EEE605 – DATA COMPRESSION HOMEWORK – 1

1)

The random variable X takes values in the alphabet $\{1, 2, 3, 4\}$. The probabilities for the different symbols are

$$P_X(1) = P_X(2) = P_X(3) = P_X(4) = 0.25$$

Calculate $H(X)$.

2)

A suggested binary code for the alphabet $\mathcal{A} = \{1, \ldots, 8\}$ has the code-word lengths $l_1 = 2$, $l_2 = 2$, $l_3 = 3$, $l_4 = 4$, $l_5 = 4$, $l_6 = 5$, $l_7 = 5$ and $l_8 = 6$. Can a prefix code with these lengths be constructed?

3)

A memoryless source has the alphabet $\mathcal{A} = \{x, y, z\}$ and symbol probabilities

$$P(x) = 0.6, \quad P(y) = 0.3, \quad P(z) = 0.1$$

a) What is the entropy of the source?

b) Construct a Huffman code for single symbols from the source and calculate the rate of the code in bits/symbol.

c) Construct a Huffman code for pairs of symbols from the source and calculate the rate of the code in bits/symbol.

4)

Determine whether the following codes are uniquely decodable:

(a) $\{0, 01, 11, 111\}$

(b) $\{0, 01, 110, 111\}$

(c) $\{0, 10, 110, 111\}$

(d) $\{1, 10, 110, 111\}$

5)

Given an alphabet $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$, find the first-order entropy in the following cases:

(a) $P(a_1) = P(a_2) = P(a_3) = P(a_4) = \frac{1}{4}$.

(b) $P(a_1) = \frac{1}{2}$, $P(a_2) = \frac{1}{4}$, $P(a_3) = P(a_4) = \frac{1}{8}$.

(c) $P(a_1) = 0.505$, $P(a_2) = \frac{1}{4}$, $P(a_3) = \frac{1}{8}$, and $P(a_4) = 0.12$.

6)

While the variance of lengths is an important consideration when choosing between two Huffman codes that have the same average lengths, it is not the only consideration. Another consideration is the ability to recover from errors in the channel. In this problem we will explore the effect of error on two equivalent Huffman codes.

a) For the source and Huffman code of the table, encode the Sequence.

$$a_2 \; a_1 \; a_3 \; a_2 \; a_1 \; a_2$$

Suppose there was an error in the channel and the first bit was received as a 0 instead of a 1. Decode the received sequence of bits. How many characters are received in error before the first correctly decoded character?

Huffman code for the original five-letter alphabet.

Letter	Probability	Codeword
a_2	0.4	1
a_1	0.2	01
a_3	0.2	000
a_4	0.1	0010
a_5	0.1	0011

b) Repeat the procedure for the given table.

Minimum variance Huffman code.

Letter	Probability	Codeword
a_1	0.2	10
a_2	0.4	00
a_3	0.2	11
a_4	0.1	010
a_5	0.1	011

c) Repeat parts (a) and (b) with the error in the third bit.

7)

Encode the message [a a r d v a r k] by using adaptive huffman coding, where our alphabet consists of the 26 lowercase letters of the English alphabet.

8)

For an alphabet $\mathcal{A} = \{a_1, a_2, a_3, a_4\}$ with probabilities $P(a_1) = 0.1$, $P(a_2) = 0.3$, $P(a_3) = 0.25$, and $P(a_4) = 0.35$, find a Huffman code

(a) using the first procedure outlined in this chapter, and

(b) using the minimum variance procedure.

Comment on the difference in the Huffman codes.

9)

For an alphabet $\mathcal{A} = \{a_1, a_2, a_3\}$ with probabilities $P(a_1) = 0.7$, $P(a_2) = 0.2$, $P(a_3) = 0.1$, design a 3-bit Tunstall code.

10)

In many communication applications, it is desirable that the number of 1s and 0s transmitted over the channel are about the same. However, if we look at Huffman codes, many of them seem to have many more 1s than 0s or vice versa. Does this mean that Huffman coding will lead to inefficient channel usage? For the Huffman code obtained in Problem 3, find the probability that a 0 will be transmitted over the channel. What does this probability say about the question posed above?

11)

A source emits letters from an alphabet $\mathcal{A} = \{a_1, a_2, a_3, a_4, a_5\}$ with probabilities $P(a_1) = 0.15$, $P(a_2) = 0.04$, $P(a_3) = 0.26$, $P(a_4) = 0.05$, and $P(a_5) = 0.50$.

(a) Calculate the entropy of this source.

(b) Find a Huffman code for this source.

(c) Find the average length of the code in (b) and its redundancy.

12)

For a source with a six-letter alphabet and a probability model $P(a_1) = P(a_3) = P(a_4) = 0.2$, $P(a_5) = 0.25$, $P(a_6) = 0.1$, and $P(a_2) = 0.05$. Generate a ternary code by combining three letters in the first and second steps and two letters in the third step.

13)

A source has symbol probabilities $p(a) = 0.4$, $p(b) = 0.1$, $p(c) = 0.3$, and $p(d) = 0.2$.

a) Find a Huffman code for the source.

b) Design a 4-bit Tunstall code for the source.

14)

Consider the following prefix code:

symbol	a	b	c	d	e	f	g	h	i
code	1010	111	00101	000	01	1011	110	0011	00100

(a) Show that the code is redundant, i.e., satisfies Kraft's inequality with strict inequality.

(b) Modify the code by deleting some bits in the codewords so that the result is a complete prefix code, i.e., satisfies Kraft's inequality with equality. You may not add or change any bits, only delete them.

15)

Let $\{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}, \mathtt{f}, \mathtt{g}, \mathtt{h}\}$ be the alphabet with the probability distribution

symbol	a	b	c	d	e	f	g	h
probability	0.10	0.15	0.15	0.05	0.25	0.10	0.15	0.05

(a) Construct a Huffman code for the above symbol distribution.

(b) Show that the code of (a)-part satisfies Kraft's inequality with equality.