STATISTICAL HORIZONS

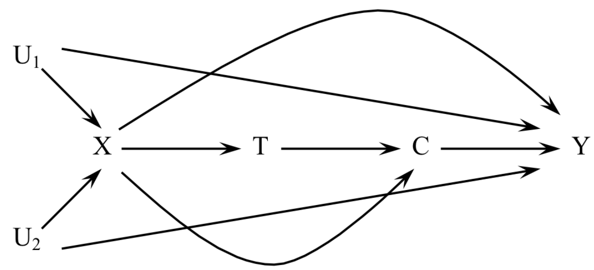
Directed Acyclic Graphs for Causal Inference

Felix Elwert, Ph.D.

Upcoming Seminar:

October 18-19, 2019, Philadelphia, Pennsylvania

What's a Direct Acyclic Graph (DAG)?



This is a DAG.

- Developed by Pearl (1988, 1995, 2009) and colleagues in computer science
- Origins in structural equation models (1920+), and Bayesian networks (1980+)
- Compatible with potential-outcomes framework of causality (Rubin 1974)

Main Uses of DAGs

Fact: All causal claims are relative to an assumed data generating process (DGP).

DAGs help with causal inference because they:

1. Graphically notate the causal assumptions about the DGP
2. Link causal assumptions to statistical associations that one can see in data
3. Support identification analysis
4. Inform key statistical topics (e.g., matching, regression, instrumental variables, mediation analysis, missing values, network analysis)

Course Procedure

Lectures interspersed with exercises

Day 1: Central concepts & understanding identification

Day 2: Link to estimation, examples, and advanced topics.

We'll handle the schedule flexibly to prioritize your interests.

Ask questions whenever you want.

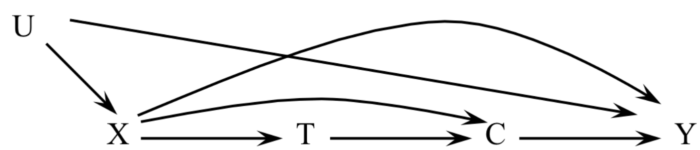
Course Outline

1. Brief intro to counterfactual causality
2. DAGs: Essential elements
3. Testable implications of a model
4. Graphical identification criteria
5. Endogenous selection bias
6. Identification by "adjustment"
7. Causal mediation analysis

1. Elements & Interpretation

Elements

DAGs encode the analyst's causal (structural) assumptions about the DGP.
→ They are a "picture of how the world works."



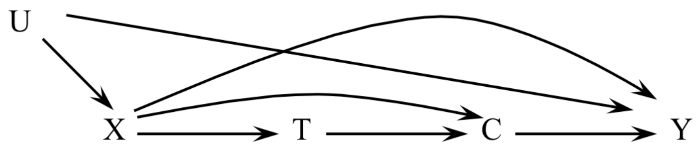
1. Nodes represent variables (observed and unobserved)

2. Arrows represent direct causal effects

3. Missing arrows represent exclusion restrictions (absent direct causal effects)

Missing arrows represent the assumptions. No exclusion, no identification!

Illustration



This DAG might capture the following qualitative DGP:

Smoking causes cancer: T→C

Cancer causes mortality: C→Y

Smoking causes mortality only via cancer: No arrow T→Y

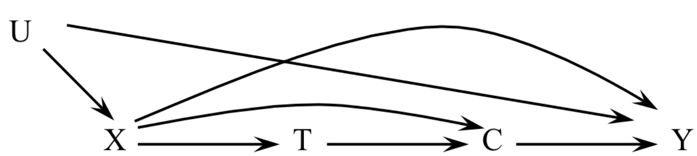
Low socio-economic status causes smoking, cancer, and mortality:
X→T, X→C, and X→Y

Childhood insults to health cause low SES and mortality: U→X and U→Y

If you do not believe this DGP, then you must change the DAG.

Remember: All causal claims are relative to the assumed DGP.

DAGs are Nonparametric



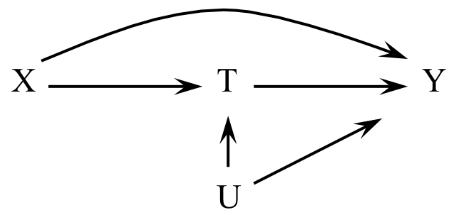
DAGs are very general tools. DAGs make no parametric assumptions about the DGP.

1. No distributional assumptions about variables (nodes)
2. No functional form assumption about causal effects (arrows)

DAGs may look like conventional linear path models, but they're in fact much more general.

Paths: Causal and Non-causal

Definition: A path between two variables is a non-self-intersecting sequence of adjacent arrows: The direction of the arrows does not matter; a given path can touch a given variable only once.



Definition: A causal path: is a path in which all arrows point away from T and toward Y. The set of causal paths comprises the total causal effect.

Definition: A non-causal path is path between T and Y in which at least one arrow points against the flow of time.

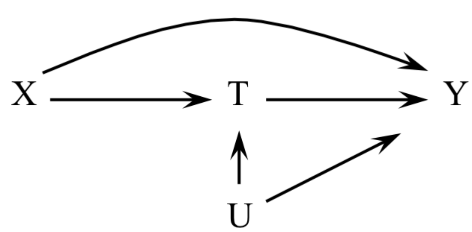
Exercise: List all paths, causal paths, and non-causal paths between T and Y.

Felix Elwert ©

Collider Variables

So-called collider variables play key role in working with DAGs

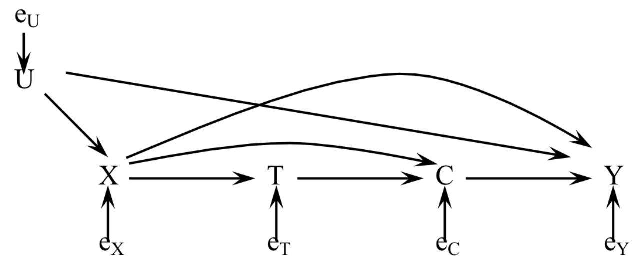
Definition: A collider variable is a variable into which two arrows point along a path.



- T is a collider on the path X→T←U.
- T is a non-collider on the path X→T→Y.

DAGs Represent Structural Equations

Graphs are equivalent to structural equations—no loss of information.



Linear (Wright 1921)

$U = e_U$

$X = a_1 + a_2U + e_X$

$T = b_1 + b_2X + e_T$

$C = c_1 + c_2T + c_3X + e_C$

$Y = d_1 + d_2C + d_3U + d_4X + e_Y$

Nonparametric (Pearl 1995)

$U = f_{U,i}(e_U)$

$X = f_{X,i}(U, e_X)$

$T = f_{T,i}(X, e_T)$

$C = f_{C,i}(T, X, e_C)$

$Y = f_{Y,i}(C, U, X, e_Y)$

2. From Causation to Association

Deriving Testable Implications

From Causation to Association

Simple rules connect the causal assumptions in the DAG to statistical associations in the data. Here's how we derive all implied associations and independences:

1. The causal effects in a DGP give rise to observable associations in data.

2. All associations travel along paths—but not all paths transmit associations.

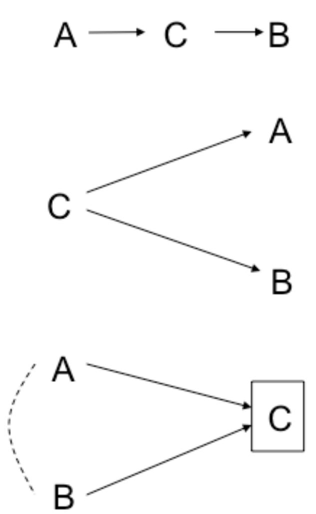
⇨ Only open paths transmit associations, while closed paths do not.

3. The three rules of association (d-separation) fully determine whether a path is open or closed.

4. Later, we will ask whether an observed association equals (identifies) the causal effect of interest.

Three Rules of Association

All marginal and conditional associations originate from 3 causal structures:



(1) Direct and indirect causation
$A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B|C$

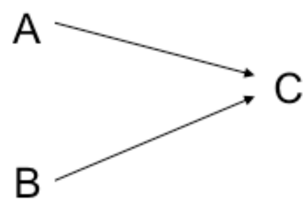
(2) Common cause confounding
$A \not\perp\!\!\!\perp B$ and $A \perp\!\!\!\perp B|C$

(3) Conditioning on a common effect ("collider"): Selection
$A \perp\!\!\!\perp B$ and $A \not\perp\!\!\!\perp B|C$

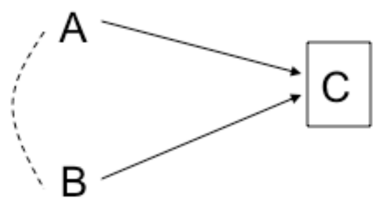
-------- : non-causal (spurious) association. ☐ : conditioning.

Conditioning on a Collider

Notice: No causal effect of A on B and no confounding



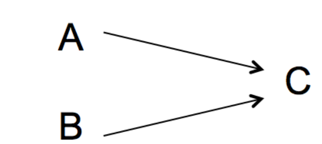
$A \perp\!\!\!\perp B$: marginally independent

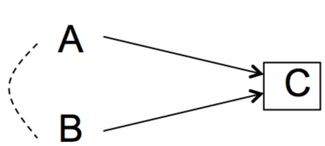


$A \not\perp\!\!\!\perp B|C$: conditionally dependent

⇒ The association is <u>biased</u> for the true causal effect.

Conditioning on a Collider





<u>Pearl's Sprinkler Example</u>
A: It rains
B: The sprinkler is on
C: The lawn is wet

<u>Hollywood Success</u>
A: Good looks
B: Acting skills
C: Fame

<u>Academic Tenure Example</u>
A: Productivity
B: Originality
C: Tenure

In all three examples, conditioning on the collider C induces a spurious association between two variables, A and B.