Introduction to causal models. Lecture 2

Tsinghua Logic Summer School

Tadeg Quillien

University of Edinburgh

2026

Reasoning under uncertainty

Beyond determinism

So far we've worked with systems that are deterministic.

- When you know the value of a variable's parents, you know its value for sure.

But uncertainty can arise, either because:

- The world is not deterministic,
- You cannot see everything.

How do we incorporate this in our framework?

Probability

The correct way to represent one's uncertainty about something is to use **probability**.
A probability is a number between 0 and 1 that represents our degree of belief in a proposition.

Example

In my office in Scotland, I look out the window. There are dark clouds. I think there is a 50% chance it will rain, so my $P(\text{Rain})$ is 0.5.

How can be manipulate our degrees of belief?

- Suppose you are 20% confident in Y and 50% confident in X; how confident should you be in 'X and Y'?

We can use the mathematical rules of probability theory to answer these questions.

A note on notation

We will mostly use Boolean variables.

$\rightarrow$ We will often use the following shortcut:

- $P(X)$ is the probability that $X = 1$.
- $P(\neg X)$ is the probability that $X = 0$.
- When necessary to avoid confusion, we use $P(\mathbf{X})$ to denote the function that takes a value of variable X as input and returns a probability $\in [0, 1]$.

Rules of probability theory

(after Pearl, 2009)

- **Rule 1.** For all X, $P(X)$ is between 0 and 1.
- **Rule 2.** If X is a sure proposition, then $P(X) = 1$.
- **Rule 3.** $P(A \vee B) = P(A) + P(B)$ if A and B are mutually exclusive.
  - By mutually exclusive, we mean that A and B can't be true at the same time.
- **Rule 4.** $P(A \wedge B) = P(A|B)P(B)$.

$P(A|B)$ is a **conditional probability**: it is the probability of A, given that we know that B is true.

Rule 4: example

(after Pearl, 2009)

- **Rule 4.** $P(A \wedge B) = P(A|B)P(B)$.

Example

What is the probability that there are dark clouds (C) and it is raining (R)?

- $P(C \wedge R) = P(R|C)P(C)$

Remark about notation

We often denote $P(A \wedge B)$ as $P(A, B)$.

A useful generalization of Rule 4

Theorem

Chain rule. For any set of variables, you can impose an arbitrary order $V_1, .., V_n$, and then write:
$P(V_1, ..., V_n) = P(V_1|V_2, ..., V_n)P(V_2|V_3, ..., V_n), ..., P(V_{n-1}|V_n)Pr(V_n)$.

Example

- C: Dark Clouds,
- R: Rain,
- W: I get Wet.

Then:
$P(W, R, C) = P(W|R, C)P(R|C)P(C)$

Some more theorems about probability

Theorems (for proof see homework)

- **Negation rule**: $P(\neg A) = 1 - P(A)$
- **Disjunction rule**: $P(A \vee B) = P(A) + P(B) - P(A, B)$.
  - Note that Rule 3 is in some sense a special case of the disjunctive rule, since when A and B are mutually exclusive $P(A, B) = 0$, and so $P(A \vee B) = P(A) + P(B)$.
- **Law of conditional probability**: $P(A|B) = \frac{P(A,B)}{P(B)}$
  - On the condition that $P(B) \neq 0$

Law of marginalization

Theorem

If B is a Boolean variable then:
$P(A) = P(A \wedge B) + P(A \wedge \neg B)$

Proof:

- From propositional logic we know that
  $A = A \wedge (B \vee \neg B) = (A \wedge B) \vee (A \wedge \neg B)$.
- $A \wedge B$ and $A \wedge \neg B$ are mutually exclusive,
- Therefore Rule 3 gives us:
  $P(A) = P((A \wedge B) \vee (A \wedge \neg B)) = P(A \wedge B) + P(A \wedge \neg B)$.

Marginalization, General form

For any two variables A and B we have: $P(A) = \sum_{b \in B} P(A, B = b)$

Probability distributions

A probability distribution is a function that assigns a probability to each possible value of a variable.

Example

Suppose I am 60% sure it is raining. Then the distribution $P(\mathbf{R})$ is:

- $P(R) = .6$
- $P(\neg R) = .4$

A **joint** probability distribution (e.g. $P(\mathbf{R,C})$) is a distribution defined over several variables.

Example

- $P(R, C) = .5$,
- $P(R, \neg C) = 0$,
- $P(\neg R, C) = .1$,
- $P(\neg R, \neg C) = .4$

Summary

- We can quantify our uncertainty using the laws of probability.
- The rules given in these slides will be very useful for the rest of the course.

Outstanding questions

How can we use this in causal models?