NAME: Maxfiled Green
DATE: 1/15/2019
Reading reflections: Week 5
STAT/CS 387

Title(s) read

- Matching Methods for Causal Inference: A review and a look forward.

Reflection

Matching Methods for Causal Inference looks at how different studies and practices have strategized in choosing treated and control subjects for comparison during replicated experiments. The paper is an overview of scattered methods used in the past in many different fields and disciplines of study.

Causal inference is very desirable as it can lead investigators to discovering how the change in one co-variate directly causes the change in another. In the literature examined the causal effect was defined as the comparison of the individuals outcome if the individual received the treatment, and the outcome if the individual did not receive the treatment, receiving the control. As discussed in class, we can only observe one of these outcomes. Thus, the goal is to predict the unobserved outcome. To implement this prediction, two groups are compared to one another, one receiving the treatment and one the control. A huge obvious requirement in these methods is that the treatment and control group must differ in an explicit random way across all variables. Methods of "matching" have developed in order to replicate randomized experiments to estimate causal effects with observational data. Observational data is inherently non randomized.The paper defines matching as an method that hopes to balance the distributions of variables present in the treatment and control groups. This is crucial in the use of observational data that are commonly used in non-experimental studies. In an experiment, there are straight forward methods that can be followed to ensure that both the treatment and control groups are balanced in co-variate distributions. However, if you are just handed a data set to perform inference on or are gathering data from observations, these notions of good experimental design cannot be implemented.

Originally, the implementation of matching methods was inspired by limiting the necessity of collecting samples from the full control group. Alternatively, matching was used to reduce bias in the estimation of the effect of a treatment given that all outcome data was available. No variables should be included in the matching process that may have affected by the treatment. To quantify how similar individuals are to one another, various "distance measures" are employed. This allows for investigators to exam whether the control and test subjects are similar enough to be compared.For example, the "propensity score" can be used to reduce all variables to a single scalar value, the probability of an individual being treated. Grouping individuals with similar propensity scores is akin to mini randomized experiment design. In order to diagnose the success of a matching scheme, several comprehensive steps are required. Primarily, all matched groups must have covariate balance. As discussed above, this ensures that the treatment is not related to the covariates. The paper suggests that multidimensional histograms of covariate distributions and marginal distributions be examined. Another common diagnostic is the difference of means divided by the standard deviation of the treated group, across treatment and control variables. This if often refered to as the standardized bias. This metric would be compared before and after matching. For continous variables, Q-Q Plots can be used to visually diagnose matching. QQ Plots compare the quantiles of variables in the treatment and control group.

Once a balanced matching has been achieved, further measures to smooth out covariate imbalance can also be taken, such as adjusted regression. Different steps need to be taken for different matching scenarios, some next steps remain controversial among scientists. The paper argues that after K:1 matching, the analysis does not need to account for any matched pairs and that conditioning on the variables that were used in the matching process's is actually sufficient for this problem. Conversely, after sub classification or full matching I thought that it was interesting that hypothesis tests and p-values should not being used as a measure of balance. This however does make a lot of sense as balance does not have a sub and super scale. At the end of the paper, some additional problems that arise are discussed. Something that I have experienced many times is missing co-variate data, making diagnosing balance and matching difficult to impossible.