CAUSAL THINKING (MATH-352)

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Final exam

Date: 01.02.2025

Name: ________________________

SCIPER: ________________________

INSTRUCTIONS TO CANDIDATES

- This exam will contribute 80% to your final grade. To obtain the maximum number of points you should be clear about your reasoning and present your arguments explicitly. You have **3 hours** to complete the exam.
- All that can be used for this exam is a pen. No notes, books, summaries, formula collections or calculators are allowed. All questions should be answered.
- The finest enumerated item in each question will be marked on a scale of $0-2$ points, indicating an incorrect, partially correct and completely correct answer respectively (half-points are not given). **The exam has 5 questions with a total of 56 points**.
- **Write the answer to every question in the booklet**. Scrap paper will be provided for rough work, but only answers written in the booklet will be marked.

Mark question 1 (TOT: 10 points):

Mark question 2 (TOT: 18 points):

Mark question 3 (TOT: 12 points):

Mark question 4 (TOT: 8 points):

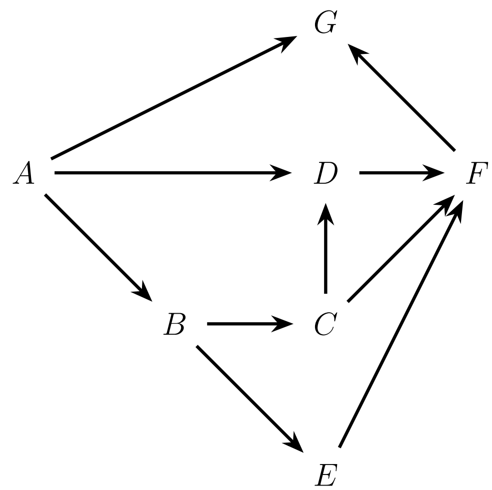
Mark question 5 (TOT: 8 points):

Variable definitions

The following variable definitions and supports apply throughout every question of the examination:

- $A \in \{0, 1\}$ is a treatment variable.
- $Y \in \mathbb{R}$ is an outcome variable.

Exercise 1 (DAG warm-up)**.** Consider the DAG below:



Use the rules of d-separation to decide whether the following independencies hold. Justify your answers.

Questions:

(1) $D \perp\!\!\!\perp B | A, C, E$,
(2) $G \perp\!\!\!\perp C | A, F$,
(3) $E \perp\!\!\!\perp G | A, D$,
(4) $B \perp\!\!\!\perp D | A, C, G, F$.
(5) $C \perp\!\!\!\perp A | B, F$

Solutions:

(1) True. All paths from D to B are blocked conditional on A, C, and E.
(2) True. All paths from G to C are blocked conditional on A and F.
(3) False, $E \to F \to G$ is open.
(4) False, $D \to F \leftarrow E \leftarrow B$ is open conditional on A, C, G, and F.
(5) False, $C \to D \leftarrow A$ is open as F is a descendent of D.

Exercise 2 (Understanding the average treatment effect)**.** We are interested in the effect of a work ethic workshop on the grades of students. Let A be a binary variable denoting attendance to the workshop ($A = 1$ for students that attend, $A = 0$ for those who do not), and let Y be their grade on the test, where Y can take any value in $[1, 6]$. Furthermore, let $C \in \{0, 1\}$ denote which class the students are in, assume that students are randomly assigned to $C = 0$ and $C = 1$ with equal probability.

Questions:

(1) Draw the DAG corresponding to this situation. Suppose that the only necessary variables are A, C and Y and that there are no relevant unmeasured variables. Suppose furthermore that participation in a workshop is directly chosen by each class's teacher (which also has a direct influence on their grades), and that a student's grades and class assignment are independent of the grades and class assignments of the other students (no interference, the students are independent and identically distributed, iid).

(2) The intervention raises the grade of the bottom 10% of students in Class $C = 0$ by 1 point on average, and does not change the grades of the rest of the class. What is the Average Treatment effect (ATE) in students of class $C = 0$?

(3) The intervention raises the grade of all students in Class $C = 1$ by 0.1 points on average. What is the ATE of students of class $C = 1$?

(4) What is the ATE of all students?

The Quantile Treatment Effect (QTE) at level α is defined as follows:

$$q_\alpha(Y^{a=1}) - q_\alpha(Y^{a=0}),$$

where $q_\alpha(X) := F_X^{-1}(\alpha)$ is the α-quantile of X.

(5) Does marginal exchangeability ($Y^a \perp\!\!\!\perp A$) imply that the QTE is identified? Give an identification formula if the QTE is identified, and justify your answer if it is not.

(6) What is the 10% QTE conditional on $C = 0$? And $C = 1$?

(7) Suppose $Y \sim \text{Unif}(1, 6)$.[1]

For $C = 0, 1$, what is the effect of the workshop on the percentage of students that pass? *We assume that the passing grade is 4.*

The following subquestions of this exercise assume that we do not know the effect of the workshop on grades, that is, the effects derived in (2)-(7) may not correspond to the effects in the following exercises.

(8) Is it possible for the ATE to be 0, but the QTE to be $\neq 0$? Give an example if you think it is possible, or justify why you think it is not possible.

(9) Suppose that a policy maker implements a randomized trial to test the effects of a policy on students. The ATE is zero. They declare:

"The average effect of the work ethic workshop on grades of the students is zero, hence there is no causal effect of the workshop on grades."

Do you agree with this statement? Justify your answer.

Solutions:

(1) See Figure 1.

(2) The ATE is $0.1 \times 1 = 0.1$.

[1] This is not necessarily a realistic distribution of grades, but it is a possible distribution of grades.

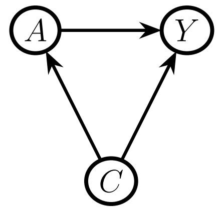


Figure 1

(3) The ATE is 0.1.
(4) As the ATE is an average of the ATEs in class 0 and in class 1, the ATE is also 0.1.
(5) Yes, exchangeability implies that the QTE is identified and equal to

$$q_\alpha(Y|A=1) - q_\alpha(Y|A=0),$$

where $q_\alpha(X|Z=z) := F^{-1}_{X|Z=z}(\alpha)$ is the α-quantile of X conditional on $Z=z$.
(6) The 10% QTE for class $C=0$ is 1, and the 10% QTE for class $C=1$ is 0.1
(7) For $C=0$, the bottom 10% of students get grades from 1 to 1.5, hence a 1 point increase does not changes the percentage of passing students.

For $C=1$, 40% of students pass without intervention. After the workshop, the 2% of students with grades between 3.9 and 4 (strictly) pass and hence the success rate is higher by 2% points.
(8) Yes, it is possible, simply take the example of class $C=0$ and say that the top 10% lose 1 point on average.
(9) This statement is misleading. The average causal effect is not informative of the quantile treatment effects. For example, it is possible that the workshop positively affects the lowest 10% of students and hence would be beneficial to implement for them.

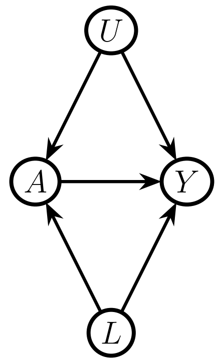


FIGURE 2

Exercise 3. Throughout this exercise, suppose that $L \in \mathcal{L}$ denotes baseline covariates in a discrete support $\mathcal{L}$, and $U \in \mathcal{U}$ denotes an unmeasured variable in a discrete support $\mathcal{U}$. Furthermore, suppose that we are considering a data-generating mechanism described by the DAG in Figure 2, and that positivity and consistency hold.

Consider the following two dynamic regimes:

$$g_1(l) := \arg\max_{a \in \{0,1\}} \mathbb{E}[Y^a | L = l],$$
$$g_2(l, a') := \arg\max_{a \in \{0,1\}} \mathbb{E}[Y^a | L = l, A = a'].$$

Suppose that ties in the arg max are broken by setting the regime to be equal to 0, that is,

$$\mathbb{E}[Y^0 | L = l] = \mathbb{E}[Y^1 | L = l] \Rightarrow g_1(l) = 0.$$

Similarly,

$$\mathbb{E}[Y^0 | L = l, A = a'] = \mathbb{E}[Y^1 | L = l, A = a'] \Rightarrow g_2(l, a') = 0.$$

Questions:

(1) Suppose there exists l and a' such that $g_1(l) \neq g_2(l, a')$. Justify whether or not exchangeability ($Y^a \perp\!\!\!\perp A | L$) can hold.
(2) Show that $\mathbb{E}[Y^{g_2}] \geq \mathbb{E}[Y^{g_1}]$.

In the remaining part of this exercise, suppose that

$$\mathbb{E}[Y^a | L = l, U = u] = \mathbb{E}[Y^a | L = l, U = u'],$$

for all $l \in \mathcal{L}$, $u, u' \in \mathcal{U}$.

(3) Can there exist l and a' such that $g_1(l) \neq g_2(l, a')$? Justify your answer.
(4) Show that

$$\mathbb{E}[Y^a] = \sum_{l \in \mathcal{L}} \mathbb{E}[Y | A = a, L = l] P(L = l).$$

(5) Show that

$$\mathbb{E}[Y^a] = \mathbb{E}\left[\frac{I(A = a)Y}{P(A = a | L)}\right].$$

(6) Deduce that, for any real-valued function $Q(l, a)$,

$$\mathbb{E}\left[\frac{I(A = g_1(L))Y}{P(A = g_1(L)|L)} + \left(1 - \frac{I(A = g_1(L))}{P(A = g_1(L)|L)}\right) Q(L, g_1(L))\right] = \mathbb{E}[Y^{g_1}].$$

Solutions:

(1) If $Y^a \perp\!\!\!\perp A|L$, then for all l and a, a',

$$\mathbb{E}[Y^a|L = l, A = a'] = \mathbb{E}[Y^a|L = l].$$

Hence, $g_1(l) = g_2(l, a')$.

(2)

$$\begin{aligned}\mathbb{E}[Y^{g_2}] &= \sum_l \sum_{a'} \mathbb{E}[Y^{g(l,a')}|A = a', L = l]P(A = a', L = l)\\ &\geq \sum_l \sum_{a'} \mathbb{E}[Y^{g_1(l)}|A = a', L = l]P(A = a', L = l) = \mathbb{E}[Y^{g_1}]\end{aligned}$$

(3) No, as

$$\begin{aligned}\mathbb{E}[Y^a|A = a', L = l] &= \sum_u \mathbb{E}[Y^a|A = a', L = l, U = u]P(U = u|A = a', L = l)\\ &= \sum_u \mathbb{E}[Y^a|L = l, U = u]P(U = u|A = a', L = l)\\ &= \sum_u \mathbb{E}[Y^a|L = l]P(U = u|A = a', L = l)\\ &= \mathbb{E}[Y^a|L = l],\end{aligned}$$

where the first equality follows by law of total expectation, the second by exchangeability conditional on L and U, the third by hypothesis, and the fourth by laws of probability.

(4) By the description of L and U, we know that $Y^a \perp\!\!\!\perp A|L, U$. Hence,

$$\mathbb{E}[Y|A = a, L = l, U = u] = \mathbb{E}[Y^a|L = l, U = u'] = \mathbb{E}[Y|A = a, L = l],$$

for all $l \in \mathcal{L}$, $u, u' \in \mathcal{U}$. Then,

$$\begin{aligned}\mathbb{E}[Y^a] &= \sum_{l\in\mathcal{L}} \sum_{u\in\mathcal{U}} \mathbb{E}[Y^a|L = l, U = u]P(L = l, U = u), \text{by the law of total probability,}\\ &= \sum_{l\in\mathcal{L}} \sum_{u\in\mathcal{U}} \mathbb{E}[Y|A = a, L = l, U = u]P(L = l, U = u)\\ &= \sum_{l\in\mathcal{L}} \sum_{u\in\mathcal{U}} \mathbb{E}[Y|A = a, L = l]P(L = l, U = u)\\ &= \sum_{l\in\mathcal{L}} \mathbb{E}[Y|A = a, L = l]P(L = l).\end{aligned}$$

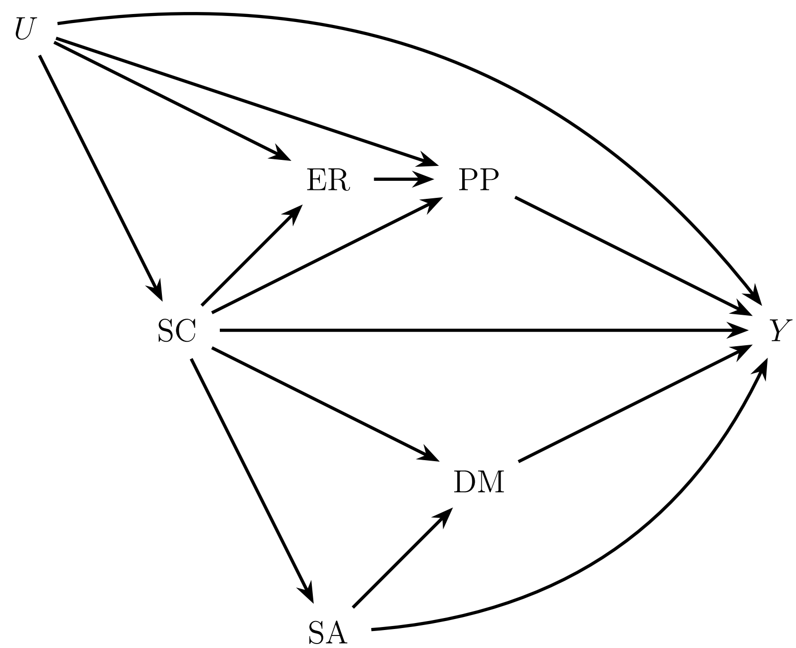
(5)
$$
\begin{aligned}
\mathbb{E}\left[\frac{I(A=a)Y}{P(A=a|L)}\right] &= \mathbb{E}\left[\frac{I(A=a)Y^a}{P(A=a|L)}\right] \\
&= \mathbb{E}\left[\frac{1}{P(A=a|L)}\mathbb{E}\left[I(A=a)Y^a|L\right]\right] \\
&= \mathbb{E}\left[\frac{1}{P(A=a|L)}\mathbb{E}\left[\mathbb{E}[I(A=a)Y^a|L,U]|L\right]\right] \\
&= \mathbb{E}\left[\frac{1}{P(A=a|L)}\mathbb{E}\left[P(A=a|L,U)\mathbb{E}[Y^a|L,U]|L\right]\right] \text{ as } Y^a \perp\!\!\!\perp A|L,U, \\
&= \mathbb{E}\left[\frac{1}{P(A=a|L)}\mathbb{E}[Y^a|L]\mathbb{E}\left[P(A=a|L,U)|L\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}[Y^a|L]\right] = \mathbb{E}[Y^a].
\end{aligned}
$$

(6) Analogously to the previous question, we can show that:
$$
\mathbb{E}\left[\frac{I(A=g_1(L))Y}{P(A=g_1(L)|L)}\right] = \mathbb{E}[Y^{g_1}].
$$

Then,
$$
\begin{aligned}
\mathbb{E}\left[\left(1-\frac{I(A=g_1(L))}{P(A=g_1(L)|L)}\right)Q(L,g_1(L))\right] &= \mathbb{E}\left[\mathbb{E}\left[1-\frac{I(A=g_1(L))}{P(A=g_1(L)|L)}|L\right]Q(L,g_1(L))\right] \\
&= \mathbb{E}\left[\left(1-\frac{\mathbb{E}[I(A=g_1(L))|L]}{P(A=g_1(L)|L)}\right)Q(L,g_1(L))\right] = 0,
\end{aligned}
$$

which proves the result.

Exercise 4. Suppose we are interested in the effect of different variables on the number of parties running for an election. Assume all variables can be intervened on and take values in $\mathbb{R}$ and assume that the data follow the distribution given by the DAG below, where SC denotes social cleavages, ER denotes electoral rules, SA denotes upper tier seat allocation, DM denotes district magnitude, PP denotes number of presidential parties and Y denotes the number of parties running for the election.



An analyst suggests the following model for estimating the expected number of parties:

$$\mathbb{E}[Y] = \beta_0 + \beta_1 \text{SA} + \beta_2 \text{DM} + \beta_3 \text{SC} + \beta_4 \text{ER} + \beta_5 \text{PP}.$$

We are interested in the causal interpretation of the parameters, assuming that the model is correct. In the context of this exercise, a causal interpretation will be defined as the expectation of the contrast of the outcome under different interventions conditional on some variables. For example, we say that β_1 has a causal interpretation if it can be written as

$$\beta_1 = \mathbb{E}[Y^s - Y^{s'} | \mathcal{Z}],$$

where s, s' are interventions on some variable in the DAG and $\mathcal{Z}$ is a set of variables in the DAG.

Suppose the model is correctly specified. For each of the parameters below, give a causal interpretation (and justify) or justify that there is no causal interpretation.

Questions:

(1) β_2
(2) β_3
(3) β_4
(4) β_5

Solutions:

(1) β_2 represents the causal effect of DM on Y, as it is identified:

$$\beta_2 = \mathbb{E}[Y^{d+1} - Y^d | SA, SC],$$

where d is an intervention on DM.

(2) β_3 does not represent a causal effect as the effect of SC on Y is unidentified due to unblocked backdoor paths through U.

(3) β_4 does not represent a causal effect for the same reasons as β_3 and furthermore can be dropped from the model.

(4) β_5 does not represent a causal effect for the same reasons as β_3 and β_4.

Exercise 5. Throughout this exercise, suppose that $L \in \mathcal{L}$ denotes baseline covariates in a discrete support $\mathcal{L}$ and that consistency holds. Furthermore, assume the following positivity conditions:

$$0 < P(A = a|L = l, Y^a = y) < 1,$$
$$0 < P(A = a|L = l) < 1,$$

for all $a \in \{0, 1\}, l \in \mathcal{L}, y \in \{0, 1\}$.

Finally, suppose that the following *sensitivity model* holds,

$$\frac{1}{\Gamma} \leq \frac{(1 - P(A = a|L = l))P(A = a|L = l, Y^a = y)}{P(A = a|L = l)(1 - P(A = a|L = l, Y^a = y))} \leq \Gamma.$$

for all $a \in \{0, 1\}, l \in \mathcal{L}, y \in \{0, 1\}$, and $\Gamma \geq 1$.

Questions:

(1) Show that, if $Y^a \perp\!\!\!\perp A|L$, the sensitivity model holds with $\Gamma = 1$.

(2) Argue that for all $a \in \{0, 1\}, l \in \mathcal{L}, y \in \{0, 1\}$,

$$\frac{1}{\Gamma}P(Y = y|L = l, A = a) \leq P(Y^a = y|L = l, A = 1 - a) \leq \Gamma P(Y = y|L = l, A = a).$$

(3) Deduce that

$$P(Y^a = y|L = l) \leq (\Gamma P(A = 1 - a|L = l) + P(A = a|L = l))P(Y = y|L = l, A = a),$$

and

$$\left(\frac{1}{\Gamma}P(A = 1 - a|L = l) + P(A = a|L = l)\right) P(Y = y|L = l, A = a) \leq P(Y^a|L = l).$$

(4) Show that, if $\Gamma = 1$, $\mathbb{E}[Y^a]$ is identified for all $a \in \{0, 1\}$. Give the identification formula.

Solutions:

(1) If $Y^a \perp\!\!\!\perp A|L$,

$$P(A = a|L = l, Y^a = y) = P(A = a|L = l).$$

Hence,

$$\frac{1}{1} \leq \frac{(1 - P(A = a|L = l))P(A = a|L = l, Y^a = y)}{P(A = a|L = l)(1 - P(A = a|L = l, Y^a = y)} = 1 \leq 1.$$

(2) We show the first inequality, the argument for the second inequality is analogous. Let $\pi(a|l) := P(A = a|L = l)$ The sensitivity model implies that

$$\frac{1 - \pi(a|l)}{\pi(a|l)} \leq \Gamma \frac{P(A = 1 - a|L = l, Y^a = y)}{P(A = a|L = l, Y^a = y)}.$$

However, by Bayes' theorem,

$$\frac{P(A = 1 - a|L = l, Y^a = y)}{P(A = a|L = l, Y^a = y)} = \frac{P(Y^a|L = l, A = 1 - a)(1 - \pi(a|l))}{P(Y^a = y|L = l, A = a)\pi(a|l)}.$$

Then, by consistency, and rearranging,

$$\frac{1}{\Gamma} \leq \frac{P(Y^a = y|L = l, A = 1-a)}{P(Y = y|L = l, A = a)}.$$

Multiplying by $P(Y = y|L = l, A = a)$ on both sides gives the result.

(3) By the law of total probability,

$$P(Y^a = y|L = l) = P(Y^a|L = l, A = 1-a)P(A = 1-a|L = l) + P(Y^a|L = l, A = a)P(A = a|L = l).$$

Then, simply applying the bounds derived in (2), we obtain the result.

(4) As Y is binary,

$$\mathbb{E}[Y^a] = P(Y^a = 1) = \sum_l P(Y^a = 1|L = l)P(L = l).$$

Then, as $\Gamma = 1$, the sensitivity model holds with equality and hence

$$\begin{aligned}\mathbb{E}[Y^a] &= \sum_l \left(P(A = 1-a|L = l) + P(A = a|L = l)\right) P(Y = y|L = l, A = a)P(L = l)\\ &= \sum_l P(Y = y|L = l, A = a)P(L = l).\end{aligned}$$