Applied Causality Reading Assignment 4

Feihan Lu

February 13, 2017

This reading response summarizes Chapter 3 Session 3.1, 3.2, 3.3 of [1] and Chapter 24 of [2]. In my opinion, the most important messages conveyed in these sessions include: (1). In **Markovian models**, the causal effect of intervention can be consistently estimated by the observed data; (2). ***Front-door criteria*** can be viewed as an extension of *Back-door criteria*: it helps to consistently estimate the causal effect under the situation where there is other variables intercepting the treatment and outcome; (3). When using back-door criteria, one can use **matched pairs** defined by the back-door variable set (or a function of it, such as **propensity scores**), along with the technique of **Monte Carlo approximation**, to consistently estimate the effects.

1 Effects of interventions in Markovian models: something we can consistently estimate

Just like what we learned in the previous lectures and readings, the major concern of estimating the causal effect of a variable X on Y is the gap between what we **want**, i.e., $Pr(Y|do(X=x))$ versus what we can **get** from the observed data,i.e., $Pr(Y|X=x)$. Chapter 3 Session 3.1, 3.2, 3.3 of [1] shows that these two distributions are the same if X is the intervention in the model and the underlying model is Markovian. It can be proved either by pruning the arrows pointing to X and plugging different values of X elsewhere in the model, or by adding arrows pointing to X and working with the augmented graph. Such model seems to be an extension of graph so far: it extends to a dynamic graph along a time dimension. But the point is, as long as the graph is Markovian, we seem to be in hope. What if the model is not Markovian?

2 Front-door criteria: an extension of Back-door criteria

The Back-door criteria tries to find the set of variables S such that X and Y are conditionally ignorable given S. The limitation is that X needs to directly point to Y without any variables in between. Front-door criteria, whose steps are based on back-door criteria, is an extended toolkit to handle the situation where there are some variables Z lying in the path from X to Y. The idea is to stratify Z and estimate the causation of X on Y given each stratum of Z by decomposing the total effect of X on Y into the effect of X on Z and the effect of Z on Y, where one uses the back-door criteria to find the last two effects consistently. The idea is really cool, and I think both back-door criteria and front-door criteria can be used as building blocks for finding conditional ignorance in complex graphs, after breaking the big graph down into small chunks.

3 Matching: a convenient follow-up of Back-door criteria

Now that one has found the back-door variable set S, the next thing is how to use it. One could of course take approaches such as regressing Y on X and S. However, this requires assuming the functional form of the regression model. Matching the data by S, on the other hand, is a non-parametric technique that makes use of the back-door variable set S. Although both methods provide consistent estimates, matching seems to outperform the regression counterpart in many situations due to its non-parametric nature. I wonder when the regression model beats the matching, though.

4 Monte Carlo approximation and Propensity Scores: two helpers for Matching

Two appealing techniques that help one during estimating the causal effects by matching are (1). Monte Carlo approximation and (2). propensity scores. The first method, supported by Law of Large Numbers, makes computation at ease, and the second method drastically reduces the dimensionality of the set S to further make the estimating convenient. Both of them, however, are not for free: Monte Carlo approximation requires large sample, and propensity scores require estimating the functional form of $f(S) = Pr(X = 1|S = s)$. While the first bottleneck is somewhat less terrifying, the second one could still be a big challenge if the S is very high-dimensional. I wonder if the propensity scores are not accurate, how much it will affect the estimation of the final causal effect of interest.

References

[1] Pearl *Causality*. Cambridge (2009).

[2] Shalizi *Advanced Data Analysis from an Elementary Point of View*. In preparation (2017).