Intro to Causal Inference

Mees de Vries (BrainCreators)

November 5, 2018

Overview

1. Motivation: why and what?
2. Causal models: the definitions and working with them.
3. Simpson's paradox: causal inference to the rescue!

Why Causal Reasoning?

Causality is not built into our normal statistical language: random variables X, Y, Z can be dependent, but there is no mathematical way to talk about which causes which.

Why Causal Reasoning?

Causality is not built into our normal statistical language: random variables X, Y, Z can be dependent, but there is no mathematical way to talk about which causes which.

"Correlation is not causation!"

Why Causal Reasoning?

Causality comes into play almost any time we want to take action based on statistical data. Consider:

- X: the frequency of high-sodium meals in my diet;
- Y: my blood pressure;
- Z: my headache frequency.

Why Causal Reasoning?

Causality comes into play almost any time we want to take action based on statistical data. Consider:

- X: the frequency of high-sodium meals in my diet;
- Y: my blood pressure;
- Z: my headache frequency.

We might very well have that (X, Y) and (Z, Y) are identically distributed. But we do not treat hypertension with painkillers!

Why Causal Reasoning?

- X: high-sodium meal count;
- Y: blood pressure;
- Z: headache count.

Why Causal Reasoning?

- X: high-sodium meal count;
- Y: blood pressure;
- Z: headache count.

If we want to model how a better diet would lower my blood pressure, we might want to estimate

$$\mathbb{P}(Y = y \mid X = x).$$

However, the quantity

$$\mathbb{P}(Y = y \mid Z = z)$$

gives us no such information. Why is this difference not in our mathematical language?

Do-calculus

Desideratum: an "operator for interventions". For X, Y random variables we want an expression

$$\mathbb{P}(Y = y \mid \mathrm{do}(X = x)).$$

Do-calculus

Desideratum: an "operator for interventions". For X, Y random variables we want an expression

$$\mathbb{P}(Y = y \mid \mathrm{do}(X = x)).$$

For X, Y, Z as before we should have

$$\mathbb{P}(Y = y \mid \mathrm{do}(X = x)) = \mathbb{P}(Y = y \mid X = x),$$

yet

$$\mathbb{P}(Y = y \mid \mathrm{do}(Z = z)) = \mathbb{P}(Y = y).$$

Example

Our causal models should be as simple as possible. We just want to encode causal relationships.

Example

Our causal models should be as simple as possible. We just want to encode causal relationships.

For our example:

$$X \rightarrow Y \rightarrow Z$$

Example

Our causal models should be as simple as possible. We just want to encode causal relationships.

For our example:

$$X \longrightarrow Y \longrightarrow Z$$

This says: X is a direct cause of Y, and Y is a direct cause of Z.

Example

Our causal models should be as simple as possible. We just want to encode causal relationships.

For our example:

$$X \longrightarrow Y \longrightarrow Z$$

This says: X is a direct cause of Y, and Y is a direct cause of Z. Implicit assumption: there are no relationships between the variables that are not in the model.

Definition

Definition

A *causal model* consists of:

- A DAG $G = (V, E)$,
- independent random variables $(U_v)_{v \in V}$,
- functions $(f_v)_{v \in V}$ of the appropriate type.

The *value* X_v in a causal model (G, U, f) at a node v is defined recursively as

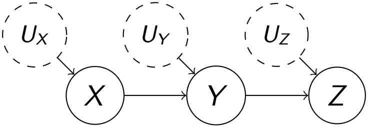
$$X_v = f_v(U_v, X_{\mathrm{par}(v)}).$$

Definition

Definition

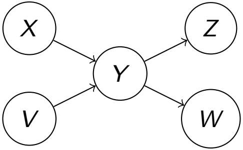
A *causal model* consists of $G = (V, E)$ with $(U_v)_{v \in V}$, $(f_v)_{v \in V}$.
The value in the causal model at v is $X_v = f_v(U_v, X_{\mathrm{par}(v)})$.

Causal model for hypertension:



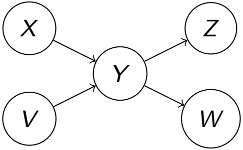
From now on, we omit the exogenous variables (the U_v) from our diagrams.

Dependence and independence



Graphical models already tell us a lot about dependence and independence.

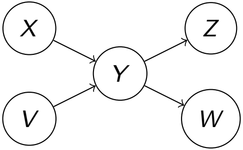
Dependence and independence



Graphical models already tell us a lot about dependence and independence.

- $X \perp V$
- $X \not\perp W$

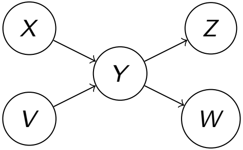
Dependence and independence



Graphical models already tell us a lot about dependence and independence.

- $X \perp V$
- $X \not\perp W$
- $Z \not\perp W$

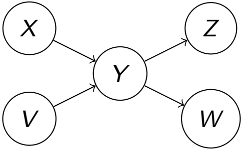
Dependence and independence



Graphical models already tell us a lot about dependence and independence.

- $X \perp V$
- $X \not\perp W$
- $Z \not\perp W$
- $Z \perp W \mid Y$

Dependence and independence



Graphical models already tell us a lot about dependence and independence.

- $X \perp V$
- $X \not\perp W$
- $Z \not\perp W$
- $Z \perp W \mid Y$
- $X \not\perp V \mid Y$

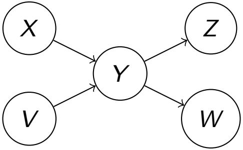
Dependence and independence

- A node v on an *undirected* path in a *directed* graph is a collider if both adjacent edges on the path go into v.
- Given $Z \subseteq V$, a node v on a path is Z-blocking if it either
  - is not a collider, and $v \in Z$, or
  - is a collider, and neither v nor any of its descendents are in Z.

Theorem

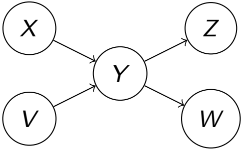
$X \perp Y \mid (Z_1, Z_2, \ldots Z_n)$ *if each undirected path* p *from* X *to* Y *has a* Z*-blocking node on it.*

Computations in Graphical Models



Graphical models allow us to simplify computations.

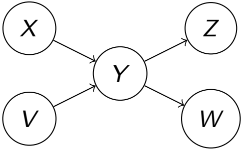
Computations in Graphical Models



Graphical models allow us to simplify computations.

$$\mathbb{P}(z, w, y, x, v) = \mathbb{P}(z \mid w, y, x, v)\mathbb{P}(w \mid y, x, v)\mathbb{P}(y \mid x, v)\mathbb{P}(x \mid v)\mathbb{P}(x)$$

Computations in Graphical Models

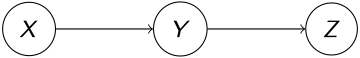


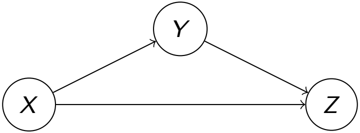
Graphical models allow us to simplify computations.

$$\begin{aligned}\mathbb{P}(z, w, y, x, v) &= \mathbb{P}(z \mid w, y, x, v)\mathbb{P}(w \mid y, x, v)\mathbb{P}(y \mid x, v)\mathbb{P}(x \mid v)\mathbb{P}(x) \\ &= \mathbb{P}(z \mid y)\mathbb{P}(w \mid y)\mathbb{P}(y \mid x, v)\mathbb{P}(x)\mathbb{P}(v)\end{aligned}$$

Caution: Intransitive

Note that models are not transitive! The following are different:





Causal Models and Bayesian Networks

So far, I have basically described Bayesian networks. What is the difference?

Causal Models and Bayesian Networks

So far, I have basically described Bayesian networks. What is the difference?

In Bayesian networks, the following two are essentially the same model:

$$X \rightarrow Y \rightarrow Z$$

$$X \leftarrow Y \leftarrow Z$$

This is obviously not what we want for causal networks!

Interventions

Recall that causality came into play with interventions. What is an intervention in our model?

Interventions

Recall that causality came into play with interventions. What is an intervention in our model?

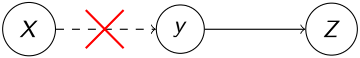
$$X \longrightarrow Y \longrightarrow Z$$

Equations:

$$\begin{aligned} X &= f_X(U_X) \\ Y &= f_Y(U_Y, X) \\ Z &= f_Z(U_Z, Y) \end{aligned}$$

Interventions

Recall that causality came into play with interventions. What is an intervention in our model?

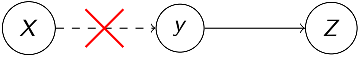


Equations:

$$
\begin{aligned}
X &= f_X(U_X) \\
Y &= f_Y(U_Y, X) \\
Z &= f_Z(U_Z, Y)
\end{aligned}
$$

Interventions

Recall that causality came into play with interventions. What is an intervention in our model?



Equations:

$$
\begin{aligned}
X' &= f_X(U_X) \\
Y' &= y \\
Z' &= f_Z(U_Z, Y') = f_Z(U_Z, y)
\end{aligned}
$$

We use the $'$ to distinguish the random variables from the original and modified networks.

That do Operator

Definition

Let (G, U, f) be a causal model with nodes X, Y. Then

$$\mathbb{P}(Y = y \mid \text{do}(X = x))$$

is the probability that $Y' = y$ in the modified network with X set to x.

That do Operator

Definition

Let (G, U, f) be a causal model with nodes X, Y. Then

$$\mathbb{P}(Y = y \mid \mathrm{do}(X = x))$$

is the probability that $Y' = y$ in the modified network with X set to x.

Example: hypertension case.

Simpson's Paradox

There is a particular medication A on the market for disease X. Statisticians want to test the effect of the medication. They randomly sample people afflicted by X, and find:

Treatment	Recovery (male)	Recovery (female)	Recovery (total)
Medication A	$9/10 = 90\%$	$20/35 = 57\%$	$29/45 = 64\%$
None	$35/40 = 88\%$	$8/15 = 53\%$	$43/55 = 78\%$
Total	$44/50 = 88\%$	$28/50 = 54\%$	$72/100 = 72\%$

Simpson's Paradox

There is a particular medication A on the market for disease X. Statisticians want to test the effect of the medication. They randomly sample people afflicted by X, and find:

Treatment	Recovery (male)	Recovery (female)	Recovery (total)
Medication A	$9/10 = 90\%$	$20/35 = 57\%$	$29/45 = 64\%$
None	$35/40 = 88\%$	$8/15 = 53\%$	$43/55 = 78\%$
Total	$44/50 = 88\%$	$28/50 = 54\%$	$72/100 = 72\%$

Question: should medication A be banned?

Simpson's Paradox

There is a particular medication A on the market for disease X. Statisticians want to test the effect of the medication. They randomly sample people afflicted by X, and find:

Treatment	Recovery (male)	Recovery (female)	Recovery (total)
Medication A	$9/10 = 90\%$	$20/35 = 57\%$	$29/45 = 64\%$
None	$35/40 = 88\%$	$8/15 = 53\%$	$43/55 = 78\%$
Total	$44/50 = 88\%$	$28/50 = 54\%$	$72/100 = 72\%$

Question: should medication A be banned?

- The medication has a bad effect on the whole population...

Simpson's Paradox

There is a particular medication A on the market for disease X. Statisticians want to test the effect of the medication. They randomly sample people afflicted by X, and find:

Treatment	Recovery (male)	Recovery (female)	Recovery (total)
Medication A	$9/10 = 90\%$	$20/35 = 57\%$	$29/45 = 64\%$
None	$35/40 = 88\%$	$8/15 = 53\%$	$43/55 = 78\%$
Total	$44/50 = 88\%$	$28/50 = 54\%$	$72/100 = 72\%$

Question: should medication A be banned?

- The medication has a bad effect on the whole population...
- but the medication has a good effect on both men and women!

Simpson's Paradox (alternative)

The statisticians decide to investigate further. They find 100 women who have the early stages of the disease for a double blind study. They additionally measure increase or decrease of a relevant protein P before and after treatment.

Treatment	Recovery (inc)	Recovery (dec)	Recovery (total)
Medication A	$1/10 = 10\%$	$28/40 = 70\%$	$29/50 = 58\%$
Placebo	$9/25 = 36\%$	$18/25 = 72\%$	$27/50 = 54\%$
Total	$10/35 = 29\%$	$46/65 = 71\%$	$56/100 = 56\%$

Simpson's Paradox (alternative)

The statisticians decide to investigate further. They find 100 women who have the early stages of the disease for a double blind study. They additionally measure increase or decrease of a relevant protein P before and after treatment.

Treatment	Recovery (inc)	Recovery (dec)	Recovery (total)
Medication A	$1/10 = 10\%$	$28/40 = 70\%$	$29/50 = 58\%$
Placebo	$9/25 = 36\%$	$18/25 = 72\%$	$27/50 = 54\%$
Total	$10/35 = 29\%$	$46/65 = 71\%$	$56/100 = 56\%$

Question: should medication A be banned (for women)?

Simpson's Paradox (alternative)

The statisticians decide to investigate further. They find 100 women who have the early stages of the disease for a double blind study. They additionally measure increase or decrease of a relevant protein P before and after treatment.

Treatment	Recovery (inc)	Recovery (dec)	Recovery (total)
Medication A	$1/10 = 10\%$	$28/40 = 70\%$	$29/50 = 58\%$
Placebo	$9/25 = 36\%$	$18/25 = 72\%$	$27/50 = 54\%$
Total	$10/35 = 29\%$	$46/65 = 71\%$	$56/100 = 56\%$

Question: should medication A be banned (for women)?

- The medication has a good effect on the whole population of women...

Simpson's Paradox (alternative)

The statisticians decide to investigate further. They find 100 women who have the early stages of the disease for a double blind study. They additionally measure increase or decrease of a relevant protein P before and after treatment.

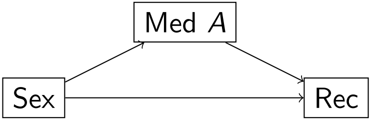
Treatment	Recovery (inc)	Recovery (dec)	Recovery (total)
Medication A	$1/10 = 10\%$	$28/40 = 70\%$	$29/50 = 58\%$
Placebo	$9/25 = 36\%$	$18/25 = 72\%$	$27/50 = 54\%$
Total	$10/35 = 29\%$	$46/65 = 71\%$	$56/100 = 56\%$

Question: should medication A be banned (for women)?

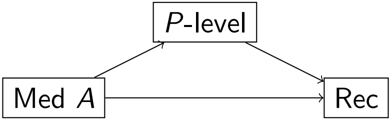
- The medication has a good effect on the whole population of women...
- but the medication has a bad effect on both people whose P-levels increase and decrease!

Causal Model for SP/SPa

The causal models are different for the two cases. Clearly, medication usage or recovery outcome cannot *cause* a difference in sex, so:

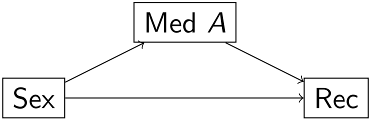


However, in our randomized trial, P-level cannot *cause* a difference in medication usage, so:



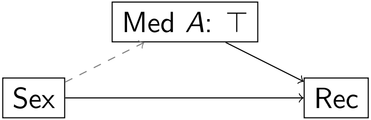
Interventions for SP

What if we make the intervention "take medication" in the original paradox?



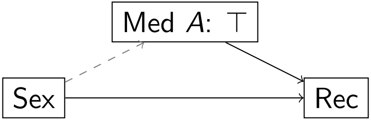
Interventions for SP

What if we make the intervention “take medication” in the original paradox?



Interventions for SP

What if we make the intervention "take medication" in the original paradox?

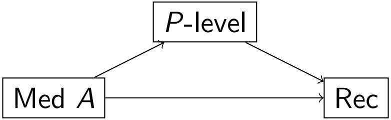


So we get:

$$
\begin{aligned}
\mathbb{P}(\text{Rec} = \top \mid \text{do}(A = \top)) &= \mathbb{P}(\text{Rec}' = \top) \\
&= \sum_{s \in \{M,F\}} \mathbb{P}(\text{Rec}' = \top \mid \text{Sex}' = s)\mathbb{P}(\text{Sex}' = s) \\
&= \sum_{s \in \{M,F\}} \mathbb{P}(\text{Rec} = \top \mid A = \top, \text{Sex} = s)\mathbb{P}(\text{Sex} = s).
\end{aligned}
$$

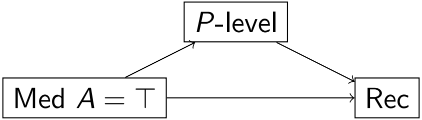
Interventions for SPa

What about the alternate paradox?



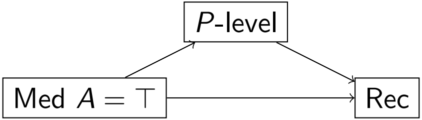
Interventions for SPa

What about the alternate paradox?



Interventions for SPa

What about the alternate paradox?



So we get:

$$\begin{aligned}\mathbb{P}(\mathrm{Rec} = \top \mid \mathrm{do}(A = \top)) &= \mathbb{P}(\mathrm{Rec}' = \top) \\ &= \mathbb{P}(\mathrm{Rec} = \top \mid A = \top)\end{aligned}$$

So do not separately take the P-level cases into account!

Take-aways

- Basic causal modeling might not directly lead to new insight, but it helps us formalize our intuitions.
- Several counterintuitive problems become nicer in the context of causal modeling.
  - Simpson's paradox
  - Berkson's paradox
  - Monty Hall problem
- Given a causal model, you can often answer questions about interventions without making interventions.
- Major drawback: we require a DAG. Many real-world complex systems have feedback, and are thus not acyclic.