3COM – Problem Sheet 3

You must submit your answers to this problem sheet on the Canvas page. The deadline is **5pm on Wednesday 6th 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.

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, 2(a), 3(a) and 4** after the first two lectures in **Week 6.** You will be able to attempt **Questions 2(b) and 3(b)** after all the lectures in **Week 6.** You will be able to attempt **Question 5** after the first two lectures in **Week 7.**

1.- Prove that the following codes are uniquely decipherable or give a string in $\{0,1\}^*$ that can be decoded in at least two ways.

(a) $C = \{01, 001, 0110, 1101\}$,

(b) $C = \{001, 10001, 110, 110001\}$.

2.- A prefix-free code C is *maximal* if for any nonempty $x \in \{0,1\}^* \setminus C$, the code $C \cup \{x\}$ is not prefix-free.

(a) Is $C = \{1, 01, 0011, 0000\}$ a maximal prefix-free code? Justify your answer.

(b) Show that a maximal prefix-free code attains Kraft's inequality; that is

$$\sum_{i=1}^{m} 2^{-\ell_i} = 1 \ .$$

(Hint: assume that $\sum_{i=1}^{m} 2^{-\ell_i} < 1$ and use the proof of McMillan's theorem to get a contradiction.)

3.- Let $C = \{c_1, \ldots, c_m\}$ be a prefix-free code and let $\ell_i := |c_i|$ and $\ell := \max_{1 \le i \le m}\{\ell_i\}$. Consider a random string $x := x_1 \ldots x_\ell$ of length ℓ; that is, for each i, we have $x_i = 0$ with probability 1/2 and $x_i = 1$ with probability 1/2.

(a) Let E_i denote the event that the codeword c_i is a prefix of x. What is the value of $\mathbb{P}(E_i)$?
(Hint: this probability should involve the term ℓ_i.)
To clarify here: In the case when $|c_i| < |x|$, E_i is the event that the codeword c_i is a prefix of x. In the case when $|c_i| = |x|$, E_i is the event that $c_i = x$.

(b) Use (a) to give an alternative proof of McMillan's theorem.

4.- The *product* code of C_1 and C_2 is the code $C_1 \times C_2$ obtained by concatenating any ordered pair of codewords from C_1 and C_2; that is $C_1 \times C_2 = \{c_1c_2 : c_1 \in C_1, c_2 \in C_2\}$. For instance, if $C_1 = \{10, 01\}$ and $C_2 = \{1\}$, then $C_1 \times C_2 = \{101, 011\}$. Is the product code of two prefix-free codes also prefix-free? Is the product of two uniquely decipherable codes also uniquely decipherable? Justify your answer.

5.- Let $S = \{s_1, \ldots, s_m\}$ be a random source with probability distribution $\mathbf{p} = (p_1, \ldots, p_m)$. For each $i \in [m]$, let ℓ_i be the smallest integer such that $2^{-\ell_i} \leq p_i$.

(a) Show that there exists a prefix-free encoding f_{SF} of S with $|f_{SF}(s_i)| = \ell_i$.

(b) Prove that $H(S) \leq \mathbb{E}(\ell(f_{SF})) < H(S) + 1$.

(c) Is f_{SF} an optimal prefix-free code? Justify your answer.

Formative questions

1.- Is $C = \{01, 10, 101\}$ uniquely decipherable? If yes, prove it; if no, give a string in $\{0,1\}^*$ that can be decoded in at least two ways.

2.- In this question we consider r-ary codes (codes over an alphabet $\Sigma_r = \{0, 1, \ldots, r-1\}$). Adapt your argument from the Question 3 in the summative section to prove the following: if C is a prefix-free r-ary code with codewords of length $\ell_1, \ldots, \ell_m$, then $\sum_{i=1}^{m} r^{-\ell_i} \leq 1$.

3.- A piece of real world – Understanding Morse Code

In 1837 Samuel F. B. Morse invented the telegraph and, to use it, developed the Morse code. The Morse code is a way to transmit messages using dots and dashes. In the following table you can find how each letter of the alphabet is encoded with Morse code.

A	· −	U	· · −
B	− · · ·	V	· · · −
C	− · − ·	W	· − −
D	− · ·	X	− · · −
E	·	Y	− · − −
F	· · − ·	Z	− − · ·
G	− − ·		
H	· · · ·		
I	· ·		
J	· − − −		
K	− · −	1	· − − − −
L	· − · ·	2	· · − − −
M	− −	3	· · · − −
N	− ·	4	· · · · −
O	− − −	5	· · · · ·
P	· − − ·	6	− · · · ·
Q	− − · −	7	− − · · ·
R	· − ·	8	− − − · ·
S	· · ·	9	− − − − ·
T	−	0	− − − − −

(a) Clearly, the Morse code (dot/dash) is not uniquely decipherable. For instance, $\cdot - -$ could be decoded as AT or as EM. To turn it into a uniquely decipherable code, a space is placed at the end of each letter. Show that the Morse code as a ternary code (dot/dash/space $\to \cdot / - / \;$), denoted by C_M, is a ternary prefix-free code.

(b) Is the Morse code C_M (dot/dash/space) a maximal prefix-free code? Justify your answer.

(c) Informally argue why the Morse code is an efficient encoding for messages in English.

Each symbol of the Morse code is transmitted using an electric wire that has two states: open (0) and closed (1). The following convention is used:

- a dot ($\cdot$) is encoded by 10.
- a dash ($-$) is encoded by 1110.
- a space between letters () is encoded by 000.

- a space between words is encoded by 0000000.

(d) Encode the message THANKS in binary code (0/1). How many bits do you need?

(e) To save money, telegraphers would have a list of frequent messages with special encodings. For instance, one could assign a non-decodable word to the message THANKS. If sending one bit cost 1p and if in 1860 there were 10 million of THANKS telegrams in the UK every year, what was the annual amount that the Electric Telegraph Company was saving?

(f) Decode the message! `https://goo.gl/kMEkuQ`