CMSI 498 – Assignment 3

[30 Points] due 28th

Instructions:
This worksheet gives you some important practice with the fundamentals of counterfactual inference, including practice with counterfactuals in both fully- and partially-specified Structural Causal Models (SCMs). Specific notes:

- Provide answers to each of the following questions and write your responses in the blanks. If you are expected to show your work in arriving at a particular solution, space will be provided for you.
- You may either turn this in under my office door before the due date or submit it on Brightspace under the Assignment 3 folder.
- Place the names of your group members below:

Group Members:

1. Jackson Watkins
2. John Scott
3. Jimmy Byrne
4. J Croucher
5. _____

25

2.2
2.3

Problem 1 – Counterfactuals in Fully-Specified, Linear SCMs

In econometrics, empirical sciences, epidemiology, and a host of other disciplines, linear regression models are popular choices for interpretable analyses of continuous data. In particular, variables *regressed upon* can be fit to some linear equation like:

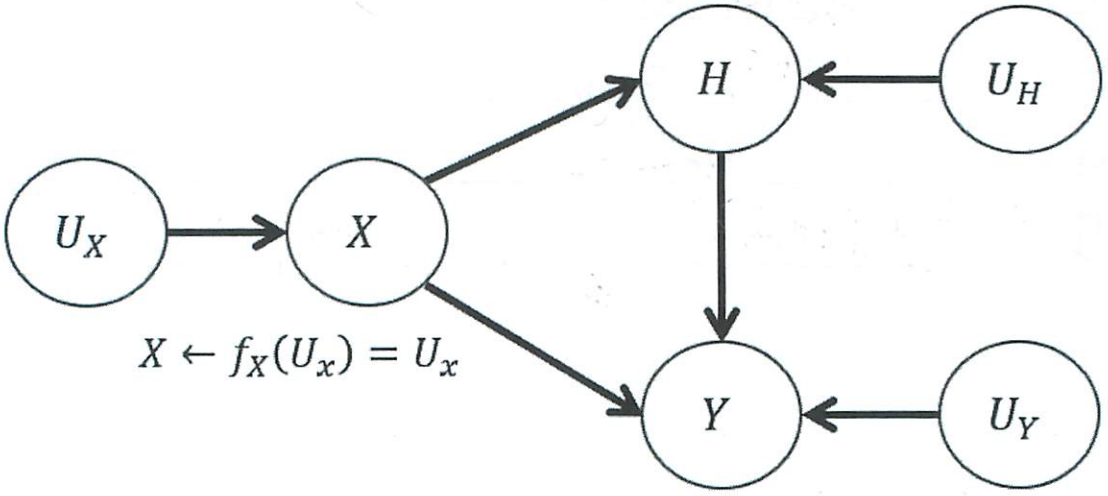
$$Y \leftarrow f_Y(X) = a * X + \epsilon$$

Above, we consider a to be some "weight" attached to X's influence on Y, with some error term ϵ to account for noise in the system. In so using linear equations in SCMs, we have a similar interpretation, except that the error terms are instead modeled as exogenous variables U_i that can account for individual differences, despite the weights being measured from population data. Let's look at a motivational example, yay!

After-School Special – Causality Edition: In an effort to improve exam scores in underprivileged communities, the government implements an after-school mentorship program in which students receive encouragement and support to help on homework and studying. To study this program's effectiveness, we create a linear SCM composed variables X (weekly hours of mentor help), H (weekly hours spent on homework), and Y (exam scores). Because this program is implemented nationally, we have population data on its effectiveness, and can standardize the units of all variables to have a mean of 0 and standard-deviation (SD) of 1, and then simply express their units in terms of SD above / below the mean. For example, a student for which $Y = 1$ is 1 standard-deviation above the average test score.

The SCM for the results of a nationwide study on these mentorship programs is given below:

$$H \leftarrow f_Y(X, U_H) = 0.5X + U_H$$



$$X \leftarrow f_X(U_x) = U_x$$

$$Y \leftarrow f_Y(X, H, U_Y) = 0.7X + 0.4H + U_Y$$

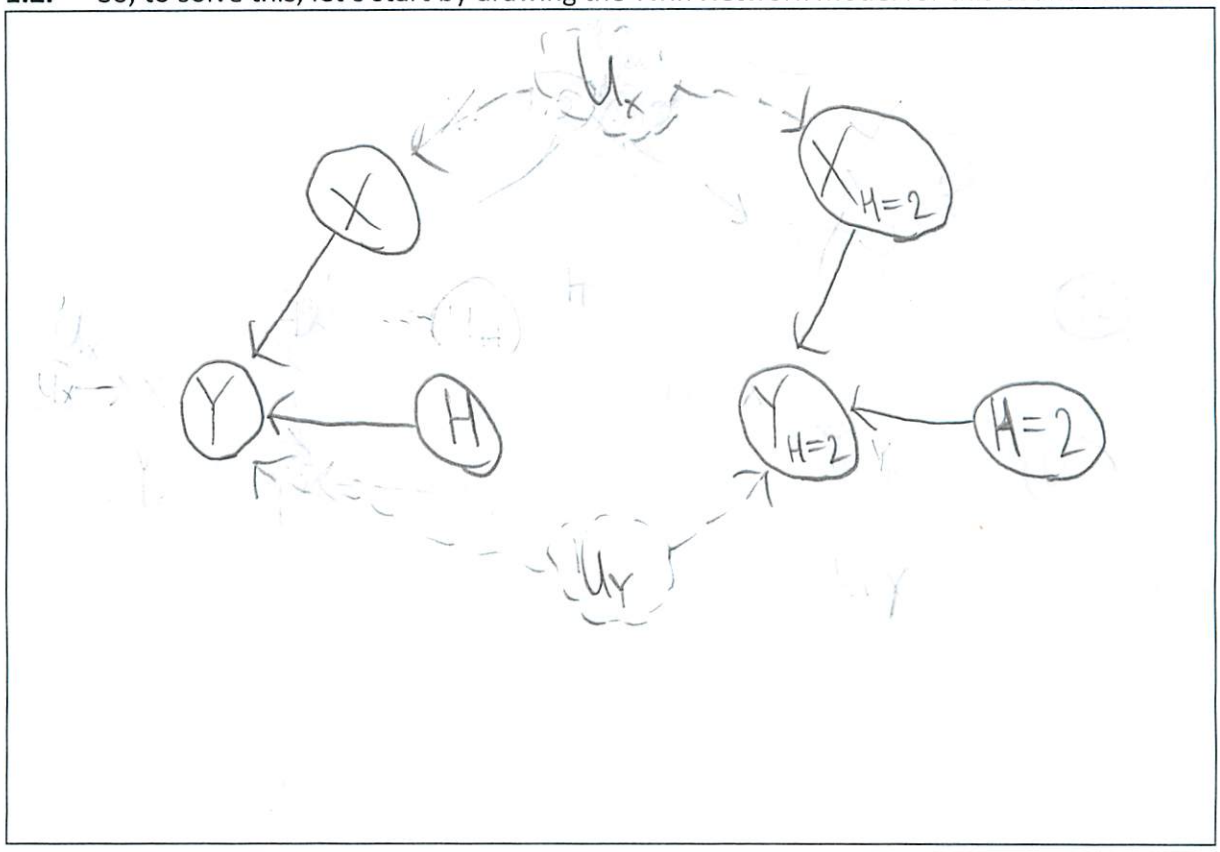
1.1. On average, how many standard deviations above the mean would a student's test scores, Y, be if they received 1 SD extra encouragement, i.e. $X = 1$? Recall: the average / expected value of all U_i is 0. If you know how to express your answer using expected value, go for it – otherwise, I won't be picky.

$$Y = 0.7X + 0.4H + U_Y$$
$$= 0.7 \cdot 1_X + 0.4(0.5(1) + U_H) + U_Y$$
$$= 0.7 + 0.4 \cdot 0.5 = 0,9$$

Suppose now we are not interested in population averages, but instead, want to know the effects of different interventions on a particular individual. To do so, we'll be employing a counterfactual query on this system, so buckle up. Our query: "For a student who received 0.5 extra SDs of encouragement ($X = 0.5$) and who performs 1 SD of homework per week above the mean ($H = 1$), and traditionally scores 1.5 SD higher than the mean on exams ($Y = 1.5$), how many SDs above the mean would she score on her exams *had she spent* 2 SDs above the mean on homework ($H = 2$)?"

In a linear SCM, this question amounts to solving for $Y_{H=2}|X = 0.5, H = 1, Y = 1.5$.

1.2. So, to solve this, let's start by drawing the Twin Network Model for this counterfactual:



1.3. With the Twin Network in hand, let's now walk through the traditional 3-step process for counterfactual computation, starting as we always do, with Abduction. In linear SCMs, this means we simply solve for the state of all exogenous terms given our evidence. So, solve for U_X, U_H, U_Y given $X = 0.5, H = 1, Y = 1.5$.

$H = 0.5X + U_H \qquad X = U_X \qquad Y = 0.7X + 0.4H + U_Y$

$U_H = H - 0.5X \qquad U_X = X \qquad U_Y = Y - 0.7X + 0.4H$

$= 1 - 0.5(0.5) \qquad = 0.5 \qquad = 1.5 - 0.7(0.5) - 0.4(1)$

$U_H = 0.75 \qquad U_X = 0.5 \qquad U_Y = 0.75$

1.4. Having solved for the state of each of these exogenous background terms for this particular student, now, let's re-run history by completing the Action and Prediction steps. Using the network in 1.2 and your Abduction from 1.3, solve for $Y_{H=2}$ in the hypothetical world of your Twin Network.

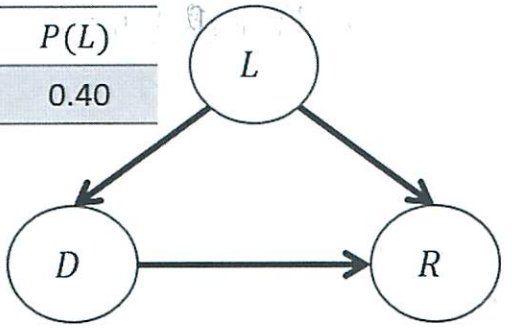
$Y = 0.7X + 0.4H + U_Y$

$= 0.7(0.5) + 0.4(2) + 0.75$

$\boxed{Y_{H=2} = 1.9}$

Problem 2 – Counterfactuals in CBNs

You are working for the FDA and have been tasked with assessing the validity of a new cold remedy that some people are concerned is medically equivalent to snake oil. You have scant information about this new remedy (Forney Industries' Flu the Coop™ tablets), but have some observational data on user characteristics when they were suffering from the flu: D = whether or not someone took the remedy, L = some binary coding of lifestyle characteristics that affect both their propensity to take (read: get scammed by) the remedy and recover from the flu, and R = whether or not they recovered from the flu shortly after taking the remedy. The following Causal Bayesian Network depicts this example:



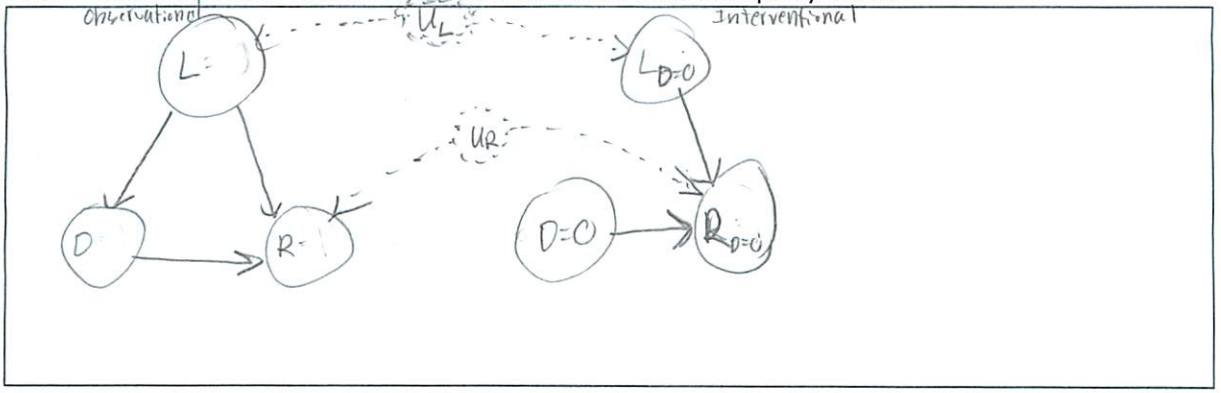
L	$P(L)$
0	0.40

L	D	$P(D \mid L)$
0	0	0.20
1	0	0.70

L	D	R	$P(R \mid L, D)$
0	0	0	0.30
0	1	0	0.10
1	0	0	0.50
1	1	0	0.80

To test the drug's validity, we wish to compute the *Probability of Necessity,* a counterfactual query that asks, "Would users who recovered using the drug have recovered *had they not taken it?*" Formally, our query is: $P(R_{D=0} | D = 1, R = 1)$

2.1. Draw the Twin Network model associated with this query below:



2.2. From the Twin Network you found, and the evidence given, perform the Abduction step of the counterfactual computation, indicating which CPTs need be updated and how.

$P(L) \rightarrow P(L \mid D=1, R=1)$

compute:

$P(L=0 \mid D=1, R=1) = \dfrac{P(L=0, D=1, R=1)}{P(D=1, R=1)}$

1. Labels: $Q = \{L\}$ $e = \{D=1, R=1\}$ $h = \{\}$

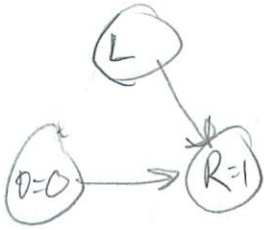
2. $P(L=0, D=1, R=1) \overset{\text{M.F.}}{=} P(L=0) \cdot P(D=1 \mid L=1) \cdot P(R=1 \mid L=0, D=1)$
$= 0.4 \cdot 0.3 \cdot 0.9$
$= 0.108$

3. $P(D=1, R=1) = \sum_l P(L=l, D=1, R=1) = P(L=0, D=1, R=1) + P(L=1, D=1, R=1)$
$.108 + P(L=1) \cdot P(D=1 \mid L=1) \cdot P(R=1 \mid L=1, D=1)$
$.108 + 0.6 \cdot 0.3 \cdot 0.2$
$= .144$

4. $P(L=0 \mid D=1, R=1) = \dfrac{.108}{.144} = 0.75$
$P(L=1 \mid D=1, R=1) = 1 - 0.75 = 0.25$

2.3. Finally, perform Action and Prediction to compute $P(R_{D=0} \mid D=1, R=1)$.

Action



$D=0$

Prediction

$P(R_{D=0}=1) = \sum_l P_{D=0}(R=1 \mid D=1, L=l) \cdot P(L=l \mid D=1, R=1)$

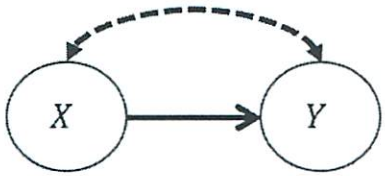
$\sum_l P_{D=0}(R=1 \mid D=1, L=l) \cdot P(L=l \mid D=1, R=1)$

$P_{D=0}(R=1 \mid D=1, L=0) \cdot P(L=0 \mid D=1, R=1) + P_{D=0}(R=1 \mid D=1, L=1) \cdot P(L=1 \mid D=1, R=1)$
$= 0.9 \cdot 0.75 + 0.20 \cdot 0.25$
$= 0.725$

Problem 3 – Empirical Counterfactuals

You're joining a new web advertising agency as a replacement for Dave. Everyone loves you by default because everyone hated Dave. He was the type to add gratuitous memes to his meeting PowerPoints, and wanted to constantly discuss the latest episode of Homeland. Worse yet, he was the one nursing a legacy advertising agent (written entirely in C) that no one else knows how to operate, what user features it employs, nor if it is particularly effective at translating ads to clickthroughs. Your first job is to recreate a more interpretable ad agent, but to make sure that the company isn't hemorrhaging money from bad advertising from Dave's lame agent.

[For simplicity, rather than realism] Suppose Dave's agent currently works by mapping some arcane user demographics to one of two ad portfolios $X \in \{0,1\}$ that will display targeted ads to users, after which it is recorded whether or not the user clicked on the ad $Y \in \{0,1\}$. Sadly, the system is affected by confounding, since you do not know how or what demographics are considered by the agent, nor how they affect the clickthrough rate. However, as a quick means of determining how good Dave's system is, you seed a small propensity for the agent to randomly choose an ad portfolio to show to the user, and collect the following data:



X	$P(X)$
1	0.40

X	Y	$P(Y \mid X)$
0	1	0.20
1	1	0.10

$do(X)$	Y	$P(Y_X)$
0	1	0.30
1	1	0.20

3.1. Suppose you now create a second agent that takes the intended output of Dave's agent as input to see if it needs to correct for its choice for X. Determine, for both $X = 0,1$ if your agent is better off *swapping from* or *agreeing with* Dave's agent's X in order to maximize clickthroughs (i.e., maximize likelihood of $Y = 1$).

$$ETT = P(Y_{x=1} = 1 \mid X=0) - P(Y=1 \mid X=0)$$

$$= \frac{P(Y_{x=1}=1) - P(Y=1 \mid X=1)P(X=1)}{P(X=0)} - P(Y=1 \mid X=0)$$

$$= \frac{0.2 - (0.1)(0.4)}{0.6} - 0.2 = \frac{0.16}{0.6} - 0.2 = \boxed{0.0\overline{6}}$$

$$P(Y_{x=0}=1 \mid X=1) - P(Y=1 \mid X=1) = \frac{0.3 - (0.2)(0.6)}{0.4} - 0.1 = \frac{0.18}{0.4} - 0.1 = \boxed{0.35}$$

Swap in both cases (X=0 and X=1)