UNIVERSITY OF CAMBRIDGE

MAMA/221, NST3AS/221, MAAS/221

MAT3

MATHEMATICAL TRIPOS Part III

Friday 13 June 2025 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

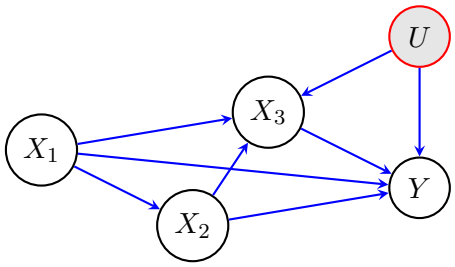
SPECIAL REQUIREMENTS
None

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

1 Causal DAGs and multiple regression

(a) Consider a linear structural equation model with respect to a directed acyclic graph (DAG). State Wright's path tracing rule and define the total causal effects in terms of path coefficients.

(b) Consider the following DAG that describes the causal relationship between Y (infant health score), X_3 (birth weight), X_2 (maternal smoking during pregnancy), X_1 (maternal education), and U (unmeasured genetic predisposition). Assume X_1, X_2, X_3 all have variance 1, and assume (X_1, X_2, X_3, Y, U) satisfies a linear structural equation model with respect to this DAG.



Consider the following linear regression problems, assuming an intercept term is always included. Among the least-squares regression coefficients in each problem (for example, there are three coefficients in problem (vii)), which have a causal effect interpretation? Justify your answer.

(i) Regress Y on X_1.

(ii) Regress Y on X_2.

(iii) Regress Y on X_3.

(iv) Regress Y on X_1 and X_2 jointly.

(v) Regress Y on X_1 and X_3 jointly.

(vi) Regress Y on X_2 and X_3 jointly.

(vii) Regress Y on X_1, X_2 and X_3 jointly.

[Hint: if you believe a regression coefficient identifies a causal effect, specify whether it reflects a total causal effect, a total causal effect given some random variable(s), or a direct causal effect. Otherwise, briefly explain why the coefficient does not have a causal interpretation.]

2 Z-estimator of the ATE

Consider a Bernoulli trial where treatment assignment Z is randomized with $P(Z = 1) = p^\star$ for some $p^\star \in (0, 1)$. The observations $\{Y_i, Z_i\}_{i=1}^n$ are i.i.d. across individuals. Assume that $\mathbb{E}[Y^2|Z = z]$ is bounded for $z = 0, 1$.

Define $\tau^\star = \mathbb{E}\big[\mathbb{E}[Y_i|Z_i = 1] - \mathbb{E}[Y_i|Z_i = 0]\big]$. Compare the following inverse probability weighted (IPW) estimators:

$$\tilde{\tau} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i Z_i}{p^\star} - \frac{Y_i(1 - Z_i)}{1 - p^\star}\right),$$

$$\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i Z_i}{\hat{p}} - \frac{Y_i(1 - Z_i)}{1 - \hat{p}}\right),$$

where $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} Z_i$.

(a) Derive the asymptotic distribution of $\tilde{\tau}$ as $n \to \infty$.

(b) By writing $(\hat{p}, \hat{\tau})$ as a Z-estimator (the solution to a system of estimating equations that you should specify), show that $\hat{\tau}$ is a consistent and asymptotically normal estimator for $\tau^\star$. [*Hint: The inverse of a* 2×2 *matrix is given by:* $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad-bc}\begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$.]

(c) Show that using $\hat{\tau}$ instead of $\tilde{\tau}$ leads to an efficiency improvement, that is, the asymptotic variance of $\hat{\tau}$ is less than or equal to that of $\tilde{\tau}$.

3 Causal DAG and conditional independence

An epidemiologist is interested in the effect of a drug (A) on the risk of heart attack (Y). The drug works partly directly, and partly indirectly via acting as a muscle relaxant (M) which in turn affects (Y). The muscle relaxing property M may also have some side effects (S).

A doctor's recommendation (Z) increases the chance of the drug being taken. The propensity of the patient to take the treatment is also a function of their age (D), which will in turn affect the muscular composition of their heart (C). It is known that muscular composition of the heart influences muscle relaxation (M).

In addition, both the likelihood of taking the treatment A and of having a heart attack Y are dependent on the patient's sex (G), and (Y) is also influenced by the patient's weight (W).

(a) Draw a minimal causal diagram that encodes all scientific assumptions above.

(b) Which of the following conditional independences hold under the graph you drew? If you believe a conditional independence holds, you need to justify your answer. Otherwise, give an open path given the corresponding conditioning set.

(i) $D \perp\!\!\!\perp G$;

(ii) $D \perp\!\!\!\perp Z \mid A$;

(iii) $Z \perp\!\!\!\perp Y \mid A, G, D$;

(iv) $C \perp\!\!\!\perp A \mid D$.

(c) Suppose we consider only the subset of patients who are known to have suffered from the side effects of the muscle relaxation effect (that is, $S = 1$). How does this change your answers for (b)?

(d) Recall that the undirected moral graph $\mathcal{G}^m$ of a DAG $\mathcal{G}$ is obtained by first adding undirected edges between all pairs of vertices that have a common child and then erasing the direction of all the directed edges.

We say that the distribution of a random vector $\boldsymbol{X} = (X_1, ..., X_p)^\top$ factorizes according to a DAG $\mathcal{G} = (V = [p], E)$ if its density function $f(\boldsymbol{x})$ satisfies

$$f(\boldsymbol{x}) = \prod_{j=1}^{p} f_{j|\mathrm{pa}(j)}(x_j \mid x_{\mathrm{pa}(j)}),$$

where $\mathrm{pa}(j)$ is the parent set of j in $\mathcal{G}$. Show if the distribution of a random vector $\boldsymbol{X} = (X_1, ..., X_p)^\top$ factorizes according to a DAG $\mathcal{G}$, it must satisfy the global Markov property with respect to $\mathcal{G}^m$. [You may assume $f(\boldsymbol{x}) > 0$ and use the Hammersley–Clifford Theorem without proof.]

4 Causal identification and doubly robust estimation of conditional average treatment effect (CATE)

Consider a standard observational study with a binary treatment $Z \in \{0,1\}$, covariates $X = (X_1, X_2)$, where X_1 denotes a discrete random variable indicating the subgroups of interest and X_2 contains the rest of the covariates, and some outcome of interest Y. Assume the propensity score $e(X) = P(Z = 1 \mid X)$ satisfies positivity ($0 < e(X) < 1$), and that strong ignorability ($Z \perp\!\!\!\perp \{Y(1), Y(0)\} \mid X$) and SUTVA hold.

(a) The parameter of interest is the subgroup causal effect:

$$\tau(x_1) = \mathbb{E}[Y(1) - Y(0) | X_1 = x_1]$$

(i) Show that $\tau(x_1)$ can be identified from observed data using outcome modeling:

$$\tau(x_1) = \mathbb{E}\left[\mathbb{E}[Y|Z=1,X]|X_1 = x_1\right] - \mathbb{E}\left[\mathbb{E}[Y|Z=0,X]|X_1 = x_1\right]$$

(ii) Show that $\tau(x_1)$ can be identified from observed data using propensity scores:

$$\tau(x_1) = \mathbb{E}\left[\frac{\mathbb{1}(X_1 = x_1)ZY}{e(X)} - \frac{\mathbb{1}(X_1 = x_1)(1-Z)Y}{1-e(X)}\right] / \mathrm{pr}(X_1 = x_1)$$

(b) More generally, define

$$\tau(X) = \mathbb{E}[Y(1) - Y(0)|X],$$
$$\tilde{\tau}^{dr}(X) = \tilde{\mu}_1^{dr}(X) - \tilde{\mu}_0^{dr}(X),$$

and

$$\tilde{\mu}_1^{dr}(X) = \mathbb{E}\left[\frac{Z\big(Y - \mu_1(X,\beta_1)\big)}{e(X;\alpha)} + \mu_1(X,\beta_1)\big|X\right],$$
$$\tilde{\mu}_0^{dr}(X) = \mathbb{E}\left[\frac{(1-Z)\big(Y - \mu_0(X,\beta_0)\big)}{1-e(X;\alpha)} + \mu_0(X,\beta_0)\big|X\right].$$

Here, $\mu_1(X,\beta_1)$ and $\mu_0(X,\beta_0)$ are the working models for the conditional mean of the outcome under $Z = 1$ and $Z = 0$, respectively, with parameters β_1 and β_0. Similarly, $e(X,\alpha)$ is a working model for the propensity score, indexed by α. For $z = 0, 1$, $\mu_z(X) = \mathbb{E}[Y|Z = z, X]$.

Show that $\tilde{\tau}^{dr}(X) = \tau(X)$ if either

(i) $e(X,\alpha) = e(X)$; or

(ii) $\mu_1(X,\beta_1) = \mu_1(X)$ and $\mu_0(X,\beta_0) = \mu_0(X)$.

END OF PAPER