Section 19

Lecture 5

Plan for today

- High-level comments about graphs
- D separation
- Examples
- The backdoor criterion

Graphs

Some things you need to know about graphs

- Graphs encode conditional independendcies
- Graphs allow us to *represent* and *organize* assumptions and prior knowledge.
- Graphs make the assumptions transparent and explicit.

So we have an algorithm for creating causal graphs

We can create a causal DAG by:

1. Draw nodes for the exposure A and the outcome Y of interest.
   - Draw an arrow from A to Y.
2. If there exists a common cause C of A and Y, write C in the graph.
   - Draw arrows from C to A and from C to Y.
     These common causes must be drawn, even if they are unmeasured.
3. If there exists a common cause C' of any pair $W, W' \in (C, A, Y)$, write C' in the graph.
   - Draw arrows from C' to W and from C' to W'.
4. Continue in this way until there are no common causes...

Properties of conditional independence

Theorem (Graphoid axioms)

Let X, Y, Z, W be random variables on a Cartesian product space. Conditional independence satsifies

1. $X \perp\!\!\!\perp Y \mid Z \implies Y \perp\!\!\!\perp X \mid Z$ *(Symmetry)*
2. $X \perp\!\!\!\perp Y, W \mid Z \implies X \perp\!\!\!\perp Y \mid Z$ *(Decomposition)*
3. $X \perp\!\!\!\perp Y, W \mid Z \implies X \perp\!\!\!\perp W \mid Y, Z$ *(Weak union)*
4. $X \perp\!\!\!\perp W \mid Y, Z$ *and* $X \perp\!\!\!\perp Y \mid Z \implies X \perp\!\!\!\perp Y, W \mid Z$ *(Contraction)*
5. *If* $p(x, y, z, w) > 0$, *then* $X \perp\!\!\!\perp W \mid Y, Z$ *and* $X \perp\!\!\!\perp Y \mid W, Z \implies X \perp\!\!\!\perp Y, W \mid Z$ *(Intersection)*

Proof of Graphoid axioms

I will not prove all of them here. I just state a brief proof of the first one here.

Proof.

1. Symmetry follows simply because

$$
\begin{aligned}
X \perp\!\!\!\perp Y \mid Z \leftrightarrow p(x \mid z)p(y \mid z) &= p(x, y \mid z) \\
&= p(y \mid z)p(x \mid z) \leftrightarrow Y \perp\!\!\!\perp X \mid Z \,.
\end{aligned}
$$

□

D separation of a path

Now we will study a beautiful graphical condition on $\mathcal{G}$ that immediately tells if $X \perp\!\!\!\perp Y \mid Z$, where X, Y, Z are disjoint sets of nodes in V, is implied by the Markov factorisation.

Definition (d-separation of a path)

A path r is d-separated by a set of nodes Z iff

1. r contains a chain $V_i \to V_j \to V_k$ or a fork $V_i \leftarrow V_j \to V_k$ such that V_j is in Z, or
2. r contains a collider $V_i \to V_j \leftarrow V_k$ such that V_j *is not* in Z and such that no descendant of V_j is in Z.

Otherwise the path is d-connected.

D separation of two nodes

Definition (d-separation of two nodes)

Nodes V_i and V_k are d-separated by a set of nodes Z if all trails between V_i and V_k are d-separated by Z. We write d-separation as

$$(V_i \perp\!\!\!\perp V_k \mid Z)_G \,.$$

If V_i and V_k are not d-separated, they are d-connected and we write

$$(V_i \not\perp\!\!\!\perp V_k \mid Z)_G \,.$$

Theorem (Soundness of d-separation)

$(V_i \perp\!\!\!\perp V_k \mid Z)_G$ *implies the statistical independence*

$$V_i \perp\!\!\!\perp V_k \mid Z.$$

A consequence of soundness is that d-separation in $\mathcal{G}$ implies conditional independence for any distribution that factorizes according to $\mathcal{G}$.

D-separation details and intuition

- D-separation can be shown solely using the Graphoid axioms (but the proof is tedious).
- d-separation allows us to determine independencies of a distribution from the structure of a statistical DAG.
- Heuristically, two variables are d-separated (independent) if there is no open path between them.

D-separation and some questions in class

$$
\begin{tikzcd}
V_0 \arrow[r] \arrow[rr, bend right] & V_1 \arrow[rr, bend left] & V_2 \arrow[r] & V_3
\end{tikzcd}
$$

Carrying a lighter A and the risk of lung cancer Y

$$
\begin{tikzcd}
L \arrow[r] \arrow[rr, bend right] & A & Y
\end{tikzcd}
$$

A gene A that causes heart disease L but not smoking Y, where C is taking aspirin (A cardiovascular drug)

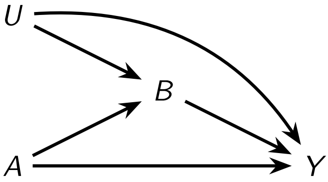
$$
\begin{tikzcd}
A \arrow[rr, bend right] & Y \arrow[r] & L
\end{tikzcd}
$$

$$
\begin{tikzcd}
A \arrow[rr, bend right] & Y \arrow[r] & L \arrow[r] & C
\end{tikzcd}
$$

Example: Birth weight paradox

- Birth weight predicts infant mortality.
- Investigators often stratify on birth weight when evaluating the effect of maternal smoking on infant mortality.
- Among infants with low birth weight, the mortality rate ratio for smoke exposed infants versus non-exposed infants is 0.79 (95% CI: 0.76, 0.82).
- This birth weight paradox has been a controversy for decades.
- One suggestion is that the effect of maternal smoking is modified by birth weight in such a way that smoking is beneficial for LBW babies.
- Is this indeed the likely explanation?

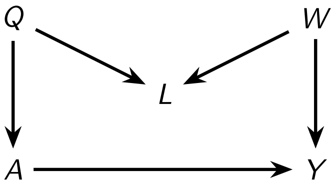
Example: Birth weight paradox



- A Smoking status of the mother
- B Birth weight
- U Unknown factor (e.g. genetic) causing low birth weight
- Y Infant mortality

PS: for this graph to be more plausible, we should also add common causes of A and Y.

Berkson bias



- A Drink a glass of red wine a day.
- Y Nausea
- L Aspirin
- Q Family history of cardiovascular disease
- W Frequency of headache

Q: We measure Aspirin. Should we adjust for Aspirin in the analysis?

Faithfulness and completeness of d-separation

Definition

A law $\mathbb{P}$ is faithful to a DAG $\mathcal{G}$ if for any disjoint set of nodes A, B, C we have that $A \perp\!\!\!\perp C \mid B$ under $\mathbb{P}$ implies $(A \perp\!\!\!\perp C \mid B)_{\mathcal{G}}$.

Theorem (Completeness of d-separation)

In a Bayesian Network with respect to a direct acyclic graph $\mathcal{G}$ there exists a faithful law $\mathbb{P}$.

We will not prove this important result[21].
The completeness of faithfulness d-separation allows us to use d-separation to represent the conditional independence structure of a multivariate distribution. You can look at the graph, and read off all independencies that hold in the entire class of distributions factorizing according to the DAG.

[21] Ann Becker, Dan Geiger, and Christopher Meek. "Perfect tree-like markovian distributions". In: *arXiv preprint arXiv:1301.3834* (2013); Pearl, *Causality: Models, Reasoning and Inference 2nd Edition*.

The causal Markov assumption and faithfulness (intuition and interpretation)

- d-separation implies statistical independence, but does not allow one to deduce that d-connection implies statistical dependence.
- However, d-connected variables will be independent only if there is an exact balancing of positive and negative causal effects.
- Because such precise balancing of effects is highly unlikely to occur, we shall henceforth generally assume that d-connected variables are dependent.

Backdoor adjustment

Definition (Backdoor path)

In a DAG $\mathcal{G}$ a backdoor path between two nodes V_i and V_j is a trail that starts in V_i and ends in V_j; and with initial edge being an arrow pointing into V_i

Example backdoor path between V_i and V_j is: $V_i \leftarrow V_k \rightarrow V_j$.

Backdoor theorem

Theorem (Backdoor theorem wrt. to a DAG)

In a causal model represented by a DAG $\mathcal{G}$ let X, Y and Z be three sets of nodes of $\mathcal{G}$, each comprised of one or more nodes. Suppose that X contains no descendants of Z and it blocks all back-door paths between any node in Z and any node in Y: Suppose that $g = (g_1, \dots, g_t)$ is a regime for $Z = (Z_1, \dots, Z_t)$ (for some $t \geq 1$) such that treatment assignments depend at most on X: Then, for any x in the support of X such that $p(Z = g(x) \mid x) \equiv Pr(Z = g(x) \mid X = x) > 0$; it holds that

$$\begin{aligned} P(Y^g = y, Z^{g+} = z, X = x) =& p^g(y, x, z) \\ =& p(y \mid x, z) I(g(x) = z) p(x) \\ =& P(Y = y \mid Z = z, X = x) I(g(x) = z) P(X = x), \end{aligned}$$

and in particular,

$$P(Y^g = y) = \sum_x \sum_z p(y \mid x, z) I(g(x) = z) p(x).$$

The backdoor theorem continues

See Pearl[22] for proof (not required for the exam etc). This theorem is very useful, because it allows us to identify causal effects even if certain nodes in the graph are unmeasured. The last part of the theorem, after "in particular", will be useful in the exercises of Lecture 5.

[22] Judea Pearl. "Causal diagrams for empirical research". In: *Biometrika* 82.4 (1995), pp. 669–688.

Implication from the Backdoor theorem

It follows from the backdoor theorem that if $Y^a \perp\!\!\!\perp A \mid L$ then

$$P(Y^a = y) = \sum_l P(Y = y \mid L = l, A = a) P(L = l).$$

We derived the identification formula above without using graphs and the Backdoor theorem earlier in this course.
However, importantly, the Backdoor theorem allows us to identify causal effects in much more complicated settings, which also involve unmeasured variables.