Algorithmic Causality with Applications

Maciej Liśkiewicz

Exercise Sheet 1 (Deadline 16.12.2022)

Sheet Objectives

- Understanding conditional independencies.
- Getting familiar with graphical representations of joint probability distributions.
- Connecting d-separation with statistical independence.
- Getting familiar with interventional distributions.

1. Conditional independence and Bayesian networks, medium, 6 points

Consider the following example: Let A, B, C be binary random variables. The random variables A and B describe the results of two independently tossed fair coins (heads and tails both have probability $1/2$). If both coin tosses give the same result, a bell is rung (this is described by random variable C).

a. Write down the probability tables for the full joint probability distribution $P(A, B, C)$ as well as the marginal distributions $P(A, B)$, $P(A, C)$ and the conditional distributions $P(A, B \mid C = 0)$, $P(A, B \mid C = 1)$.

b. Given those tables, analyze which of the following independencies hold:

- $(A \perp\!\!\!\perp B)_P$
- $(A \perp\!\!\!\perp C)_P$
- $(A \perp\!\!\!\perp B \mid C)_P$

c. Represent the joint probability distribution by a Bayesian network. It is sufficient to draw the graphical representation as a DAG G.

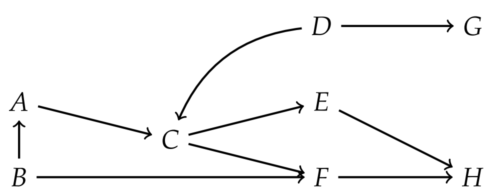
d. Given this graphical representation and using d-separation, analyze which of the following independencies hold:

- $(A \perp\!\!\!\perp B)_G$
- $(A \perp\!\!\!\perp C)_G$
- $(A \perp\!\!\!\perp B \mid C)_G$

e. In the theorem d-separation vs. conditional independence, Verma and Pearl state that $(X \perp\!\!\!\perp Y|Z)_G \implies (X \perp\!\!\!\perp Y|Z)_P$. Using your results from this exercise, what can you conclude for the opposite direction?

2. A closer look at d-separation, medium, 6 points

a. For the DAG below and the sets $\mathbf{X} = \{A, B\}$, $\mathbf{Y} = \{G, H\}$, find all d-separators $\mathbf{Z}$ relative to $(\mathbf{X}, \mathbf{Y})$.



b. Is there, for every DAG G and disjoint sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$, a set $\mathbf{Z} \subseteq \mathbf{V}$, which d-separates $\mathbf{X}$ and $\mathbf{Y}$? What if the sets $\mathbf{X}$ and $\mathbf{Y}$ are not connected by a direct edge, i.e. there exist no incident nodes $A \rightarrow B$ or $A \leftarrow B$ such that $A \in \mathbf{X}$ and $B \in \mathbf{Y}$.

c. A d-separator $\mathbf{Z}$ relative to $(\mathbf{X}, \mathbf{Y})$ is minimal if, for every nonempty $\mathbf{W} \subseteq \mathbf{Z}$, the set $\mathbf{Z} \setminus \mathbf{W}$ does not d-separate $\mathbf{X}$ and $\mathbf{Y}$. Show for the graph from 1. which d-separators are minimal for $\mathbf{X} = \{A, B\}$ and $\mathbf{Y} = \{G, H\}$.

d. Propose an algorithm which, for a given DAG G and disjoint sets $\mathbf{X}, \mathbf{Y} \subseteq \mathbf{V}$, finds a d-separator $\mathbf{Z}$. What is the time complexity of your algorithm?

e. Bonus (without additional points): For the problem in 4., find a linear time algorithm (in the size of the DAG).

3. Observation vs Intervention, medium, 6 points

We consider the *recommendation letter example* (Example 3.2.1 from Koller, Friedman (2009), see also lecture slide 16 on the topic *Causal Inference: Backgrounds*).

The goal of this task is to compare the effect of observations and interventions on the chance of getting a strong recommendation letter.

We begin with the SAT score and consider, on the one hand, the setting that this score is observed as low/high and, on the other hand, that this score is set to low/high by an intervention.

Compute the corresponding probabilities

1. $P(L = \text{strong} \mid S = \text{low})$,
2. $P(L = \text{strong} \mid S = \text{high})$,
3. $P(L = \text{strong} \mid do(S = \text{low}))$,
4. $P(L = \text{strong} \mid do(S = \text{high}))$.

Consider now that we observe/intervene on the difficulty instead of the SAT score. Compute the probabilities

5. $P(L = \text{strong} \mid D = \text{easy})$,
6. $P(L = \text{strong} \mid D = \text{hard})$,
7. $P(L = \text{strong} \mid do(D = \text{easy})$,
8. $P(L = \text{strong} \mid do(D = \text{hard})$.

Compare the results above for the setting of observing/intervening on the SAT score with those for the difficulty and discuss your findings.

Note: Recall that a strong recommendation was encoded as l^1, a hard difficulty as d^1 and high SAT score as s^1.