CSC 310, Spring 2004 — Assignment #1

Due at **start** of lecture on February 11. Worth 6% of the course grade.

Note that this assignment is to be done by each student individually. You may discuss it in general terms with other students, but the work you hand in should be your own.

Question 1 (15 marks):

a) Explain why the code with codewords $\{0,\ 0010,\ 0001100\}$ is uniquely decodable. A convincing informal explanation is OK.

b) Is the code in part (a) instantaneous? Show why or why not.

c) Show that the code with codewords $\{0,\ 0010,\ 000100\}$ is not uniquely decodable, by giving an example of a string of bits that can be decoded in more than one way.

d) Does the code in part (c) satisfy the Kraft-McMillan inequality?

Question 2 (25 marks): Consider a source with source alphabet $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ in which the symbols probabilities are as follows:

$$p_1 = 0.27,\ p_2 = 0.09,\ p_3 = 0.23,\ p_4 = 0.11,\ p_5 = 0.15,\ p_6 = 0.15$$

a) Compute the entropy of this source.

b) Find a Huffman code for this source. Show your work.

c) Compute the expected codeword length for the Huffman code you found in part (b).

d) Find an optimal instantaneous code for this source that is *not* a Huffman code. (Your code should not just be different from the Huffman code you found in part (b); it must also be different from any other Huffman code that could be obtained by changing the way arbitrary choices are made in the Huffman code algorithm.)

Question 3 (30 marks): Let C be a uniquely-decodable binary code for a source with symbol probabilities $p_1, \ldots, p_I$ in which the codewords for these symbols have lengths $l_1, \ldots, l_I$. Suppose that for some distinct i, j, and k, $l_i = l_j = l_k$. Prove that if C is optimal (ie, has minimal expected codeword length), then $p_i \leq p_j + p_k$. Caution: C is not necessarily a Huffman code; it might not even be instantaneous.

Question 4 (30 marks total): Suppose a source produces independent symbols from the alphabet $\{a_1,\ a_2,\ a_3\}$, with probabilities $p_1 = 0.4999999$, $p_2 = 0.4999999$, and $p_3 = 0.0000002$.

a) Compute the entropy of this source.

b) Find an optimal code for this source, and compute its expected codeword length.

c) Find an optimal code for the second extension of this source (ie, for blocks of two symbols), and compute its expected codeword length, and the expected codeword length divided by two.

d) Prove (without any tedious calculations) that in order to compress to within 1% of the entropy by encoding blocks of size N from this source, N will have to be at least 5.