ASSIGNMENT 2

Exercise 1 (Bad codes)**.** Which of the following binary codes cannot be a Huffman code for any distribution? Why?

a. 0, 10, 111, 101

b. 00, 010, 011, 10, 110

c. 1, 000, 001, 010, 011

Exercise 2 (Huffman codes)**.** For the distribution $(p_1, ..., p_n)$, where

$$p_1 > p_2 > \cdots > p_n > 0,$$

we have an optimal binary prefix code. Show that

a. If $p_1 > 2/5$ then the corresponding codeword has length 1.

b. If $p_1 < 1/3$ then the corresponding codeword has length at least 2.

Exercise 3 (Huffman code for a wrong source)**.** The purpose of this problem is to see what happens when you design a code for the wrong set of probabilities. Consider a Huffman code that is designed for a symbol source whose probability is given by P. Suppose that we use this code for the source with distribution Q. Find the average number of binary code symbols per source symbol and compare it with the entropy of the source for the following.

1. $P = (0.5, 0.3, 0.2)$, $Q = (0.65, 0.2, 0.15)$
2. $P = (0.5, 0.3, 0.2)$, $Q = (0.15, 0.2, 0.65)$
3. $P = (0.5, 0.3, 0.1, 0.1)$, $Q = (0.3, 0.2, 0.3, 0.2)$

Can the optimal codes for P and Q be the same?

Exercise 4 (Shannon code, divergence)**.** Suppose we wrongly estimate the probability of a source of information, and that we use a Shannon code for a distribution Q whereas the true distribution is P. Show that

$$H(P) + D(P||Q) \leq L(C) \leq H(P) + D(P||Q) + 1\,.$$

So $D(P||Q)$ can be interpreted as the increase in descriptive complexity due to incorrect information.

Exercise 5 (Huffman Codes)**.** The sequence of six independent realizations of source X is encoded symbol-by-symbol using a binary Huffman code. The resulted string is 10110000101. We know that the alphabet of X has five elements and that its distribution is either $\{0.4, 0.3, 0.2, 0.05, 0.05\}$ or $\{0.3, 0.25, 0.2, 0.2, 0.05\}$. Which of them is the distribution of X?

Exercise 6 (Entropy bound)**.** Let $p(x)$ be a probability mass function of random variable X. Prove that

$$\log(1/d) Pr\{p(X) \leq d\} \leq H(X)$$

for any $d \geq 0$. *Hint* – Use Markov's inequality.

Exercise 7 (Entropy)**.** Let X and Y be the outcomes of a pair of dice thrown independently (hence each independently takes on values in $\{1, 2, 3, 4, 5, 6\}$ with equal probabilities). Let $Z = X + Y$ and let $Q = Z \mod 2$. Compute the following entropies: $H(X)$, $H(Y)$, $H(Z)$, $H(Q)$.

Exercise 8 (Entropy)**.** Let X be a random variable taking values in M points $a_1, \ldots, a_M$ and let $p_X(a_M) = \alpha$. Show that

$$H(X) = -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) + (1 - \alpha) H(Y)$$

where Y is a random variable taking values in $M-1$ points $a_1, \ldots, a_{M-1}$ with probabilities $P_Y(a_j) = P_X(a_j)/(1 - \alpha)$ for $1 \leq j \leq M - 1$. Show that

$$H(X) \leq -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) + (1 - \alpha) \log(M - 1)$$

and determine the condition for equality.