MAMA/224, NST3AS/224, MAAS/224

MAT3

MATHEMATICAL TRIPOS Part III

Friday 6 June 2025 9:00 am to 11:00 am

PAPER 224

INFORMATION THEORY

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

(*a*) State the log-sum inequality.

(*b*) Prove that the relative entropy $D(P\|Q)$ between two probability mass functions P, Q on the same finite alphabet A is jointly convex in (P, Q).

(*c*) Recall the notation $D_e(P\|Q) = (\log_e 2)D(P\|Q)$. Let P, Q be two arbitrary probability mass functions on the same finite alphabet A. Show that, for any function $g : A \to \mathbb{R}$

$$D_e(P\|Q) \geqslant \mathbb{E}[g(X)] - \log_e \mathbb{E}[e^{g(Y)}],$$

where X and Y have probability mass functions P and Q, respectively.

(*d*) Show that, if $D(P\|Q) < \infty$ then, in fact,

$$D_e(P\|Q) = \sup_g \Big\{ \mathbb{E}[g(X)] - \log_e \mathbb{E}[e^g(Y)] \Big\},$$

where the supremum is over all functions $g : A \to \mathbb{R}$.

2

(*a*) State Sanov's theorem, taking care to include all the necessary definitions.

(*b*) Suppose $\{X_n\}$ is a sequence of independent and identically distributed (i.i.d.) random variables with values in a finite alphabet A and with common probability mass function Q. Let $P \neq Q$ be another probability mass function on A, and suppose both P and Q have full support.

Show that the probability $\mathbb{P}(P^n(X_1^n) \geqslant Q^n(X_1^n))$ decays exponentially with n, i.e.,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(P^n(X_1^n) \geqslant Q^n(X_1^n)) \leqslant -C(P,Q),$$

and identify the exponent $C(P,Q)$ as the solution to an optimization problem involving relative entropies. Explain why $C(P,Q)$ is strictly positive.

(*c*) Let $\{P_\theta\ ;\ \theta \in \Theta\}$, be a finite parametric family of distinct probability mass functions P_θ of full support on a finite alphabet A, with $\Theta = \{\theta_1, \theta_2, \ldots, \theta_k\}$. Suppose $\{X_n\}$ are i.i.d. with common probability mass function P_{θ^*} for some $\theta^* \in \Theta$, and let

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} P_\theta^n(X_1^n),$$

denote the maximum likelihood estimate, with ties broken arbitrarily.

Show that the "error probability" $\mathbb{P}(\hat{\theta}_n \neq \theta^*)$ decays exponentially to zero, i.e.,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}(\hat{\theta}_n \neq \theta^*) \leqslant -C^*(\Theta).$$

Give a formula for $C^*(\Theta)$ again in terms of relative entropies, and explain why it is strictly positive.

3

Let X be a random variable with probability mass function (PMF) P on the alphabet $A = \{0, 1, \ldots, m-1\}$. Let $d(x, y) = \mathbb{I}_{\{x \neq y\}}$ denote the *Hamming distance* between $x, y \in A$. For $d \in (0, 1)$, we will consider the minimisation problem

$$R(d) = \inf_{(X,Y):\mathbb{E}[d(X,Y)] \leqslant d} I(X;Y), \tag{1}$$

where the infimum is over all pairs of random variables (X, Y), such that both X and Y take values in A, X has PMF P, and $\mathbb{E}[d(X, Y)] \leqslant d$.

(*a*) For $d \in (0, 1)$, let

$$\phi(d) = \sup \left\{ H(Q) \ : \ \text{PMFs } Q \text{ on } A \text{ s.t. } \sum_{i=1}^{m-1} Q(i) \leqslant d \right\}.$$

Show that $\phi(d)$ is a concave function of d.

(*b*) For $d \in (0, 1)$, let (X, Y) denote any pair of random variables satisfying the constraints in the minimisation in (1). Show that

$$I(X;Y) \geqslant H(X) - \phi(d).$$

Carefully justify all your steps. *Hint.* It may be helpful to consider the values $d_y = \mathbb{E}[d(X, Y)|Y = y]$ for all $y \in A$.

(*c*) Show that, if X is uniformly distributed on A, then $R(d) = \log m - \phi(d)$. *Hint.* Create a pair (X, Y) by letting Y be uniformly distributed on A and then selecting an appropriate conditional distribution for X given Y.

4

(*a*) State both the direct and converse parts of the codes-distributions correspondence.

(*b*) Using the result in (*a*) show that, for any random variable X with values in a finite alphabet A and any prefix-free code C on A with length function L:

$$\mathbb{E}[L(X)] \geqslant H(X).$$

(*c*) Consider the collection of all, *not* necessarily prefix-free, variable-rate codes C on a finite alphabet A. These codes are simply invertible maps C from A to the set B^* of all finite-length binary strings, including the empty set:

$$B^* = \{\varnothing, 0, 1, 00, 01, 10, 11, 000, 001, \ldots\}.$$

(*i*) Suppose X is a random variable with values in A. For $R \geqslant 0$, consider the problem of determining the code that achieves the optimal "excess-rate probability"

$$\inf_C \mathbb{P}(L(X) \geqslant R),$$

where the infimum is over all variable-rate codes C with corresponding length function L.

Argue that, regardless of the value of R, this infimum is achieved by a code (C^*, L^*) that operates as follows: First all $x \in A$ are ordered by decreasing probability, and then each x is assigned a binary description $C^*(x)$ in lexicographical order, so that the k^{th} most likely x has $L^*(x) = \lfloor \log k \rfloor$.

(*ii*) Suppose X has probability mass function P on $A = \{1, 2, \ldots, m\}$, such that $P(1) \geqslant P(2) \geqslant \cdots \geqslant P(m)$. Show that, for all $x \in A$:

$$L^*(x) \leqslant -\log P(x).$$

(*iii*) Suppose X has probability mass function P on an arbitrary finite alphabet A, where P does not necessarily have ordered probabilities. Show that, for any R,

$$\mathbb{P}(L^*(X) \geqslant R) \leqslant \mathbb{P}(-\log P(X) \geqslant R),$$

and that:

$$\mathbb{E}[L^*(X)] \leqslant H(X).$$

Compare this last bound with the result you proved in part (*b*).

END OF PAPER