(a) Construct a binary Huffman code

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.58</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.23</td>
<td>0.35</td>
<td>0.42</td>
</tr>
<tr>
<td>2</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.22</td>
<td>0.23</td>
</tr>
<tr>
<td>3</td>
<td>0.07</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td>0.10</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.06</td>
<td>0.07</td>
<td>0.07</td>
<td>0.07</td>
<td>0.10</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>0.06</td>
<td>0.06</td>
<td>0.06</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

\[ L = 4 \times 0.23 + 3 \times 0.35 + 0.42 = 0.92 + 1.05 + 0.42 = 2.39 \text{ bits} \]

\[ H(X) \approx 2.333 \text{ bits} \]

\[ L - H(X) = 0.057 \text{ bits} \]

(b) Construct a ternary Huffman code

<table>
<thead>
<tr>
<th>Symbol</th>
<th>Probability</th>
<th>Code</th>
<th>Code</th>
<th>Code</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
<td>0.42</td>
</tr>
<tr>
<td>1</td>
<td>0.22</td>
<td>0.22</td>
<td>0.22</td>
<td>0.36</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>0.13</td>
<td>0.16</td>
<td>0.16</td>
<td>0.22</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>0.07</td>
<td>0.13</td>
<td>0.13</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.06</td>
<td>0.07</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.05</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ L = 3 \times 0.16 + 2 \times 0.2 + 0.64 = 1.52 \]

\[ H(X) = \sum_{i=1}^{7} p_i \log_3 \frac{1}{p_i} = 1.472 \]

\[ L - H(X) = 0.048 \]
Problem 2 (Cover-Thomas, prob. 5.27)

(a) Let $S$ be the target set. Clearly, the code itself is in $S$, so take $S_1 = C$. Then a string $x \in X^*$ is in $S$ if either there is $y \in S_1$ s.t. $xy \in S$ or there is a $z \in S$ s.t. $zx \in S_1$.

Construct $S$ iteratively: $S = U_{i \geq 1} S_i$, where for $i \geq 2$

$$S_i = \{ x \in X^* : \exists y \in S_1, yx \in S_{i-1} \} \cup \{ x \in X^* : \exists z \in S_{i-1}, zx \in S_i \}$$

(b) For all $i \geq 1$, $l(x) \leq l_{\text{max}}$ for all $x \in S_i$, so $|S| \leq 2^{l_{\text{max}}}.$

(c) $\{0, 10, 11\}$ is prefix, so of course it is U.D.

$\{0, 01, 11\}$ we have $S_1 = \{0, 01, 11\}, S_2 = \{1\}$ (a suffix of 01), $S_3 = S_2$, $S_{n+1} = S_n$ for all $n \geq 3$.

$\{0, 01, 10\}$ we have $S_2 = \{1\}$ (a suffix of 01), $S_3 = \{0\}$ because 01 is a codeword, but 00 $\in C$, so this code is not U.D.

$\{0, 01\}$ This code is U.D. $S_2 = \{1\}, S_3 = S_4 = \ldots = \emptyset$

$\{00, 01, 10, 11\}$ is prefix, so U.D.

$\{110, 11, 10\}$ we have $S_2 = \{0\}, S_3 = S_4 = \ldots = \emptyset$, so this code is U.D.

$\{110, 11, 100, 00, 10\}$ we have $S_2 = \{0\}$ (since 100 $\in C$), $S_3 = \{0\}$ (for the same reason, etc.)

(d) An infinite string of codewords is parsed uniquely if the code is prefix, so the property in question cannot occur in a prefix code, or in any instantaneous code.

At the same time, for the code $\{0, 01, 11\}$ the infinite string $01111...$ can begin either with 0, or with 01.

A similar example is possible for the last code: the sequence $10000...$ is not uniquely decodable.
Problem 3

(a) We have \( I(U; \hat{U}) = \log M - H(U|\hat{U}) \geq \log M - p \log M - h(p) \),

where \( p = P(U \neq \hat{U}) \).

By the Data Processing Inequality we have

\[ I(X; Y) \geq I(U; \hat{U}) \geq \log M - p \log M - 1, \]

or \( p \log M \geq \log M - I(X; Y) - 1 \).

Conclude by dividing by \( \log M \).

(b) Let \( \mathbf{p}_x = (p_1, \ldots, p_m) \); form the pmf \( \mathbf{p}' = (p_{\text{max}}, \frac{1-p_{\text{max}}}{m-1}, \ldots, \frac{1-p_{\text{max}}}{m-1}) \).

Compute \( D(P_x \parallel \mathbf{p}') = p_{\text{max}} \).

Compute

\[ D(P_x \parallel \mathbf{p}') = \sum_{i=1}^{m} p_i \log \frac{p_i}{p_{\text{max}}} (m-1) = \sum_{i \neq j} p_i \log p_i + (1-p_{\text{max}}) \log \frac{m-1}{1-p_{\text{max}}} \]

\[ = -H(X) + p_{\text{max}} \log \frac{1}{p_{\text{max}}} + (1-p_{\text{max}}) \log \frac{1}{1-p_{\text{max}}} + (1-p_{\text{max}}) \log (m-1) \]

Observe that \( 1-p_{\text{max}} \) is the probability of not guessing \( X \) correctly in the absence of any information about \( X \), so \( p = 1-p_{\text{max}} \).

We have

\[ D(P_x \parallel \mathbf{p}') = -H(X) - h(p) + p \log (1|X| - 1) \geq 0 \]

with equality if and only if \( P_x = \mathbf{p}' \).
Problem 4

(a) We have \( P(X = i) = \binom{n}{i} p^i (1 - p)^{n-i} \) and if \( i < np \), then
\[ P(X = i - 1) < P(X = i) \]

Then
\[ P(X \leq \lambda n) \leq (\lambda n + 1) P(X = \lambda n) \leq (\lambda n + 1) 2^{-nD(\lambda llp)} \tag{1} \]
where the last inequality follows from Theorem 11.1.4.

Now write
\[ (\lambda n + 1) 2^{-nD(\lambda llp)} = 2^{-nD(\lambda llp)} = -nD(\lambda llp) \tag{2} \]

Now let us use the lower bound in Theorem 11.1.4:
\[ P(X \leq \lambda n) \geq P(X = \lambda n) \geq \frac{1}{n+1} 2^{-nD(\lambda llp)} \tag{3} \]

Estimates (1) - (3) imply that \( P(X \leq \lambda n) \sim 2^{-nD(\lambda llp)} \).

(b) Use symmetry as follows: Let \( \tilde{p} = 1 - p \); \( j = n - i \), then we have \( 1 - \lambda < 1 - \tilde{p} \)

\[ \sum_{i=\lambda n}^{n} \binom{n}{i} p^i (1 - p)^{n-i} = \sum_{i=\lambda n}^{n} \binom{n}{i} (1 - p)^{n-i} \tilde{p}^i = \sum_{j=0}^{\tilde{p}^{n-i}} \binom{n}{j} \tilde{p}^j (1 - p)^{n-i} = 2^{-nD(\tilde{p} llp)} \]

(c) In this case use the fact that \( P(X \leq \lambda n) \) is a complement of the tail of the distribution:
\[ P(X \leq \lambda n) = 1 - P(X > \lambda n) \sim 1 - 2^{-nD(\lambda llp)} \]

\[ \uparrow \text{part (b) take } n \geq \frac{1}{D(\lambda llp)} \log \frac{1}{\varepsilon} \]

(d) We have \( H(Y) \approx 0.8813 \).

(please see next page)
Problem 4 part (d).

We have $H(Y) = 0.8813$, so the typical set is formed of sequences $y^n$, $n=1000$ such that

$$-931.3 \leq \log_2 P_{yn}(y^n) \leq 2 = 931.3$$

We find (by computer), that sequences $y^n$ that satisfy (i) have $m$ ones and $n-m$ zeros, where

$$260 \leq m \leq 340$$

Thus $A_{n}^{(n)} = \{ y^n \text{ such that } y^n \text{ contain at least 260 and at most 340 ones} \}$

$$P(A_{n}^{(n)}) = \sum_{m=260}^{340} \binom{n}{m} p^m (1-p)^{n-m} \approx 0.994 ;$$

$$\log_2 |A_{n}^{(n)}| \approx 920.62$$

At the same time, $\frac{p^m (1-p)^{n-m}}{p^{m+1} (1-p)^{n-m-1}} = \frac{1-p}{p} > 1$ for all $m$, so $P_{yn}(y^n)$ is maximum when $y^n$ contains only 0's and no 1's.

The set $B_{n}^{(n)}$ is therefore formed of all sequences with $m$ zeros, $m=1000, 999, \ldots$ as long as required to obtain $P_{yn}(B_{n}^{(n)}) \geq 0.95$.

Answer: $B_{n}^{(n)} = \text{all } y^n \text{ with } 0 \leq \# \text{ones} \leq 632.3 \text{ and 46\% of } y^n \text{ with 324 ones}.$

$P_{yn}(B_{n}^{(n)}) \approx 0.950065.$

$\log_2 |B_{n}^{(n)}| \approx 903.92$

The most probable $y^n = 0^{1000} \notin A_{n}^{(n)}$.
Problem 5.

Take a large-probability set of source sequences $B = X^k$ and a sufficiently large $k$ so that

$$P_{X^k}(B) \geq 1 - \varepsilon$$  \hspace{1cm} (1)

Suppose moreover that $B$ has the smallest size among all such subsets. Let $A_{\delta}^{(k)}$ be a $\delta$-typical set of $X^k$ and suppose that $k$ is such that

$$P(A_{\delta}^{(k)}) > 1 - \varepsilon$$

(omitting $X^k$ from $P_{X^k}$)

From (1) and (2) we have

$$P(B \cap A_{\delta}^{(k)}) \geq 1 - 2\varepsilon.$$  \hspace{1cm} (3)

Using $|A_{\delta}^{(k)}|$ the estimate for $P(x^k)$, where $x^k \in A_{\delta}^{(k)}$

$$P(x^k) \leq 2^{-k(H(x) - \delta)}$$

We obtain from (3)

$$|A_{\delta}^{(k)} \cap B| \geq (1 - 2\varepsilon) \frac{2^k}{2}$$  \hspace{1cm} (4)

Now suppose that $k$ is large enough $|B| \leq 2^n$. We will encode only sequences in $B$, mapping them on binary vectors of length $n = n(k, \varepsilon)$. The error probability $p_e \leq \varepsilon$. We have

$$\limsup_{k \to \infty} \frac{\log |B|}{k} \leq \limsup_{k \to \infty} \frac{\log |A_{\delta}^{(k)}|}{k} \leq H(x) + \delta$$  \hspace{1cm} (5)

At the same time, from (4)

$$\liminf_{k \to \infty} \frac{\log |B|}{k} \geq H(x) - \delta + \frac{2}{k} \liminf_{k \to \infty} \left(\frac{\log (1 - 2\varepsilon) - \frac{2}{k}}{k}\right) = H(x) - \delta.$$  

The last two equations imply that

$$\lim_{k \to \infty} \frac{\log |B|}{k} = H(x).$$