Exercise 1: Huffman Code

Given is a discrete iid process $X$ with the alphabet $\mathcal{A} = \{a, b, c, d, e, f, g\}$. The pmf $p_X(x)$ is specified in the following table.

<table>
<thead>
<tr>
<th>$x$</th>
<th>$p_X(x)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a$</td>
<td>$1/3$</td>
</tr>
<tr>
<td>$b$</td>
<td>$1/9$</td>
</tr>
<tr>
<td>$c$</td>
<td>$1/27$</td>
</tr>
<tr>
<td>$d$</td>
<td>$1/27$</td>
</tr>
<tr>
<td>$e$</td>
<td>$1/27$</td>
</tr>
<tr>
<td>$f$</td>
<td>$1/9$</td>
</tr>
<tr>
<td>$g$</td>
<td>$1/3$</td>
</tr>
</tbody>
</table>

(a) Develop a Huffman code for the given pmf $p_X(x)$.

(b) Calculate the average codeword length of the developed Huffman code.

(c) Calculate the absolute and relative redundancy for the developed Huffman code.
Exercise 2: Huffman Codes and Entropy Measures

Let \( Z = \{Z_n\} \) be a binary iid process with alphabet \( \{0, 1\} \) and pmf \( \{0.5, 0.5\} \) (e.g., coin toss).

Let \( X = \{X_n\} \) be a random process given by \( X_n = Z_{n-1} + Z_n \).

(a) Determine the marginal pmf \( p_X(x) \) and the marginal entropy \( H(X) \).

(b) Develop a scalar Huffman code and calculate its average codeword length.

(c) Determine the conditional pmf \( p_{X_n|X_{n-1}}(x_n|x_{n-1}) \) and the conditional entropy \( H(X_n|X_{n-1}) \).

(d) Develop a conditional Huffman code and calculate its average codeword length.

(e) Develop a block Huffman code for \( N = 2 \) symbols and calculate its average codeword length.

(f) Optional (more difficult):
   - Derive a formula for the \( N \)-th order block entropy \( H_N(X_n, \ldots, X_{n+N-1}) \).
   - Determine the entropy rate \( \bar{H}(X) \).
   - Is \( X \) a Markov process?
Exercise 3: Estimate Entropy Measures (Implementation Task)

Write a program (in a programming language of your choice) that estimates the following entropy measures based on the statistics of a given input file:

- Marginal entropy:  \( H(S_n) \)
- 1-st order conditional entropy:  \( H(S_n | S_{n-1}) \)
- Block entropy of size \( N = 2 \):  \( H(S_n, S_{n+1}) \) [calculate also \( H(S_n, S_{n+1})/2 \) ]

Assume that all files represent a sequence of 8-bit samples (i.e., each byte represents a symbol).

Test your program for the following sample files (from course website):

- white uniform noise:  “whiteUniformNoise.raw”
- white Gaussian noise:  “whiteGaussianNoise.raw”
- correlated Gaussian noise:  “correlatedGaussianNoise.raw”
- English text file:  “englishText.txt”
- 8-bit audio data:  “audioData.raw”
- 8-bit image data:  “imageData.raw”

What can you conclude about the potential to compress these files?
Exercise 4: Geometric Pmf (Optional)

Given is a Bernoulli process (binary iid process) $\mathbf{B} = \{B_n\} = \{B\}$ with the alphabet $\mathcal{A}_B = \{0, 1\}$ and the pmf $p_B(0) = p$, $p_B(1) = 1 - p$ with $0 < p < 1$.

Consider the random variable $X$ that specifies the number of successive random variables $B_n$ that have to be observed to get exactly one “1”.

(a) Determine the pmf for $X$ as function of $p$.

(b) Determine the entropy $H(B)$ as function of $p$.

(c) Determine the entropy $H(X)$ as function of $H(B)$ and $p$.

(d) What structure has an optimal scalar variable-length code for $X$ and $p \leq 0.5$?
   Calculate its average codeword length as function of $p$.
   Calculate its relative redundancy as function of $p$.

Hints:

$$\forall |a| < 1, \sum_{k=0}^{\infty} a^k = \frac{1}{1-a} \quad \text{and} \quad \forall |a| < 1, \sum_{k=0}^{\infty} k a^k = \frac{a}{(1-a)^2}$$