Review of *Causal inference in statistics: An overview*

Grace Yi Chen

This article gives an overview of some general assumptions and theories for causal inference. Causal relationship is different from association because its goal is to infer the dynamics of treatment effects under changing interventions. Simply adjusting confounders in statistical models could not obtain unbiased estimates of causal effects without causal assumption in observational studies. Pearl unifies several causal theories and notations by introducing Structural Causal Model (SCM), which provides a coherent mathematical basis as well as precise language for communicating assumptions in causal inference. SCM was first introduced by Wright who augmented the causal linear equation with a causal diagram. Later, Structural equation modeling (SEM) was used as a basis for nonparametric causal models and a mathematical operator called do(*) was used for actual implementation. A directed acyclic graphical procedure was included and used to identify causal effect assuming the causal models are Markovian. Finally, the author compares SCM with PO framework and combine the strong features of both.

I think Pearl gives a great review of the theories and methods in causal inference and introduces the general theory and analysis procedures based on SCM. Some of the common causal notations in the statistical literature include potential-outcome (PO) notation, Pearl's do(*) operators and Graphical models. I previous learnt causal inference based on PO framework so the notation here using SCM is unfamiliar to me. But there are a lot in common. Both of these two frameworks require us to think about assumptions about the causal relationship and model first before doing analysis. Causal assumptions usually could only be verified using a randomized experimental study. Thus, it is very important that the notation for causal assumptions is meaningful and unambiguous. SCM pays a special attention to using graphical models, as Pearl thinks it is more straightforward to show the causal relationship and assumptions. Once we make causal assumptions through graphics, we could estimate the intervention effect accordingly. In addition, the assumptions in SCM seems to be stronger than PO framework as it also requires Bayesian subjective assumptions that the models are Markovian with conditional independence. One thing that I found very interesting is that SCM derives a range of possible values for the quantities of interest when conditions for identification are not met, while PO framework assumes there is no unmeasured confounders.

Questions:

1. Rubin stated that a model should control for all pre-treatment variables. However, Pearl mentioned that this may increase or decrease confounding bias. I am wondering how we could decide whether to include a variable in causal models or not.
2. What else are the differences between PO and SCM framework in causal inference other than the points I mentioned above?