Problem Set 1

6.S091: Causality

IAP 2024

Due: Thursday, January 18th at 3pm EST

- Problem sets **must** be done in LaTeX.
- Problem sets are to be submitted at the following link: `https://bit.ly/causality-iap24-pset1`
- You may use any programming language for your solutions. You **are not required** to turn in your code.

New grading policy:

- There will be **100** possible points (instead of 30).
- **60** points will be required to pass (instead of 18).
- This problem set will be worth **30** points (instead of 10).

Problem 1: Interventions and Adjustment [12 points]

Background

Mickey and Minnie have come up with a new business venture. Their fellow mice love receiving cheese, and humans often award cheese to mice who can quickly run a maze, so they will begin a training program aimed at reducing the time taken to run a maze. Naturally, Mickey and Minnie care about truth in advertising. Before beginning their venture, they want to estimate the efficacy of their training program:

What is the effect of receiving training ($\texttt{do}(A = 1)$) on the probability that a mouse is rewarded cheese ($Y = 1$)?

The deadline for VC funding is soon, and they don't have the time to run a randomized trial. However, they do already have some past experience that might be useful. For the past few years, Mickey and Minnie have been offering this training to their friends, free of charge. Having taken 6.S091 at the Mouse Institute of Technology, they know they need to be careful when interpreting this past experience. In particular, Mickey and Minnie suspect that gender might be a possible confounding factor. Their male mouse friends are more likely to sign up for the training program, and they cannot rule out sexism on the part of the humans in terms of how much cheese is rewarded.

Mickey and Minnie's SCM

To help answer their questions, Mickey and Minnie have thought long and hard to develop a structural causal model $\mathcal{M}$ over the following variables:

- $U = 1$ for male, $U = 0$ for female
- $A = 1$ if training is received, $A = 0$ otherwise
- $M = 1$ if the maze is solved "quickly" (in less than one minute), $M = 0$ otherwise
- $Y = 1$ if cheese is rewarded, $Y = 0$ otherwise

The structural causal model $\mathcal{M}$ is as follows:

$$\begin{aligned} U &= \varepsilon_u & \varepsilon_u &\sim \mathsf{Ber}(0.5) \\ A &= \mathbb{1}_{\{\varepsilon_a + U/4 \geq 1/2\}} & \varepsilon_a &\sim \mathsf{Unif}(0,1) \\ M &= \mathbb{1}_{\{\varepsilon_m + 10(1-A) \leq 60\}} & \varepsilon_m &\sim \mathsf{Unif}(0,100) \\ Y &= \mathbb{1}_{\{\varepsilon_y + M/2 + U/4 \geq 1\}} & \varepsilon_y &\sim \mathsf{Unif}(0,1) \end{aligned} \tag{1}$$

Let $\mathbb{P}_{\mathcal{X}}$ denote the distribution over $\mathcal{X}$ entailed by $\mathcal{M}$.

Preliminaries [4 points]

(a) [1 point] Draw the causal graph over U, A, M and Y.

Note: You may use the tikz package in LaTeX to create the figure, or use your favorite drawing software and input the figure as an image.

(b) [1 point] Derive the following conditional distributions, as they are implied by the structural causal model $\mathcal{M}$.

(i) $\mathbb{P}_{\mathcal{X}}(U)$

(ii) $\mathbb{P}_{\mathcal{X}}(A \mid U)$

(iii) $\mathbb{P}_{\mathcal{X}}(M \mid A)$

(iv) $\mathbb{P}_{\mathcal{X}}(Y \mid M, U)$

Note: You may write the distribution of a binary variable that is 1 with probability p as $\mathsf{Ber}(p)$.

(c) [2 points] Let $\mathbb{P}_{\mathcal{X}}$ be an arbitrary distribution over binary-valued variables $\mathcal{X}$. Write a (`Python/Julia/other`) method, called `compute_conditional`, which takes the following inputs:

- An arbitrary distribution $\mathbb{P}_{\mathcal{X}}$ over binary-valued variables $\mathcal{S}$,
- two disjoint sets $\mathbf{A}, \mathbf{B} \subseteq \mathcal{X}$, with $\mathbf{B}$ possibly empty, and
- two values $\mathbf{a} \in \{0,1\}^{|\mathbf{A}|}$, $\mathbf{b} \in \{0,1\}^{|\mathbf{B}|}$,

and returns the value $\mathbb{P}(\mathbf{A} = \mathbf{a} \mid \mathbf{B} = \mathbf{b})$. Use `compute_conditional` to compute the following probabilities:

(i) $\mathbb{P}_{\mathcal{X}}(Y = 1)$

(ii) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid M = 0, A = 0)$

(iii) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid M = 0, A = 1)$

Interventional [3 points]

Let I_a be the intervention with target $T(I) = \{A\}$, which changes the causal mechanism of A into $A = a$. In other words, I_a is the intervention $\mathtt{do}(A = a)$.

(d) [1 point] Let $\mathcal{M}^{I_a}$ be the structural causal model formed by applying the intervention I_a on the SCM $\mathcal{M}$. Write down the interventional SCM $\mathcal{M}^{I_a}$, using the same format as in Equation (1).

(e) [1 point] Let $\mathbb{P}_{\mathcal{X}}(\mathcal{X} \mid \mathtt{do}(A = a))$ denote the distribution over $\mathcal{X}$ entailed by $\mathcal{M}^{I_a}$. Derive the following conditional distributions:

(i) $\mathbb{P}_{\mathcal{X}}(U \mid \mathtt{do}(A = a))$

(ii) $\mathbb{P}_{\mathcal{X}}(A \mid U, \mathtt{do}(A = a))$

(iii) $\mathbb{P}_{\mathcal{X}}(M \mid A, \mathtt{do}(A = a))$

(iv) $\mathbb{P}_{\mathcal{X}}(Y \mid M, U, \mathtt{do}(A = a))$

(f) [1 point] Use `compute_conditional`, with $\mathbb{P}_{\mathcal{X}}(\mathcal{X} \mid \mathtt{do}(A = 1))$ and $\mathbb{P}_{\mathcal{X}}(\mathcal{X} \mid \mathtt{do}(A = 0))$ as input, to report the following probabilities:

(i) $\mathbb{P}_{\mathcal{X}}(Y \mid \mathtt{do}(A = 0))$

(ii) $\mathbb{P}_{\mathcal{X}}(Y \mid \mathtt{do}(A = 1))$

Backdoor Adjustment [2 points]

(g) [2 points] Since U is an adjustment set for $\mathbb{P}_{\mathcal{X}}(Y \mid \mathtt{do}(A = a))$, we have

$$\mathbb{P}_{\mathcal{X}}(Y \mid \mathtt{do}(A = a)) = \sum_{u \in \{0,1\}} \mathbb{P}_{\mathcal{X}}(Y \mid A = a, U = u) \cdot \mathbb{P}_{\mathcal{X}}(U = u) \tag{2}$$

Use `compute_conditional` to compute the following probabilities:

(i) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid A = 0, U = 0)$

(ii) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid A = 0, U = 1)$

(iii) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid A = 1, U = 0)$

(iv) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid A = 1, U = 1)$

Plugging these values into the right-hand side of Equation (2), confirm numerically that the result matches your answers from **(f)**. Show your work.

Frontdoor Adjustment [3 points]

Suppose that Mickey and Minnie only have access to the joint distribution over A, M, and Y, i.e., the distribution

$$\mathbb{P}_{\mathcal{X}}(A, M, Y) = \sum_{u \in \{0,1\}} \mathbb{P}_{\mathcal{X}}(U, A, M, Y).$$

This will occur, for example, if they have only recorded samples of A, M, and Y, and they use those samples to estimate their joint distribution. This is a common situation where the confounder U is an *unobserved* variable.

(h) [3 points] Luckily, Mickey and Minnie recall the frontdoor adjustment formula for the causal graph from **(a)**:

$$\mathbb{P}_{\mathcal{X}}(Y \mid \texttt{do}(A = a)) = \sum_{m \in \{0,1\}} \mathbb{P}_{\mathcal{X}}(M = m \mid A = a) \sum_{a' \in \{0,1\}} (\mathbb{P}_{\mathcal{X}}(Y \mid M = m, A = a')\mathbb{P}_{\mathcal{X}}(A = a')) \tag{3}$$

Use `compute_conditional` to compute the following eight probabilities:

(i) $\mathbb{P}_{\mathcal{X}}(M = 0 \mid A = 0)$

(ii) $\mathbb{P}_{\mathcal{X}}(M = 1 \mid A = 0)$

(iii) $\mathbb{P}_{\mathcal{X}}(A = 0)$

(iv) $\mathbb{P}_{\mathcal{X}}(A = 1)$

(v) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid M = 0, A = 0)$

(vi) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid M = 0, A = 1)$

(vii) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid M = 1, A = 0)$

(viii) $\mathbb{P}_{\mathcal{X}}(Y = 1 \mid M = 1, A = 1)$

Plugging these values into the right-hand side of Equation (3), confirm numerically that the result matches your answers from **(f.i)**. Show your work.

Problem 2: Instrumental Variables [6 points]

Background

In this problem, you will observe how the accuracy of the *instrumental variables* (IV) estimator is sensitive to the "strength" of the instrumental variable. You will receive 1,000 samples of (W, A, Y) generated from the following SCM, for varying values of the coefficient β_{wa}:

$$\begin{aligned} U &= \varepsilon_u & \varepsilon_u &\sim \mathcal{N}(0,1) \\ W &= \varepsilon_w & \varepsilon_w &\sim \mathcal{N}(0,1) \\ A &= 8U + \beta_{wa} W + \varepsilon_a & \varepsilon_a &\sim \mathcal{N}(0,1) \\ Y &= 3U + 6A + \varepsilon_y & \varepsilon_y &\sim \mathcal{N}(0,1) \end{aligned}$$

Downloading data

In particular, for $\beta_{wa} = 0.1$, the samples can be downloaded from:

`https://github.com/kmatton/6.S091-causality-2024/tree/main/psets/pset1/samples_0.01.csv`,

for $\beta_{wa} = 1$, the samples can be downloaded from:

`https://github.com/kmatton/6.S091-causality-2024/tree/main/psets/pset1/samples_1.csv`,

and for $\beta_{wa} = 100$, the samples can be downloaded from:

`https://github.com/kmatton/6.S091-causality-2024/tree/main/psets/pset1/samples_100.csv`,

In each file, the first column corresponds to W, the second column corresponds to A, and the third column corresponds to Y. Each row corresponds to one sample.

Estimating the effect of A on Y using instrumental variables

For each value of β_{wa}, report the following quantities:

- $\widehat{\beta}_{aw}$, the coefficient of a linear regression (with intercept) of A onto W.
- $\widehat{\beta}_{yw}$, the coefficient of a linear regression (with intercept) of Y onto W.
- The ratio $\widehat{\beta}_{yw}/\widehat{\beta}_{aw}$

Do so in the form of a table:

β_{wa}	$\hat{\beta}_{aw}$	$\hat{\beta}_{wy}$	$\hat{\beta}_{aw}/\hat{\beta}_{wy}$
0.01			
1			
100			

Problem 3: The PC Algorithm [12 points]

Background

You have observed 10,000 samples from the distribution $\mathbb{P}_{\mathcal{X}}$ over variables $X_1, X_2, X_3, X_4, X_5, X_6$, and X_7. The distribution is Markov to some unknown DAG $\mathcal{G}$. Your goal is to recover the DAG $\mathcal{G}$ from samples, using the PC algorithm.

You are welcome to use packages for helper functions. For example, in Python, you might find the following packages helpful:

- `networkx` for manipulating graphs, e.g. the `graph.neighbors(node)` method to find the neighbors of a node, and
- `itertools` for listing all subsets $\mathbf{S} \subseteq \mathcal{X}$ of size k.

Downloading data

10,000 samples of $(X_1, X_2, X_3, X_4, X_5, X_6, X_7)$ are located at

https://github.com/kmatton/6.S091-causality-2024/tree/main/psets/pset1/pcalg_samples.csv,

with the i-th column corresponding to X_i, and each row corresponding to one sample.

Testing conditional independence from data

The PC algorithm removes edge $X_i - X_j$ by checking whether $X_i \perp\!\!\!\perp_{\mathbb{P}_{\mathcal{X}}} X_j \mid \mathbf{S}$. However, you cannot directly check this statement, since you do not have access to the distribution $\mathbb{P}_{\mathcal{X}}$ itself, but only to samples from $\mathbb{P}_{\mathcal{X}}$. Therefore, you will need to replace these checks with a *conditional independence test*.

In this problem, you will use a conditional independence test based on the partial correlation between X_i and X_j given $\mathbf{S}$. The p-value of the partial correlation test is implemented in the function `compute_pvalue`, found in the file

https://github.com/kmatton/6.S091-causality-2024/tree/main/psets/pset1/pcalg_helper_functions.py,

Given a significance level α, we say that we *reject* the null hypothesis of conditional independence if

`compute_pvalue(samples, i, j, S)` $\leq \alpha$

Otherwise, we say that the test passes, i.e., we find it plausible that $X_i \perp\!\!\!\perp_{\mathbb{P}_{\mathcal{X}}} X_j \mid \mathbf{S}$, and delete the edge $X_i - X_j$.

Skeleton phase [8 points]

Write a function `pcalg_skeleton(samples, alpha)` which perform the skeleton phase of the PC algorithm using the conditional independence test described above, with significance level $\alpha =$ `alpha`.

Sanity Check: You can check the correctness of your function by checking that `pcalg_skeleton(pcalg_samples, 0.05)` outputs the graph in Figure 1.

(a) Report the number of edges in the estimated skeleton when $\alpha = 0.2$ and using only the first 500 rows of `samples`, i.e., in Python, the number of edges in the skeleton output by

`pcalg_skeleton(samples[:500], 0.2)`

(b) Report the number of edges in the estimated skeleton when $\alpha = 0.001$ and using only the first 500 rows of `samples`, i.e., in Python, the number of edges in the skeleton output by

`pcalg_skeleton(samples[:500], 0.001)`

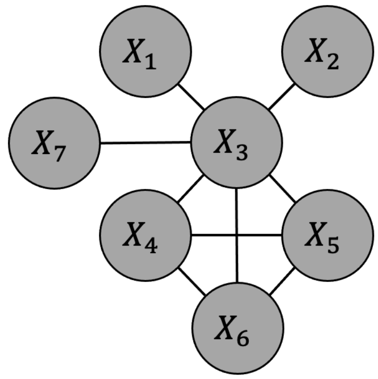

Figure 1: The skeleton that should be output by `pcalg_skeleton(pcalg_samples, 0.05)`.

Orientation phase [4 points]

Write a function `pcalg_orient(skeleton, separator_function)` which outputs any unshielded colliders according to the orientation phase of the PC algorithm. *You do **not** need to implement the final step of the orientation phase which applies the Meek rules.*

(c) Let `estimated_skeleton`, `estimated_separator_function` be the output of `pcalg_skeleton(pcalg_samples, 0.05)`, i.e., the output that you used to check your implementation of `pcalg_skeleton`. What are the unshielded colliders output by `pcalg_orient(estimated_skeleton, estimated_separator_function)`?

(d) Show that, by the Meek rules, you can orient $X_3 \to X_4$, $X_3 \to X_5$, $X_3 \to X_6$, $X_3 \to X_7$. Can you orient any more edges? If not, explain why.