MATHEMATICAL TRIPOS PART II (2024-2025)

Coding and Cryptography - Example Sheet 2 of 4

1 In a binary symmetric channel (BSC) we usually take the probability p of error to be less than 1/2. Why do we not consider $1 \geq p > 1/2$? What if $p = 1/2$?

2 Suppose we connect two BSCs with error probabilities p and q in series or in parallel. How are the channel matrices related? (Note, in parallel, the answer should be a 4×4 matrix.)

3 A BSC 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?

4 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%.

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

6 If there is a perfect e-error correcting binary code of length n, show that $V(n, e)$ divides 2^n. This condition is not sufficient for such a code to exist. We prove this by establishing the following results.

(i) Verify that $\frac{2^{90}}{V(90,2)} = 2^{78}$.

(ii) Suppose that C is a perfect 2-error correcting binary code of length 90 and size 2^{78}. Explain why we may suppose, without loss of generality, that the zero word $\mathbf{0} \in C$.

(iii) Let C be as in (ii) with $\mathbf{0} \in C$. Consider the set

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

Show that, corresponding to each $\mathbf{x} \in X$, we can find a unique $\mathbf{c}(\mathbf{x}) \in C$ such that $d(\mathbf{c}(\mathbf{x}), \mathbf{x}) = 2$. Show that $d(\mathbf{c}(\mathbf{x}), \mathbf{0}) = 5$.

(iv) Continuing with the argument of (iii), show that $c_i(\mathbf{x}) = 1$ whenever $x_i = 1$. If $\mathbf{y} \in X$, find the number of solutions to the equation $\mathbf{c}(\mathbf{x}) = \mathbf{c}(\mathbf{y})$ with $\mathbf{x} \in X$ and, by considering the number of elements of X, obtain a contradiction.

This result, obtained by Marcel Golay, shows that there is no perfect $(90, 2^{78})$-code. He found another case when $2^n/V(n, e)$ is an integer and there *does* exist an associated perfect code (now called the *Golay code.*)[1].

[1] The deep connections between the Golay code and certain Mathieu groups (a class of sporadic finite simple groups) is beyond the scope of this course. See the great little book *From error correcting codes through sphere packings to simple groups* by Thomas Thompson (Carus Mathematical Monographs, 1983).

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

8 Prove the *Singleton bound* for $A(n, d)$, namely,

$$A(n, d) \leq 2^{n-d+1}.$$

9 (i) Show that $H(X|Y) \geqslant 0$ 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 $H(X|Y) \leq H(X)$.

10 Let $(X_n)_{n \geqslant 1}$ be a source with letters drawn from an alphabet $\mathcal{A}$. Let $N \geqslant 1$ be an integer and put $Y_i = (X_{(i-1)N+1}, X_{(i-1)N+2}, \ldots, X_{iN})$. Show that the information rate of the source $(Y_n)_{n \geqslant 1}$ is N times that for $(X_n)_{n \geqslant 1}$.

11 Show that the binary channel with channel matrix

$$\begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

has capacity $\log 5 - 2$.

12 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

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

Further Problems

13 Let C be an $[n, m, d]$-code. Show that

$$m(m-1)d \leqslant \sum\sum d(\mathbf{c}_i, \mathbf{c}_j) \leqslant \frac{1}{2}nm^2$$

where the sum is over all codewords $\mathbf{c}_i$ and $\mathbf{c}_j$ of C. Use this to give an upper bound on $A(n, d)$ in the case $n < 2d$.

14 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 $I(X; Y)$.

15 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.

. *Comments & corrections should be sent to Rachel Camina (rdc26).*