UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Monday, 14 June, 2021 12:00 pm to 2:00 pm

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

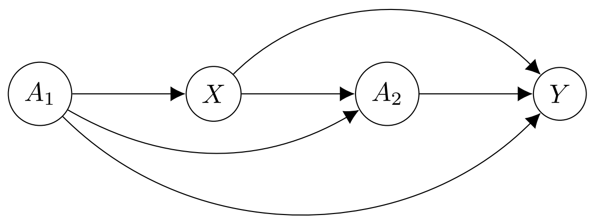
SPECIAL REQUIREMENTS
None

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

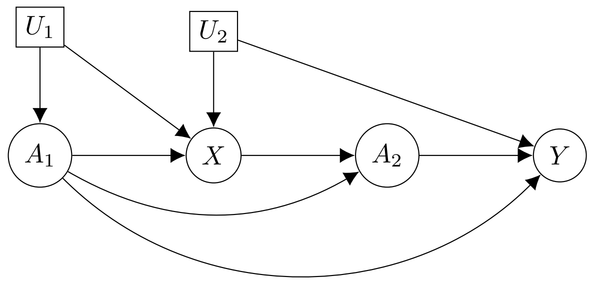
1

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

(a) Suppose the random variables (A_1, X, A_2, Y) satisfy a causal model defined by the graph below which represents a sequentially randomised experiment. Let $Y(a_1, a_2)$ be the counterfactual of Y under the intervention $(A_1, A_2) = (a_1, a_2)$. Explain what *causal identification* of $\mathbb{E}[Y(a_1, a_2)]$ means. Then obtain an identifying formula for $\mathbb{E}[Y(a_1, a_2)]$. [You may directly apply the g-computation formula in the lectures.]



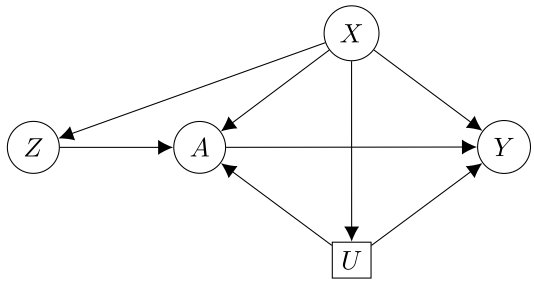
(b) Now consider the causal graphical model in the next diagram with two unmeasured variables U_1 and U_2. Show that the formula you obtained in part (a) still identifies $\mathbb{E}[Y(a_1, a_2)]$. Do you think your formula still holds when there is a directed edge from X to Y in this graph? Justify your answer.



(c) Consider the more general setting where we observe $T \geqslant 2$ sequential treatments $A_1, A_2, \ldots, A_T$ and T intermediate measures $X_1, X_2, \ldots, X_T$. The variables are measured in the temporal order: $A_1, X_1, A_2, X_2, \ldots, A_T, X_T$. Let $Y = X_T$ be the final outcome we are interested in.

Suppose there are unmeasured variables that are causal ancestors of some of the observed variables. Extend your proof in part (b) to give a sufficient condition such that $\mathbb{E}[Y(a_1, a_2, \ldots, a_T)]$ is identifiable. Your condition should be general enough to include part (b) as a special case. [Your condition can either be graphical or a series of conditional independences for counterfactuals.]

2



(a) Consider the setting in the diagram above where the causal effect of A on Y is still confounded by an unmeasured variable U after conditioning on the scalar variable X. An instrumental variable (Z in the diagram) can help to remove the unmeasured confounding bias. In words, describe the assumptions in this causal model in order for Z to be a valid instrumental variable. Then state these assumptions mathematically.

Throughout the rest of this question, suppose the assumptions in part (a) are satisfied and the causal effect of A is homogeneous, so $Y(a) - Y(0) \equiv \beta_0$ where $Y(a)$ is the counterfactual outcome under the intervention $A = a$. Assume $\mathbb{E}[X] = \mathbb{E}[Z] = \mathbb{E}[A] = \mathbb{E}[Y] = 0$.

Consider the following parametric assumptions on the distribution of (X, Z, A, Y):

- Assumption 1: $\mathbb{E}[Y - \beta_0 A \mid X] = \alpha_0 X$;
- Assumption 2: $\mathbb{E}[Z \mid X] = \gamma_0 X$.

These assumptions might not be true but can still help us to estimate β_0. Let (Z_i, X_i, A_i, Y_i), $i = 1, \dots, n$ be an i.i.d. sample from this distribution and $\hat{\gamma}$ be the least squares estimator of γ_0 defined in Assumption 2, $\hat{\gamma} = (\sum_{i=1}^n Z_i X_i)/(\sum_{i=1}^n Z_i^2)$.

Let $(\hat{\alpha}, \hat{\beta})$ be the solution of the following equations:

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta A_i - \alpha X_i)(Z_i - \hat{\gamma} X_i) = 0,$$

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \beta A_i - \alpha X_i)X_i = 0.$$

(b) Without using Assumption 1 or Assumption 2, give a sufficient condition on the distribution of (X, Z, A, Y) so that the solution exists with probability tending to 1 as $n \to \infty$. [You need not state regularity conditions for the law of large numbers in your proof.]

[QUESTION CONTINUES ON THE NEXT PAGE]

(c) Show that $\hat{\beta}$ is doubly robust in the sense that $\hat{\beta}$ consistently estimates β_0 when $n \to \infty$ if at least one of Assumptions 1 and 2 is true.

(d) Suppose both Assumptions 1 and 2 are true. Let $V_i = Y_i - \beta_0 A_i - \alpha_0 X_i$. Derive the asymptotic variance of $\hat{\beta}$ under two additional assumptions: $\mathrm{Var}(V_i \mid A_i, X_i) = \sigma^2$ and $\mathbb{E}[A_i \mid X_i, Z_i] = \lambda(Z_i - \gamma_0 X_i)$. [You need not state regularity conditions for the central limit theorem and the Z-estimation theory covered in the lectures.]

3

Suppose random variables $A, Y, X_1, X_2, \ldots, X_p$ and their counterfactuals satisfy the causal model with respect to a directed acyclic graph $\mathcal{G}$. We are interested in identifying the expectation of $Y(a)$ where $Y(a)$ is the counterfactual of Y under the intervention $A = a$. Suppose all the random variables are measured.

(a) State the definition of *faithfulness* of a graphical model.

For the rest of this question, assume the joint distribution of $(A, Y, X_1, X_2, \ldots, X_p)$ is faithful to $\mathcal{G}$.

(b) Let $I \subseteq \{1, 2, \ldots, p\}$ be an index set and $\mathbf{X}_I$ be the corresponding entries in $\mathbf{X} = (X_1, \ldots, X_p)$. State the definition of a *back-door path* from A to Y and the *back-door criterion* for $\mathbf{X}_I$ that allows one to conclude that $Y(a)$ is conditionally independent of A given $\mathbf{X}_I$.

$$Y(a) \perp\!\!\!\perp A \mid \mathbf{X}_I. \tag{1}$$

(c) In the rest of this question, we examine how to formalise the concept that a random variable X_i is a "confounder" for the causal effect of A on Y (condition (1) only means there are"no unmeasured confounders"). It is expected that a good definition of a confounder should satisfy the following two properties:

Property 1: If $\mathbf{X}_J$ contains all the confounders among $X_1, \ldots, X_p$, then condition (1) holds for $I = J$.

Property 2: For any X_i that is a confounder, there exists a possibly empty subset $J \subseteq \{1, 2, \ldots, p\} \setminus \{i\}$ such that condition (1) holds for $I = J \cup \{i\}$ but not for $I = J$.

Consider the following candidate definitions of a confounder. For each of the following definitions, say whether or not the definition satisfies Property 1 and whether or not the definition satisfies Property 2. Give a proof or counterexample in each case as appropriate.

Definition 1: X_i is a confounder if X_i is not a descendant of A and blocks a back-door path from A to Y in $\mathcal{G}$.

Definition 2: X_i is a confounder if X_i is not a descendant of A and belongs to every "minimal sufficient adjustment set". ($I \subseteq \{1, 2, \ldots, p\}$ is called a minimal sufficient adjustment set if (1) holds for I but not any strict subset of I.)

Definition 3: X_i is a confounder if X_i is not a descendant of A and there exists $J \subseteq \{1, 2, \ldots, p\} \setminus \{i\}$ (J can be empty) such that $X_i \not\perp\!\!\!\perp A \mid \mathbf{X}_J$ and $X_i \not\perp\!\!\!\perp Y \mid (A, \mathbf{X}_J)$.

[You may choose to work on any order of the Definitions and Properties.]

4

This question concerns an applied study of the causal effect of Catholic high school attendance on educational attainment and test scores in the United States. This study uses a survey of 11,839 students in 8th grade who are then followed up after their 10th grade and 12th grade.

A wide variety of information was collected. The investigator of this study defines treatment as a binary indicator of whether the survey participant was in a Catholic high school or a non-Catholic school in their 10th grade (denoted as `CH10`). This question considers two outcome variables: the 12th grade maths standardised score (`MATHS12`) and a binary indicator of whether the survey participant enrolled in a four-year college (`COLL12`). Several covariates are adjusted for in this study, including log of family income (`INCOME8`), whether the student lived in an urban area (`URBAN8`), and the 8th grade maths score (`MATHS8`). The real study actually adjusts for many other covariates; for the purpose of this question, we will just assume these are the only three covariates that are adjusted for.

(a) The investigator also constructs a subsample of the survey participants who went to a Catholic middle school in their 8th grade. The following table compares the covariate means in the the full sample (11,839 participants) with the Catholic 8th grade subsample (1,006 participants):

	Full sample			Catholic 8th grade		
	`CH` = 1	`CH` = 0	Difference	`CH` = 1	`CH` = 0	Difference
Sample size	11,167	672		366	640	
`INCOME`	10.23	10.72	.49***	10.47	10.66	.19***
`URBAN`	.19	.46	.27***	.47	.51	.04
`MATHS8`	51.19	55.05	3.86***	54.12	55.59	1.47

*** means the difference is statistically significant at level 0.01.

Give one advantage and one disadvantage of using the Catholic 8th grade subsample to study the causal effect of Catholic high school attendance on educational outcomes.

(b) The investigator then uses ordinary least squares to estimate the "effect" of `CH10` on `MATHS12`. The results (estimated coefficients of `CH10` and standard errors) are shown in the next table, adjusting for different sets of covariates using both the full sample and the Catholic 8th grade subsample.

	Full sample			Catholic 8th grade		
Adjust for	None	`INCOME` & `URBAN`	All	None	`INCOME` & `URBAN`	All
Coefficient	.73	.37	.32	.60	.48	60
Standard error	.08	.09	.09	.13	.15	.15

[QUESTION CONTINUES ON THE NEXT PAGE]

(b.i) Support your conclusions in part (a) using observations from the new table.

(b.ii) State a sufficient condition that allows the investigator to interpret the estimated coefficients under the "All" columns in this table as causal effects. The sufficient condition should be in terms of the distribution of $A = \texttt{CH10}$, $Y = \texttt{MATHS12}$, $\mathbf{X} = (\texttt{INCOME8}, \texttt{URBAN8}, \texttt{MATHS8})$ and their relevant counterfactuals.

(b.iii) Let $(A_i, Y_i, \mathbf{X}_i)$, $i = 1, \ldots, n$ be the dataset under analysis ($n = 11839$ in the full sample and $n = 1006$ in the Catholic 8th grade subsample). Give the algebraic expression for the inverse probability weighting estimator of the average treatment effect. Does this estimator consistently estimate the average treatment effect under your condition in part (b.ii)?

(c) The investigator now turns to the binary outcome COLL12. Due to concerns about unmeasured confounding, the investigator proposes to use the following model as a sensitivity analysis. Let $A = \texttt{CH10}$, $Y = \texttt{COLL12}$ and $\mathbf{X} = (\texttt{INCOME8}, \texttt{URBAN8}, \texttt{MATHS8})$. The investigator assumes that the observations are an iid sample from the following model

$$\begin{aligned} A &= I(\mathbf{X}^T \boldsymbol{\alpha} + U > 0), \\ Y &= I(A\beta + \mathbf{X}^T \boldsymbol{\gamma} + V > 0), \end{aligned} \tag{1}$$

where I is the indicator function and the unobserved variables (U, V) have a bivariate normal distribution

$$\begin{pmatrix} U \\ V \end{pmatrix} \mid \mathbf{X} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right).$$

The investigator proposes to treat ρ as a sensitivity parameter.

(c.i) Suppose we interpret (1) as a nonlinear structural equation model. Explain why the no unmeasured confounders assumption no longer holds when $\rho \neq 0$.

(c.ii) Let $(A_i, Y_i, \mathbf{X}_i)$, $i = 1, \ldots, n$ be the dataset under analysis. Suppose the observations are iid. For any given value of ρ, suggests an estimator $\hat{\beta}_\rho$ of β assuming the model above holds. [You may use the distribution function $F_\rho(u, v) = \mathbb{P}(U \leqslant u, V \leqslant v)$ without giving its closed form expression.]

(c.iii) The investigator suggests that by varying ρ between -1 and 1, one may obtain causal effect estimators with different strengths of the unmeasured confounder. Explain why this interpretation is logically flawed.

END OF PAPER