Reading Response

Susanna Makela

February 1, 2017

Chapter 4 of Winship & Morgan introduces the conditioning strategy for estimating causal effects by thinking about back-door paths and how to properly block them. When I first read this chapter (about a year ago), I remember being surprised to read that conditioning on all potential confounders and covariates could actually lead to inadvertently inducing more spurious correlation instead of removing it. I think for many social scientists and statisticians who first learn about causal inference in the potential outcomes framework (like myself), this statement seems impossible, but the examples in this chapter clearly showed when it can occur. As the authors explain, the goal of conditioning is "not to adjust for any particular confounder but rather to remove the portion of the total association ... that is noncausal." I wonder, though, for a case like the example in section 4.2, how similar the estimates of the causal effect would be in practice when using each of the three possible conditioning sets $\{C, O, \{C, O\}\}$. Mathematically, of course, any of the three would give a consistent estimate, but I'm curious as to how each of them would play out with actual data.

My difficulty with causal graphs has always been in constructing them for a given application. For my project on the household- and community-level effects of open defecation on child health, the number of factors known and debated to affect child health are vast. The data I am using measure many variables at the child, mother, and household levels. It's not clear to me how to construct a plausible causal graph from the available data; the possibilities seem endless, and I imagine that any one causal graph could be reasonably criticized by experts in epidemiology, public health, and medicine. On the other hand, just because thinking about the potential relationships between determinants of both child health and household/community-level sanitation is difficult, doesn't mean it shouldn't be done. Now that I think about it, though, I wonder how causal graphs encode variables that are hierarchically related. For example, the religion of a given household is nested in the neighborhood religious composition. Is it important to account for this hierarchy in the causal graph? Or can we simply use arrows to denote that neighborhood religious composition is a function of

household religion?