Introduction to Causal Inference

Solutions to Quiz 3

Teppei Yamamoto

Monday, July 18 2016

Instructions:

- Write your name in the space provided below before you begin.
- You have **20 minutes** to complete the quiz.
- You may answer questions either in English or Japanese.
- The exam is closed book and calculators are not allowed.
- Please turn off your phone before you begin.
- Note that I have allocated more space than we anticipate you will need for each problem. Just answer the questions as best you can, don't try to fill the available space.
- Good luck!

Your Name: _____

Problem 1

Are the following statements true or false? Why? For each question, state whether you think the statement is true or false, and explain why you think so in a few sentences.

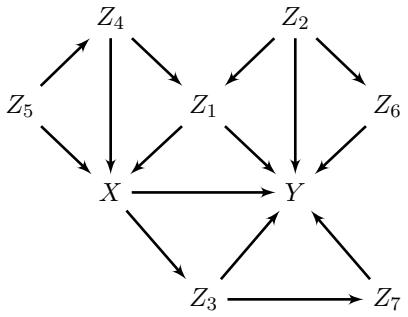
a) Under selection on observables with discrete X_i, the ATT can be unbiasedly estimated using subclassification as follows: 1) defining strata by the values of X_i; 2) within strata, taking the difference in the means of the observed outcome between treated and untreated units; 3) averaging over the within-stratum differences, weighting by the marginal distribution of X_i.

False. Because we are estimating the ATT, we should weight by the probability of X_i *conditional* on $D_i = 1$

b) When matching on continuous X_i in observational studies, finding more than one match for each treated unit improves precision by averaging over multiple control units, but may incur bias in the estimate.

True. Retaining more observations in the matched sample improves precision, but for a continuous variable, using more matches potentially means including less similar units as matches, risking bias.

c) Considering the DAG below, $X \to Z_3 \to Z_7 \to Y$ is a back door path from X to Y.



False. $X \to Z_3 \to Z_7 \to Y$ is not a back door path from X to Y because it does not begin with a directed edge that points to the first variable.

Problem 2

Imbens' approach to sensitivity analysis uses two sensitivity parameters to quantify the degree of unconfoundedness in an observational study. What do those two parameters represent? How are they used to analyze sensitivity? Explain.

Imbens' approach uses two sensitivity parameters, γ and δ. δ represents the difference in average unobserved confounder between treatment conditions and γ represents the effect of the unobserved confounder on our outcome variable. We set those parameters to various values and calculate the implied causal effects and investigate when the original conclusion based on conditional ignorability would substantively change. If the change would occur only with severe degrees of confounding, the study is judged to be insensitive (i.e. robust) to the violation of the assumption.

Problem 3

Consider an observational study with 8 observations, of which four were treated. We use $D_i \in \{0, 1\}$ to denote the treatment (1 for treatment and 0 for control) and Y_i the observed outcome for unit i. We use X_i to denote an observed covariate.

1. The observed data is presented in the Table below. By using one-to-one, nearest-neighbor matching (with replacement) on the covariate X_i, fill in the missing counterfactuals and calculate the missing τ_i *for the treated*.

i	D_i	Y_i	Y_{1i}	Y_{0i}	X_i	τ_i
1	1	8	8	6	6	2
2	1	4	4	2	3	2
3	1	6	6	8	12	-2
4	1	12	12	2	1	10
5	0	3		3	19	
6	0	6		6	6	
7	0	8		8	12	
8	0	2		2	2	

2. Using your table, estimate the ATT by one-to-one, nearest-neighbor matching (with replacement) on the covariate X.

$ATT = (2 + 2 - 2 + 10)/4 = 3$