UNIVERSITY OF CAMBRIDGE

MAMA/224, NST3AS/224, MAAS/224

MAT3

# MATHEMATICAL TRIPOS Part III

Wednesday 5 June 2024 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 and prove the two data processing inequalities for mutual information. State all the properties you use clearly.

($b$) Suppose $X, Y$ are integer-valued, independent and identically distributed random variables. Show that:

$$\frac{H(X+Y)-H(X)}{H(X-Y)-H(X)} \leqslant 2.$$

State all the basic properties you use clearly.

Suppose $\{X_n\}$ is a Markov chain on a finite state space $A$.

($c$) Show that, if the chain is stationary, then, for all $n$,

$$H(X_{n+1}|X_1) \geqslant H(X_n|X_1).$$

($d$) Show that:

$$I(X_1;X_3) + I(X_2;X_4) \leqslant I(X_1;X_4) + I(X_2;X_3).$$

State all the basic properties you use clearly.

**2**

($a$) Let $Y$ be a discrete random variable with probability mass function $Q$ on a finite alphabet $A$. For any $\alpha \in (0,1)\cup(1,\infty)$, the *Rényi entropy of order* $\alpha$ *of* $Y$ is defined as:

$$H_\alpha(Y) = \frac{1}{1-\alpha} \log \sum_{y \in A} Q(y)^\alpha.$$

Show that: $\lim_{\alpha \to 1} H_\alpha(Y) = H(Y)$.

Let $X_1^n = (X_1, X_2, \ldots, X_n)$ be a vector of $n$ discrete random variables with values in a finite alphabet $A$ and with joint probability mass function $P_n$ on $A^n$.

($b$) State the codes-distributions correspondence and show that, for any prefix-free code $(C_n, L_n)$, we have $\mathbb{E}[L_n(X_1^n)] \geqslant H(X_1^n)$.

($c$) Now suppose that instead of minimising the expected description length we wish to minimise the exponential moment $\mathbb{E}\big[2^{\rho L_n(X_1^n)}\big]$ for some $\rho > 0$. Show that for any prefix-free code $(C_n, L_n)$, we have,

$$\frac{1}{\rho} \log \mathbb{E}\big[2^{\rho L_n(X_1^n)}\big] \geqslant H_\alpha(X_1^n),$$

for $\alpha = \frac{1}{1+\rho}$. *Hint.* You may find it useful to employ Jensen's inequality for the convex map $x \mapsto x^{-\rho}$, $x > 0$.

($d$) Use ($a$) and ($c$) to give an alternative proof of ($b$).

**3**

($a$) State and prove the limiting version of Sanov's theorem in the following form:

$$\text{“}\lim_{n\to\infty} -\frac{1}{n} \log Q^n(\hat{P}_n \in E) = D(P^* \| Q).\text{”}$$

Give all the necessary definitions and assumptions for this statement.

($b$) Let $\{X_n\}$ be independent and identically distributed random variables with probability mass function $Q$ that has full support on a finite alphabet $A$. Let $f : A \to \mathbb{R}$ have mean $\mu = \mathbb{E}[f(X_1)]$. Use the result of ($a$) to prove the weak law of large numbers in this case: For any $\epsilon > 0$:

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mu\Big| \geqslant \epsilon\Big) \to 0 \qquad \text{as } n \to \infty.$$

**4**

($a$) Roughly speaking, the "error exponents for fixed-rate compression" theorem says that a memoryless source can be compressed to any rate above the entropy with error probability decaying exponentially with some exponent $D^* > 0$ that depends on the rate and the source distribution.

State this theorem precisely and prove its direct part.

Let $Q$ be a probability mass function with entropy $H(Q) < \log |A|$ on a finite alphabet $A$. For each $\alpha \in [0,1]$ define a new random variable $X_\alpha$ with probability mass function $Q_\alpha$ on $A$:

$$Q_\alpha(x) = \frac{Q(x)^\alpha}{\sum_{y \in A} Q(y)^\alpha}, \quad x \in A.$$

($b$) Show that, for any $\alpha \in (0,1)$:

$$\frac{d}{d\alpha} H(X_\alpha) = -\alpha(\log_e 2)\mathrm{Var}(\log Q(X_\alpha)).$$

($c$) Show that, for any $H(Q) < R < \log |A|$, the exponent $D^*$ in part ($a$) equals $D^* = D(Q_{\alpha^*} \| Q)$ where $\alpha^*$ is the unique $\alpha \in (0,1)$ that achieves $H(Q_\alpha) = R$.

**END OF PAPER**