Chapters 8 to 10

Josh Bollar

November 18, 2018

8 Counterfactuals

Counterfactual reasoning is essential for:

1. Assigning blame (e.g., court room case)
2. Affecting future policy (e.g., climate change) – requires an inductive assumption (even if we show that CO2 increasing caused climate change last year we have to still make that claim for next year).

8.1 Types of Counterfactual Analysis

Their are three different types of causes:

1. Necessary - A necessarily causes B if without A, B would not have happened
2. Sufficient - A sufficiently causes B if A happening results in B happening
3. Necessary and sufficient - both

Both *necessary* and *sufficient* have obvious numerical extensions, describing the strength to which degree A causes B, but we can't phrase these expressions in terms of probability alone because we need a counterfactual language.

Example. Necessary not sufficient cause. See below piano example.

Example. Sufficient not necessary cause. Consider the firing squad with two soldiers. If both soldiers fire, then *in that world* neither is necessary (since the other would have killed the prisoner) but both are sufficient.

Note. Directness and causation are orthogonal. For instance, Joe can block the fire exit with a large object, causing someone to die. Even though we have a mediator (large object blocking a pathway), we still have sufficient indirect causation.

It's too weak to have notions of necessary and sufficient – we can only talk about degrees of necessity and sufficiency (and perhaps the semantics of the binary words are that we have a threshold of necessity and sufficiency). Consider the following example:

> Alice attempts to shoot Bob, but misses. Bob, in an attempt to flee, runs under a building where someone is moving in with a piano. The piano falls by accident, killing Bob.

Was the gunshot sufficient to kill Bob? Not really, the gunshot missed and the probability of the piano falling on Bob is low. It would be unreasonable for Alice to plan to kill Bob by scaring him with a gunshot into running under a piano. We wouldn't charge Alice with manslaughter, just attempted manslaughter.

Was the gunshot necessary to kill Bob? This shows that the "necessary" definition is underspecified. In general, Bob can die from natural causes, so a gunshot is not strictly necessary. However, we probably want to hold Bob's general health constant when evaluating whether the gunshot was necessary to kill Bob. In that case, the gunshot was necessary to kill Bob, since Bob only fled due to the gunshot.

This tells us we need a richer language to express what we mean by "strength of a cause" as well as sufficiency and necessity because causality is dependent on "what world" we're analyzing the causal effect in.

Note. Why do we need to bother with sufficiency? Consider using a match to burn down a house. Is the arson the match's fault or the oxygen's? Both are necessary, but only one is useful, and that's the one that has stronger sufficient causal strength.

8.2 Expressing Causality Formally

For simplicy, we're querying the causal effect of the indicator X being activated on the indicator Y.

The probability of necessity, PN, is:

$$\mathrm{PN} = \mathbb{P}\left\{Y_{X=0} = 0 \,|\, X = 1, Y = 1\right\} .$$

The probability of sufficiency, PS, is:

$$\mathrm{PS} = \mathbb{P}\left\{Y_{X=1} = 1 \,|\, X = 0, Y = 0\right\} .$$

What does $\mathbb{P}\left\{Y_{X=a} = b|E\right\}$ mean? Why is it not $\mathbb{P}\left\{Y = b \,|\, \mathrm{do}(X = a), E\right\}$? $Y_{X=a}$ is the variable Y after we update our belief about the nature of Y and its relationship to X given that we observe E had occured (Pearl calls this "hindsight"). The original paper gives a full elaboration: in particular, the do version calculates the probability that $Y = b$ removing all arrows into X in the original causal model. $Y_{X=a} = b$ considers the causal *submodel* (literally the subgraph of the original causal model with a fixed to X), and uses the induced conditional probability tables from the subgraph to compute the behavior of Y, but with respect to the missing nodes taking on values from the original causal model (letting us ask questions that seemingly contradict themselves like $\mathbb{P}\left\{Y_{X=1} = 1, X = 0\right\}$).

9 Mediation

There exist causal paths that are not one edge long. Wherever this happens a natural question how much of the root cause is the direct effect or indirect effects? This affects future treatments.

For example, the mediator of "Department you apply to" explains Simpson's paradox in where gender effects the department applied to and both of those affect the admissions outcome.

We consider a couple of proposed measurements of direct/indirect effect (controlled vs natural). We consider binary variables again for simplicity and as for the effect of X on Y with a mediator M.

$$\begin{aligned}
\text{CDE}(m) &= \mathbb{P}\left\{Y = 1 \,|\, \text{do}(X=1), \text{do}(M=m)\right\} - \mathbb{P}\left\{Y = 1 \,|\, \text{do}(X=0), \text{do}(M=m)\right\} \\
\text{NDE}(m) &= \mathbb{P}\left\{Y_{M=m} = 1 \,|\, \text{do}(X=1)\right\} - \mathbb{P}\left\{Y_{M=m} = 0 \,|\, \text{do}(X=0)\right\}
\end{aligned}$$

The CDE controls M and asks how X increases the probability of Y. The NDE considers how M changes naturally due to changes in X, and lets those affect the probability of $Y_{M=m} = 0$, but only keeping M constant in a counterfactual.

$$\text{NIE}(m) = \mathbb{P}\left\{Y_{M=1} = 1 \,|\, \text{do}(X=1)\right\} - \mathbb{P}\left\{Y_{M=0} = 0 \,|\, \text{do}(X=0)\right\}$$

This section needs elaboration/paper link (TODO).

CDE + CIE does not equal total effect, but NDE - NIE does! TODO look at an example here.

Question. This chapter helps assess proposed mediators versus their lack of mediators, but there's a more metaphiscally dubious question at hand here where we wonder "where are we done?" Why is the model $X \to Y$ vs $X \to A \to Y$? Why can't it be $X \to A \to B \to Y$ or $X \to A \to B \to C \to Y$? When do you stop being fine-grained enough?

9.1 AI and other Grandstanding

TODO transduction

TODO why is causality essential for strong ai