ECE 564/645 - Digital Communications, Spring 2014
Homework #1
Due: February 12, 2014 (in class)

1. Consider an independent and identically distributed (IID) sequence (X_i), where each X_i is drawn from the alphabet $\mathcal{X} = \{A, B, C\}$ and the probability mass function of each of the X_i is given by:

$$p_{X_i}(x) = \begin{cases} 0.5, & x = A \\ 0.3, & x = B \\ 0.2, & x = C \end{cases} .$$

   (a) Find the entropy $H(X)$ of this source. (You will need your calculator for this.)

   (b) Design a Huffman code that takes blocks of length 2 characters (i.e. $N = 2$) and find its rate (in output bits per input character). Show that the rate falls between easily obtained upper and lower bounds based on the source entropy $H(X)$.

   (c) Design a Shannon-Fano code that takes blocks of length 2 characters (i.e. $N = 2$) and find its rate (in output bits per input character). Show that the rate falls between easily obtained upper and lower bounds based on the source entropy $H(X)$ and your answer to part (b).

   (d) Without finding the Huffman code (this would take far too much time!), find lower and upper bounds to the rate (in output bits per input character) of a Huffman code that takes symbols eight at a time (i.e. $N = 8$). Full credit goes to the tightest bounds.

2. A Huffman code finds the codewords for the optimal lossless code to assign to a given block of N source symbols.

   (a) Show that the following codes could not be the binary strings produced by a Huffman coder for any N for any source distribution where every string to be coded has non-zero probability:

   - $\{11, 001, 000, 100, 101, 111, 0100, 0101, 0110, 0111\}$
   - $\{01, 1001, 1011, 1010, 1000, 1110, 1111, 1101, 1100\}$

   (b) Suppose a source produces an independent and identically distributed (IID) sequence of discrete random random variables. Can there exist a Huffman code on blocks of length $N = 2$ that has smaller rate than a Huffman code on blocks of length $N = 4$ for this source? (Be sure to justify your answer.)