ELEC 535 Homework 5

Due date: In class on Friday, February 28, 2003

Instructor: Rudolf Riedi Rice University, Spring 2003

Problem 5.1 (Optimal codes for uniform distributions)

Consider a random variable with m equiprobable outcomes. Recall that the entropy of this information source is $\log_2 m$ bits.

(a) For what value of m does the average codeword length L_m equal the entropy $H = \log_2 m$?

(b) We know that $L < H + 1$ for any probability distribution. The *redundancy* of a code is defined to be $\rho = L - H$. For what value(s) of m, where $2^k \le m \le 2^{k+1}$, is the redundancy of the optimal code maximized? What is the limiting value of this worst case redundancy as $m \to \infty$?

Problem 5.2 (The Sardinas-Patterson test for unique decodability)

A code is not uniquely decodable iff there exists a finite sequence of code symbols which can be resolved in two different ways into sequences of codewords. That is a situation such as

$$
\begin{array}{l}
| \qquad A_1 \qquad\qquad | \qquad A_2 \qquad | \qquad A_3 \qquad \ldots \qquad A_m \quad | \\
-\,-\,-\;\; -\,-\,-\,-\,-\;\; -\,-\;\; -\,-\,-\,-\,-\;\; -\,-\,-\,-\;\; -\,-\,-\,-\;\; -\,-\,-\,-\;\; -\,-\,-\,-\;\; -\,-\,-\,-\;\; - \\
| \qquad B_1 \qquad | \qquad B_2 \qquad | \qquad B_3 \qquad\qquad \ldots \qquad B_n \quad |
\end{array}
$$

must occur where each A_i and each B_i is a codeword. Note that B_1 must be a prefix of A_1 with some resulting "dangling suffix." Each dangling suffix must in turn be either a prefix of a codeword or have another codeword as its prefix, resulting in another dangling suffix. Finally, the last dangling suffix in the sequence must also be a codeword. Thus one can set up a test for unique decodability in the following way:

- Construct a set S of all possible dangling suffixes.
- The code is uniquely decodable iff S contains no codeword.

(a) Suppose the codeword lengths are l_i, $i = 1, 2, \ldots, m$. Find a good upper bound on the number of elements in the set S.

(b) Determine which of the following codes is uniquely decodable:

i. $\{0, 10, 11\}$
ii. $\{0, 01, 11\}$
iii. $\{0, 01, 10\}$
iv. $\{0, 01\}$
v. $\{00, 01, 10, 11\}$
vi. $\{110, 11, 10\}$
vii. $\{110, 11, 100, 00, 10\}$

(c) For each uniquely decodable code above, construct, if possible an infinite encoded sequence with a known starting point, such that it can be resolved into codewords in two different ways. This shows, that while finite code words are uniquely decodable, infinite code words might not be.

(d) Prove that such a sequence cannot arise in a prefix code.

Problem 5.3 (Shannon code)

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

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

the sum of the probabilities of all symbols less than i. Then the codeword 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 prefix free and the average length satisfies

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

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