Q1 a) The source has 7 symbols. The minimum worst-case length of a variable length code is attained by a code which assigns symbols \( \{\emptyset, 0, 1, 00, 01, 10, 11\} \) to symbols with decreasing probabilities. Thus, \( L(X) = 2 \).

In fact, the same assignment as above also attains the minimum average codeword length. Thus,

\[
L(X) = 0 \times 1/2 + 2 \times 1 \times 1/8 + 4 \times 2 \times 1/16 = 3/4.
\]

Note that since \( -\log p(x) \) is an integer for every \( x \), \( L^u_0(X) = L^p_0(X) = H(X) = 9/4 \) and can be attained by the Huffman code or the Shannon-Fano code (see, also, Q2(a)).

b) In the manner of \( L(X) \) and \( \bar{L}(X) \), \( L_\epsilon(X) \) and \( \bar{L}_\epsilon(X) \) are also attained by the code which assigns shortest available codewords in decreasing order of probabilities to the symbols till the probability adds up to greater than \( 1 - \epsilon \); all the remaining symbols are assigned to the shortest codeword, the empty sequence. Thus, the optimal lengths for \( \epsilon = 0.25 \) are attained by the following code:

<table>
<thead>
<tr>
<th>symbol</th>
<th>codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>\emptyset</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>{3, 4, 5, 6}</td>
<td>\emptyset</td>
</tr>
</tbody>
</table>

Thus \( L_{0.25}(X) = 1/4 \).

c) From (b) we also have, \( L_{0.25}(X) = 1 \).

Q2 In view of the result, \( \bar{L}^{\text{prefix}}_0(X) > H(X) \) if there is a symbol \( x \) where \( -\log P(x) \) is not an integer. Consider the following example with \( \mathcal{X} = \{0, 1, 2\} \):

\[
P_X(0) = 0.5, \quad P_X(1) = 0.5 - P_X(2) = 0.2.
\]

The Huffman code for this example has the following codeword assignment

<table>
<thead>
<tr>
<th>symbol</th>
<th>codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
</tr>
</tbody>
</table>
Since Huffman code is optimal and the average length of the Huffman code equals $3/2$, 
\[
L_0^p(X) = 3/2.
\]
On the other hand
\[
H(X) = 1/2 + 2/10 \log 10/2 + 3/10 \log 10/3 = 1 + (1/2)h(2/5) < 3/2.
\]

Q3 a) **Construction of Huffman code.**

(a) Assign symbols 2 and 3 as the left and the right child of a node (2, 3). We now have three symbols, 1, (2, 3), and 4, each with the same probability 1/3.

(b) Assign symbols 1 and (2, 3) as the left and the right child of a node (1, (2, 3)). We now have two symbols, (1, (2, 3)) and 4, the first with probability 2/3 and the second with probability 1/3.

(c) Assign (1, (2, 3)) and 4 as left and right child of the root.

The Huffman code is obtained by moving from root to the leaves and associating each left child with 0 and the right child with 1. We get the following codewords

<table>
<thead>
<tr>
<th>symbol</th>
<th>codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>00</td>
</tr>
<tr>
<td>2</td>
<td>010</td>
</tr>
<tr>
<td>3</td>
<td>011</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

**Construction of Shannon-Fano code.** Let $x_1 = 1, x_2 = 4, x_3 = 2, x_4 = 3$ and let $p_i = P(x_i)$, $i = 1, 2, 3, 4$. Thus, $(p_1, p_2, p_3, p_4)$ is a decreasing sequence. Let $F_1 = 0, F_2 = p_1 = 1/3, F_3 = p_1 + p_2 = 2/3, F_4 = F_3 + 1/6 = 5/6$. The codeword for the symbol $x_i$ is obtained by taking the $[−\log p_i]$ most significant bits of binary representation of $F_i$. Thus, $x_1$ and $x_2$ is represented using 2 bits each and $x_3$ and $x_4$ using 3 bits each. We get the following code.

<table>
<thead>
<tr>
<th>symbol</th>
<th>codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_1$</td>
<td>00</td>
</tr>
<tr>
<td>$x_2$</td>
<td>01</td>
</tr>
<tr>
<td>$x_3$</td>
<td>101</td>
</tr>
<tr>
<td>$x_4$</td>
<td>110</td>
</tr>
</tbody>
</table>

**Construction of Elias code.** We work with the original ordering of the symbols. Then,
\[
\begin{align*}
F(1) &= 0 + P(1)/2 = 1/6, \\
F(2) &= P(1) + P(2)/2 = 5/12, \\
F(3) &= P(1) + P(2) + P(3)/2 = 7/12, \\
F(4) &= P(1) + P(2) + P(3) + P(4)/2 = 5/6.
\end{align*}
\]

The Elias code is obtained by representing the symbol $i$ with the $[−\log P(i)]+1$ most significant bits in the binary representation of $F(i)$. Thus, we represent 1, 2, 3, and 4 with codewords of length 3, 4, 4, and 3, respectively. We get the following code.
### Question 4

We will give the implementation of the arithmetic coding with infinite precision. Thus, we will compute the limits of intervals as real numbers. At each step we will maintain the starting point and the width of the interval.

#### Encoding:

(a) \( X_1 = 3 \Rightarrow C_1 = F(X_1) = 1/2, \ W_1 = P(X_1) = 1/6. \)

(b) \( X_2 = 1 \Rightarrow C_2 = C_1 + W_1 \times F(X_2) = 1/2, \ W_2 = W_1 \times P(X_2) = 1/18. \)

(c) \( X_3 = 4 \Rightarrow C_3 = C_2 + W_2 \times F(X_3) = 29/54, \ W_3 = W_2 \times P(X_3) = 1/54. \)

(d) \( X_4 = 2 \Rightarrow C_4 = C_3 + W_3 \times F(X_4) = 88/162, \ W_4 = W_3 \times P(X_4) = 1/324. \)

Thus, we can encode the sequence 3142 by the shortest binary sequence such that the real number in \([0,1]\) represented by the sequence belongs to the interval \([176/324, 177/324) = [0.543209..., 0.54629...]; it suffices to find a binary number in the interval [0.544, 0.546]. Clearly, the first bit is 1. For the next three bits, consider \(2 \times 0.444, 0.046 = [0.088, 0.092], \ 4 \times [0.444, 0.046] = 0.176, 0.184\), and \(8 \times [0.444, 0.046] = 0.352, 0.368\). Thus, the second, the third and the fourth bits are 0. For the fifth bit, consider \(2 \times 0.352, 0.368 = [0.704, 0.736]\). Thus, the fifth bit is 1. For the sixth bit, consider \(2 \times 0.204, 0.236 = [0.408, 0.472]\). Thus, sixth bit is also 0. For the seventh bit, consider \(2 \times 0.408, 0.472 = [0.816, 0.944]\). Thus, the seventh bit is 1. For the eighth bit, consider \(2 \times 0.316, 0.444 = [0.632, 0.888]\). Thus, the eighth bit is 1. For the ninth bit, consider \(2 \times 0.132, 0.388 = [0.264, 0.776]\), which contains 1/2. Thus, the ninth bit is 1 and we can stop. The resulting codeword is 10010111.

#### Decoding: The received codeword 10010111 corresponds to the number \(c = 0.544921875\).

(a) Let \(c_1 = c\). Since \(c_1 \in [F(3), F(4)) = [1/2, 1/2 + 1/6), \ X_1 = 3. \)

(b) Consider \(c_2 = (c_1 - F(3))/P(3) = 0.26953125. \) Since \(c_2 \in [F(1), F(2)) = [0, 1/3), \ X_2 = 1. \)

(c) Consider \(c_3 = (c_2 - F(1))/P(1) = 0.80859375. \) Since \(c_3 \in [F(4), 1) = [2/3, 1), \ X_3 = 4. \)

(d) Consider \(c_4 = (c_3 - F(4))/P(4) = 0.42578125. \) Since \(c_3 \in [F(2), 1) = [1/3, 1/2), \ X_3 = 2. \)

Thus, the first encoded symbol was decoded first and the FIFO property holds.

### Question 5

(a) Let \(c : X \to \{0, 1\}^*\) denote the encoder for the Huffman code for \(P_X\) and let \(l_c(x)\) be the length of the codeword assigned to the symbol \(x \in X\). Consider the encoder \(f(x, y) = c(x), x \in X, y \in Y\). Note that the length \(l(x, y)\) assigned to the symbol \((x, y)\) by this code is
then, the average length of this encoder is given by

\[
\bar{L}_f(X|Y) = \sum_{x,y} P_{XY}(x,y) l(x,y)
\]

\[
= \sum_{x,y} P_{XY}(x,y) l_c(x)
\]

\[
= \sum_x P_X(x) l_c(x)
\]

\[
= I^p(X),
\]

where the final equality holds since the Huffman code attains optimal average length for \(X\). Since this is a particular choice of source code, the best code can only be better and therefore

\[
\bar{L}_p(X|Y) \leq \bar{L}^p(X).
\]

b) The statement is not true. We first note that

\[
\bar{L}_p(X|Y) = \sum_y P_Y(y) \bar{L}^p(X|Y = y),
\]

where, with an abuse of notation, \(X|Y = y\) denotes the random variable with pmf \(P_{X|Y = y}\). Indeed for any code \(f(x,y)\) with the lengths \(l(x,y)\) form a prefix free code for every \(y \in \mathcal{Y}\). Therefore, for every \(y \in \mathcal{Y}\),

\[
\sum_x P_{X|Y = y}(x) l(x,y) \geq \bar{L}^p(X|Y = y),
\]

which upon averaging over \(y\) gives,

\[
\bar{L}_f(X|Y) = \sum_y P_Y(y) \sum_x P_{X|Y = y}(x) l(x,y) \geq \sum_y P_Y(y) \bar{L}^p(X|Y = y).
\]

Since this is true for every \(f\),

\[
\bar{L}^p(X|Y) \geq \sum_y P_Y(y) \bar{L}^p(X|Y = y)
\]

(1)

On the other hand, denote by \(f_y, y \in \mathcal{Y}\), the Huffman code for the pmf \(P_{X|Y = y}\) on \(X\). For this code, let \(l_y(x)\) be the length of the codeword for \(x \in \mathcal{X}\). Consider the encoder \(f(x,y) = f_y(x)\). For this code the average length is given by

\[
\bar{L}_f(X|Y) = \sum_y P_Y(y) \sum_{x} P_{X|Y = y}(x) l_y(x) = \sum_y P_Y(y) \bar{L}^p(X|Y = y).
\]

Therefore this encoder meets the lower bound in (1) and so

\[
\bar{L}^p(X|Y) = \sum_y P_Y(y) \bar{L}^p(X|Y = y).
\]

(2)

Observe that if \(X\) and \(Y\) are independent we have \(\bar{L}^p(X|Y = y) = \bar{L}^p(X), \forall y \in \mathcal{Y}\). Thus from (2), we have \(\bar{L}^p(X|Y) = \bar{L}^p(X)\). We shall now show that the opposite direction is not true.

Consider the following joint pmf \(P_{XY}\):

\[
P_{XY}(1,1) = P_{XY}(2,1) = 1/6, P_{XY}(3,1) = P_{XY}(4,1) = 1/12
\]
\[ P_{XY}(3, 2) = P_{XY}(4, 2) = \frac{1}{6}, P_{XY}(1, 2) = P_{XY}(2, 2) = \frac{1}{12}. \]

Clearly \( X \) and \( Y \) are not independent. Also, \( P_X(x) = \frac{1}{4}, \forall x \in \mathcal{X} \). Thus \( \mathcal{L}^p(X) = \log |\mathcal{X}| = 2 \). Observe that since the pmfs \( P_{X|Y=1} \) and \( P_{X|Y=2} \) are just permutations of each other we have from (2),

\[ \mathcal{L}^p(X|Y) = \mathcal{L}^p(X|Y = 1) = \mathcal{L}^p(X|Y = 2). \]

One possible Huffman code construction for the pmf \( P_{X|Y=1} = \{ \frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6} \} \) is as follows:

(a) Assign symbols 3 and 4 as the left and the right child of the node \((3, 4)\), a node with probability \( \frac{1}{3} \).
(b) Assign symbols 1 and 2 as the left and the right child of the node \((1, 2)\).
(c) Assign \((1, 2)\) and \((3, 4)\) as the left and the right child of the root.

<table>
<thead>
<tr>
<th>symbol</th>
<th>codeword</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>00</td>
</tr>
<tr>
<td>2</td>
<td>01</td>
</tr>
<tr>
<td>3</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>11</td>
</tr>
</tbody>
</table>

Thus the resulting code is as follows

Thus the Huffman code for the pmf \( P_{X|Y=1} \) equals 2. Thus \( \mathcal{L}^p(X|Y) = 2 \)