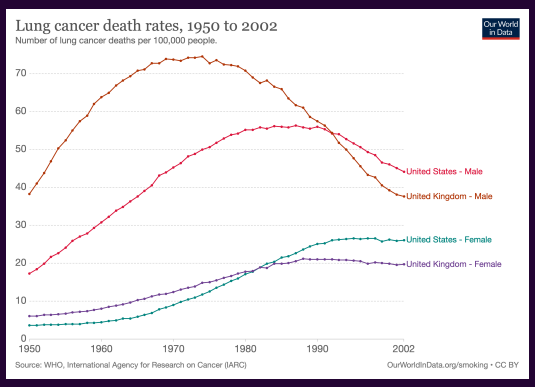
Machine learning

Causal inference

Joshua Loftus

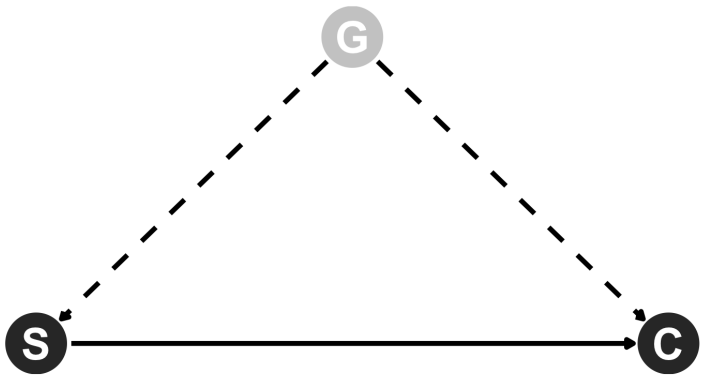


Hey kids... don't smoke

R. A. Fisher on smoking and lung cancer (in 1957)

> ... the B.B.C. gave me the opportunity of putting forward examples of the two classes of alternative theories which **any statistical association, observed without the predictions of a definite experiment**, allows--namely, (1) that the supposed effect **is really the cause**, or in this case that incipient cancer, or a pre-cancerous condition with chronic inflammation, is a factor in inducing the smoking of cigarettes, or (2) that cigarette smoking and lung cancer, though not mutually causative, are **both influenced by a common cause**, in this case the individual genotype ...

Graphical notation for causality

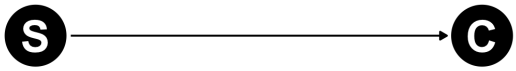


Variables: vertices (or nodes)

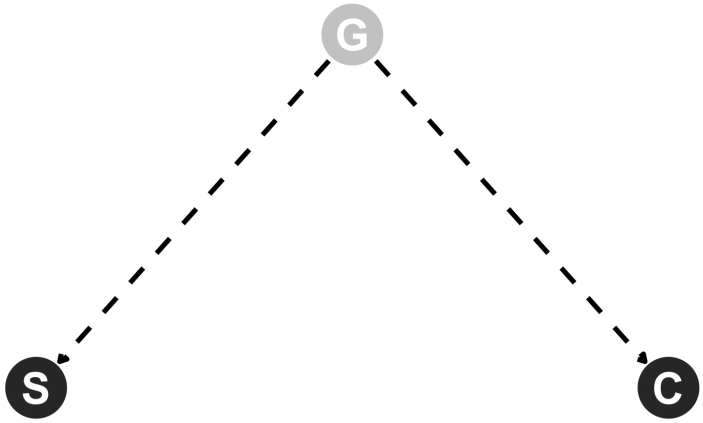
Relationships: directed edges (arrows)

Shaded node / dashed edges: unobserved variable

Smoking causes cancer?



Genotype is a common cause?



Fisher: correlation is not causation

He did not use graphical notation like this

But the graphs can be very useful

Explaining an observed correlation

We find a statistically significant correlation between X and Y

What does it mean?

1. False positive (spurious correlation)
2. X causes Y
3. Y causes X
4. Both have common cause U [possibly unobserved]

Statistically indistinguishable cases (without "experimental" data)

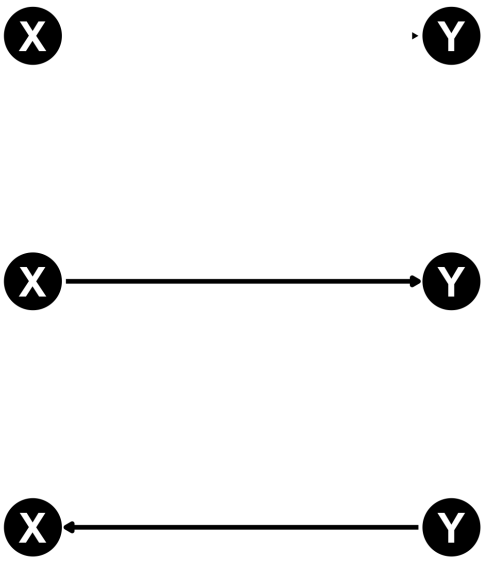
Importantly different consequences!

A simple mathematical model of causality

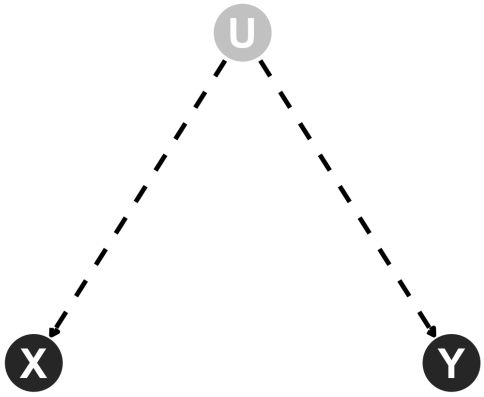
Think about **interventions** that change some target variable T

- Forget about the arrows pointing into T (intervention makes them irrelevant)
- Change T, e.g. setting it to some arbitrary new value $T = t$
- This change propagates along directed paths out of T to all descendant variables of T in the graph, causing their values to change

All of these changes could be deterministic, but most likely in our usage they are probabilistic



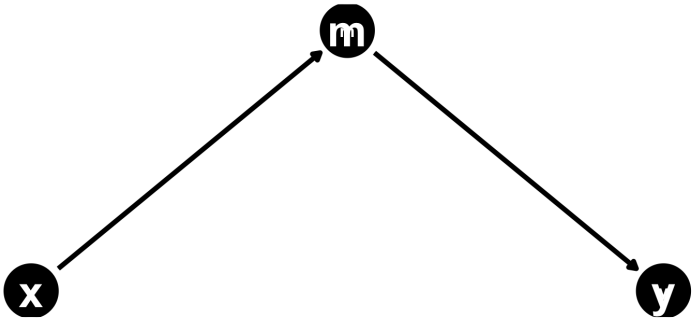
Exercise: in each of these cases, if we intervene on X which other variable(s) are changed as a result?



Computing counterfactuals

If we know/estimate *functions* represented by edges, we can simulate/compute the consequences of an intervention

$$x = \text{exogeneous}, \quad m = f(x) + \varepsilon_m, \quad y = g(m) + \varepsilon_y$$



$$x \leftarrow x', \quad m \leftarrow f(x') + \varepsilon_m, \quad y \leftarrow g(f(x') + \varepsilon_m) + \varepsilon_y$$

Causal inference: much more difficult

Predictive machine learning is about

$$p_{Y|X}(y|x)$$

and regression--conditional expectation, conditional quantile, etc. If we passively observe some value of x, what would we observe about y?

Causal inference is about (various notations)

$$p(y|\mathrm{do}[X = x]), \quad \text{i.e.} \quad p(y|X \leftarrow x)$$

i.e. what happens to Y when we actually **intervene** on X

If we actively change x, what would we observe about y?

Experiments

Actually do interventions while collecting data

Observational studies

Try to infer causal relationships without interventions, by using ~~dark arts~~ more/specialized assumptions/methods that require careful interpretation

Scientific progress: be wrong in more interesting/specific ways

Potential outcomes: another causal framework

Relative strengths/weaknesses compared to DAGs

- Narrow focus: goal is to estimate one edge in a graph
- Difficult to express more complex relationships

This is not a course on causal inference

Covering a few basics for interesting connections to ML!

So when is ML (and e.g. regression) useful for causal inference?

Idea: adjusting for confounders

Confounders: other variables that obscure the (causal) relationship from X to Y, e.g.

- Y: health outcome
- X: treatment dose
- Z: disease severity

Without considering Z, it might seem like larger doses of X correlate with worse health outcomes

Solution: add more variables to the model

Make model complex enough to capture important factors

Similarly, might need to model *non-linear causal relationships*

Strategy: two staged regression

Two-stage least squares (2SLS)

Suppose we want to learn the causal relationship of D on Y, but

$$Y = D\theta + X\beta + \varepsilon_Y$$

$$D = X\alpha + \varepsilon_D$$

In words: X is confounding the relationship

- First stage: regress out X
- Second stage: using residuals from first stage,

$$\text{regress } Y - X\hat{\beta} \text{ on } D - X\hat{\alpha}$$

Powerful, intuitive idea

Orthogonal projection

We'll come back to this

(Think about fitting the relationships using ML instead of regression)

Propensity

General theme: applying ML strengths to causality

Special case: if the "treatment" (causal) variable is categorical

Many causal methods for this case involve predicting the treatment itself

i.e. prediction with categorical outcome, **classification**

We'll also come back to this (propensity methods) after covering classification

Summary: a key ingredient in many causal inference methods involves classification, can leverage ML tools

Guiding ideas / warnings

- More complex models (ML) do not guarantee better causal inference

Might even make things worse (just like with prediction)

- Models with better predictions may be worse for causal inference

Even if prediction accuracy is measured on test data!

- Inference = causal inference...?

Or, at least, causal interpretations can be special case of trade-off between prediction and inference/interpretation

Causal inference

An exciting interdisciplinary field

Practically important, connections to ML

"Data scientists have hitherto only predicted the world in various ways; the point is to change it" - Joshua Loftus

Mixtape... remix? 🤔