Introduction to causal models: Lecture 3

Tsinghua Logic Summer School

Tadeg Quillien

University of Edinburgh

2026

Causal reasoning under uncertainty

Injecting Stochasticity into SCMs

Two ways we will use probability in SCMs:

- 'Generate' the value of exogenous variables (variables that have no parents).
- Make the causal link between two variables non-deterministic.

$P(\mathcal{U})$

Our definition of SCMs contained something called $P(\mathcal{U})$.

- This is a joint distribution over all exogenous variables in the model.

Example

The pavement can be wet (W) because of the Rain (R), or the Sprinkler (S).
We can represent this using the following SCM (suppose we live in a very dry place):

- $P(R) = .05$,
- $P(S) = .4$,
- $W := R \vee S$.

Note: we typically assume that the probabilities of the exogenous variables don't depend on each other.

Making the link between variables non-deterministic

- Suppose we have a causal model with clouds (C) and Rain (R).
- One possible structural equation is $R := C$,
- But clouds don't always cause rain!
- We need a way to say something like 'Clouds cause rain 40% of the time'.

Solution: unobserved variables

- We add an unobserved exogenous variable to the model:
- $P(U_{CR}) = .4$
- We write the equation for R as: $R := C \wedge U_{CR}$.

Variables like U_{CR} are sometimes called 'noise' variables and are often omitted from Graphs.

Making the link between variables non-deterministic

- In the Rain example, rain only happens when there are clouds.
- But many variables can also be caused by unobserved factors.

In general, we will write the equation for a link between C and E as:

- $E := (C \wedge U_{CE}) \vee U_E$

where U_E is the 'base rate' of E.

Remark

- By doing things this way we implicitly assume that the world is deterministic.
- I.e. Causal relations seem stochastic because we can't see everything.
- This might not always be true (cf. quantum mechanics), but is a good approximation at the macro level.
- Assuming determinism will be useful when modeling counterfactual reasoning.

Probabilities and interventions

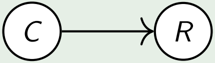
$P(A|B)$ is the probability of A given that we have observed B.
Another important probability is $P(A|\text{Do}(B))$,

- The probability of A given that we intervened to set B to 1.

We say that $P(\mathbf{A}|do(\mathbf{B}))$ is an **interventional distribution**.

Example

C: Clouds, R: Rain.



- $P(C|\text{Do}(R)) = P(C)$
- But $P(C|R) > P(C)$,
  - Because observing Rain tells us that Clouds are probably present.

One limitation of SCMs

Ideally we would like to guess the correct SCM for a system by collecting experimental data. Can we do this?

Example

We test a drug D by giving it to participants and giving a placebo to a control group. We record how many people in each group get headaches H. We estimate the following conditional probability table:

- $P(H|D) = 4/5$
- $P(\neg H|D) = 1/5$
- $P(H|\neg D) = 1/5$
- $P(\neg H \neg D) = 4/5$

Can we use these data to infer the structural equation linking D to H?

One limitation of SCMs

Example

- $P(H|D) = 4/5$
- $P(H|\neg D) = 1/5$

Equation 1

$H := U_H \vee (D \wedge U_{DH})$; With $P(U_H) = 1/5$ and $P(U_{DH}) = 3/4$. We have:

- $P(H|D) = Pr(U_H \vee U_{DH}) = 1/5 + 3/4 - 1/5 \times 3/4 = 4/5$,
- $Pr(H|\neg D) = Pr(U_H) = 1/5$.

Equation 2

$H :=$ Same(D, U_H); With $Pr(U_H) = 4/5$. We have:

- $Pr(H|D) = Pr(U_H) = 4/5$,
- $Pr(H|\neg D) = 1 - Pr(U_H) = 1/5$.

Causal Bayes Nets

The same conditional probability table can be generated by several different SCMs.
$\rightarrow$ It is difficult to estimate SCMs from data (even interventional data!) without making strong assumptions.

- But we should not give up the goal of representing causal knowledge.
- Since we have a conditional probability table, let's just use this!
- This is the idea behind Causal Bayes Nets.

Causal Bayes Nets

Causal Bayes Nets (CBNs) work the same way as SCMs, except that we use conditional probability tables instead of structural equations.
The conditional probability table for variable V expresses the distribution $P(V|pa_V)$, where pa_V is the set of direct parents of V.

Example

Conditional probability table for our $D \rightarrow H$ model:

- $P(H|D) = 4/5$
- $P(\neg H|D) = 1/5$
- $P(H|\neg D) = 1/5$
- $P(\neg|H\neg D) = 4/5$

Causal Bayes Nets

Causal Bayes Nets (CBNs) work the same way as SCMs, except that we use conditional probability tables instead of structural equations.
The conditional probability table for variable $\mathbf{V}$ expresses the distribution $P(\mathbf{V}|\mathbf{pa_V})$, where $\mathbf{pa_V}$ is the set of direct parents of $\mathbf{V}$.

Example

$$R \rightarrow W \leftarrow S$$

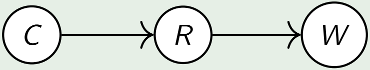
Conditional probability table:

- $P(W|R, S) = .9$
- $P(W|R, \neg S) = .8$
- $P(W|\neg R, S) = .8$
- $P(W|\neg R, \neg S) = .1$

Causal Bayes Nets

Causal Bayes Nets (CBNs) work the same way as SCMs, except that we use conditional probability tables instead of structural equations.
The conditional probability table for variable $\mathbf{V}$ expresses the distribution $P(\mathbf{V}|\mathbf{pa_V})$, where $\mathbf{pa_V}$ is the set of direct parents of $\mathbf{V}$.

Example



Here we would have two conditional probability tables:

- One for R,
- One for W.

Interventions in Causal Bayes Nets

Definition

In a CBN, an **intervention** Do(X=x) consists in replacing the conditional probability table for X with a new distribution putting all its mass at $X = x$.

Example

Conditional probability table for our $D \to H$ model:

- $P(H|D) = 4/5$
- $P(H|\neg D) = 1/5$

After intervention Do($H = 0$):

- $P(H) = 0$

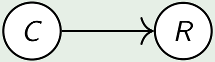
Relationship between CBNs and SCMs

- A CBN can be viewed as an abstracted, simplified version of an SCM.
- For each SCM, you can construct a single CBN that describes the conditional probabilities induced by the SCM.
  - But the opposite is not true: there are many SCMs that can correspond to the same CBN.

Joint distributions

A causal model (SCM or CBN) implicitly defines a joint probability distribution over its variables.

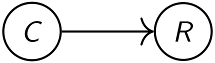
Example



C	R	$P(C, R)$
0	0	0.50
0	1	0.00
1	0	0.20
1	1	0.30

Table: Joint distribution $P(\mathbf{C}, \mathbf{R})$ for the model $C \to R$. Causal model parameters are $P(C) = .5$, $P(R|C) = .6$, $P(R|\neg C) = 0$.

How did we construct the joint distribution?



Causal model parameters are $P(C) = .5$, $P(R|C) = .6$, $P(R|\neg C) = 0$.
Using the chain rule, we have:

- $P(\mathbf{C}, \mathbf{R}) = P(\mathbf{R}|\mathbf{C})P(\mathbf{C})$
- $P(\mathbf{C}, \mathbf{R}) = P(\mathbf{C}|\mathbf{R})P(\mathbf{R})$

But the first expression is easier to compute!

- We can use the model parameters directly!

We call this expression the **factorization** of the causal model.

Factorization in bigger models

$$C \longrightarrow R \longrightarrow W$$

From the chain rule, we can write $P(\mathbf{C}, \mathbf{R}, \mathbf{W}) = P(\mathbf{W}|\mathbf{R}, \mathbf{C})P(\mathbf{R}|\mathbf{C})P(\mathbf{C})$.
But wait! $P(\mathbf{W}|\mathbf{R}, \mathbf{C})$ is just $P(\mathbf{W}|\mathbf{R})$.

- By definition the probability of W is entirely determined by the value of R.

So we simplify and write $P(\mathbf{C}, \mathbf{R}, \mathbf{W}) = P(\mathbf{W}|\mathbf{R})P(\mathbf{R}|\mathbf{C})P(\mathbf{C})$.
In general, we have:

Factorization (informal)

We can express a joint distribution over the variables in a causal model as a product of the conditional distributions and exogenous probabilities for that model.

Factorization

Factorization (informal)

We can express a joint distribution over the variables in a causal model as a product of the conditional distributions and exogenous probabilities for that model.

Theorem

Factorization. The joint distribution over the variables in a causal model can be expressed as:
$P(X_1, ..., Xn) = \prod_i P(X_i | \mathrm{pa}_{X_i})$,
where pa_{X_i} is the set of direct parents of X_i.

- (Note that we use the convention that $P(X_i | \emptyset) = P(X_i)$).

Causal models and compression

Factorization allows us to encode joint probability distributions in a compact way.

Example

You have eight binary variables $A_1, A_2, ..., A_8$.
Writing down the joint distribution table would take $2^8 = 256$ entries!
If the variables form a causal chain $A_1 \rightarrow A_2, \rightarrow, ..., \rightarrow A_8$, you only need $1 + 2 \times 7 = 15$ parameters.

$\rightarrow$ Something like a Causal Bayes Net can be useful even if you are not interested in causality per se...
$\rightarrow$ This is the idea behind the **Bayes Net**.

Bayes Nets

Bayes Nets (informal)

A Bayes Net is like a Causal Bayes Net except that it cannot be used to predict the consequences of intervention.

Definition

A Bayes Net is a joint probability distribution over variables $X_1, ..., X_n$, along with a DAG, such that the distribution can be factorized as: $P(X_1, ..., Xn) = \prod_i P(X_i | \text{pa}_{X_i})$, where pa_{X_i} are the direct parents of X_i in the DAG.

- (Reminder: DAG means 'Directed Acyclyc Graph')

Bayes Nets

Definition

A Bayes Net is a joint probability distribution over variables $X_1, ..., X_n$, along with a DAG, such that the distribution can be factorized as: $P(X_1, ..., Xn) = \prod_i P(X_i|\mathrm{pa}_{X_i})$, where pa_{X_i} are the direct parents of X_i in the DAG.

- (Reminder: DAG means 'Directed Acyclyc Graph')

Example

Take a joint distribution $P(\mathbf{A}, \mathbf{B})$. There are two different Bayes Nets that encode this distribution:

- $P(\mathbf{A}|\mathbf{B})P(\mathbf{B})$ with DAG $B \rightarrow A$.
- $P(\mathbf{B}|\mathbf{A})P(\mathbf{A})$ with DAG $A \rightarrow B$.

Arrows in a Bayes Net do not necessarily encode causal direction!

A historical note

- Historically, **Bayes Net** came first: they were invented as a way to efficiently encode joint distributions.
- But researchers soon realized that the arrows often flowed from cause to effect.
- **Causal Bayes Nets** were invented as a way to formalize this intuition.
- **Structural Causal Models** came even later.

We have taught these models in reverse order relative to their chronology.

- We think it makes more sense that way, but other sources follow the chronological order.

Back to the ladder

- **Bayes Nets**: can only answer **observational** (Level-1) queries.
  - If we see $X = x$, what is the value of Y?
- **Causal Bayes Nets**: does everything a BN does; can also answer **interventional** (Level-2) queries.
  - If we $\mathrm{Do}(X = x)$, what is the value of Y?
- **Structural Causal Models**: does everything a CBN does; can also answer **counterfactual** (Level-3) queries.
  - If X had been x (instead of its current value), what would have been the value of Y?

Why do we need SCMs for counterfactuals?

Answer on day 5!

Next lecture

How do we use causal models to make predictions?
How can we learn causal models?