

Assignment 1

CS6480: Causal Inference and Learning
IIT-Hyderabad
Feb-Apr 2021

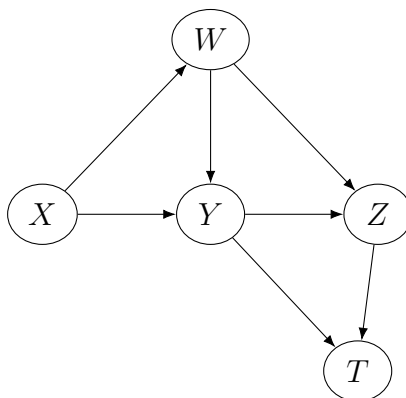
Max Marks: 40
Due: 21st Mar 2021 11:59 pm

Instructions

- Please use Piazza to upload your submission by the deadline mentioned above. Your submission should comprise of a single file, named `<Your_Roll_No>_Assign1`, with all your solutions.
- For late submissions, 10% is deducted for each day (including weekend) late after an assignment is due. Note that each student begins the course with 5 grace days for late submission of assignments, of which upto 3 grace days can be used for a single assignment. Late submissions will automatically use your grace days balance, if you have any left. You can see your balance on the Marks and Grace Days document, soon to be shared under the course Google drive.
- Please read the [department plagiarism policy](#). Do not engage in any form of cheating - strict penalties will be imposed for both givers and takers. Please talk to instructor or TA if you have concerns.

Questions

1. **(15 points)** Answer the following questions using the causal graph shown below. First 7 questions carry 1 mark each. Last 4 questions carry 2 marks each.



- (a) Name all the parents of Z
- (b) Name all the ancestors of Z

- (c) Name all the children of W
- (d) Name all the descendants of W
- (e) Identify all the paths(not necessarily direct) between X and T
- (f) Identify all directed paths from X to T
- (g) Are W and T marginally independent? Explain.
- (h) Use the product decomposition of probability distributions to identify the joint probability distribution of the five variables represented as nodes (T, W, X, Y, Z) .
- (i) Are W and T conditionally independent given Z ? If yes, prove that $p(W, T|Z)=p(W|Z)p(T|Z)$. If not, explain why.
- (j) Are W and T conditionally independent given Z, Y ? If yes, prove that:

$$p(W, T|Z, Y) = p(W|Z, Y)p(T|Z, Y)$$

If not, explain why.

- (k) Find a set of variables that d-separates X and T . Explain your reasoning. What does this tell us about the conditional independence of X and T in the data (i.e., when are they conditionally independent)?
2. **(6 points)** The table below gives data for 8 fast food restaurants that were part of a study of the effect of raising the minimum wage in New Jersey (NJ). The “treatment” group is the 2 restaurants in NJ and the “control” group is a set of 6 restaurants in Pennsylvania (PA) (where the minimum wage was not raised). The outcome Y^{obs} is the number of people employed (including part-time employees) at the end of the year. There are two covariates – X_{i1} , the identity of the fast food chain (Burger King or Kentucky Fried Chicken) and X_{i2} , duration of employment at the end of the year prior to the increase in the minimum wage.

Observation i	Treatment W_i	Restaurant Chain X_{i1}	Initial Employment X_{i2}	Final Employment Y_i^{obs}
1	NJ	BK	22.5	30.0
2	NJ	KFC	14.0	12.5
3	PA	KFC	13.8	17
4	PA	BK	26.5	18.5
5	PA	BK	20.0	19.5
6	PA	BK	13.5	21
7	PA	BK	32.5	26.5
8	PA	KFC	21.0	23.0

Now, answer the following questions:

- (a) We want to use matching to estimate the effect of raising the minimum wage assuming that unconfoundedness holds. We will match a single control unit with each treatment unit (without replacement). Our distance measure is $D(i, j) = 100 \times \mathbb{I}(X_{i1} \neq X_{j1}) + |X_{i2} - X_{j2}|$ where the indicator \mathbb{I} is 1 if the two units are different chains and 0 if they are the same chain. Identify the matches for the 2 treatment units. **(2 points)**

- (b) Define Average Treatment Effect on the Treated (ATT) (*Hint*: This is ATE restricted to the treated!). Under which causal inference assumption is ATT equal to Average Treatment Effect (ATE)¹? Explain (2 points)
- (c) Find ATT from the above table. (2 points)
3. (12 points) In an observational study, when we observe all the relevant features/attributes, we can use adjustment formula to identify the causal effect (e.g., using regression analysis as we saw in class). In this question, we assume that we have all the information required for identifying the causal effect. Now, consider the following equations that generate the data. All noise random variables are normally distributed and are independent of each other. T is the treatment variable and Y is the outcome/target variable.

$$W = \mathcal{N}_w(65, 5)$$

$$T = \frac{W}{18} + \mathcal{N}_t(0, 1)$$

$$Y = 1.05T + 2W + \mathcal{N}_y(0, 1)$$

$$Z = 0.4T + 0.3Y + \mathcal{N}_z(0, 1)$$

- Draw the causal graph looking at the equations. Show noise variables in the graph explicitly in the causal graph. Represent noise variables as nodes and draw an edge connecting noise variables and the features(W, T, Y, Z) (1 point).
 - Generate 10,000 data points using these set of equations. Make the continuous treatment variable T binary by re-assigning its value to 1 if its original value is greater than 3.5 else to 0 (2 points).
 - Write code to implement linear regression which returns regression coefficients (2 points).
 - Using the regressor implemented above, write code to evaluate $\mathbb{E}(Z|do(T = t)) - \mathbb{E}(Z|do(T = t'))$ (1 point)
 - Now, use your code to evaluate the following quantities. Report the numbers and a discussion on your findings (6 points).
 - $\mathbb{E}(Y|T = 1) - \mathbb{E}(Y|T = 0)$
 - $\mathbb{E}(Y|do(T = 1)) - \mathbb{E}(Y|do(T = 0))$ with $\{W\}$ as adjustment set(i.e., regress Y on T, W with $do(T = t)$)
 - $\mathbb{E}(Y|do(T = 1)) - \mathbb{E}(Y|do(T = 0))$ with $\{W, Z\}$ as adjustment set(i.e., regress Y on T, W, Z with $do(T = t)$)
4. (5 points) Answer the questions below:
- (a) What is the mean independence assumption?(1 point)
- (b) How it is useful in claiming that the regression coefficient of x in $y = \beta_0 + \beta_1 x + u$, β_1 is the same as the causal parameter? u in the equation above represents an error term that is usually assumed to include unobserved variables(2 points).
- (c) Explain your reasoning by relating with Conditional Average Treatment Effect (CATE) and Exchangeability (2 points).
5. (2 points) If we assume that the distribution of covariates X is same across treatment groups T , prove that $p(Y = y|do(T = t)) = p(Y = y|T = t)$.

¹We use Average Treatment Effect (ATE) and Average Causal Effect (ACE) synonymously