Review of *The central role of the propensity score in observational studies for causal effects*

Grace Yi Chen

This article introduces the definition, properties and application of propensity score when doing causal inference in observational studies. The treatment assignment in observational studies is not random which makes it hard to naively assess the treatment effect between treatment and control groups. Propensity score is the conditional probability of assignment to the treatment group given confounding variables. It plays an important role when doing causal inference in observational studies. There are several assumptions. For example, Stable Unit Treatment Variance Assumption (SUTVA) assumes that treatment assignment for one unit should not affect the outcome for another. The authors first introduced balancing score b(x), which is a function of the observed confounding covariates and the conditional distribution of x given b(x) is the same for treated and controls. The balancing score leads to the additional assumption of strong ignorability that the treatment assignment is strongly ignorable given the balancing score. Propensity score is a balancing score with the coarsest function of x so the strong ignorability also hold for propensity score. The authors later proved and demonstrated that the unbiased estimation of average treatment effect is the difference between treatment and control means at each value of a propensity score. Consequently, we could apply propensity score in pair matching, subclassification and covariance adjustment to produce unbiased estimates of the average treatment effect.

I think Rosenbaum and Rubin gave a good overview and discussion about the properties and applications of propensity score. Randomized trials can give unbiased estimation of treatment effect but are usually expensive and not always possible. Also, experiments often lack external validity while we need to generalize experimental results to a target population. We should make use of the observational data to inspire new ideas and hypothesis. Propensity score is very important to draw causal conclusion by adjusting for confounding factors. Also, propensity score could serve as a dimension reduction tool when we have a lot of confounding factors to adjust for. As the authors mentioned, propensity scores could be used for matching, subclassification, weighting and covariance adjustment in the regression model. However, it is possible that the propensity score model between treatment and confounders could be mis-specified, especially for non-binary treatments. Propensity score methods can be sensitive to model misspecification, but I am aware that there are some methods trying to solve this problem. For example, doubly-robust estimator allows for model misspecification of the propensity score model or the outcome model.

Questions:

1. A recent paper by Gary King mentioned that propensity score matching may increase imbalance and bias because it intends to approximate an RCT instead of a blocked RCT (King & Nielsen, 2019). When would propensity score matching be the most appropriate?

2. Can we add the confounding variables in the regression model instead of modeling them as a propensity score?