EE5585: HOMEWORK 1

(1) Suppose, $\mathcal{X} = \{A, B, C, D\}$. A source produces i.i.d. X symbols from this source, with $\Pr(X = A) = p_A, \Pr(X = B) = p_B, \Pr(X = C) = p_C, \Pr(X = D) = p_D$. You are given a file $\mathcal{F} = AACDDBBBBBCAABCDAABAADCB$ generated by this source.

    (a) What is your best guess for p_A, p_B, p_C, p_D? Reason.   3

    (b) What is the entropy (in bits) of the probability distribution you guessed?   3

    (c) What is the Huffman code for the probability distribution that you guessed? What is the average number of bits per symbol?   5

    (d) Encode the file $\mathcal{F}$ with the Huffman code you have designed. What is the length of the encoded binary file? What is the average number of bits that have been used for a symbol in this file?   5

(2) Consider the code $\{0, 01\}$. Is this code uniquely decodable? Why? Is it instantaneous?   2

(3) 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?   2

(4) Write the Lempel-Ziv parsing for the file $\mathcal{F}$ of Problem 1. What is the number of bits that you need to write the entire compressed file (with LZ algorithm).   5