UNIVERSITY OF CAMBRIDGE

MAMA/221, NST3AS/221, MAAS/221

MAT3
MATHEMATICAL TRIPOS Part III

Friday 7 June 2024 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

Consider a randomized clinical trial to test the efficacy of a new drug. Let the treatment received by patient $1 \leqslant i \leqslant n$ be denoted as Z_i, which takes one of three values: 0 (placebo), 1 (low dosage), or 2 (high dosage). Let the outcome of patient i be denoted as Y_i, which takes one of two values: 0 (death after 1 year) or 1 (death before 1 year). We will assume no interference effect and use $Y_i(z)$ to denote the potential outcome of patient i with treatment $Z_i = z$, $z = 0, 1, 2$. Let $N_{zy} = \sum_{i=1}^{n} 1_{\{Z_i=z, Y_i=y\}}$ be the number of patients with treatment $z = 0, 1, 2$ and outcome $y = 0, 1$.

(i) Write down the joint distribution of $Z_1, \ldots, Z_n$ for the following two treatment assignment mechanisms:

(a) The treatments are completely randomized with the only constraint that n_0 patients receive the placebo, n_1 patients receive the low dosage of the new drug, and n_2 patients receive the high dosage ($n_0 + n_1 + n_2 = n$).

(b) Each patient receives the three treatments independently with probabilities π_0, π_1, and π_2 ($\pi_0 + \pi_1 + \pi_2 = 1$).

Under the treatment assignment mechanism in (b), derive the conditional distribution of $Z_1, \ldots, Z_n$ given $N_{z\cdot} = \sum_{i=1}^{n} 1_{\{Z_i=z\}}, z = 0, 1, 2$.

(ii) Describe Fisher's exact test in the analysis of 2×2 contigency tables.

(iii) Suppose Statistician A assumes no dosage effect and groups the low dosage and high dosage patients together before analyzing the data. In other words, Statistician A uses the exposure $A_i = 1_{\{Z_i \geqslant 1\}}$. Explain how the data can then be summarized by a 2×2 contigency table. Show that under both treatment assignment mechanisms in part (i), Fisher's exact test for the sharp null hypothesis $H_A : Y_i(0) = Y_i(1) = Y_i(2),\ i = 1, \ldots, n$ is valid in the sense that its p-value P_A satisfies

$$\mathbb{P}(P_A \leqslant \alpha) \leqslant \alpha \text{ for all } 0 \leqslant \alpha \leqslant 1 \text{ under } H_A.$$

(iv) Statistician B is interested in testing dosage effect and discards the placebo patients in her data analysis. She then summarizes her data using another 2×2 contigency table and uses Fisher's exact test to test the sharp null hypothesis $H_B : Y_i(1) = Y_i(2),\ i = 1, \ldots, n$; denote the p-value as P_B. Show that under both treatment assignment mechanisms in part (i), this test is still valid and is nearly independent of the test used by Statistician A in the sense that

$$\mathbb{P}(P_A \leqslant \alpha_1, P_B \leqslant \alpha_2) \leqslant \alpha_1 \alpha_2 \text{ for all } 0 \leqslant \alpha_1, \alpha_2 \leqslant 1 \text{ under } H_A \text{ and } H_B.$$

[*Hint:* Use the law of iterated expectation by conditioning on $A_1, \ldots, A_n$.]

2

Let $\boldsymbol{V} = (V_1, \ldots, V_p)$ be a random vector. Let $\mathcal{G}^u = (\mathcal{V}, \mathcal{E})$ be an undirected graph and $\mathcal{G}^d = (\mathcal{V}, \mathcal{D})$ be an directed acyclic graph with vertex set $\mathcal{V} = \{1, \ldots, p\}$. Let $f(\boldsymbol{v})$ denote the probability density function of $\boldsymbol{V}$ at $\boldsymbol{v}$.

(i) Explain what it means for the distribution of $\boldsymbol{V}$ to

(a) factorize according to $\mathcal{G}^u$;

(b) satisfies the global Markov property with respect to $\mathcal{G}^u$.

Then state the Hammersley-Clifford theorem.

(ii) Suppose $\boldsymbol{V} = (V_1, \ldots, V_p)$ follow the multivariate normal distribution with mean $\boldsymbol{\mu}$ and a positive definite covariance matrix $\boldsymbol{\Sigma}$. Recall that the probability density function of $\boldsymbol{V}$ at $\boldsymbol{v} = (v_1, \ldots, v_p)$ is given by

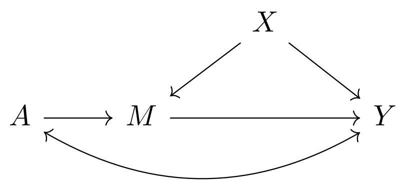
$$f(\boldsymbol{v}) = (2\pi)^{-p/2} |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{v} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{v} - \boldsymbol{\mu})\right\}.$$

For any $j, k \in \{1, \ldots, p\}$, $j \neq k$, show that $V_j \perp\!\!\!\perp V_k \mid \boldsymbol{V}_{\{1,\ldots,p\} \setminus \{j,k\}}$ if and only if $(\boldsymbol{\Sigma}^{-1})_{jk} = 0$, where $(\boldsymbol{\Sigma}^{-1})_{jk}$ is the (j, k)-entry of $\boldsymbol{\Sigma}^{-1}$.

(iii) Explain what it means for the distribution of $\boldsymbol{V}$ to factorize according to $\mathcal{G}^d$. Given $\mathcal{G}^d$, find the undirected graph $\mathcal{G}^u$ with the fewest possible edges such that whenever the distribution of $\boldsymbol{V}$ factorizes according to $\mathcal{G}^d$, it must also factorize according to $\mathcal{G}^u$.

3

Consider the causal model represented by the following acyclic directed mixed graph (ADMG).



(i) Use nonparametric structural equations to define the basic potential outcomes in this model, then describe the conditional independence relations between them.

(ii) Define what it means for a vertex to be fixable in a causal ADMG. Is A fixable in the above graph?

(iii) Which of the following conditional independences are implied by the causal graph? Justify your answer using the m-separation criterion.

(a) $A \perp\!\!\!\perp X \mid Y$.

(b) $M \perp\!\!\!\perp Y(m) \mid X$;

(c) $A \perp\!\!\!\perp Y(a) \mid X$;

(d) $M \perp\!\!\!\perp Y(m) \mid A, X$;

(e) $A \perp\!\!\!\perp M(a) \mid X$.

(iv) Assuming all random variables are discrete, derive an identification formula for the probability distribution of $Y(a)$. [*Hint:* Extend the front-door formula.]

4

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 $a = 0, 1$. The following assumptions are made in this question: (1) consistency of potential outcomes; (2) no unmeasured confounders: $A \perp\!\!\!\perp Y(a) \mid X$, $a = 0, 1$; (3) positivity: $0 < \pi(X) < 1$ where $\pi(X) = \mathbb{P}(A = 1 \mid X)$.

Consider the expected conditional covariance between A and Y:

$$\beta = \mathbb{E}\{\operatorname{Cov}(A, Y \mid X)\}.$$

(i) Show that β identifies a weighted average treatment effect in the sense that there exists a function $w(X)$ such that

$$\beta = \mathbb{E}[w(X)\{Y(1) - Y(0)\}].$$

Find an expression of $w(X)$.

(ii) Assuming X is discrete, show that the influence function for the statistical functional β is given by

$$\psi(X, A, Y) = \{A - \pi(X)\}\{Y - \mu(X)\} - \beta, \tag{1}$$

where $\mu(X) = \mathbb{E}(Y \mid X)$.

[You may use any results given in the lectures. 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) Suppose we have an i.i.d. sample (X_i, A_i, Y_i), $i = 1, \ldots, n$ and two estimators $\hat{\pi}_m(X)$ and $\hat{\mu}_m(X)$ obtained from an independent, external dataset with sample size m. The plug-in estimator of β is given by

$$\hat{\beta} = \frac{1}{n}\sum_{i=1}^{n}\{A_i - \hat{\pi}_m(X_i)\}\{Y_i - \hat{\mu}_m(X_i)\}.$$

Suppose Y has bounded support. Suppose as $n, m \to \infty$, $\operatorname{MSE}(\hat{\pi}_m) \to 0$, $\operatorname{MSE}(\hat{\mu}_m) \to 0$, and $\sqrt{n}\operatorname{MSE}(\hat{\pi}_m) \cdot \operatorname{MSE}(\hat{\mu}_m) \to 0$, where

$$\operatorname{MSE}(\hat{\mu}_m) = \mathbb{E}[\{\hat{\mu}_m(X) - \mu(X)\}^2] \text{ and } \operatorname{MSE}(\hat{\pi}_m) = \mathbb{E}[\{\hat{\pi}_m(X) - \pi(X)\}^2].$$

Show that

$$\sqrt{n}(\hat{\beta} - \beta) \to \mathrm{N}\big(0, \operatorname{Var}(\psi(X, A, Y))\big) \text{ in distribution.}$$

END OF PAPER