EEE 431: Telecommunications I
Homework 2

1) Consider an 8-ary discrete memoryless source with output alphabet $\{a, b, c, d, e, f, g, h\}$ and PMF $\{1/2, 1/4, 1/8, 1/16, 1/64, 1/64, 1/64, 1/64\}$.

a) What is the entropy of the source in bits/source output?

b) Design a binary Huffman code for the source.

c) Determine the average length of the code in the previous part in bits/source output. Can a different (lossless) source code have a smaller average length? Why or why not?

2) Consider a DMS with a probability mass function $\{1/2, 1/3, 1/6\}$.

a) Design a binary Huffman code for the source, compute the average codeword length and compare it with its entropy.

b) Design a binary Huffman code considering pairs of source outputs. What is the average codeword length (per source output)? How does the result compare with that of the previous part.

3) Consider a discrete memoryless source (DMS) $\mathcal{S}_1$ with outputs $X_1, X_2, \ldots$. Assume that the source samples take on the possible values $a_1, a_2, \ldots, a_6$ with probabilities $1/4, 1/4, 1/4, 1/8, 1/16, 1/16$, respectively.

a) Design a Huffman code for the source $\mathcal{S}_1$ assuming that each source sample is encoded separately. Compute the entropy of the source, and determine whether the code can be improved in terms of average length or not.

b) Consider a second discrete memoryless source $\mathcal{S}_2$ with outputs $Y_1, Y_2, \ldots$. Assume that the samples take on the values b_1, b_2, b_3. Assume that the ith source output Y_i is correlated with the ith source output X_i of the first DMS $\mathcal{S}_1$ (but, it is independent of all other source samples of $\mathcal{S}_1$). The conditional PMF of Y_i given X_i is as follows:

$$\begin{gathered}
P(Y = b_1|X = a_1) = 1, \\
P(Y = b_2|X = a_2) = 1/2, \quad P(Y = b_3|X = a_2) = 1/2, \\
P(Y = b_2|X = a_3) = 1/2, \quad P(Y = b_3|X = a_3) = 1/2, \\
P(Y = b_1|X = a_k) = 1, \quad \text{for } k = 4, 5, 6.
\end{gathered}$$

Determine the PMF of the source sample Y_i. Design a Huffman code for $\mathcal{C}_2$ assuming that the samples are encoded separately. Can the code be improved in terms of average length or not?

c) We now encode the outputs of the two sources $\mathcal{S}_1$ and $\mathcal{S}_2$ together, that is, we encode the pairs (X_i, Y_i) jointly. Determine the possible values taken by (X_i, Y_i), and the respective probabilities. Also, design a Huffman code (taking one such pair at a time). What is the average length of the code? Is this lower than the sum of the entropies your have obtained in the previous two parts? If so, how can you explain this? Can the code be improved in terms of the average length by applying Huffman coding to multiple pairs of source outputs together? Why or why not?

4) Encode the following binary sequence using Lempel-Ziv coding

$$1101\,\underbrace{000\cdots0}_{10\text{ many}}\,\underbrace{111\cdots1}_{20\text{ many}}\,\underbrace{000\cdots0}_{12\text{ many}}\,\underbrace{111\cdots1}_{20\text{ many}}\,\underbrace{000\cdots0}_{20\text{ many}}.$$

5) Assume that a binary sequence has been compressed using Lempel-Ziv-Welch coding where the initial dictionary is simply $0 \to 1$, and $1 \to 2$. The end of file character # is denoted by the 0-th dictionary entry.

Determine the original text (sequence of 0's and 1's) if the encoded version is given by:

$$1, 3, 1, 2, 4, 7, 4, 6, 8, 9, 11, 3, 10, 13, 0.$$

6) Consider a source with alphabet $\{a, b, c, d\}$. Use Lempel-Ziv-Welch (LZW) coding to encode the sequence

$$\underbrace{aaaaaaaaaa}_{10\ a\text{'s}}\,\underbrace{bbbbbbbbbb}_{10\ b\text{'s}}\,\underbrace{cccccccccc}_{10\ c\text{'s}}.$$

Assume that the initial dictionary is $a \to 1$, $b \to 2$, $c \to 3$ and $d \to 4$.

Compare the number of bits required to encode the original text and the compressed versions.

7) The signal $s(t) = 2\sin(5000\pi t) - \cos(2200\pi t)$ is sampled at a rate $f_s = 2000$ Hz. The resulting samples $s[n]$ are impulse modulated, i.e.,

$$x(t) = \sum_{n=-\infty}^{\infty} s[n]\delta(t - nT_s)$$

is formed, where T_s is the sampling period ($T_s = 1/f_s$). We pass the signal $x(t)$ through an ideal low-pass filter with bandwidth $W = 1000$ Hz. What is the resulting signal?

8) A source output X, which is modeled as a random variable with the probability density function (PDF) given by

$$f_X(x) = \begin{cases} \frac{2-|x|}{4}\,, & if\ -2 < x < 2, \\ 0\,, & \text{otherwise}, \end{cases}$$

is being quantized using a 4-level quantizer with the quantization regions $[-2, -1]$, $(-1, 0]$, $(0, 1]$, and $(1, 2]$.

a) Determine the optimal reconstruction levels for the four quantization regions that minimize the mean squared error $D = E[(X - Q(X))^2]$ (where $Q(\cdot)$ denotes the quantization function).

b) Now assume that a new distortion function $D' = E[(X - Q(X))^4]$ is being used instead of the mean squared error distortion. Determine an equation whose solution is the optimal reconstruction level for the quantization interval $(0, 1]$ (do not solve the equation).

9) A source output is modeled as a random variable X with probability density function

$$f_X(x) = c \cdot \Lambda\left(\frac{t}{3}\right)$$

is quantized using a 6−level quantizer (over the input range of $[-3, 3]$).

a) Determine the constant c.

b) Determine the signal to quantization ratio assuming that the quantizer is uniform.

c) Assume that the quantization region boundaries are the same as those of the uniform quantizer. Determine the optimal reconstruction level for the quantization interval $[2, 3]$. Is it the midpoint of the interval?