Bryce Besler
2019.08.23

Stastical Learning Group - Causal Modelling

Objectives

- Understand the three building blocks of causality: mediation, confounding, causality
- Define a Structural Causal Model
- Describe applications of SCMs

Causality

We have intuition about causality but have struggled for centuries to provide a precise definition. Often you will read papers that talk about "associations", "relations", or the completely absorbent "correlation". However, we can immediately call causality if we don't think too hard.

Does the rooster crow cause the sun to rise?
Does advil cause blindness?
Did our dance anger the Gods, causing a drought?

Graphically, we are trying to define the following:

$$X \longrightarrow Y$$

X causes Y

How would you define cause?

We are going to use the definition by David Lewis based in the language of counterfactuals

"Y would not have occured if not for X"

Examples   Had I made that shot our team would have won the game
Had I put on winter tires I wouldn't have rearended that guy.

When in doubt, ask: If not for X, would Y happen?

<u>Basics of Causality</u>
We are going to explore three basic structures that elucidate a way of thinking about causalition.

Example: Mediation   How does a fire alarm work?

$$\text{Fire} \rightarrow \text{Alarm}$$

But of course this is wrong as the alarm must detect a product of fire

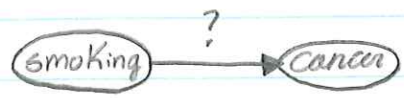
$$\text{Fire} \rightarrow \text{Smoke} \rightarrow \text{Alarm}$$

Really, we have a smoke detector! Now ask how we could fool the alarm (say with Bryce's cooking)

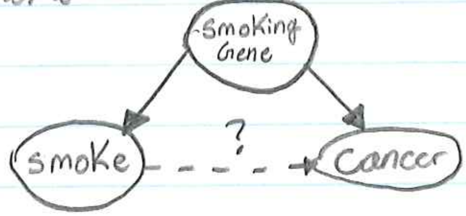
Mediators define the mechanism of causation

$$X \rightarrow Z \rightarrow Y$$

Example: Confounder There was a great debate in the 60's about whether smoking caused lung cancer.

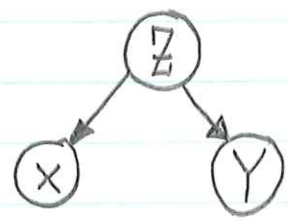


The issue was that Ronald Fisher postulated a gene that increased one's chance of cancer and propensity to smoke



Fisher was concerned that we could not randomize genetics. Could we ever establish if smoking caused cancer?

Confounders cause both X and Y

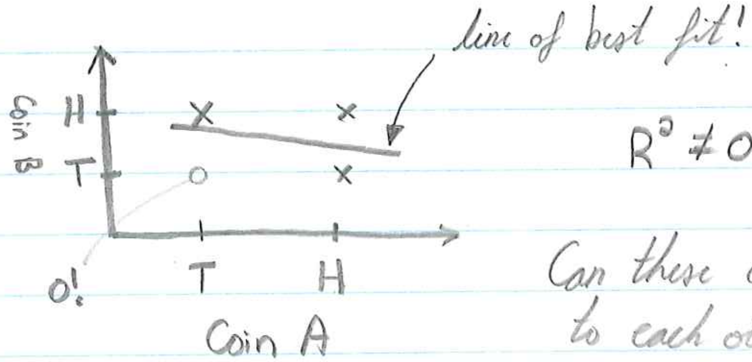


Example: Collider

Consider a game where you flip two coins and record if they are heads or tails. However, there is a bell in the room which rings if either coin is a heads. Now, you only record the coins if the bell rings. What is the relationship between the coins?

Coin A	Coin B	# Occurances
H	H	25
H	T	25
T	H	25
~~T~~	~~T~~	~~25~~

However, we don't record here!

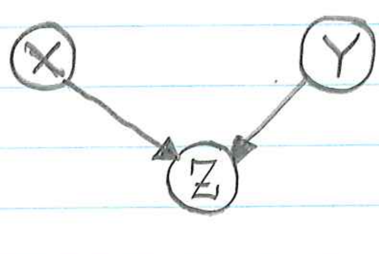


$R^2 \neq 0$!

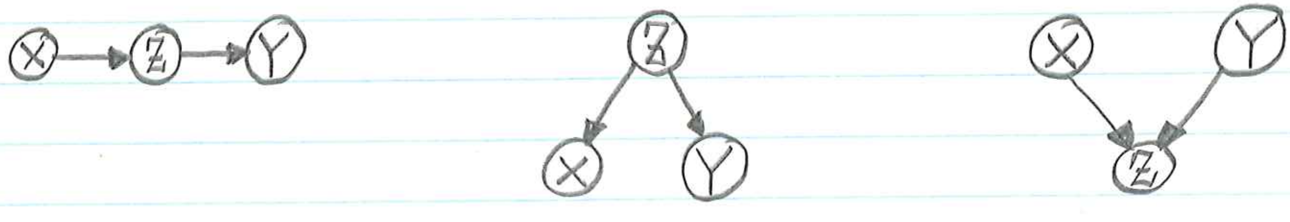
Can these coins talk to each other?

The issue is we conditioned on the Bell ringing. Here, the bell is a collider.

A collider has two or more causes



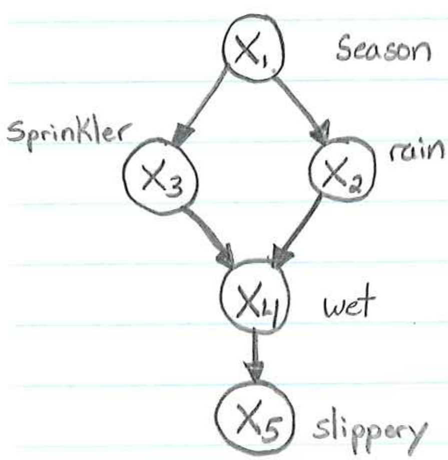
We can now define our three building blocks



Mediator Confounder Collider

These allow us to build complex models of the causal world. They help in experimental design, logistics, interpretation, etc. We work through some simple examples.

Example: Slippery Sidewalk



Mediators: Sprinkler, Rain, (Wet)
Collider: Wet
Confounder: Season

If I wanted to know the effect of Season on Slippery, should I control for Wet?

What variable should I control for if I want to know the effect of Rain on Slippery?

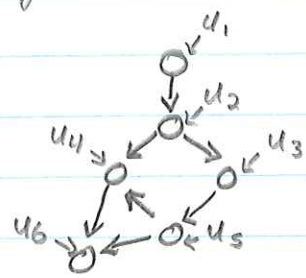
Do you need a randomized experiment to understand the effect of Wet on Slippery? What about rain on slippery?

Structural Causal Model

We are now ready for the definition of a Structural Causal Model (SEM).

SEMs are a triple $M = \langle U, V, F \rangle$

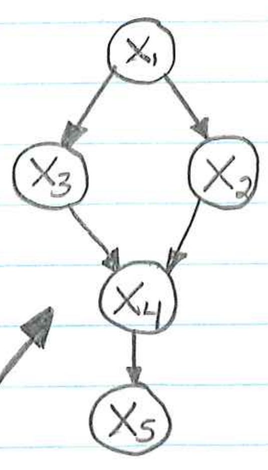
$v_i = f(PV_i, u_i)$



U background (exogenous) variables determined by factors outside the model.
V endogenous variables determined by variables inside the model
F set of functions mapping from parents of V_i to V_i.

If exogenous variables are not known, a probability function $P(u)$ can be added to each U, similar in spirit to a prior.

Example: Slippery again



Directed graph

set of functions

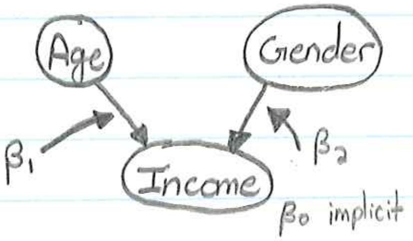
$X_1 = u_1$
$X_2 = [(X_1 = \text{winter}) \vee (X_1 = \text{fall}) \vee u_2] \wedge \neg u_2'$
$X_3 = [(X_1 = \text{summer}) \vee (X_1 = \text{spring}) \vee u_3] \wedge \neg u_3'$
$X_4 = [X_2 \vee X_3 \vee u_4] \wedge \neg u_4'$
$X_5 = [X_4 \vee u_5] \wedge \neg u_5'$

Example: Linear Regression

You read an academic paper where they build the following linear model from data.

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

Y = Income
X_1 = Age
X_2 = Gender

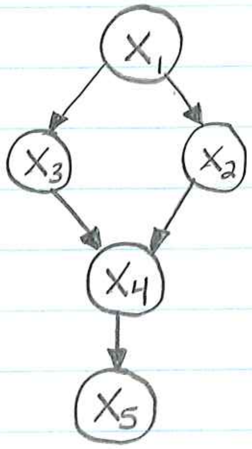


$$\overset{\varepsilon}{\downarrow}$$
$$u_{\text{Income}} \sim N(0, \sigma^2)$$

We can very easily create continuous outcome situations just like we do in standard statistics. We can go further, having nested models with non-linear functions at each vertex.

Computing with SEMs

Now, we would like to do something with these models. Lets see how that is done.



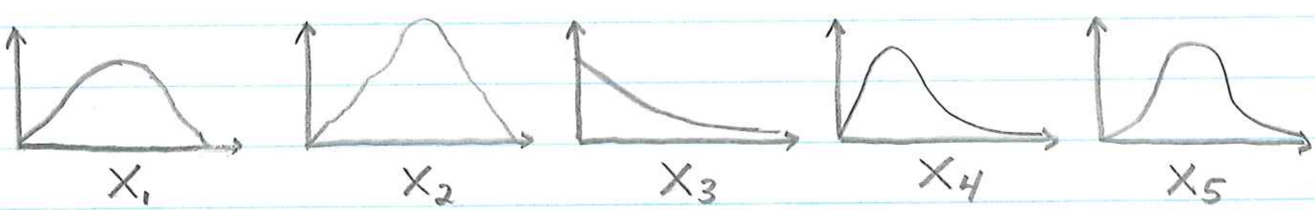
Consider a probabilistic SEM

$$M = \langle U, V, F, P(u) \rangle$$

We are given the following mappings

$$
\begin{aligned}
X_1 &= u_1 & u_1 &\sim N \\
X_2 &= \beta_0 + \beta_1 X_1 + u_2 & u_2 &\sim N \\
X_3 &= X_1 \exp(u_3) & u_3 &\sim U \\
X_4 &= X_2 X_3 u_3 & u_4 &\sim \text{Bernoulli} \\
X_5 &= \beta_0 + \beta_1 X_4 + u_5 & u_5 &\sim N
\end{aligned}
$$

We could simply sample U, pump through the equations, and see what we get for V. Doing this many times, we get a sampling distrubution.

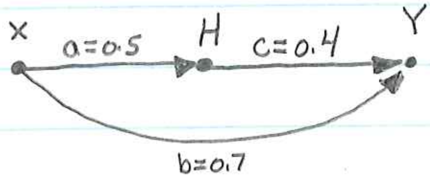


We could ask questions

$$E[X_5] \; ?$$
$$Var[X_4] \; ?$$
$$P(X_3) \; ?$$

Finally, we can climb the "Ladder of Causation" as coined by Judea Pearl.

Example: Exam scores.



X = Encouragment
H = Homework
Y = Exam Score

$$\left.\begin{aligned} X &= u_X \\ H &= aX + u_H \\ Y &= bX + cH + u_Y \end{aligned}\right\} \text{ Linear System}$$

$\sigma_{u_i u_j} = 0 \quad \forall\, i,j \in \{X, H, Y\}$ (exogeneous vars uncorrelated)

Prediction What is the expected exam score if you observe that Johnny studied for 1 hour?

$$E[Y \mid H=1] = b\, E[X \mid H=1] + c \cdot 1 + E[u_Y \mid H=1]$$

Having observed H, we gain information about X

$$E[X \mid H=1] = 1/a - u_H$$

Knowing $\sigma_{u_i u_j} = 0$, this is independant

Action What is the expected exam mark if study hours are controlled to be one?

$$E[Y \mid do(H=1)] = b\, E[X] + c \cdot 1 + E[u_Y \mid H=1]$$

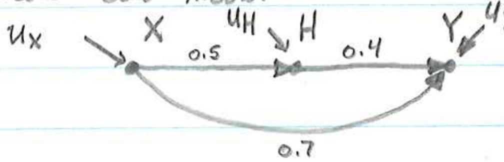
intervention "breaks" the connection between X and H.

Counter Faction

Given that Johnny was moderatly encouraged ($X=0.5$), studied hard ($H=1$) and scored $Y=1.5$, what would Johnny's score have been if he had doubled his homework?

$$E[Y_{H=2} \mid H=1] = ?$$

1) Update our model



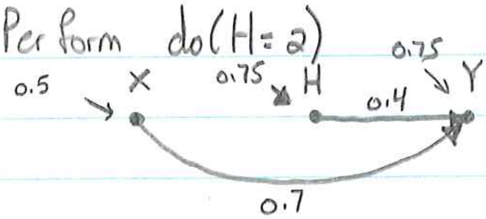
$X = u_X = 0.5$

$H = aX + u_H \Rightarrow u_H = 1 - 0.5 \cdot 0.5 = 0.75$

$Y = bX + cH + u_Y \Rightarrow u_Y = 1.5 - 0.7 \cdot 0.5 - 0.4 \cdot 1 = 0.75$

$U = (u_X, u_Y, u_Z)$ defines Johnny's exogeneous state.

2) Perform $do(H=2)$



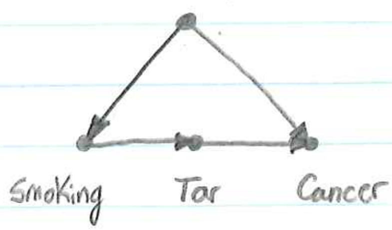
$$E[Y] = \overbrace{0.7 \cdot 0.5}^{u_X} + \overbrace{0.4 \cdot 2}^{H} + \overbrace{0.75}^{u_Y}$$

$$E[Y] = 1.90$$

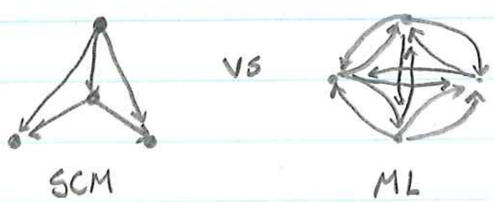
Had Johnny studied two hours, we would have expected a score of 1.90

Applications

Causal modelling has implications for experimental design, statistical analysis, artificial intelligence and computers.

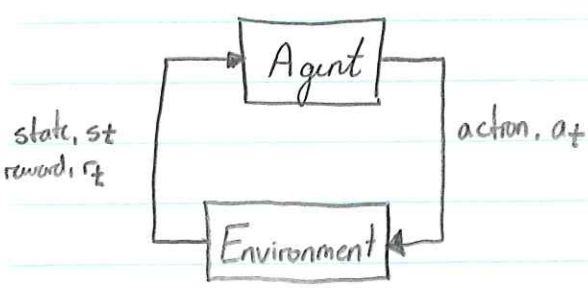


The smoking gene was later discovered. However, application of SCM allows us to still assess the effect of smoking on cancer by using tar to block the confounder.



These models greatly help with dimensionality reduction because of the sparsity of the models.

In experimental design, analyzing SCMs allows one to maximize resource utilization and know whether an effect needs a randomized experiment or can be estimated from observational data.



One large area of exploration is Reinforcement Learning. A machine with an understanding of causality of actions and environment appears close in spirit to a thinking machine.

References

Judea Pearl   Causality
"   Book of Why
"   Causal Inference in Statistics: A Primer