3COM – Problem Sheet 4

You must submit your answers to this problem sheet on the Canvas page. The deadline is **5pm on Wednesday 20th March 2024**.

The summative questions on this problem sheet will be marked and contribute to your final grade in the course. The formative questions do not need to be submitted: they are designed to give you extra practice, and will also be useful for exam preparation. Note that you won't be able to answer some of the formative questions until the end of Week 10.

The examples classes on Fridays 11am in Nuffield G13 are the primary times to seek support for this problem sheet. Both personalised and generic feedback will be provided as soon as all scripts have been marked (you will be notified of this via a Canvas message).

Summative questions

You will be able to attempt **Questions 1–4** already. You will be able to attempt **Questions 5–7** after the first two lectures in **Week 9.**

1.- Let $S := \{s_1, s_2, s_3, s_4, s_5\}$ be a random source with probability distribution so that $p_1 = 1/2$ and $p_2 = p_3 = p_4 = p_5 = 1/8$. Run the Huffman encoding algorithm to obtain an optimal prefix-free encoding of S. What is the expected length of this encoding? You must justify your answer, including drawing the associated binary tree obtained when running the algorithm.

2.- For each of the following codes C, provide a random source S such that the Huffman encoding f_H of S satisfies $C = C(f_H)$ or, otherwise, justify why C cannot be a Huffman code of any random source. In the former case you should run the Huffman encoding algorithm to justify your answer. (In this question you can assume that we only consider random sources S with probability distribution $\mathbf{p} = (p_1, ..., p_m)$ such that each $p_i > 0$.)

(a) $C = \{0, 11, 100, 110, 101\}$

(b) $C = \{10, 11, 01, 000, 001\}$

(c) $C = \{11, 01, 101, 000, 100\}$

3.- Consider the code $C = \{11111111, 10000011, 01010100, 00101000\}$. How many errors can C detect? and correct?

4.- Consider the repetition code where each bit is transmitted n times where $n \geq 3$ is odd. Show that if the error probability p of the transmission channel satisfies $0 < p < 1/2$, then the probability of erroneous decoding can be bounded by a decreasing function of n. (You may use without proof Chebyschev's inequality: for a random variable X with $\mathrm{Var}(X) < \infty$ we have $\mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$. You may also use without proof the formulae for the expected value and the variance of the binomial distribution.)

5.- Show that $A(5, 3) \leq 4$ (Hint: use $A(4, 3) \leq 2$, Ex. 5.6 from Lecture Notes).

6.- Show that $A(n, 6) \leq 2^{n-5}$ for all $n \geq 6$.

7.- Show that for any integer $n \geq 2$ and $1 \leq d \leq n$ we have $A(n, d) \leq 2A(n - 1, d)$.

Formative questions

1.- Let $S = \{s_1, \ldots, s_m\}$ be a random source with probability distribution $\mathbf{p} = (p_1, \ldots, p_m)$ and $p_i > 0$. Let f_H be its Huffman encoding. Show that $C = C(f_H)$ is a maximal prefix-free code (recall the definition of maximal prefix-free was given in Question 2 in the summative section of Problem Sheet 3).

2.- Prove the triangle inequality for d_H. That is, given $n \in \mathbb{N}$ and any $x, y, z \in \{0,1\}^n$, prove that $d_H(x,y) \le d_H(x,z) + d_H(z,y)$.

3.- The first error-correcting/detecting encoding was introduced by Hamming in 1950 and it is known as the Hamming $[7,4]$-code. Let $n = 7$ and $k = 4$. Given $x = x_1x_2x_3x_4 \in \{0,1\}^4$, we construct $f(x) = y_1 \ldots y_7 \in \{0,1\}^7$ where $y_i = x_i$ if $i \in [4]$ and

$$\begin{aligned} y_5 &= x_1 + x_2 + x_3 \pmod 2 \\ y_6 &= x_1 + x_2 + x_4 \pmod 2 \\ y_7 &= x_1 + x_3 + x_4 \pmod 2 \end{aligned}$$

(a) Write all the codewords of the Hamming $[7,4]$-code C.

(b) Show that $d_H(C) = 3$.

(c) Decode the words 1001010 and 1111111 using a minimum distance decoder.

4.- Show that $A(8,4) \ge 16$.

5.- Prove that $A(3m,d) = 2$ for every $2m < d \le 3m$.

6.- Show that $A(2d,d) \le 4d$ for all $d \in \mathbb{N}$.

7- Let C be a linear code of length $n \ge 2$. Let

$$C^\perp := \{v \in \mathbb{F}_2^n \;:\; v \cdot c = 0, \text{ for every } c \in C\}.$$

(Here, $\cdot$ denotes the usual dot product of vectors. So e.g. $(1,1)\cdot(1,1) = 1+1 = 0$.)

(a) Show that $C^\perp$ is also a linear code.

(b) Suppose that C consists precisely of the zero vector and the all-one vector. What is $C^\perp$?

8.- Consider the linear code C with the following generator matrix:

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

(i) What is the length n, dimension k and size $|C|$ of C?

(ii) Write down a parity check matrix of C.

(iii) What is the value of $d_H(C)$?

(iv) Decode both $y = (111100)$ and $z = (011011)$ using syndrome decoding.

To justify your answers you may quote any results necessary from the course.