Homework Set #1

Due 30 September 2008

Problem 1 (Uniquely Decodable/Instantaneous Codes)

Let $L = \sum_{i=1}^{m} p_i l_i^{100}$ be the expected value of the 100^{th} power of the word lengths associated with an encoding of the random variable X. Let $L_1 = \min L$ over all instantaneous codes and let $L_2 = \min L$ over all uniquely decodable codes. What inequality relationship exists between L_1 and L_2? Prove your answer.

Problem 2 (Slackness in Kraft Inequality)

An instantaneous code has word lengths $l_1, l_2, \cdots, l_m$, which satisfy the *strict* inequality

$$\sum_{i=1}^{m} D^{-l_i} < 1.$$

The code alphabet is $\mathcal{D} = \{0, 1, 2, \cdots D-1\}$. Show that there exist arbitrarily large sequence of code symbols which cannot be decoded into sequence of codewords.

Problem 3 (Data Compression)

Find an optimal set of binary codeword **lengths** $l_1, l_2, \ldots$ (minimizing $\sum p_i l_i$) for an instantaneous code for each of the following probability mass functions:

(a) $\mathbf{p} = \left(\frac{10}{41}, \frac{9}{41}, \frac{8}{41}, \frac{7}{41}, \frac{7}{41}\right)$

(b) $\mathbf{p} = \left(\frac{9}{10}, (\frac{9}{10})(\frac{1}{10}), (\frac{9}{10})(\frac{1}{10})^2, (\frac{9}{10})(\frac{1}{10})^3, \ldots\right)$

Problem 4 (Suffix Free Codes)

Consider codes that satisfy the suffix condition, which says that no codeword is a suffix of any other codeword. Assume that the alphabet size of the code is D and let l_i be the length of the i^{th} codeword, $i \in \{1, .., N\}$, where N is the number of codewords.

(a) Show that a suffix condition code is uniquely decodable

(b) Prove that a suffix code satisfies

$$\sum_{i=1}^{N} D^{-l_i} \leq 1$$

(i.e. prove that the Kraft inequality is valid for suffix free codes)

Problem 5 (Huffman Codes)

(a) Find the binary Huffman code for the source with probabilities $(\frac{1}{3}, \frac{1}{5}, \frac{1}{5}, \frac{2}{15}, \frac{2}{15})$.

(b) Use the code designed for the source of (a) to encode the source with probabilities $(\frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})$. What is the average length of this code?

(c) Which of these codes cannot be a Huffman code for any probability distribution?

1. $\{0, 10, 11\}$
2. $\{00, 01, 10, 110\}$
3. $\{01, 10\}$

(d) Consider the source with probabilities $(\frac{1}{16}, \frac{1}{16}, \frac{3}{16}, \frac{5}{16}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8})$.

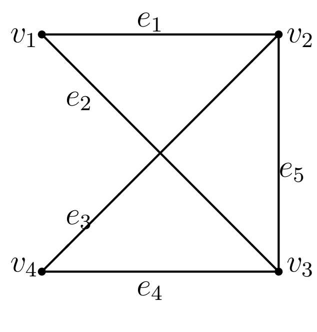
1. Find a 3-ary Huffman code for this source.
2. How many symbols did you merge at each step?

(e) Consider the source with probabilities $(\frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{5}{16}, \frac{1}{4}, \frac{1}{4})$.

1. Find a 3-ary Huffman code for this source.
2. How many symbols did you merge at each step? At what step did you merge two symbols instead of three?
3. Argue that it is better in terms of average length of the code to merge two symbols in the first step and not in the last step (compare the average length of the two codes to answer).

Problem 6 (Random Walk on a Graph)

Consider the following graph. A graph is a set of vertices ($\{v_1, v_2, v_3, v_4\}$ in our example) and a set of edges ($\{e_1, e_2, e_3, e_4, e_5\}$ in our example). A particle starts walking randomly on this graph (i.e. at each vertex v_i the particle picks one of the adjacent edges uniformly at random and walks the edge to the other end). The random walk $\{X_n\}$, $X_n \in \{v_1, v_2, v_3, v_4\}$, is a sequence of vertices of the graph.



(a) Show that this stochastic process is a Markov chain.

(b) Find the transition matrix.

(c) Find the stationary distribution of this Markov chain.

(d) Now consider a general graph with m vertices and consider that each vertex v_i has d_i adjacent edges (so the number of edges of the graph would be $\frac{1}{2}\sum_{i=1}^{m} d_i$). answer to (a), (b), and (c) for this general graph.
(*Hint: To answer (c), Try to guess the stationary distribution using the one you calculated for the simple graph and verify that this distribution is indeed the stationary distribution of the Markov chain*)