UNIVERSITY OF CAMBRIDGE

MAMA/221, NST3AS/221, MAAS/221

MAT3

MATHEMATICAL TRIPOS Part III

Monday, 5 June, 2023 9:00 am to 11:00 am

PAPER 221

CAUSAL INFERENCE

Before you begin please read these instructions carefully

Candidates have TWO HOURS to complete the written examination.

Attempt no more than **THREE** questions.
There are **FOUR** questions in total.
The questions carry equal weight.

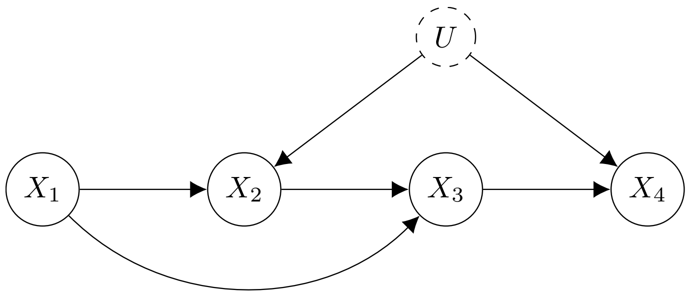
STATIONERY REQUIREMENTS
Cover sheet
Treasury tag
Script paper
Rough paper

SPECIAL REQUIREMENTS
None

You may not start to read the questions printed on the subsequent pages until instructed to do so by the Invigilator.

1

Suppose (X_1, X_2, X_3, X_4, U) satisfies the causal model corresponding to the following graph. Suppose $X_1, \ldots, X_4$ are observed (indicated by solid circles) and U is unobserved (indicated by the dashed circle).



(i) State the d-separation criterion, and then use it to conclude that the graph implies no conditional independence relationships between the observed variables, that is, for all disjoint subsets $\mathcal{I}, \mathcal{J}, \mathcal{K} \subset \{1, 2, 3, 4\}$, $X_{\mathcal{I}} \perp\!\!\!\perp X_{\mathcal{J}} \mid X_{\mathcal{K}}$ is generally not true.

(ii) Suppose $X_1, \ldots, X_4, U$ are all discrete. Show that

$$\sum_{x_2} p(x_4 \mid x_1, x_2, x_3) p(x_2 \mid x_1) \text{ does not depend on } x_1,$$

where $p(x_4 \mid x_1, x_2, x_3)$ is the conditional probability of $X_4 = x_4$ given $X_1 = x_1, X_2 = x_2, X_3 = x_3$, and $p(x_2 \mid x_1)$ is the conditional probability of $X_2 = x_2$ given $X_1 = x_1$.

(iii) Now suppose X_1, X_2, X_3 are all binary. Derive identification formulas for the average treatment effects

$$\mathbb{E}\left(X_4(X_j = 1) - X_4(X_j = 0)\right) \text{ for } j = 1, 2, 3,$$

where $X_4(X_j = x_j)$ is the potential outcome of X_4 when X_j is set to x_j.

2

Consider the problem of inferring the causal effect of a binary treatment variable A on a real-valued outcome variable Y. Let X be some observed covariate. Let $Y(a)$ be the potential outcome when A is set to $a = 0, 1$.

(i) State the no unmeasured confounders assumption. Show that undet this assumption, and additionally other sensible assumptions that you should state carefully, the average treatment effect on the treated (ATT) is identified by

$$\mathbb{E}(Y(1) - Y(0) \mid A = 1) = \mathbb{E}(Y \mid A = 1) - \frac{\mathbb{E}(\pi(X)\mu(X))}{\mathbb{E}(\pi(X))}, \tag{1}$$

where $\pi(X) = \mathbb{P}(A = 1 \mid X)$ and $\mu(X) = \mathbb{E}(Y \mid A = 0, X)$.

(ii) Assuming X is discrete, show that the influence curve for the functional $\beta = \mathbb{E}(\pi(X)\mu(X))$ is given by

$$(1 - A)\frac{\pi(X)}{1 - \pi(X)}(Y - \mu(X)) + A\mu(X) - \beta.$$

[You may use any results given in the lectures. In particular, it may be useful to know that the influence curve of $\mathbb{E}(Y \mid X = x)$ (for fixed x) is given by

$$\frac{1_{\{X=x\}}}{\mathbb{P}(X = x)} \cdot (Y - \mathbb{E}(Y \mid X = x)).]$$

(iii) Given an i.i.d. sample $(X_i, A_i, Y_i), i = 1, \ldots, n$, use the above results to suggest an estimator of the right hand side of (1).

3

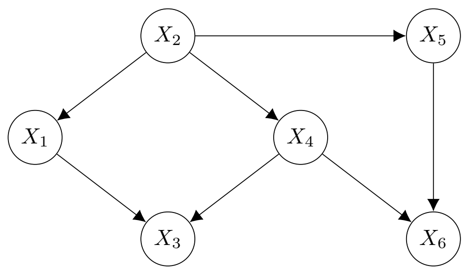
You may assume all random variables in this question are discrete.

Consider a random vector $X = (X_1, \ldots, X_p)$ and the subvector

$$S = \begin{pmatrix} X_1 \\ X_{\mathcal{M}} \\ X_{\mathcal{N}} \end{pmatrix},$$

where $\mathcal{M}$ and $\mathcal{N}$ are disjoint subsets of $\{2, \ldots, p\}$. We say $X_{\mathcal{M}}$ is a "Markov blanket" of X_1 in S if $X_1 \perp\!\!\!\perp X_{\mathcal{N}} \mid X_{\mathcal{M}}$. In this context, we refer to the cardinality of $\mathcal{M}$ as the size of Markov blanket.

(i) Suppose X factorizes according to the directed acyclic graph below. Show that a Markov blanket of X_1 in X is (X_2, X_3, X_4).



(ii) Suppose X factorizes according to and is faithful to a directed acyclic graph $\mathcal{G}$. Describe, with justification, the smallest Markov blanket of X_1 in X.

(iii) State the definition of conditional independence for random variables and use it to prove the contraction axiom: if A, B, C, D are random vectors that satisfy $A \perp\!\!\!\perp B \mid D$ and $A \perp\!\!\!\perp C \mid (B, D)$, then $A \perp\!\!\!\perp (B, C) \mid D$.

(iv) Suppose we are interested in estimating the causal effect of an intervention on a random variable A on another variable Y. Suppose X is a suffcient adjustment set in the sense that $Y(a) \perp\!\!\!\perp A \mid X$ where $Y(a)$ is the potential outcome of Y under the intervention $A = a$. Show that any Markov blanket of A in (A, X) is also a sufficient adjustment set.

4

Consider the following randomized encouragement trial to evaluate the causal effect of smoking cessation on blood pressure. The experimenter randomly assigns each smoker who participate in this trial to the treatment group or the control group by flipping a fair coin. Subjects in both groups receive a leaflet about the health risks of smoking, but the subjects in the treatment group receive an incentive of £1,000 if they quit smoking. After a year, we measure the blood pressure of all trial participants. (For this question, we assume that subjects who quit smoking do not resume smoking in this period.)

Let Z denote the treatment assignment: $Z = 1$ means the subject is in the treatment group and $Z = 0$ means the control group. Let A be the indicator for quitting smoking and Y be the blood pressure. You may treat the random variables corresponding to each subject as i.i.d.

(i) Often, the randomized encouragement Z is used as an *instrumental variable* for the exposure of interest A. Draw the causal graph for what this means and write down the assumptions using potential outcomes. Explain what these assumptions mean in the context of the trial desribed above.

For the rest of this question you may assume Z is a valid instrumental variable for A.

(ii) Use potential outcomes to describe the monotonicity assumption that the possibility of receiving £1,000 should only motivate the subjects to quit smoking. Then use it to prove that the complier average treatment effect of A on Y is identified by the Wald ratio

$$\frac{\mathbb{E}(Y \mid Z = 1) - \mathbb{E}(Y \mid Z = 0)}{\mathbb{E}(A \mid Z = 1) - \mathbb{E}(A \mid Z = 0)}.$$

(iii) Suppose U is an unobserved, common cause of A and Y, and there are no other confounders apart from U. Show that the (overall) average treatment effect of A on Y is also identified by the Wald ratio if there is no confounder-instrument interaction in determining the exposure in the sense that

$$\mathbb{E}(A \mid Z = 1, U) - \mathbb{E}(A \mid Z = 0, U)$$

does not depend on U.

END OF PAPER