Indian Institute of Technology Kharagpur
Information and Coding Theory (MA41024/MA60020)
Assignment - 2, Spring Semester 2021-22

Last date of submission ($\star$-marked questions): March 7, 2022

1. Consider an alphabet of 8 symbols whose probabilities are as follows:

(a)

A	B	C	D	E	F	G	H
1/2	1/4	1/8	1/16	1/32	1/64	1/128	1/128

(b)

A	B	C	D
1/4	1/8	1/2	1/8

What is the entropy of the each of the above symbol set? Construct a uniquely decodable prefix codes for both the symbol sets and explain why those are uniquely decodable and have the prefix property.

2. A.Which of the following codes are
   (a) uniquely decodable?
   (b) instantaneous?

$C1 = \{00, 01, 0\}$
$C2 = \{00, 01, 100, 101, 11\}$
$C3 = \{0, 10, 110, 1110, ...\}$
$C4 = \{0, 00, 000, 0000\}$

3. $\star$ One is given 6 plates of food. It is known that precisely one plate has gone bad. The probability that the i-th plate is bad, is given by $p_i, 1 \leq i \leq 6$ as follows:

$$(p_1, p_2, ..., p_6) = (8/23, 6/23, 4/23, 2/23, 2/23, 1/23).$$

Tasting will determine the bad food. Suppose you taste the food one at a time. Choose the order of tasting to minimize the expected number of tastings required to determine the bad food. Remember, if the first 5 plates pass the test you don't have to taste the last.

   (a) What is the expected number of tastings required?

(b) Which plate should be tasted first?

Now suppose you get smart. For the first sample, you mix some of the food on a fresh plate and sample the mixture. You proceed, mixing and tasting, stopping when the bad food has been determined,

(a) What is the minimum expected number of tastings required to determine the bad food?

(b) What mixture should be tasted first?

(Hint: Use Huffman Coding)

4. Let $S = \begin{bmatrix} a_1, \ldots, a_m \\ p_1, \ldots, p_m \end{bmatrix}$. Now suppose a_is are encoded into strings from a set of D code letters in a uniquely decodable manner. If $m = 6$ and the codeword lengths are $(l_1, l_2, \ldots, l_6) = (1, 1, 2, 3, 2, 3)$. Then find a good lower bound on D.

5. ⋆ Consider the random variable $X = \begin{bmatrix} a_1 & a_2 & a_3 & a_4 & a_5 & a_6 & a_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{bmatrix}$

   (a) Find a binary Huffman code for X.

   (b) Find the expected code length for this encoding.

   (c) Find a ternary Huffman code for X.

6. ⋆ A bin has two biased coins, one with probability of head p and the other with probability of head $1-p$. One of these coins is chosen at random (i.e., with probability 1/2), and is then tossed n times. Let X denote the identity of the coin that is picked, and let Y_1 and Y_2 denote the results of the first two tosses.

   (a) Calculate $I(Y_1; Y_2|X)$.

   (b) Let Y_i denote the resut of i-th coin toss. Calculate $\lim_{n\to\infty} \frac{1}{n} H(X, Y_1, Y_2, \ldots, Y_n)$

7. ⋆ Consider the following method for generating a code for a random variable X which takes on m values $\{1, 2, ..., m\}$ with probabilities $p_1, p_2, ..., p_m$ .Assume that the probabilities are ordered so that $p_1 \geq p2 \geq \ \geq p_m$. Define

$$F_i = \sum_{k=1}^{i-1} p_k$$

   the sum of the probabilities of all symbols less than i . Then the code-word for i is the number $F_i \in [0, 1]$ rounded off to l_i bits, where $l_i = \lceil \log \frac{1}{p_i} \rceil$

   (a) Show that the code constructed by this process is a prefix code and the average length satisfies

$$H(X) \leq L < H(X) + 1$$

   (b) Construct the code for the probability distribution $(0.5, 0.25, 0.125, 0.125)$.

8. $\star$ Consider the following table:

Symbol (x)	$p(x)$	$q(x)$	C_1	C_2
a_1	1/2	1/2	0	0
a_2	1/4	1/8	10	100
a_3	1/8	1/8	110	101
a_4	1/16	1/8	1110	110
a_5	1/16	1/8	1111	111

where the first column list the source letters, 2nd and 3rd columns are two pmfs for the letters, and last two columns are two codes. Check if the codes C_1 and C_2 are optimal codes corresponding to p and q. Also calculate $D(p\|q)$ and $D(q\|p)$ and note if you can relate the relative entropy with the codes.

9. Consider a source random variable X with three source letters and the probabilities are given by $0.6, 0.3$ and 0.1. Then answer the following:

   What are the lengths of the binary Huffman and Shannon-Fano-Elias codewords for X? What is the smallest value of source code letters D such that the expected Shannon-Fano-Elias codeword length equals the expected Huffman codeword length?

All the Best!!