UNIVERSITY OF CAMBRIDGE

MAMA/224, NST3AS/224, MAAS/224

MAT3

MATHEMATICAL TRIPOS Part III

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 and prove the data processing inequalities for mutual information.

All random variables in the remainder of this question take values in finite subsets of $\mathbb{Z}$.

(b) For any three independent random variables X, Y, Z, show that:

$$H(X+Y+Z) + H(Y) \leqslant H(X+Y) + H(Y+Z).$$

(c) For any two random variables X, Y, the *Ruzsa distance* $d_R(X, Y)$ between X and Y is defined as

$$d_R(X,Y) = H(X' - Y') - \frac{1}{2}H(X') - \frac{1}{2}H(Y'),$$

where X', Y' have the same marginal distributions as X, Y, respectively, and X' is independent of Y'. Using only part (a) and basic properties of the entropy and mutual information (but not the Ruzsa triangle inequality), show that, for any three random variables X, Y, Z,

$$d_R(X, Z) \leqslant d_R(X, Y) + d_R(Y, Z).$$

(d) Show that $d_R(X, -Y) \leqslant 3d_R(X, Y)$ for any X, Y.

(e) Show that for any pair of independent X, Y, the entropies of their sum and their difference satisfy:

$$H(X-Y) + H(X) + H(Y) \leqslant 3H(X+Y).$$

2

(a) State and prove Stein's lemma for hypothesis testing. Define all the terms you use carefully.

(b) State without proof the Neyman-Pearson region and its equivalent form in terms of relative entropy.

(c) Let $\hat{P}_{x_1^n}$ denote the type of a string x_1^n with values in a finite alphabet A, and let P, Q denote two distinct probability mass functions of full support on A. In a hypothesis test between P and Q, consider the decision region

$$B_n = \{x_1^n : D(\hat{P}_{x_1^n} \| P) \leqslant \delta\},$$

for $\delta > 0$. Show that the error probability $e_1^{(n)}$ associated with B_n satisfies

$$e_1^{(n)} \leqslant 2^{-nD(\delta)}, \qquad \text{for all } n \geqslant 1,$$

and identify $D(\delta)$ as an optimization problem in terms of relative entropy.

(d) Show that the error probability $e_2^{(n)}$ associated with B_n satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \log e_2^{(n)} \leqslant -\delta.$$

3

(a) State the log-sum inequality.

(b) State and prove the data processing property of relative entropy.

(c) Let P, Q be two probability mass functions on the same finite alphabet A. Show that

$$D(P\|Q) \leqslant (\log e)\chi^2(P\|Q),$$

where *the* χ^2 *divergence* between P and Q is defined as

$$\chi^2(P\|Q) = \sum_{x\in A} \frac{(P(x) - Q(x))^2}{Q(x)}.$$

Let Q be a probability mass function of full support on a finite alphabet A, and $g : A \to \mathbb{R}$ an arbitrary function.

(d) Recall the notation $D_e(P\|Q) = (\log_e 2)D(P\|Q)$. Show that, for any pair of random variables X, Y with probability mass functions P, Q on A, respectively,

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

(e) By identifying the maximizer explicitly, or otherwise, show that in fact

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

where the supremum is over all random variables X on A.

4

(a) Let X be a random variable on $\{0, 1, 2, \ldots\}$, with mean $\mu \in [0, \infty)$. Show that

$$H(X) \leqslant (1+\mu)h\Big(\frac{1}{1+\mu}\Big),$$

where $h(p) = -p\log p - (1-p)\log(1-p)$ denotes the binary entropy function.

Hint. You may find it useful to consider the random variable Y with probability mass function Q given by $Q(k) = (\frac{1}{1+\mu})(\frac{\mu}{1+\mu})^k$, for $k \geqslant 0$.

Let X be an arbitrary discrete random variable with values in a finite alphabet A.

(b) Let (C, L) be a one-to-one (not necessarily prefix-free) code. That is, C is an injective map from A to the set of all finite-length binary strings

$$\{\emptyset, 0, 1, 00, 01, 10, 11, 000, 001, ...\},$$

including the empty string $\emptyset$ of length zero.

Explain the existence of a code (C^*, L^*) with the property that the ith most likely $x \in A$ has $L^*(x) = \lfloor \log i \rfloor$.

(c) Show that $\mathbb{E}[L^*(X)] \leqslant H(X)$ and that $H(X|L^*(X)) \leqslant \mathbb{E}[L^*(X)]$.

(d) Using the results of the previous three parts, show that

$$\mathbb{E}[L^*(X)] \geqslant H(X) - \log[H(X)+1] - \log e.$$

END OF PAPER