Problem 1 (AEP and mutual information ★) Let \( \{X_i, Y_i\}_{i=1}^n \) be an i.i.d. sequence drawn according to \( p_{XY}(x, y) \). We form the log-likelihood ratio of the hypothesis that \( X \) and \( Y \) are independent versus the hypothesis that \( X \) and \( Y \) are dependent. Find the limit of

\[
\frac{1}{n} \log_2 \frac{p_{X^n}(X^n) p_{Y^n}(Y^n)}{p_{X^n=Y^n}(X^n Y^n)} \quad \text{as} \quad n \to \infty.
\]

Problem 2 (Product and AEP ★) Let \( X \) be defined as

\[
X = \begin{cases} 
1 & \text{with probability } \frac{1}{2}, \\
2 & \text{with probability } \frac{1}{4}, \\
3 & \text{with probability } \frac{1}{4}.
\end{cases}
\]

Let \( \{X_i\}_{i=1}^n \) be an iid sequence drawn according to this distribution. Find the limiting behavior of

\[
Y = \left( \prod_{i=1}^n X_i \right)^{\frac{1}{n}}.
\]

Problem 3 (Erasure entropy rate ★★) Recall that the entropy rate of a stationary process \( X \) is

\[
H(X) = \lim_{n \to \infty} H(X_0 | X_{-1}, \ldots, X_{-n}).
\]

We define the erasure entropy rate as

\[
H^{-}(X) = \lim_{n \to \infty} H(X_0 | X_n, \ldots, X_1, X_{-1}, \ldots, X_{-n}),
\]

that is we condition not only on the past but also on the future. Now, let \( X \) be a first order homogeneous Markov chain. We denote the two states of the Markov chain by 0 and 1, and we assume that \( p_{X_1 | X_0}(0 | 1) = p_{X_1 | X_0}(1 | 0) = p \).

(a) Show that \( H(X) = H_b(p) \).

(b) Show that \( H^{-}(X) = 2H_b(p) - H_b(2p(1-p)) \). (Hint: show that \( H(X_2 | X_0) = H_b(2p(1-p)) \))

(c) Sketch \( H^{-}(X) \) and \( H(X) \) as a function of \( p \).

Problem 4 (Huffman coding 2 ★ [OPTIONAL]) Construct a binary Huffman code for the random variable \( X \) and compute the average length

\[
X \quad 1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \\
p_X(x) \quad 0.25 \quad 0.05 \quad 0.1 \quad 0.13 \quad 0.2 \quad 0.12 \quad 0.08 \quad 0.07
\]

Problem 5 (Huffman codes with costs ★★) Words like Run!, Help! and Fire! are short, not because they are frequently used, but perhaps because time is precious in the situations in which these words are required. Suppose that \( X = i \) with probability \( p_i \) for \( i \in \{1, \ldots, m\} \). Let \( l_i \) be the number of binary symbols in the codeword associated to \( X = i \), and let \( c_i \) denote the cost per letter of the codeword when \( X = i \). Therefore, the average cost \( C \) of the description of \( X \) is

\[
C = \sum_{i=1}^m p_i c_i l_i.
\]
(a) Minimize $C$ over all $l_i$ such that $\sum_{i=1}^{m} 2^{-l_i} \leq 1$. Ignore any implied integer constraint on $l_i$. Exhibit the minimizing $l_i^*$ and the associated minimum value $C^*$.

(b) How would you use the Huffman code procedure to minimize $C$ over all uniquely decodable codes? Let $C_{Huffman}$ denote this minimum.

(c) Can you show that

$$C^* \leq C_{Huffman} \leq C^* + \sum_{i=1}^{m} p_i c_i?$$

Problem 6 (Optimal Codeword Length ★★★) Although the codeword lengths of an optimal variable length code are complicated functions of the message probabilities $p_1, \ldots, p_m$, it can be said that less probable symbols are encoded into longer codewords. Suppose that the message probabilities are given in decreasing order $p_1 > p_2 \geq \cdots \geq p_m$.

(a) Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 > \frac{2}{5}$, then that symbol must be assigned a codeword of length 1.

(b) Prove that for any binary Huffman code, if the most probable message symbol has probability $p_1 < \frac{1}{3}$, then that symbol must be assigned a codeword length at least 2.

Problem 7 (Shannon code ★☆ [optional]) Consider the following method for generating a code for a random variable $X$ which takes on $m$ values $\{1, \ldots, m\}$ with probabilities $p_1, \ldots, p_m$. Assume that the probabilities are ordered so that $p_1 \geq p_2 \geq \cdots \geq p_m$. Define

$$F_i = \mathbb{P}_X[X \leq i - 1] = \sum_{k=1}^{i-1} p_i.$$ 

The codeword for $i$ is the number $F_i$ rounded off to $\ell_i = \lceil \log \frac{1}{p_i} \rceil$ bits.

(a) Show that the code constructed by this process is prefix-free and that the average length satisfies

$$\mathbb{H}(X) \leq L \leq \mathbb{H}(X) + 1.$$

(b) Construct the code for the probability distribution $(0.5, 0.25, 0.125, 0.125)$

Problem 8 (Huffman code for uniform source ★★★) Consider a source $X$ with 100 equally likely symbols (that is $p_X(x) = 10^{-2}$ for all $x = 1 \ldots 100$). Find the average length of a Huffman code for this source.