ELEN 161
A. I. Zecevic

Homework 1

1. Consider a source that can send ten different messages, $\{x_1, x_2, \ldots, x_{10}\}$, whose probabilities are:

$$p(x_1) = 0.35$$
$$p(x_2) = 0.20$$
$$p(x_3) = 0.15$$
$$p(x_4) = 0.10$$
$$p(x_5) = 0.05$$
$$p(x_6) = 0.05$$
$$p(x_7) = 0.04$$
$$p(x_8) = 0.03$$
$$p(x_9) = 0.02$$
$$p(x_{10}) = 0.01$$

(a) Use the Huffman algorithm to produce an optimal encoding of these messages (show the labeled graph explicitly).

(b) Compute the average number of bits per message that corresponds to your encoding scheme, and compare this value to the information entropy of the system. How similar are these two quantities?

2. Consider a source that can send three messages, whose probabilities are $p(x_1) = 0.6$, $p(x_2) = 0.25$ and $p(x_3) = 0.15$.

(a) Find an optimal encoding for these messages, and compute the average number of bits per message.

(b) Suppose that you are now asked to aggregate these messages and send them in packets of 2. Find an optimal encoding scheme for these packets, and calculate the average number of bits for the aggregated message.

(c) Compute the information entropy for this system, and explain how it relates to the answers you obtained in parts (a) and (b).

3. Consider a source that can send two messages, whose probabilities are $p(x_1) = 0.8$ and $p(x_2) = 0.2$.

(a) Find an optimal encoding for these messages, and compute the average number of bits per message.

(b) Suppose that you are now asked to aggregate these messages and send them in packets of 2. Find an optimal encoding scheme for these packets, and calculate the average number of bits for the aggregated message.

(c) Repeat part (b) in case the packets contain 3 messages.

(d) Compute the information entropy for this system, and explain how this value is related to the average number of bits per message calculated in parts (b) and (c).

4. Consider a system that consists of 3 molecules A, B and C, which are combined in *pairs*. We will assume that each of the 9 possible combinations triggers a process that produces one of 3 possible organic compounds, which are denoted R_1, R_2 and R_3. These compounds are related to the pairs of molecules in the following way:

R_1 is produced if pairs AA, AB, AC, or BB are sent.

R_2 is produced if pairs BA, BC, or CB are sent.

R_3 is produced if pairs CA or CC are sent.

The relative frequencies with which these pairs of molecules occur in nature are:

$$p(AA) = 0.10$$
$$p(AB) = 0.10$$
$$p(AC) = 0.05$$
$$p(BA) = 0.25$$
$$p(BB) = 0.20$$
$$p(BC) = 0.10$$
$$p(CA) = 0.05$$
$$p(CB) = 0.05$$
$$p(CC) = 0.10$$

For compounds R_1, R_2 and R_3, these frequencies are $p(R_1) = 0.45$, $p(R_2) = 0.4$ and $p(R_3) = 0.15$.

(a) Find the information entropy associated with the 9 pairs of "messenger" molecules.

(b) Assuming that these pairs remain unchanged in the transmission process, calculate the mutual entropy $I(X|Y)$. What does this value tell you about the loss of information in the system? Explain.

(c) Suppose we now allow for the possibility that one of the two molecules in the pair can be modified in the transmission process (due to environmental factors, or other external influences). If we assume that the probability of each such modification is equal to α, show how the conditional probabilities $p(y_j|x_i)$ change.

Note: You should present your results in the form of tables such as the one shown below:

$$
\begin{aligned}
p(R_i|AA) &= xxx \\
p(R_i|AB) &= xxx \\
&\vdots \\
p(R_i|CC) &= xxx
\end{aligned}
$$

(d) How is the calculation of $I(X|Y)$ affected when we take into account possible modifications in the molecule pairs? Hint: Compare the number of required operations in case $\alpha = 0$ and $\alpha > 0$.