Causal Directed Acyclic Graphs

Kosuke Imai

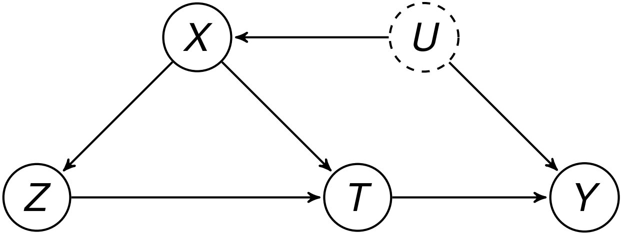
Harvard University

STAT186/GOV2002 CAUSAL INFERENCE

Fall 2019

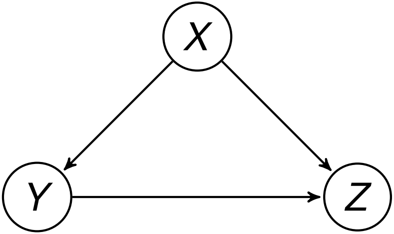
Elements of DAGs (Pearl. 2000. *Causality*. Cambridge UP)

- $\mathcal{G} = (E, V)$
  1. V: nodes or vertices $\rightsquigarrow$ variables (observed and onobserved)
  2. E: directed arrows $\rightsquigarrow$ *possibly non-zero* direct causal effects



- Acyclic: no simultaneity, the future does not cause the past
- Encoded assumptions
  - Absence of variables: all common (observed and unobserved) causes of any pair of variables
  - Absence of arrows: zero causal effect

DAG Terminology



- chain: $X \to Y \to Z$
- fork: $Y \leftarrow X \to Z$
- inverted fork: $X \to Z \leftarrow Y$

- Parents (Children): directly causing (caused by) a vertex $i \to j$
- Ancestors (Descendents): directly or indirectly causing (caused by) a vertex $i \to \cdots \to j$
- Path: an acyclic sequence of adjacent nodes
  - Causal path: all arrows pointing away from T and into Y
  - Non-causal path: some arrows going against causal order
- Collider: a vertex on a path with two incoming arrows

Nonparametric Structural Equation Models (NPSEM)

- Equivalence to the nonparametric structural equation models:

$$
\begin{aligned}
Y &= f_1(T, U, \epsilon_1) \\
T &= f_2(X, Z, \epsilon_2) \\
Z &= f_3(X, \epsilon_3) \\
X &= f_4(U, \epsilon_4)
\end{aligned}
$$

- NPSEM allows:
  1. any functional form
  2. any form of heterogenous effects
  3. any form of interaction effects
  4. LSEM as a special case

- Likelihood function:

$$
P(X_1, X_2, \ldots, X_J) = \prod_{j=1}^{J} P(X_j \mid \mathrm{pa}(X_j))
$$

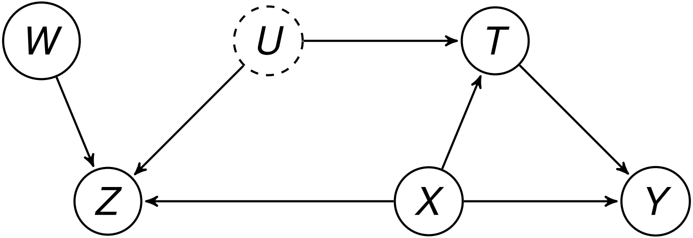
D-separation

- Does the conditional independence, $A \perp\!\!\!\perp B \mid C$, hold where A, B, C are sets of vertices?
  1. Identify all paths from any vertex in A to any vertex in B
  2. Check if each path is blocked
  3. If all paths are blocked, then A is d-separated from B by C

- Path is blocked,
  1. if it includes a noncollider vertex that is in C, or
  2. if it includes a collider that is not in C and no descendant of any collider is in C

- If A and B are d-separated, $A \perp\!\!\!\perp B \mid C$ holds
- If A and B are d-connected (i.e., not d-separated), $A \not\perp\!\!\!\perp B \mid C$ in at least one distribution compatible with DAG

D-separation Example

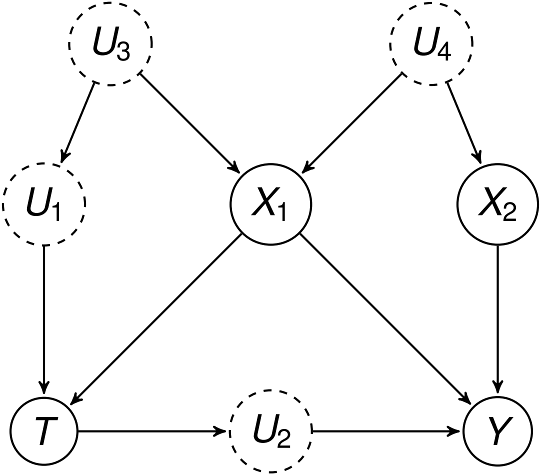


1. Are W and Y marginally independent of each other?
2. What happens if we condition on Z, X, T, or any combination of them?

Backdoor Criterion

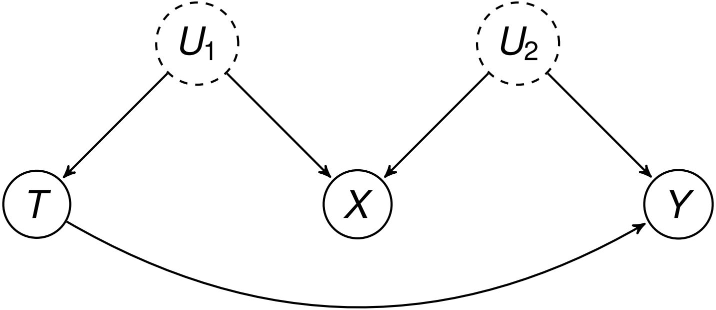
- Can we nonparametrically identify the average effect of T on Y given a set of variables X?
- Backdoor criterion for X:
  1. No vertex in X is a decendent of T, and
  2. X d-separates every path between T and Y that has an incoming arrow into T (backdoor path)
- Need to block all non-causal paths
- In the previous example, does X satisfy the backdoor criterion?
- Backdoor criterion implies the confounder selection criterion:
  (VanderWeele and Shpitser. 2011. *Biometrics*)
  If there exist a set of observed covariates that meet the backdoor criterion, it is sufficient to condition on all observed pretreatment covariates that either cause treatment, outcome, or both.
- Estimation: $P(Y_i(t)) = \sum_x P(Y \mid T = t, X = x) P(X = x)$

Example of Backdoor Criterion



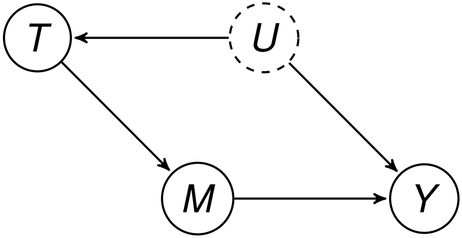
- Can we identify the causal effect of T on Y by conditioning on X_1?
- What about conditioning on X_1 and X_2?

M-Structure and M-Bias



- Should we condition on X or not?
- Conditioning on too many variables can induce bias
- Pearl's smoking and lung cancer example:
  - $X =$ wearing seatbelt
  - $U_1 =$ attitudes towards social norms
  - $U_2 =$ attitudes towards safety and health measures

Frontdoor Criterion (Pearl. 1995. *Biometrika*)



- U = unobserved confounders
- M = mediator $\rightsquigarrow$ causal mechanism
- Frontdoor criterion for M:
  1. M intercepts all directed paths from T to Y
  2. No backdoor path from T to M
  3. All backdoor paths from M to Y are blocked by T

$$P(Y(t)) = \sum_m \left\{ P(M_i = m \mid T_i = t) \sum_{t'} P(Y \mid T = t', M_i = m) P(T_i = t') \right\}$$

Evaluating Backdoor and Frontdoor Criteria

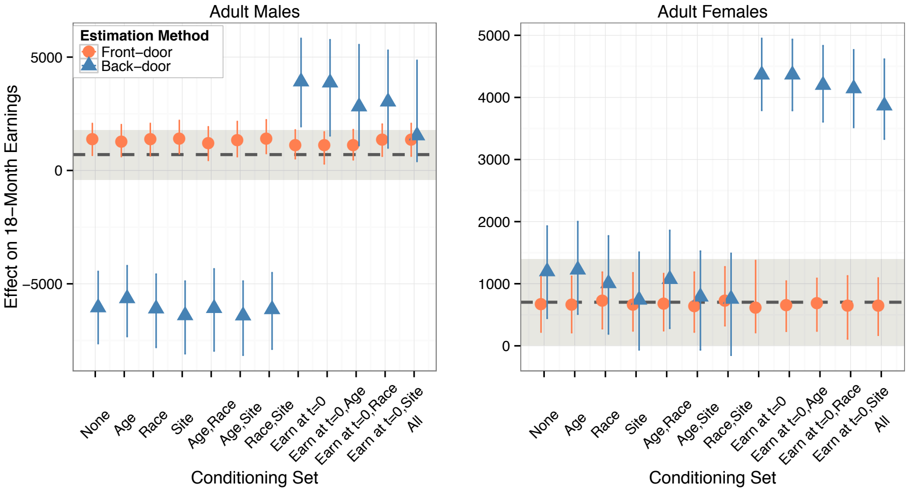
(Glynn and Kashin. 2018. *J. Am. Stat. Assoc.*)

- National Job Training Partnership Act (JTPA) study
- Randomized experiment: ATT on the wage after 18 months
  - adult female: \$702 (participation rate 55%)
  - adult male: \$700 (participation rate 57%)

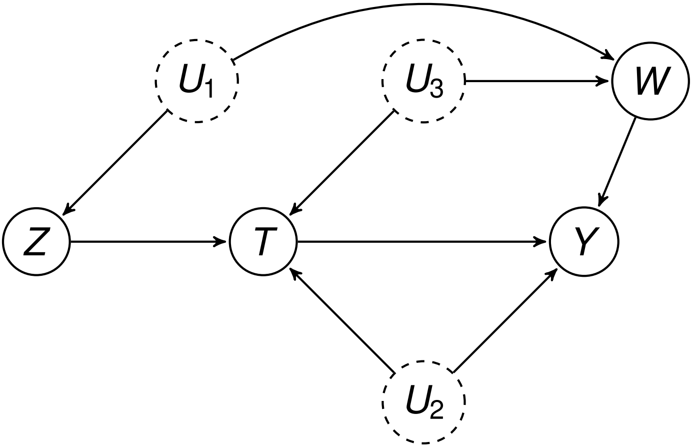
- Non-experimental control group
  - T: encouragement to participate in the program,
  - M: actual participation
  - Y: wage after 18 months

- Comparison group for actual participants
  - backdoor criterion: those assigned to the control group
  - frontdoor criterion: those who chose not to participate

Results

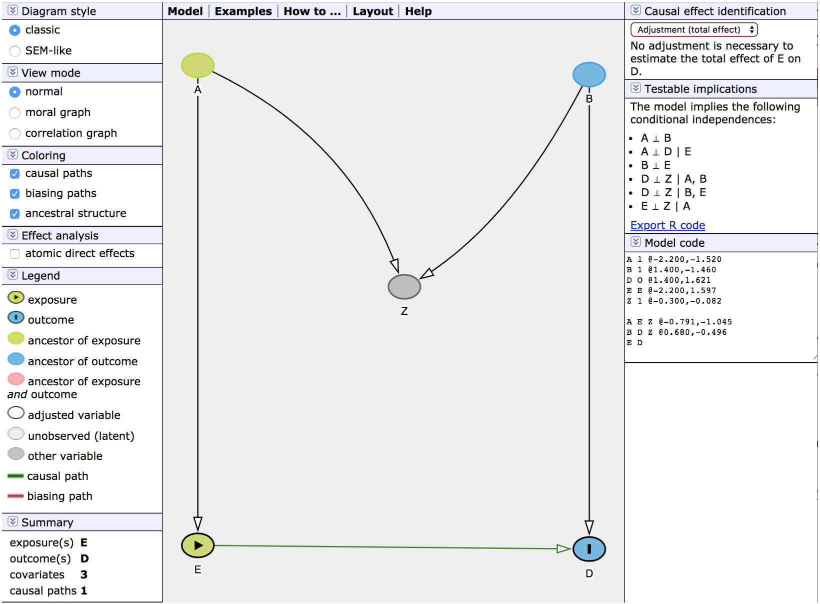

Adult Males
Adult Females
Estimation Method
Front–door
Back–door
Effect on 18–Month Earnings
5000
0
–5000
4000
3000
2000
1000
None
Age
Race
Site
Age,Race
Age,Site
Race,Site
Earn at t=0
Earn at t=0,Age
Earn at t=0,Race
Earn at t=0,Site
All
Conditioning Set

Instrumental Variables (Brito and Pearl. 2002. UAI.)



- Z is a valid instrumental variable conditional on W if
  1. W contains only non-descendants of Y
  2. W d-separates Z from Y in the subgraph G_s obtained by removing edge $T \rightarrow Y$
  3. W does not d-separate Z from T in G_s

DAGitty (http://dagitty.net/)



Potential Outcomes vs. DAGs Controversy

- Imbens and Rubin (2015):
  Pearl's work is interesting, and many researchers find his arguments that path diagrams are a natural and convenient way to express assumptions about causal structures appealing. In our own work, perhaps influenced by the type of examples arising in social and medical sciences, we have not found this approach to aid drawing of causal inferences.

- Pearl's blog post:
  So, what is it about epidemiologists that drives them to seek the light of new tools, while economists seek comfort in partial blindness, while missing out on the causal revolution? Can economists do in their heads what epidemiologists observe in their graphs? Can they, for instance, identify the testable implications of their own assumptions? Can they decide whether the IV assumptions are satisfied in their own models of reality? Of course they can't; such decisions are intractable to the graph-less mind.

Concluding Remarks

- Potential outcomes are useful when thinking about treatment assignment mechanism $\rightsquigarrow$ experiments, quasi-experiments
- DAGs are useful when thinking about the causal structure $\rightsquigarrow$ complex causal relationships, causal mechanisms
- Growing literature on causal discovery
- Readings:
  - Pearl. (2009). *Causality*. Cambridge UP
  - Elwert. (2013). Chapter 13: Graphical Causal Models in *Handbook of Causal Analysis for Social Research*
  - Peters, Janzing, and Schölkopf. (2018). *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press.