Inference and learning in causal models: Lecture 1

Tsinghua Logic Summer School

Tadeg Quillien

University of Edinburgh

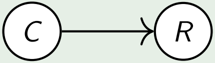
2026

Inference in causal models

Sometimes inferences can be transparently ‘read off’ from the model.

Example

Consider the model: C: Clouds, R: Rain.



$R := C \wedge U_{CR}$

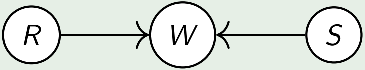
It is easy to see that $P(R|C) = P(U_{CR})$.

- In a CBN, the inference is easier: $P(R|C)$ is already a parameter of the model!

What about more complex inferences?

Inference in causal models

Example



Assume that we have a Causal Bayes Net that gives us the probability $P(\mathbf{W}|\mathbf{R}, \mathbf{S})$.
We know that $R = 1$ but we don't know $\mathbf{S}$.
How do we compute $P(W|R)$?

- We have to marginalize over possible values of $\mathbf{S}$.
- The desired probability is given by: $P(W|R) = P(W|S, R)P(S) + P(W|\neg S)P(\neg S)$
  - Proof: See homework.

Diagnostic inferences

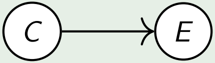
So far we've inferred the value of a variable, given the value of its parents.

- so-called **predictive** inferences.

What about the reverse?

- We call them '**diagnostic** inferences'.

Example



What is $P(\mathbf{C}|\mathbf{E})$?
Our causal model only has $P(\mathbf{E}|\mathbf{C})$ and $P(\mathbf{C})$.

Bayes' rule

Our goal is to derive a formula that gives us $P(C|E)$ as a function of $P(E|C)$ and $P(C)$.

- From Rule 3, we can expand the joint distribution $P(C,E)$ in two different ways:
- $P(C,E) = P(E|C)P(C)$,
- $P(C,E) = P(C|E)P(E)$.
- Dividing by $P(E)$ on each side:
- $P(C|E) = \frac{P(E|C)P(C)}{P(E)}$.
- We can re-write $P(E)$ using the law of marginalization:
- $P(C|E) = \frac{P(E|C)P(C)}{P(E|C)P(C) + P(E|\neg C)P(\neg C)}$.

$\rightarrow$ All these terms are available in the causal model!

Bayes' rule

Theorem

Bayes' rule: For any two variables C and E we have:

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)}$$

$$= \frac{P(E|C)P(C)}{P(E|C)P(C) + P(E|\neg C)P(\neg C)}$$

Remark

Bayes' rule is sometimes succinctly written as:

$P(C|E) \propto P(E|C)P(C)$,

where $\propto$ means 'proportional to'.

Bayes' rule: Example

Example

Consider the model: C: Clouds, R: Rain.

$$C \longrightarrow R$$

with $P(C) = .3$, $P(R|C) = .6$, $P(R|\neg C) = 0$.

- What is $P(C|\neg R)$, i.e. how likely is it that there are clouds, given that there is no rain?
- Answer:

$$P(C|\neg R) = \frac{P(\neg R|C)P(C)}{P(\neg R)}$$

$$= \frac{0.4 \times 0.3}{0.4 \times 0.3 + 1 \times 0.7} = \frac{.12}{.12 + 0.7} \approx 0.15$$

Statistical (in)dependence

Of course sometimes there are inferences where you already know the answer without having to calculate.

Example

$$\begin{tikzcd} A \arrow[r] & C & B \arrow[l] \end{tikzcd}$$

Intuitively, $P(A|B) = P(A)$, and $P(B|A) = P(A)$. Variables A and B just don't 'depend' on each other.

Statistical dependence (informal)

X and Y are statistically dependent if learning the value of X gives you information about the value of Y, and vice-versa.

Statistical (in)dependence

Definition

Two variables X and Y are **statistically dependent** if $P(\mathbf{X}|\mathbf{Y}) \neq P(\mathbf{X})$.

- I.e. there is at least an x and a y such that $P(X = x|Y = y) \neq P(X = x)$.
- Statistical dependence is denoted as $X \not\perp Y$.

X and Y are independent if they are no dependent.

- Denoted as $X \perp Y$.

Remark

Statistical dependence is not the same as causal dependence.

- Statistical dependence is symmetrical: $X \not\perp Y$ implies $Y \not\perp X$.

Statistical (in)dependence

Example

$$A \longrightarrow C \longleftarrow B$$

with $C := A \wedge B$, and $P(B) > 0$.

- $A \perp B$, because they are both exogenous.
- $P(C|A) > P(C)$, so $A \not\perp C$.

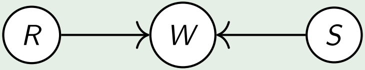
Conditional (in)dependence

Just as we have probabilities and conditional probabilities, we can also introduce a **conditional** version of dependence.

Conditional dependence (informal)

X and Y are **dependent, conditional** on W, if when you know the value of W, learning the value of X gives you information about the value of Y.

Example



- R does not depend on S.
- But if we know $W = 1$, then learning S makes R less likely.

Conditional (in)dependence

Definition

X and Y are dependent, conditional on W, if $Pr(\mathbf{Y}|\mathbf{X},\mathbf{W}) \neq Pr(\mathbf{Y}|\mathbf{W})$.

- Denoted as $X \not\perp Y|W$.

Conditional (in)dependence

Example

$$C \longrightarrow R \longrightarrow W$$

- C and W are **dependent** (Clouds make it more likely you get Wet).
- But $P(W|R, C) = P(W|R)$.
  - If you already know if is Raining, learning there are Clouds does not teach you anything more about whether you're Wet.
- So conditional on R, W and C are **independent**.

Patterns of (conditional) dependence are implied by the structure of the network

Example

$$C \longrightarrow R \longrightarrow W$$

We know that $C \perp W | R$ even if we don't know the structural equations or probabilities.

Remark

- How to infer (conditional) dependence from the network structure is well-understood,
- but outside the scope of this course.
- See e.g. the discussion of 'd-separation' in Pearl, 2009.

Next lecture

- Inference by sampling
- Learning causal effects and causal models