MATHEMATICAL TRIPOS PART II (2005–06)

Coding and Cryptography - Example Sheet 2 of 4 **T.A. Fisher**

Note: These questions are in the order of the course – not in order of difficulty.

16) A Binary Symmetric Channel with error probability $p = \frac{1}{3}$ is used to send codewords 1100, 0110, 0001, 1111 with probabilities $\frac{1}{4}, \frac{1}{2}, \frac{1}{12}, \frac{1}{6}$. How would you decode 1001 using (i) ideal observer decoding, or (ii) maximum likelihood decoding?

17) Suppose we use eight hole tape with the standard paper tape code (i.e. the simple parity check code of length 8) and the probability that an error occurs at a particular place on the tape (i.e. a hole occurs where it should not or fails to occur where it should) is 10^{-4}. A program requires about 10 000 lines of tape (each line containing eight places) using the paper tape code. Using the Poisson approximation, direct calculation (possible with a hand calculator but really no advance on the Poisson method) or otherwise show that the probability that the tape will be accepted as error free by the decoder is less than .04%.

    Suppose now that we use the Hamming scheme (making no use of the last place in each line). Explain why the program requires about 17 500 lines of tape but that any particular line will be correctly decoded with probability about $1 - (21 \times 10^{-8})$ and the probability that the entire program will be correctly decoded is better than 99.6%.

18) Determine the set of integers n for which the repetition code of length n is perfect.

19) We show that, even if $2^n/V(n,e)$ is an integer, no perfect code may exist.
(i) Verify that $2^{90}/V(90,2) = 2^{78}$.
(ii) Suppose that C is a perfect 2-error correcting code of length 90 and size 2^{78}. Explain why we may suppose without loss of generality that $0 \in C$.
(iii) Let C be as in (ii) with $0 \in C$. Consider the set

$$X = \{x \in \mathbb{F}_2^{90} : x_1 = 1,\ x_2 = 1,\ d(0,x) = 3\}.$$

Show that corresponding to each $x \in X$ we can find a unique $c(x) \in C$ such that $d(c(x), x) = 2$.
(iv) Continuing with the argument of (iii) show that $d(c(x), 0) = 5$ and that $c(x)_i = 1$ whenever $x_i = 1$. If $y \in X$ find the number of solutions to the equation $c(x) = c(y)$ with $x \in X$ and, by considering the number of elements of X, obtain a contradiction.
(v) Conclude that there is no perfect $[90, 2^{78}]$-code.

20) (i) Construct a $[7, 8, 4]$-code from Hamming's code.
(ii) Prove that $A(n,d) \le 2A(n-1,d)$.
(iii) Prove that if d is even then $A(n-1,d-1) = A(n,d)$.
(iv) Hence compute $A(6,4)$.

21) Let C be an $[n,m,d]$-code. Show that $m(m-1)d \le \sum\sum d(c_i,c_j) \le \frac{1}{2}nm^2$ where the sum is over all codewords c_i and c_j of C. Use this to give an upper bound on $A(n,d)$ in the case $n < 2d$.

22) Let $0 < \delta < 1/2$ and write

$$\alpha(\delta) = \limsup_{n\to\infty} \frac{\log A(n, \lfloor n\delta \rfloor)}{n}.$$

Use the Hamming and GSV bounds to show that $1 - H(\delta) \le \alpha(\delta) \le 1 - H(\delta/2)$.

23) Let $X_1, X_2, \ldots$ be a Bernoulli source with letters drawn from an alphabet Σ. Let $f_n : \Sigma^n \to \{0,1\}^*$ be an optimal binary code for $(X_1, \ldots, X_n)$ with word length random variable S_n.
(i) Use the noiseless coding theorem to show that $\frac{1}{n}E(S_n) \to H(X_1)$ as $n \to \infty$.
(ii) Let $\varepsilon > 0$. By (i) there exists $N \ge 1$ with $E(S_N) < N(H(X_1)+\varepsilon)$. By considering the sets

$$A_n = \{\, x \in \Sigma^n \mid f_N^*(x) \text{ has length } \le n(H(X_1) + 2\varepsilon) \,\}$$

for n a multiple of N, show that the source is reliably encodable at rate $H(X_1) + 2\varepsilon$. What does this tell you about the information rate of the source?

24) (i) Show that $H(X|Y) \ge 0$ with equality if and only if X is a function of Y.
(ii) Give an example where $H(X|Y = y) > H(X)$, even though $H(X|Y) \le H(X)$.

25) Players A and B play a (best of) 5 set tennis match. Let X be the number of sets won by A, and let Y be the total number of sets played. Assuming that the players are equally matched and the outcome of each set is independent, compute the conditional entropies $H(X|Y)$, $H(Y|X)$ and the mutual information $I(X;Y)$.

26) Show that $H(X,Z|Y) \le H(X|Y) + H(Z|Y)$ and deduce that $H(X|Y,Z) \le H(X|Y)$. Can you find the condition for equality?

27) Consider two DMC's of capacity C_1 and C_2 with each having input alphabet Σ_1 and output alphabet Σ_2. Connecting in parallel gives the product channel with input alphabet $\Sigma_1 \times \Sigma_1$, output alphabet $\Sigma_2 \times \Sigma_2$, and channel probabilities given by

$$P(y_1 y_2 \text{ received} \,|\, x_1 x_2 \text{ sent}) = P(y_1 \text{ received} \,|\, x_1 \text{ sent}) P(y_2 \text{ received} \,|\, x_2 \text{ sent}).$$

Show that the product channel has capacity $C = C_1 + C_2$.

28) Show that the capacity of the DMC with channel matrix

$$\begin{pmatrix} 1-\alpha-\beta & \alpha & \beta \\ \alpha & 1-\alpha-\beta & \beta \end{pmatrix}$$

is $C = (1-\beta)(1 - \log(1-\beta)) + (1-\alpha-\beta)\log(1-\alpha-\beta) + \alpha \log \alpha$.

29) Codewords 00 and 11 are sent with equal probability through a BSC with error probability p. Compute the mutual information between the codeword sent and the first digit received as output. Show that the extra mutual information to accrue on receipt of the second digit is $H(2p(1-p)) - H(p)$ bits.

30) For random variables X and Y we define $\Delta(X,Y) = H(X|Y) + H(Y|X)$. Show that Δ satisfies the triangle inequality. Is it a metric?