Applied Causality Reading Assignment 5

Feihan Lu

February 21, 2017

This reading response summarizes Chapter 8 of [?]. The most important messages conveyed include: (1). The causal estimand of interest based on potential outcomes in randomized experiments can be translated into functions based on observed and missing data, so the whole causal inference becomes a *missing data imputation* problem; (2). To model the distribution of the missing random variable, one can use Naive methods that overly simply the distribution, or infer the posterior conditional distribution of the missing values given the observed data in a Bayesian framework.

1 Naive missing data imputation of missing potential outcomes

The goal is to impute the missing potential outcomes or estimate the underlying distribution. The naive methods introduced in the first several paragraphs of this chapter rely on overly simplified assumptions of the distribution of the missing potential outcomes. For the first method of mean imputation, the key assumption is

$$D(Y_i^{mis}(w)) = \bar{Y}^{obs}(w)$$

where $w = 0, 1$, and for the second method, the assupmtion is

$$D(Y_i^{mis}(w)) = D(Y_i^{obs}(w)) = U(Y_i^{obs}(w))$$

where U is uniform distribution. When they are very easy to calculate, they are too simple that ignore uncertainty among the missing random variable (the equality in the first assumption) or the uncertainty about the form of the distribution (the last equality in the second assumption). However, in many practice, I think people do use these methods. The way they do such kind of imputation is to iteratively impute the missing values given the observed values and the values imputed in last the iteration. I wonder how this differentiates the Bayesian method.

2 Bayesian missing data imputation for missing potential outcomes

The key thing in the Bayesian framework is

$$P(Y^{mis}|Y^{obs}, W) \propto \int_\theta P(Y^{mis}|Y^{obs}, W, \theta) P(Y^{obs}, W|\theta) P(\theta) d\theta$$

where for each subject i,

$$\begin{bmatrix} Y_i^{mis} \\ Y_i^{obs} \end{bmatrix} = \begin{bmatrix} W_i & 1 - W_i \\ 1 - W_i & W_i \end{bmatrix} \times \begin{bmatrix} Y_i(0) \\ Y_i(1) \end{bmatrix}$$

One can assume the likelihood $P(Y_i(0), Y_i(1), W|\theta)$ and prior $P(\theta)$. Once you put everything in Bayesian framework, it becomes a lot easier. One thing to notice is that, the model still requires the assumption of missing completely at random, does it?

References

[1] Imbens, Rubin *Causal Inference for Statistics, Social, and Biomedical Sciences.* Cambridge (2012).