Problem 4.8 (Davide Basilio Bartolini)
Since the sequence $X$ comes from a discrete memoryless source, the samples $X_i$ are independent random variables and $p_2 = 1 - p_1$. First, we need to compute the estimated duration for a sample, which is (due to the fact that the value of a symbol is equal to its duration):

$$E[T] = E[X_i] = p_1 + 2p_2 = 2 - p_1$$

Then, we compute the entropy for a sample of the process:

$$H(X_i) = -(p_1 \log_2 p_1 + p_2 \log_2 p_2) = -(p_1 \log_2 p_1 - (1 - p_1) \log_2(1 - p_1))$$

So, the expression for the entropy per unit time is:

$$\mathcal{H}(\mathcal{X}) = \frac{H(X_i)}{E[T]} = \frac{-(p_1 \log_2 p_1 - (1 - p_1) \log_2(1 - p_1))}{2 - p_1}$$

To find the value for $p_1$ that maximizes $\mathcal{H}(\mathcal{X})$, we can compute its partial derivative w.r.t. $p_1$ and find the value of $p_1$ that makes it zero (i.e. the value for which its numerator is zero):

$$\frac{\delta \mathcal{H}(\mathcal{X})}{\delta p_1} = \frac{T \frac{\delta H(X_i)}{\delta p_1} - H(X) \frac{\delta T}{\delta p_1}}{T^2}$$

$$= \frac{\log_e(1 - p_1) - 2 \log_e p_1}{T^2}$$

(using base $e$ for the logs in entropies)

$$\log_e(1 - p_1) - 2 \log_e p_1 = 0 \iff 1 - p_1 = p_1^2 \iff p_1^2 + p_1 - 1 = 0$$

The last equation which gives the results for $p_1$: $-\frac{1}{2}(\sqrt{5} + 1)$ and $\frac{1}{2}(\sqrt{5} - 1)$. Of the two found values, the second is a maximum for $\mathcal{H}(\mathcal{X})$ (moreover, the first solution is less than zero, so it cannot be a probability value) and the required value of $p_1$ which maximizes $\mathcal{H}(\mathcal{X})$ is $p_1 = \frac{1}{2}(\sqrt{5} - 1) = 0.61803$.

The value of the maximum $\mathcal{H}(\mathcal{X})$ may be found simply by substituting the value of $p_1$ just computed and it is $\mathcal{H}(\mathcal{X})_{\text{max}} = 0.69424$ bits.

Problem 4.11 (Davide Basilio Bartolini)

1. True:

$$H(X_n|X_0) = H(X_n, X_0) - H(X_0)$$

$$H(X_{n-n}|X_0) = H(X_{n-n}, X_0) - H(X_0)$$

and $H(X_n, X_0) = H(X_{n-n}, X_0)$ due to the stationarity of the process.
2. False (in general):

\[ H(X_n|X_0) = H(X_n, X_0) - H(X_0) \]
\[ H(X_{n-1}|X_0) = H(X_{n-1}, X_0) - H(X_0) \]

but nothing can be said in general about the comparison between \( H(X_n, X_0) \) and \( H(X_n, X_0) \).

For example, take a sequence of binary random variables \( X_1, X_2, \ldots, X_n, X_{n+1}, \ldots \) i.i.d. \(~p(0) = p(1) = \frac{1}{2}\) such that the relation \( X_i = X_{kn+i} \) holds (i.e. the sequence is periodic of period \( n \)). This sequence is a stationary process (the conjunct probability of the samples does not vary with a time shift) and we have \( H(X_n|X_0) = 0 \) and \( H(X_{n-1}|X_0) = H(X_{n-1}) = 1 \). So, in this case, \( H(X_n|X_0) < H(X_{n-1}|X_0) \) and this is a counterexample to the proposed inequality.

3. True:

\[ H(X_n|X_1, X_2, \ldots, X_{n-1}, X_{n+1}, \ldots, X_{2n}) = H(X_{n+1}|X_2, \ldots, X_n, X_{n+2}, \ldots, X_{2n+1}) \]  
(stationarity)

\[ (\text{conditioning reduces entropy}) \geq H(X_{n+1}|X_1, \ldots, X_n, X_{n+2}, \ldots, X_{2n+2}) \]

4. True:

\[ H(X_n|X_1, \ldots, X_{n-1}, X_{n+1}, \ldots, X_{2n}) = H(X_{n+1}|X_2, \ldots, X_n, X_{n+2}, \ldots, X_{2n+1}) \]  
(stationarity)

\[ (\text{conditioning reduces entropy}) \geq H(X_{n+1}|X_1, \ldots, X_n, X_{n+2}, \ldots, X_{2n+2}) \]

**Problem 4.18 (ohnson Jonaris GadElkarim)**

For the first coin \( P_r \{Head\} = p \), For the second coin \( P_r \{Head\} = 1 - p \)

For choosing the coin: \( P_r \{X = 1^{st} \text{coin}\} = P_r \{X = 2^{nd} \text{coin}\} = 1/2 \)

<table>
<thead>
<tr>
<th>X=First Coin</th>
<th>X=Second Coin</th>
<th>( P(Y_1, Y_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Y_1 = H, Y_2 = H )</td>
<td>( 1/2p^2 )</td>
<td>( 1/2(1 - p)^2 )</td>
</tr>
<tr>
<td>( Y_1 = T, Y_2 = T )</td>
<td>( 1/2(1 - p)^2 )</td>
<td>( 1/2p^2 )</td>
</tr>
<tr>
<td>( Y_1 = H, Y_2 = T )</td>
<td>( 1/2p(1 - p) )</td>
<td>( 1/2p(1 - p) )</td>
</tr>
<tr>
<td>( Y_1 = T, Y_2 = H )</td>
<td>( 1/2p(1 - p) )</td>
<td>( 1/2p(1 - p) )</td>
</tr>
</tbody>
</table>

a) Since \( Y_1, Y_2, \ldots, Y_n \) are i.i.d with the knowledge of \( X \), hence:

\[ I(Y_1; Y_2|X) = 0 \]

b)

\[ I(X; Y_1, Y_2) = H(X) - H(X|Y_1, Y_2) \]

We have \( H(X) = -2 * 1/2 \log(1/2) = 1 \)

Also \( H(X|Y_1, Y_2) = H(X, Y_1, Y_2) - H(Y_1, Y_2) \)

\[ H(X, Y_1, Y_2) = -2 * 1/2 * p^2 \log(p^2) - 2 * 1/2 * (1 - p)^2 \log(1 - p)^2 - 4 * 1/2 * p(1 - p) \log p(1 - p) \]

\[ H(Y_1, Y_2) = -2 * (1/2 * p^2 + 1/2 * (1 - p)^2) \log((1/2 * p^2 + 1/2 * (1 - p)^2)) - 2 * 1/2 * P(1 - p) \log P(1 - p) \]

c) Since \( Y_1, Y_2, \ldots, Y_n \) are i.i.d with the knowledge of \( X \), we can define \( H(p) \) as \( H(p) = -p \log(p) - (1 - p) \log(1 - p) \)

\[ H(Y_1, \ldots, Y_n, X) = H(X) + H(Y_1, \ldots, Y_n|X) = 1 + n * H(p) \]
Hence

\[ H(p) = \lim_{n \to \infty} \frac{1}{n} H(Y_1, \ldots, Y_n | X) \]

\[ \leq \lim_{n \to \infty} \frac{1}{n} H(Y_1, \ldots, Y_n) \]

\[ \leq \lim_{n \to \infty} \frac{1}{n} H(Y_1, \ldots, Y_n, X) = H(p) \]

Hence

\[ H(Y) = \lim_{n \to \infty} \frac{1}{n} H(Y_1, \ldots, Y_n) = H(p) \]

**Problem 5.1 (Kenneth S. Palacio Baus)**

We know that all instantaneous codes are uniquely decodable (which leads to \( L_2 \leq L_1 \)). \( L_2 \) is the minimum \( L \) over all uniquely decodable codes, these lengths must satisfy the Kraft inequality. We have also that from any lengths satisfying the Kraft inequality, an instantaneous code can be constructed \([?] pp.102\) and must exist and have the same expected \( L \). From here \( L_1 \leq L_2 \). From both inequalities it can be concluded that \( L_1 = L_2 \).

**Problem 5.2 (Kenneth S. Palacio Baus)**

Given the information provided, it’s possible to use the Kraft inequality to estimate \( D \):

<table>
<thead>
<tr>
<th>( i )</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>( \sum_{i=1}^{6} D^{-l_i} )</th>
<th>( \leq 1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( l_i )</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>( \sum_{i=1}^{6} D^{-l_i} )</td>
<td>( \leq 1 )</td>
</tr>
<tr>
<td>( D = 2 )</td>
<td>1/2</td>
<td>1/2</td>
<td>1/4</td>
<td>1/8</td>
<td>1/4</td>
<td>1/8</td>
<td>7/4</td>
<td>no</td>
</tr>
<tr>
<td>( D = 3 )</td>
<td>1/3</td>
<td>1/3</td>
<td>1/9</td>
<td>1/27</td>
<td>1/9</td>
<td>1/27</td>
<td>26/27</td>
<td>yes</td>
</tr>
<tr>
<td>( D = 4 )</td>
<td>1/4</td>
<td>1/4</td>
<td>1/16</td>
<td>1/64</td>
<td>1/16</td>
<td>1/64</td>
<td>21/32</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 1: Kraft inequality for some values of \( D \), problem 5.2

Hence, \( D = 3 \). Since although Kraft inequality holds for \( D = 3 \) and \( D = 4 \), it’s more convenient to choose the minimum amount of symbols for the alphabet \( D \).

To explain the title of the problem, recall the binary numeric system which has only two symbols 0 and 1, widely associated with the two possible states a transistor can have when working as a controlled switch and the reason why computers use binary numbers. As well the decimal system is known to be originated on the number of fingers human beings have, 10. So, we could relate the ternary notation of this problem with a possible Martian race of beings having three fingers.