Introduction to Causal Inference

Conor Mayo-Wilson

1 Big Picture

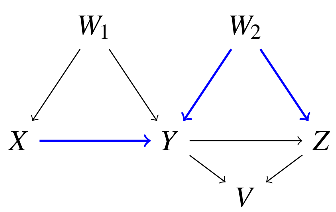
- *Goal:* Predicting the effects of an intervention. This requires:
  - Discovering causal structure
  - Using causal structure (and data) to estimate intervention effects

2 Terminology

- Parent, Child, Ancestor, Descendant, Unshielded collider,
- Directed path, undirected path
- Definition[Blocking]: A collection of variables $\mathscr{C} \subseteq \mathscr{V} \setminus \{V_1, V_2\}$ **blocks** a path from V_1 to V_2 if and only if either (1) the path contains a variable X in $\mathscr{C}$, and X is *not* a collider on the path or (2) the path contains a collider X that is *not* in $\mathscr{C}$, **and** none of X's descendants are in $\mathscr{C}$ (or both).
- Definition[d-Separation]: V_1 and V_2 are **d-separated** given $\mathscr{C}$ if every path between V_1 and V_2 is blocked by $\mathscr{C}$.
- Definition[Causal sufficiency]: A set of variables $\mathscr{V}$ is **causally sufficient** if for any pair of variables $V_1, V_2 \in \mathscr{V}$, if U is a common cause of V_1 and V_2, then $U \in \mathscr{V}$.
- Definition[Causal Markov Condition]: If G is a causal graph over a causally sufficient set of variables $\mathscr{V}$ and P is the joint distribution of $\mathscr{V}$, then the pair $\langle G, P \rangle$ is said to satisfy the **Causal Markov Condition** (CMC) if every variable in V is conditionally independent of its non-effects given its direct causes.
- Definition[Causal Faithfulness Condition]: If G is a causal graph over a causally sufficient set of variables $\mathscr{V}$ and P is the joint distribution of $\mathscr{V}$, then the pair $\langle G, P \rangle$ satisfies the **Causal Faithfulness Condition** if two variables V_1 and V_2 are conditionally independent given $\mathscr{C}$ if and only if they are entailed to be independent by the CMC.

3 Exercises

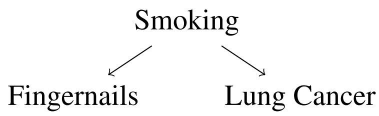
Exercise 1: Consider the graph below:



- Consider the path $\langle X, Y, W_2, Z \rangle$. Which of the following sets block the path? The empty set, $\{W_2\}, \{W_2, V\}, \{Y\}$, and/or $\{Y, W_2\}$.
- Find all paths between X and Z that are active (i.e., not blocked) given the empty set and the set $\{Y\}$.

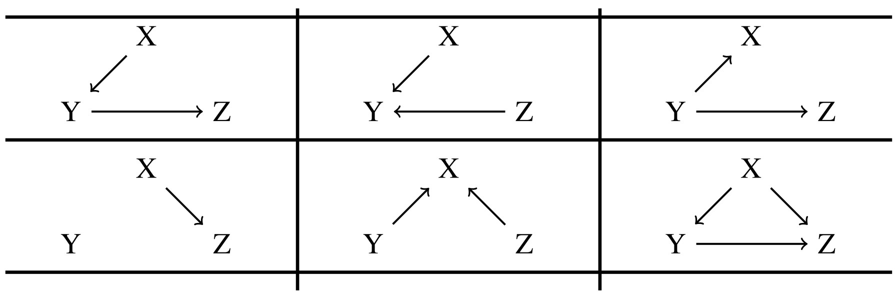
Exercise 2: Consider the following

$$P(Smoker) = .178$$
$$P(Yellow|Smoker) = .12$$
$$P(Yellow|Non-Smoker) = .001$$
$$P(Cancer|Smoker) = .076$$
$$P(Cancer|Non-Smoker) = .02$$



- Calculate the probability that a randomly selected member of the population has lung cancer given she does not have yellow fingernails.
- In an experimental group in which yellow stains are removed (with probability one) from subjects' fingernails, Calculate the probability that a randomly selected member of the population has lung cancer given she does not have yellow fingernails.

Exercise 3: Suppose each of the following is a causal graph over a causally sufficient set of variables. Suppose P is a probability distribution that is Markov and faithful to the graph. Write down all of the conditional independences in P.



Exercise 4: I gather data concerning three variables among applicants to MIT: gender (G), admission to MIT (A), and SAT score (S). My findings are below. Assume there are no latent common causes of A, G, and S. What causal graphs are best supported by my findings?

- Women are admitted to MIT at a higher rate then are men.
- But gender and SAT score are independent of one another.
- SAT score is positively correlated with admission to MIT.

Exercise 5: Which of the following graphs are Markov equivalent?

$$X \longleftarrow Y \longrightarrow W \longrightarrow Z$$

$$X \longleftarrow Y \longleftarrow W \longrightarrow Z$$

$$X \longrightarrow Y \longleftarrow W \longrightarrow Z$$

$$X \longrightarrow W \longleftarrow Y \longrightarrow Z$$