MATHEMATICAL TRIPOS PART II (2004–05)

Coding and Cryptography - Sample Tripos Questions **T.A. Fisher**

Note: This course is an extension of the Part IIA Coding and Cryptography course. So there are plenty of past tripos questions available. (See over for some comments on these.) This sheet gives some sample tripos style questions on the topics *not* covered in the old course

1) What does it mean to say that $f : \Sigma_1 \to \Sigma_2^*$ is a decipherable code. State the Kraft inequality and prove that it is satisfied by decipherable codes. Show that the expected word length $E(S)$ of a binary code satisfies $E(S) \geq H(p_1, \ldots, p_m)/\log a$ where $p_1, \ldots, p_m$ are the source probabilities and $a = |\Sigma_2|$. Given an example in the case $m = 5$ and $a = 2$ to show that this lower bound may be attained.
[*You may quote Gibbs' inequality without proof.*]

2) A binary Huffman code is used for encoding letters $1, 2, \ldots, m$ emitted with respective probabilities $p_1 \leq \ldots \leq p_m$. Prove that, if $p_1 < 1/3$, all codewords have length at least 2. Prove that, if $p_1 > 2/5$, letter 1 has a codeword of length one. Which of these codes is a binary Huffman code:
(a) 0,10,110,111,
(b) 00,01,10,11.
Find a probability distribution for which both codes are optimal.

3) Define Huffman's encoding rule for binary codes. Calculate the codeword lengths for the symbol-probabilities $\frac{1}{5}$, $\frac{1}{5}$, $\frac{1}{6}$, $\frac{1}{10}$, $\frac{1}{10}$, $\frac{1}{10}$, $\frac{1}{10}$, $\frac{1}{30}$.
Prove, or provide a counterexample to the assertion that if the length of a codeword from a Huffman code equals ℓ then, in the same code, there exists another codeword of length ℓ' such that $|\ell - \ell'| \leq 1$.

4) Define the information rate H of a source. What does it means to say that a source satisfies the Asymptotic Equipartition Property (AEP)? Calculate the information rate of a Bernoulli source, carefully stating any properties of entropy that you use. State Shannon's First Coding Theorem, without proof.

5) Define the entropy $H(X)$ and the mutual information $I(X;Y)$ of random variables X and Y. State and prove Gibbs' inequality. Prove that

$$0 \leq I(X;Y) \leq \min\{H(X), H(Y)\}.$$

6) Let X, Y be random variables taking values in an alphabet Σ of size m. Define the conditional entropy $H(X|Y)$ and quote a formula for it in terms of $H(X)$, $H(Y)$ and $H(X,Y)$. State an upper bound on $H(X)$ and indicate the probability distribution for which this upper bound is attained. Prove that if $p = P(X \neq Y)$ then

$$H(X|Y) \leq H(p) + p\log(m-1).$$

7) Define the information capacity of a discrete memoryless channel. State Shannon's Seconding Coding Theorem. Compute the capacity of the discrete memoryless channel with channel matrix

$$\begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}.$$

Comments on past tripos questions

The past tripos questions for the old IIA Coding and Cryptography course are suitable for this years course, with the following exceptions:

Questions (2003/I/10F(ii)) and (1999/I/10A(ii)) were set as bookwork questions, but this year only the lower bound was covered in lectures. In fact the upper bound is proved using Hamming's bound and Stirling's formula.

Question (2000/I/10A(i)) asks for the definition of the weight enumerator of a linear code C of length n. It is the polynomial $W_C(s,t) = \sum A_j s^j t^{n-j}$ where A_j is the number of codewords of weight j.

Question (1999/II/9A(i)) asks for the definition of a commutative public key system. In lectures we noted that RSA has this property, but did not give a formal definition.