GI12/4C59 – Homework #3 (Due 12am, November 4, 2004)

Aim: To get familiarity with the elements of coding theory. Presentation, clarity, and synthesis of exposition will be taken into account in the assessment of these exercises.

This document is available at `http://www.cs.ucl.ac.uk/staff/M.Pontil/courses/IT-homework3.pdf`

1. [**30 pts**] Consider the code $\{0, 01, 11\}$

   (a) Is it nonsingular?

   (b) Is it uniquely decodable?

   (c) Is it instantaneous?

   Explain your observation.

2. [**40 pts**] Consider the random variable

$$X = \begin{pmatrix} a & b & c & d & e \\ \frac{1}{2} & \frac{7}{32} & \frac{1}{8} & \frac{3}{32} & \frac{1}{16} \end{pmatrix}$$

   (a) Find a binary Huffman code for X.

   (b) Verify that the expected codelength $\bar{L}$ of the above code is consistent with the formula

$$H(X) \leq \bar{L} < H(X) + 1.$$

   (c) Can you modify two values of $p(x)$ so that the above code is still optimal and now has average length equal to the entropy of X? Explain your observation.

   (d) Find a ternary Huffman code for X.

3. [**30 pts**] *Shannon codes and Huffman codes.* Consider a random variable X which takes on four values with probabilities $(\frac{1}{3}, \frac{1}{3}, \frac{1}{4}, \frac{1}{12})$.

   (a) Construct a Huffman code for this random variable.

   (b) Show that there exist two different sets of optimal lengths for the codewords, namely, show that codeword length assignments $(1, 2, 3, 3)$ and $(2, 2, 2, 2)$ are both optimal.

   (c) Conclude that there are optimal codes with codeword lengths for some symbols that exceed the Shannon code length $\lceil \log \frac{1}{p(x)} \rceil$.

GI12/4C59 – Solutions to Homework #3

1. (a) Yes

   (b) Yes, if a sequence start with 1 the next symbol will be 1 and we decode the first two symbols as x_3 and then iterate. If the sequence start with k 0's and the next two symbols are "11" all those zeros are decoded as x_1 and 11 as x_3. Instead, if the next two symbols are "10" we decode the first $k-1$ zeros as x_1, 01 as x_2 and start decoding from the last 1.

   (c) No, 0 is a the prefix of 01.

2. (a) An optimal Huffman code is $(0, 11, 101, 1000, 1001)$.

   (b) $\bar{L} = 2 - \frac{1}{16}$. The inequality follows from the property

$$\log t \leq \lceil \log t \rceil < \log t + 1, \quad t > 0$$

   where, for every real number z, $\lceil z \rceil$ is the smallest integer which is greater than or equal to z.

   (c) If p is the dyadic distribution $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16})$, the above code is still optimal and now $\bar{L} = H(X)$.

   (d) A ternary Huffman code is $(0, 2, 10, 11, 12)$.

3. *Shannon codes and Huffman codes.*

   (a) Applying the Huffman algorithm gives us the following table

Code	Symbol	Probability			
0	1	1/3	1/3	2/3	1
11	2	1/3	1/3	1/3	
101	3	1/4	1/3		
100	4	1/12			

   which gives codeword lengths of 1,2,3,3 for the different codewords.

   (b) Both set of lengths 1,2,3,3 and 2,2,2,2 satisfy the Kraft inequality, and they both achieve the same expected length (2 bits) for the above distribution. Therefore they are both optimal.

   (c) The symbol with probability 1/4 has an Huffman code of length 3, which is greater than $\lceil \log \frac{1}{p} \rceil$. Thus the Huffman code for a particular symbol may be longer than the Shannon code for that symbol. But on the average, the Huffman code cannot be longer than the Shannon code.