Problem 1 (Weak Law of Large Numbers).

(a) (Markov’s Inequality) Let the probability density function of $X$ be $p_X(x)$. As $X$ is non-negative and $a > 0$,

$$E[X] = \int_0^a x p_X(x) dx + \int_a^{\infty} x p_X(x) dx \geq a \int_a^{\infty} p_X(x) dx = a \Pr\{X \geq a\}.$$

(b) (Chebyshev’s Inequality) Set $X = (Y - \mu)^2$. Then,

$$E[X] = E[(Y - \mu)^2] = \sigma^2.$$

Using Markov’s inequality,

$$\Pr\{|Y - \mu| \geq b\} = \Pr\{(Y - \mu)^2 \geq b^2\} = \Pr\{X \geq b^2\} \leq \frac{E[X]}{b^2} = \frac{\sigma^2}{b^2}.$$

(c) (Weak Law of Large Numbers) As $X_1, \cdots, X_n$ are IID, $E[X_n] = E[\frac{1}{n} \sum_{i=1}^{n} X_i] = \frac{1}{n} \sum_{i=1}^{n} E[X_i] = \mu$ and

$$Var[X_n] = E[(X_n - E[X_n])^2] = E\left[\left(\frac{1}{n} \sum_{i=1}^{n} X_i - \mu\right)^2\right]$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{n} (X_i - \mu)^2 + \sum_{j=1}^{n} \sum_{k=1}^{n} (X_j - \mu)(X_k - \mu)\right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} E[(X_i - \mu)^2] + \sum_{j=1}^{n} \sum_{k=1}^{n} E[(X_j - \mu)(X_k - \mu)]$$

$$= \frac{\sigma^2}{n}.$$

Using Chebyshev’s inequality for $X_n$, we get

$$\Pr\{|X_n - \mu| \geq \epsilon\} \leq \frac{Var[X_n]}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2}, \quad \forall \epsilon > 0.$$

Problem 2 (Jensen’s Inequality). When $n = 2$, $E[f(X)] = \sum_{i=1}^{n} p_i f(x_i) + p_2 f(x_2) \geq f(\sum_{i=1}^{n} p_i x_i) = f(E[X])$, because $f$ is convex. Assume Jensen’s inequality is true for $n = k \geq 2$. If we show that it is true for $(k + 1)$, we will have proven Jensen’s inequality using mathematical induction. Let $X$ take values $x_1, \cdots, x_{(k+1)}$ with probabilities $p_1, \cdots, p_{(k+1)}$, such
that $\sum_{i=1}^{(k+1)} p_i = 1$. Let $p'_i = p_i / (\sum_{j=1}^{k} p_j)$ so that $\sum_{i=1}^{k} p'_i = 1$. From our assumption, we know that for $n = k$

$$E[f(X)] \geq f(E[X]) \iff \sum_{i=1}^{k} f(x_i)p'_i \geq f\left(\sum_{i=1}^{k} x_i p'_i\right).$$

Now, for $n = k + 1$

$$E[f(X)] = \sum_{i=1}^{k+1} f(x_i)p_i = \sum_{j=1}^{k} p_j \sum_{i=1}^{k} f(x_i) \frac{p_i}{\sum_{j=1}^{k} p_j} + f(x_{(k+1)})p_{(k+1)}$$

$$= (1 - p_{(k+1)}) \sum_{i=1}^{k} f(x_i)p'_i + f(x_{(k+1)})p_{(k+1)}$$

$$\geq (1 - p_{(k+1)}) f\left(\sum_{i=1}^{k} x_i p'_i\right) + f(x_{(k+1)})p_{(k+1)} \quad \text{(by assumption)}$$

$$\geq f\left((1 - p_{(k+1)}) \sum_{i=1}^{k} x_i p'_i + p_{(k+1)}x_{(k+1)}\right) \quad \text{(because f is convex)}$$

$$= f\left(\sum_{i=1}^{k+1} x_i p_i\right) = f(E[X]).$$

**Problem 3 (Huffman Coding).** (i) Given $Y = y_1$, the conditional probability distribution of $X$ is

$$X \quad \Pr\{X|Y = y_1\} \quad | x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6$$

$$1/3 \quad 1/3 \quad 1/12 \quad 1/12 \quad 1/12 \quad 1/12$$

A possible binary Huffman tree for the above conditional distribution looks like Figure 1. From the tree, the expected codeword length $W_1$ of a binary Huffman code for $X$ given $Y = y_1$ is $W_1 = 1/3 + 2/3 + 0 + 4/3 = 7/3$ bits.

(ii) Given $Y = y_2$, the conditional probability distribution of $X$ is

$$X \quad \Pr\{X|Y = y_2\} \quad | x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6$$

$$1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6 \quad 1/6$$

A possible binary Huffman tree for the above conditional distribution looks like Figure 2. From the tree, the expected codeword length $W_2$ of a binary Huffman code for $X$ given $Y = y_2$ is $W_2 = 0 + 2/3 + 6/3 = 8/3$ bits.

(iii) When $Y$ is random, the probability distribution of $X$ is

$$X \quad \Pr\{X\} \quad | x_1 \quad x_2 \quad x_3 \quad x_4 \quad x_5 \quad x_6$$

$$7/24 \quad 7/24 \quad 5/48 \quad 5/48 \quad 5/48 \quad 5/48$$

2
Figure 1: A possible binary Huffman code construction for Problem 3(i).

Figure 2: A possible binary Huffman code construction for Problem 3(ii).
A possible binary Huffman tree for the above distribution looks like Figure 3. From the tree, the expected codeword length $W$ of a binary Huffman code for $X$ is $W = 0 + \frac{14}{12} + \frac{15}{12} = \frac{29}{12}$ bits. $W = \frac{29}{12} = 3/4 \times 7/3 + 1/4 \times 8/3 = \Pr\{Y = y_1\}W_1 + \Pr\{Y = y_2\}W_2$. So yes, $W = \Pr\{Y = y_1\}W_1 + \Pr\{Y = y_2\}W_2$ for this particular joint distribution.

![Figure 3: A possible binary Huffman code construction for Problem 3(iii).](image)

(iv) No, it is not true in general. Consider the following joint distribution:

<table>
<thead>
<tr>
<th>$Y \setminus X$</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$y_1$</td>
<td>$1/4$</td>
<td>$1/24$</td>
<td>$1/24$</td>
</tr>
<tr>
<td>$y_2$</td>
<td>$1/6$</td>
<td>$1/3$</td>
<td>$1/6$</td>
</tr>
</tbody>
</table>

For this distribution, it can be verified that $W_1 = \frac{5}{4}$ bits, $W_2 = \frac{3}{2}$ bits, and $W = \frac{19}{12}$ bits. However, $\Pr\{Y = y_1\}W_1 + \Pr\{Y = y_2\}W_2 = \frac{1}{3} \times \frac{5}{4} + \frac{2}{3} \times \frac{3}{2} = \frac{17}{12}$ bits.

(v) Consider $Z = (X, Y)$ as a random variable with $\Pr\{Z = (x_i, y_j)\} = \Pr\{x_i, y_j\}$, $i = 1, \ldots, 6$, $j = 1, 2$. Then one possible Huffman tree looks like Figure 4. From the tree, the expected codeword length for $Z$ is $0 + 1 + 0 + \frac{4}{3} + \frac{5}{6} = \frac{19}{6}$ bits.

**Problem 4** (Code Mismatch). The expected codeword length is $L = \sum_{i=1}^{n} p_i l_i$. As $\log(1/q_i) \leq l_i \leq \log(1/q_i) + 1$, we get

$$
\sum_{i=1}^{n} p_i \log \frac{1}{q_i} \leq L \leq \sum_{i=1}^{n} p_i \left( \log \frac{1}{q_i} + 1 \right)
$$

$$
\sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} + \sum_{i=1}^{n} p_i \log \frac{1}{p_i} \leq L \leq \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} + \sum_{i=1}^{n} p_i \log \frac{1}{p_i} + \sum_{i=1}^{n} p_i
$$

$$
H(p) + D(p||q) \leq L \leq H(p) + D(p||q) + 1.
$$
Figure 4: A possible binary Huffman code construction for Problem 3(v).