UNIVERSITY OF CAMBRIDGE

M. PHIL. IN STATISTICAL SCIENCE

Friday 1 June 2007 9.00 to 11.00

INFORMATION AND CODING

*Attempt **THREE** questions.*

*There are **FOUR** questions in total.*

The questions carry equal weight.

STATIONERY REQUIREMENTS
Cover sheet
Treasury Tag
Script 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 an alphabet with m letters each of which appears with probability $1/m$. A binary Huffman code is used to encode the letters, in order to minimise the expected codeword-length $(s_1 + \ldots + s_m)/m$ where s_i is the length of the codeword assigned to letter i. Set $s = \max[s_i : 1 \leqslant i \leqslant m]$, and let n_ℓ be the number of codewords of length ℓ.

(a) Show that $2 \leqslant n_s \leqslant m$.

(b) For what values of m is $n_s = m$?

(c) Determine s in terms of m.

[Hint: *You may find it useful to write* $m = a2^k$ *where* $1 \leqslant a < 2$.]

(d) Prove that $n_{s-1} + n_s = m$, i.e. any two codeword-lengths differ by at most 1.

(e) Determine n_{s-1} and n_s.

(f) Describe the codeword-lengths for an idealised model of English (with $m = 27$).

2 Consider an information source emitting a sequence of letters (U_n) which are independent identically distributed random variables taking values $1, \ldots, m$ with probabilities $p_1, \ldots, p_m$. Let $u^{(n)} = (u_1, \ldots, u_n)$ denote a sample string of length n from the source. Given $0 < \epsilon < 1$, let $M(n, \epsilon)$ denote the minimal size of a set of strings $u^{(n)}$ of total probability at least $1 - \epsilon$. Show the existence of the limit

$$\lim_{n \to \infty} \frac{1}{n} \log_2 M(n, \epsilon)$$

and determine its value. Comment on the significance of this result for coding theory.

3 State and prove the Hamming and Gilbert–Varshamov bounds for codes. State and prove the corresponding asymptotic bounds.

4 Define a cyclic code of length N.

Show how codewords can be identified with polynomials in such a way that cyclic codes correspond to ideals in the polynomial ring with a suitably chosen multiplication rule.

Prove that any cyclic code $\mathcal{X}$ has a unique *generator*, i.e. a polynomial $c(X)$ of minimum degree, such that the code consists of the multiples of this polynomial. Prove that the rank of the code equals $N - \deg c(X)$, and show that $c(X)$ divides $X^N + 1$. Describe all cyclic codes of length 16.

A *check polynomial* $h(X)$ of a cyclic code $\mathcal{X}$ of length N is defined by the condition: $a(X) \in \mathcal{X}$ if and only if $a(X)h(X) = 0 \mod (1+X^N)$. How is the check polynomial related to the generator of $\mathcal{X}$? Given $h(X)$, construct the parity-check matrix and interpret the cosets $\mathcal{X} + y$ of $\mathcal{X}$. Justify your answers.

Find the generators and the check polynomials of the repetition and parity-check codes. Find the generator and the check polynomial of Hamming's code of length 7.

END OF PAPER