Reader Report

Reading: Pearl 2009a 3.4 and chapter 5

Claudia Shi (js5334)

May 6, 2019 A counter-factual claim could take form like this

"Y would be y had X been x in situation $U = u$". We could interpret the phrase "had X been x" as an instruction to modify the structural equation model. With that as our intuition, we formally define unit level counter-factual as this:

$$Y_x(u) \triangleq Y_{M_x}(u)$$

where M is the structural model, M_x is the modified version of the structual model had $X = x$, and u is a unit.

There's something subtle about defining the term in a unit level. "unit" here is interpreted as "individual". As the example in the footnote, "Drinking hemlock causes death" and "Socrates drinking hemlock caused his death" are the distinction between population-level causes and unit-level causes. Pearl points out that, the law of nature, would remain invariant as the individual causes reflected in function. I am not sure if I understood this sentence. What is the "law of nature"? Is it the inherent causality of the world? Is it the population effect?

Okay, dialing this down a bit, this is a question of attribute query. When is it identifiable then? In a linear system, the result does not depend on the distribution of U, so it can just be a additive relationship that's measured by the coefficient. In nonlinear systems, we are not so lucky. (Do we really know what's linear relationship and what's not in observational data?

)

Chapter 5 discussed a few cases of "counterfactuals at work" It covered a) mediating effects and b) causes of effects and probabilities of causation.
Direct effect can be broken down into two categories, controlled direct effect and natural direct effect. NDE is some sort of nested counterfactual. I understand the difference between CDE and NDE is that in an ideal world, CDE would be sufficient! But we live in a would where people are fickle, things are unpredictable, so just fixing a mediating variable is not sufficient, as it might influence other variables that are nearly impossible to measure.

probability of necessity is cool. It is allows us to conditional on the outcome. Moreover, assuming a monotonic relationship between Y and X

$$PN = \frac{P(y|x) - P(y|x')}{P(y|x)} + \frac{P(y|x') - P(y|do(x'))}{P(x,y)}$$

The second term is clever in that it is a correction for confounding bias!