Practice Midterm Examination (Spring 2016)
(Total: 80 points)

There are 3 problems, each problem with multiple parts, each part worth 10 points. Your answer should be as clear and readable as possible. Please justify any claim that you make.

1. Binary codes (30 points). Consider a source that emits five symbols \{A, B, C, D, E\} with probabilities 0.3, 0.3, 0.2, 0.1, and 0.1, respectively.

(a) Construct a binary Huffman code for this source, taking one source symbol at a time. What is the average codeword length for this code?

(b) Repeat part (a) for a binary Shannon–Fano code taking one source symbol at a time.

(c) Construct a probability distribution \(p = (p_A, p_B, p_C, p_D, p_E)\) on \{A, B, C, D, E\}, for which the code that you constructed in part (a) has an average length equal to its binary entropy \(H(p)\).

2. One-bit quantizer (30 points). Let \(X\) be drawn according to the pdf

\[
f_X(x) = \begin{cases} 
2x, & 0 \leq x \leq 1, \\
0, & \text{otherwise.}
\end{cases}
\]

(a) Given the quantization regions \(R_1 = \{x : 0 \leq x < b\}\) and \(R_2 = \{x : b \leq x \leq 1\}\), find the quantization points \(a_1 \in R_1\) and \(a_2 \in R_2\) that minimize the mean squared error (MSE) in terms of \(b\).

(b) Given the quantization points \(a_1 < a_2\), find the quantization regions that minimize the MSE in terms of \(a_1\) and \(a_2\).

(c) Using parts (a) and (b), find the optimal quantizer by specifying \(a_1, a_2,\) and \(b\) that minimize the MSE.
3. *Source coding for emergency (20 points).* Consider a source that emits six symbols \{A, B, C, D, E, F\} with probabilities 0.3, 0.2, 0.2, 0.1, 0.1, and 0.1, respectively. Here the symbol \(F\) only has the probability 0.1, but it is meant for an emergency message, say, “Fire!”, and hence we would like to spend few bits in encoding \(F\).

(a) Design a binary instantaneous code that takes one source symbol at a time with the shortest possible average codeword length under the constraint that \(F\) should be mapped to the codeword “0” of length 1.

(b) *(Difficult.)* Now assume that each source symbol \(i\) has some associated cost per letter \(c_i\) of the codeword. The normal symbols \(A, B, C, D, E\) has a cost of \(c_i = 1\) per codeword letter, while the emergency symbol \(F\) has a cost of \(c_i = 4\) per code letter. Hence, if the codeword for symbol \(i\) has the length \(l_i\), then the cost spent for the symbol is \(c_i l_i\). For example, the code that maps

\[
\begin{align*}
A & \rightarrow 0001 \\
B & \rightarrow 00000 \\
C & \rightarrow 1100 \\
D & \rightarrow 111 \\
E & \rightarrow 01 \\
F & \rightarrow 10
\end{align*}
\]

spends the cost of 4, 5, 4, 3, 2, and 8, respectively, for the symbols \(A, B, C, D, E,\) and \(F\). Design a binary instantaneous code that minimizes the average cost

\[
\sum_i p_i c_i l_i
\]

of sending a symbol.
There are 3 problems, each problem with multiple parts, each part worth 10 points. Your answer should be as clear and readable as possible.

1. *A dyadic source (70 pts).* Consider a source that produces an i.i.d. sequence of symbols from the alphabet \{A, B, C, D, E, F\} with probabilities \(1/2, 1/8, 1/8, 1/8, 1/16, 1/16\).

   (a) Find the (binary) entropy of the source.
   (b) Find the Huffman code that takes 1 source symbol at a time.
   (c) Find the average codeword length of the code constructed in part (b).
   (d) Can you improve the compression rate (measured by the average codeword length per source symbol) by taking 2 source symbols at a time? Justify your answer.
   (e) Find the Shannon–Fano code that takes 1 source symbol at a time.
   (f) Find the average codeword length of the code constructed in part (e).
   (g) (Difficult.) Is the Shannon–Fano code always optimal for sources with dyadic probabilities (that is, the probabilities of the form \(2^{-k}\))? Prove it or provide a counterexample.

2. *Hierarchical Huffman coding (40 pts).*

   Consider a source that produces an i.i.d. sequence of symbols from the alphabet \{A, B, C, D, E, F, G\} with probabilities \(1/28, 2/28, 3/28, 4/28, 5/28, 6/28, 7/28\).

   A clever engineer would like to construct the binary Huffman code, but would like to find a shortcut.

   (a) Find the quaternary Huffman code with code alphabet \{\(\alpha, \beta, \gamma, \delta\)\} that takes 1 source symbol at a time.
   (b) From the quaternary Huffman code found in part (a), construct a binary code by mapping

   \[
   \begin{align*}
   \alpha &\to 00, \\
   \beta &\to 01, \\
   \gamma &\to 10, \\
   \delta &\to 11.
   \end{align*}
   \]

   (c) Is the code in part (b) instantaneous? Justify your answer.
   (d) Is the code in part (b) optimal (that is, does it achieve the shortest average codeword length among all binary instantaneous codes)? Prove it or provide a counterexample.
3. Scalar quantization (50 pts). Consider a source with the probability density function

\[ f(x) = \begin{cases} 
 1/7, & -5 \leq x \leq 1 \text{ or } 4 \leq x \leq 5, \\
 0, & \text{otherwise.} 
\end{cases} \]

A clever engineer would like to construct the optimal 1-bit scalar quantizer by specifying two quantization regions \([b_0 = -5, b_1]\) and \((b_1, b_2 = 5]\), and two quantization points \(a_1 \leq b_1\) and \(a_2 > b_1\).

(a) Starting with \(b_1 = 0\), find the optimal \(a_1\) and \(a_2\) that minimize the mean squared error.

(b) Given \(a_1\) and \(a_2\) found in part (a), find the optimal \(b_1\).

(c) Given \(b_1\) found in part (b), find the optimal \(a_1\) and \(a_2\).

(d) What is the mean squared error of the quantizer constructed in part (c).

(e) Is the quantizer constructed in part (c) optimal? Prove it or provide a counterexample.