for X we can estimate $D \longrightarrow Y$
  - e.g. calculate size of effect $D \longrightarrow Y$ for each value of X and average.
- More generally we should close (control for) any backdoor paths.

Collider

- Again we are interested in measuring $D \longrightarrow Y$.
- But now suppose the arrows for X are reversed.

$$\begin{tikzcd} D \arrow[rr] \arrow[dr] & & Y \arrow[dl] \\ & X & \end{tikzcd}$$

- X is called a **collider**
  - when two variables cause a third variable along a path, the third variable is called a collider.
- If we try to control for the collider we introduce bias
  - because we have both direct $D \longrightarrow Y$ plus the effect through X.
- We should ignore X and just estimate $D \longrightarrow Y$.

Unobservables

- To date we have ignored unobservables (errors)
  - Let U denote an unobservable.
- The following causes no problems and is the usual assumption for OLS

$$\begin{tikzcd} D \arrow[rr] & & Y \\ & U \arrow[ur] & \end{tikzcd}$$

- The following leads to biased estimation

$$\begin{tikzcd} D \arrow[rr] & & Y \\ & U \arrow[ul] \arrow[ur] & \end{tikzcd}$$

- Now there is a backdoor via the confounder U
  - but we can't control for U as it is not observed.

Instrumental Variables

- Continuing the previous example we suppose there is a variable Z (called an instrument) with

$$\begin{array}{ccccccc} Z & \longrightarrow & D & & \longrightarrow & & Y \\ & & & \nwarrow & & \nearrow & \\ & & & & U & & \end{array}$$

- So Z does not directly effect Y (exclusion restriction)
- And Z does directly effect D (relevance).
- Then we can estimate by instrumental variables regression.

More complicated examples

- In practice examples are more complicated.
- It is very helpful to write out a full model with arrows for your particular applications.
- Any backdoor paths then need to be closed.
- Any collider variables should not be included in the analysis.

More complicated example

- Based on Cunningham (2021) pages 100-103.
  - $X1 \longrightarrow D$ and $X1 \longrightarrow Y$
  - the additional variable $X2 \longrightarrow D$ and $X2 \longrightarrow X1$
  - the unobservable U does not cause problems as only path is $U \longrightarrow Y$.
  - there are no colliders

$$
\begin{tikzcd}
X2 \arrow[r] \arrow[dr] & X1 \arrow[d] \arrow[dr] & & \\
& D \arrow[r] & Y & U \arrow[l]
\end{tikzcd}
$$

- The paths between D and Y are
  - $D \longrightarrow Y$
  - $D \longleftarrow X1 \longrightarrow Y$
  - $D \longleftarrow X2 \longrightarrow X1 \longrightarrow Y$
  - $D \longleftarrow U \longrightarrow X2 \longrightarrow X1 \longrightarrow Y$
- There are three backdoor paths (none involving a collider).
- These can all be closed by controlling for $X1$.

References for DAGs

- These books are given in approximate order of increasing difficulty.
- Cunningham, Scott (2021), Causal Inference: The MixTape, Yale University Press, especially chapter 3.
- Stephen L. Morgan and Christopher Winship (2014), Counterfactuals and Causal Inference: Methods and Principles for Social Research, Second edition, Cambridge University Press, especially chapters 1.4, 3 and 4.
- Judea Pearl (2009), *Causality*, Cambridge University Press,