UNIVERSITY OF CAMBRIDGE

MATHEMATICAL TRIPOS Part III

Tuesday, 22 June, 2021 12:00 pm to 2:00 pm

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 Let P, Q be two probability mass functions on the same finite alphabet A.

(a) State Stein's lemma for a hypothesis test between P and Q.

(b) State the Neyman-Pearson lemma for the same hypothesis test as in part (a), and write down the alternative form of the Neyman-Pearson region in terms of relative entropy.

(c) Give a proof of the direct part of Stein's lemma using the Neyman-Pearson region instead of the decision region based on likelihood ratio-typical strings. Specify the value of the threshold you need for the Neyman-Pearson region.

(d) Prove the converse part of Stein's lemma.

2 In your proofs of the following three inequalities, justify each step in your arguments. All random variables are assumed to take values in finite alphabets.

(a) Let $\{X_n\}$ be a sequence of independent, discrete random variables, and let Z be another discrete random variable. Show that:

$$H(Z) \geqslant \sum_{i=1}^{\infty} I(X_i; Z).$$

(b) Let $X_1, X_2, \ldots, X_n$ be arbitrary discrete random variables. Prove that

$$H(X_1^n) \leqslant \frac{1}{n-1} \sum_{i=1}^{n} H(X_1^{i-1}, X_{i+1}^n),$$

where, for $j \geqslant i$, X_i^j denotes the block of random variables $(X_i, \ldots, X_j)$, while for $j < i$ X_i^j can be trivially assumed to be the "empty" random variables $X_i^j = 0$ with probability one.

(c) Now let $X_1^n = (X_1, X_2, \ldots, X_n)$ be independent random variables with values in a finite alphabet A, write P_i for the probability mass function (PMF) of each X_i, $i = 1, 2, \ldots, n$, and let $P = P_1 \times P_2 \times \cdots \times P_n$ denote their joint PMF. Let Y_1^n be arbitrary random variables with values in A with joint PMF Q. Write $P^{(i)}$ for the PMF of (X_1^{i-1}, X_{i+1}^n) and $Q^{(i)}$ for the PMF of (Y_1^{i-1}, Y_{i+1}^n), for each $i = 1, 2, \ldots, n$. Using part (b) or otherwise, prove that:

$$D(Q\|P) \leqslant \sum_{i=1}^{n} \Big(D(Q\|P) - D(Q^{(i)}\|P^{(i)}) \Big).$$

[*Hint: Expand* $D(Q\|P)$ *and use part* (b) *on* $H(Y_1^n)$.]

3

(a) State and prove the Pythagorean identity for relative entropy.

(b) Let E be a closed, convex set of probability mass functions (PMFs) on a finite alphabet A. Let P be a PMF of full support on A, and suppose that $Q^* \in E$ achieves the infimum, $\inf_{Q\in E} D(P\|Q)$. Here you will show that:

$$D(P'\|Q') + D(P'\|P) \geqslant D(P'\|Q^*), \qquad \text{for all } P' \text{ and all } Q' \in E. \tag{1}$$

i. Show that for any $Q' \in E$:

$$\sum_{a\in A} P(a)\Big[1 - \frac{Q'(a)}{Q^*(a)}\Big] \geqslant 0.$$

[*Hint: Use* $Q_t := (1-t)Q^* + tQ'$, *for* $0 \leqslant t \leqslant 1$.]

ii. Show that for any $Q' \in E$ and any P'

$$\sum_{a\in A'} P'(a)\Big[1 - \frac{P(a)Q'(a)}{P'(a)Q^*(a)}\Big] \geqslant 0,$$

where $A' = \{a \in A \,:\, P'(a) > 0\}$ denotes the support of P'.

iii. Prove (1).

4

(a) State Kraft's inequality.

(b) State and prove the competitive optimality property of the Shannon code.

Suppose $C : A \to B^*$ is a one-to-one code on a finite alphabet A, that is, C is an injective map from A to the set of all finite-length binary sequences

$$B^* := \{\lambda\} \cup \left[\bigcup_{n \geqslant 1} \{0,1\}^n \right],$$

including the empty string λ of length zero. Let $L : A \to \{0, 1, \ldots\}$ denote the length function of C.

(c) Suppose X is a random variable with probability mass function P on A. Show that there is always a code C with a length function L such that,

$$\mathbb{E}[L(X)] \leqslant H(X) \quad \text{bits},$$

where $H(X)$ is the entropy of X. [*Hint: Explain why you can assume without loss of generality that* $A = \{1, 2, \ldots, m\}$ *and that the probabilities of* $P(i)$ *are non-increasing. Think of an efficient way to assign codewords from* B^* *to the elements* $i \in A$.]

(d) Let X be uniformly distributed on $A = \{1, 2, \ldots, m\}$ for some $m \geqslant 3$. Give an example of a one-to-one code C with length function L such that $\mathbb{E}[L(X)]$ is strictly less than $H(X)$.

END OF PAPER