Exercises

1	2	3	4	5	6

Surname, First name

1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9
0	0	0	0	0	0	0	0

Causal Data Science (5294CADS6Y)
Exam

Please make sure you enter YOUR NAME and SURNAME on the top left of this page, as well as YOUR STUDENT NUMBER on the top right of this page. Note: you also need to mark each digit of the student number in the matrix.

The exam has 100 points in total, you need **55 points to PASS.**

There are two types multiple choice questions**:**

- single correct answer: these have round bullet points
- multiple correct answers: these have square bullet points and are indicated by SELECT ALL THAT APPLY

The exam consists of six parts, where each part represents the material in each of the week of classes:

1. Week 1: basics, probability recap (11 points)
2. Week 2: graphical models, d-separation, causal graphs, structural causal models (18 points)
3. Week 3: simpler identification strategies for causal effect estimation (14 points)
4. Week 4: estimating causal effects: matching, propensity scores, IPW (13 points)
5. Week 5: causal discovery (28 points)
6. Week 6: do-calculus, transportability, advanced topics (16 points)

Each part has some simple multiple choice questions, some longer exercises and some more advanced discussion questions marked as difficult, so plan your time accordingly.

Week 1: Explaining the potential and limitations of causal inference, basics

1p **1a** The number of pool drownings in US for each year in the last 10 years is strongly correlated with the number of films Nicolas Cage has appeared in in the same time period. Does Nicolas Cage appearing in a movie cause people to drown more? ________________

1p **1b** If a variable Z is a common cause of the treatment T and the outcome Y, we say this variable is a ________________ for the effect of T on Y .

1p **1c** What is the name of the apparent paradox in which, if we estimate a total average treatment effect with or without adjusting for some covariates, we get a different result?

- (a) Simpson's paradox
- (b) Monty hall paradox
- (c) Barber paradox

1p **1d** We want to estimate the causal effect of a treatment T on a set of patients. In the data that we have, our patients have various income levels, and richer patients get assigned the treatment T more often. We also know that in general our richer patients (even without receiving the treatment T) have usually better outcomes. How can we classify the type of data that we have?

- (a) observational
- (b) randomized controlled trial
- (c) double-blind clinical trial

1p **1e** What is the highest level of Pearl's Causal Hierarchy (Level 3)?

- (a) counterfactuals
- (b) association
- (c) interventions

1p **1f** In a given distribution P, X is independent of Y given Z. Which of the following also always holds in P?

- (a) X is independent of Y
- (b) Y is independent of X given Z
- (c) X is independent of Z

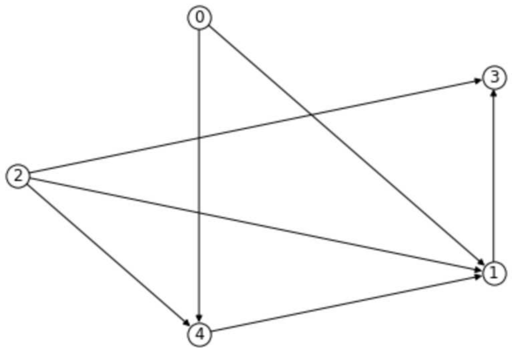
5p **1g** We have 3 random variables X, Y, Z, for which $Y \perp\!\!\!\perp Z || X$
Write down all 5 simplest factorizations that you can get from applying the chain rule and then applying the conditional independence simplification, when it is possible.

Week 2: Graphical models, d-separation and causal graphs

1p **2a** A directed path from i to j is a path in which all edges point towards ________

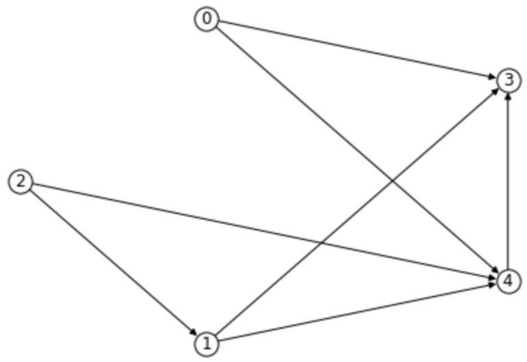
1p **2b** A non-collider on a path is an node in which the edges from its ____________ nodes are not both incoming.

1p **2c** In the following graph, is the path (0,4, 2, 3) blocked when we condition on 1?



ⓐ Yes ⓑ No

10p **2d** Which are the d-separations that hold in this graph? List **ALL 5** d-separations that hold and for each explain briefly why it is holds in this graph.



5p **2e** Consider the following linear Gaussian structural causal model:

$$\begin{cases} X = -Y - \epsilon_X \\ Y = 2\epsilon_Y \\ Z = 2Y + 3\epsilon_Z \\ \epsilon_X, \epsilon_Y, \epsilon_Z \sim N(0,1) \\ \epsilon_X \perp\!\!\!\perp \epsilon_Y, \epsilon_X \perp\!\!\!\perp \epsilon_Z, \epsilon_Y \perp\!\!\!\perp \epsilon_Z \end{cases}$$

Hint: keep in mind that for two random variables A and B

$$\mathbb{E}[aA + bB] = a\mathbb{E}[A] + b\mathbb{E}[B]$$
$$\mathrm{Var}[aA + bB] = a^2\mathrm{Var}[A] + b^2\mathrm{Var}[B] \quad \text{if} \quad A \perp\!\!\!\perp B$$

Compute the following distributions using the above rules:
1. P(X) = N(________, __________)
2. P(X|do(Z=-1)) = N(________, __________)
3. P(X|do(Y=1)) = N(________, __________)

We now perform a soft intervention on Y that changes its mechanism to

$$Y \leftarrow 3 + \epsilon_Y$$

Compute P(X| soft intervention on Y) and explain how you computed it (2 points):

Week 3: Identification strategies for causal effect estimation

1p **3a** What is a complete criterion for identifying all valid adjustment sets (the sets that can be used for the adjustment formula)? ____________________

1p **3b** A variable M that blocks all directed paths from treatment to outcome is a ____________________

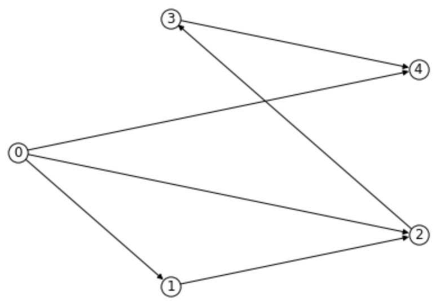
1p **3c** According to the **backdoor criterion**, a valid adjustment set Z for treatment i and outcome j needs to satisfy which of the following criteria (SELECT ALL THAT APPLY)?

- ☐ Z does not contain any descendant of the treatment i
- ☐ Z can contain some descendants of the treatment i, as long as it does not contain any descendant of any other node r on the directed path from treatment i to outcome j
- ☐ Z blocks all backdoor paths from treatment i to outcome j

1p **3d** Which of the following assumptions is one of the requirements for a variable I to be considered a valid instrument?

- (a) I is a descendant of the treatment
- (b) I causes the treatment
- (c) I mediates all effects from the treatment to the outcome

5p **3e** Given the following graph, apply the backdoor criterion and find **at least 5 valid adjustment sets** for the effect of X3 on X4, explaining why they are valid.



5p **3f** **Difficult:** Discuss the applicability of the backdoor criterion, adjustment criterion and the frontdoor criterion. Which are the cases in which one can apply only some of these criteria?

Week 4: Causal effect estimation - Matching, propensity scores, IPW

1p **4a** In potential outcome notation, what is the formula for the average causal effect of treatment on the control (ATC)?

1p **4b** What is the **Stable Unit Treatment Value Assumption (SUTVA)?**

- (a) The treatment assignment of a unit i should not affect the treatment assignment of any other unit.
- (b) The treatment assignment of a unit i is completely random.
- (c) The treatment assignment of a unit i depend on the treatment assignment of all other units in a stable way.
- (d) The treatment assignment of a unit i is stable in time.

1p **4c** In Inverse probability weighting we:

- (a) Reweigh the outcome of each unit by the inverse of the probability that they were assigned to whichever treatment option they were actually assigned to.
- (b) Reweigh the outcome of each unit by the probability of being in the treatment group.
- (c) Reweigh the outcome of each unit by the probability of being in the control group.

5p **4d** Given the SCM below and the realization for a specific unit i, compute the unit-level counterfactuals below for this unit explaining every step:

$$
\begin{array}{ll}
X \leftarrow \epsilon_X & \epsilon_X \sim \mathcal{N}(0,1) \\
Y \leftarrow 3 + \epsilon_Y & \epsilon_Y \sim \mathcal{N}(0,0.5) \\
Z \leftarrow X + 2Y - 2 + \epsilon_Z & \epsilon_Z \sim \mathcal{N}(0,1) \\
V \leftarrow X + Y - Z + \epsilon_V & \epsilon_V \sim \mathcal{N}(0,2)
\end{array}
$$

Realization for a specific unit i:

$$(x^i = 0.25, \quad y^i = 2.9, \quad z^i = 4.1, \quad v^i = -1)$$

1. Compute the counterfactual for this unit for the assignment X <- 1 (while everything else is the same)
2. Compute the counterfactual for this unit for the assignment for Y <- 1 (while everything else is the same)

5p **4e** **Difficult:** Discuss the differences between exact matching and inverse probability weighting, and their relative benefits or drawbacks.

Week 5: Causal Discovery

1p **5a** What type of structure does the GES algorithm output? ____________________

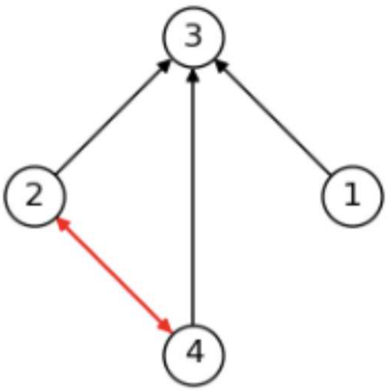
5p **5b** Apply SGS on this set of conditional independences and show the results of each of the three phases, explaining why you remove or orient edges at each step.

$$2 \perp\!\!\!\perp 3 | 1$$
$$1 \perp\!\!\!\perp 5 | 2, 3$$
$$1 \perp\!\!\!\perp 5 | 2, 3, 4$$
$$4 \perp\!\!\!\perp 5 | 2, 3$$
$$4 \perp\!\!\!\perp 5 | 1, 2, 3$$

2p **5c** In score-based causal discovery, the BIC score encourages sparsity of the graph.

(a) True (b) False

10p **5d** Starting from this CPDAG, which are the other four CPDAGs in the phase 1 neighbours of GES of this CPDAG?
Explain how you computed the results step by step.



(a) Start

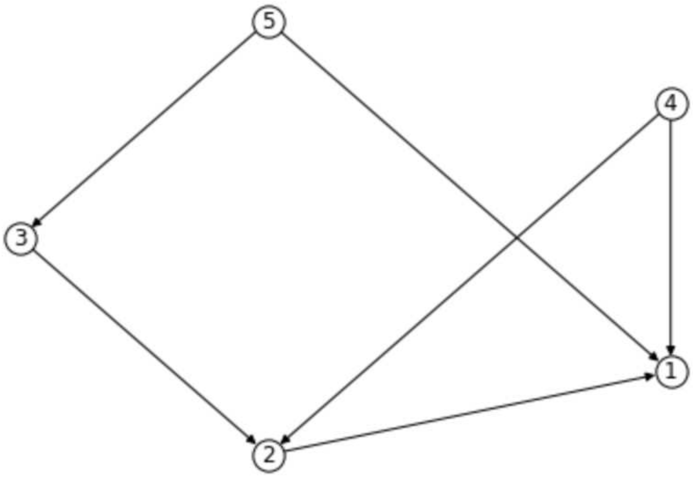
10p **5e** **Difficult:** Discuss the applicability and trade-offs between constraint-based causal discovery methods and restricted models like Additive noise models or LINGAM.

Week 6: do-calculus, transportability, advanced topics

1p **6a** Invariant Causal Prediction finds a subset of the ______________________________ of a variable Y.

2p **6b** If we use transportability and we have multiple datasets, can we in some cases identify more than we can with do-calculus? _______________

3p **6c** In this graph and in any distribution P that is Markov and faithful to it, $P(X_1|do(X_3))$ is :



- (a) P(X1|X3)
- (b) P(X1)
- (c) Neither P(X1) nor P(X1|X3)

10p **6d** **Difficult:** Discuss how can causality help machine learning based on the examples discussed in the classes.

This page is left blank intentionally