EE5585: HOMEWORK 1

INSTRUCTOR: A. MAZUMDAR

All problems carry equal points

(1) A source emits letters from an alphabet $\mathcal{X} = \{1, 2, 3, 4, 5\}$, with $p(1) = 0.15, p(2) = 0.04, p(3) = 0.26, p(4) = 0.05, p(5) = 0.50$.
  - Calculate the entropy of this source.
  - Find a Huffman code for this source.
  - Find the average length of the Huffman code.

(2) Answer the following short questions.
  - Consider the code $\{0, 01\}$. Is this code uniquely decodable? Why? Is it instantaneous?
  - Suppose $\mathcal{X} = \{0, 1\}$. The random variable (source) X takes value in $\mathcal{X}$, with $\Pr(X = 0) = \frac{3}{4}$ and $\Pr(X = 1) = \frac{1}{4}$. What is the probability that the source produce a sequence 0000011111?

(3) Write the Lempel-Ziv parsing for the file

$$\mathcal{F} = 000001011100010000101000010001110010101010100100.$$

What is the number of bits that you need to write the entire compressed file (with LZ algorithm).

(4) Consider a random variable X that takes on four values with probabilities $(1/3, 1/3, 1/4, 1/12)$.
  - Construct a Huffman code for this random variable.
  - 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.
  - Conclude that there are optimal codes with codeword lengths for some symbols that exceed the Shannon code length $\lceil \log \frac{1}{\log p(a)} \rceil$.

(5) Although the codeword lengths of an optimal variable-length code are complicated functions of the message probabilities $\{p_1, p_2, \ldots, p_m\}$, it can be said that less probable symbols are encoded into longer codewords. Suppose that the message probabilities are given in decreasing order, $p_1 > p_2 \geq \cdots \geq p_m$.
  - Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 > 2/5$, that symbol must be assigned a codeword of length 1.
  - Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 < 1/3$, that symbol must be assigned a codeword of length ≥ 2.