The Logic of Structure-Based Counterfactuals

Lorenzo Lagos
March 28, 2017

Pearl makes a good effort at pointing out the differences between the structural approach and the Neyman-Rubin framework. He points out that the counterfactual quantities $Y_x(u)$ in the Neyman-Rubin framework are taken as primitive. This means that this approach lacks a mathematical model that would allow one to decide, for example, whether a collection of potential-outcome expressions is contradictory. It would be helpful to provide an example of such a contradiction. In the following graphical model $Z \rightarrow X \rightarrow Y$, writing down $Y_{xz}(u)$ may be contradictory if $X_z(u) = x'$ instead of x because under the same background conditions u having $Z = z$ implies that $X \neq x$ but rather $X_z(u) = x'$. My guess is that Pearl is not referring to this type of contradiction, because the do notation would allow for such a counterfactual. Isn't this then a weakness of the do notation? That is, that any sort of intervention is allowed in graphical models, even if it is contradictory to how the world works reality.

I agree with Pearl that structural equations and their associated graphs are particularly useful for expressing assumptions about cause-effect relationships Meanwhile, in the Neyman-Rubin framework, assumptions are articulated as conditional independence relationships involving counterfactual variables. According to Pearl, this means that in the potential-outcomes framework, 1) assumptions and implications are formulated *after* the collection of data; and 2) causal assumptions are then communicated in terms of the absence of correlations. I disagree with the former, because even economists have to put a lot of effort in finding an instrument that is *a priori* reasonable. This means that they do have a model of how the world works in the back of their heads(perhaps not as clearly expressed as a graphical model). The latter statement is true to some extent. Any applied work has to provide the evidence supporting evidence for their assumptions. Simply drawing a cute graph doesn't get rid of he fact that the assumptions implied in that graph are true. Hence, it is always useful to communicate causal assumptions in terms of the absence of correlations because that is what is done in practice to check the assumptions. However, I agree that a graphical model puts us in a situation that forces us to think of the substantive knowledge of our field as to ascertain whether such an independence condition holds.

The coolest point of the chapter was footnote 29. Pearl shows that Angrist's assumptions for an instrument (Z is ignorable and Z affects X) are too strong. Nonetheless, both 7.8(b) and 7.8(c) a bi-directed arrow from Z to X that is separate from the bi-directed arrow from X to Y. I wish the best of luck to any empiricist that proposes such an instrument, given that it will be impossible to provide any convincing empirical evidence that those two bi-directed arrows are not connected.