**Problem 5.1:** Consider a variable-length code $C$ for a DMS with alphabet $\mathcal{X} = \{1, 2, \ldots, m\}$ and distribution $P\{X = i\} \triangleq p_i$, $i = 1, \ldots, m$. Let $l_1, \ldots, l_m$ be the lengths of the codewords in $C$. Define

$$L = \sum_{i=1}^{m} p_i l_i^{100}, \qquad L_1 \triangleq \min_{C \in \mathcal{P}} L \quad \text{and} \quad L_2 \triangleq \min_{C \in \mathcal{U}} L,$$

where $\mathcal{P}$ denotes the set of all prefix codes, and $\mathcal{U}$ is the set of all uniquely decodable (UD) codes. Since $\mathcal{P} \subset \mathcal{U}$ we obviously have that $L_1 \geq L_2$.
Now let $C^*$ (with codeword lengths $l_1^*, l_2^*, \ldots, l_m^*$) be the optimal UD code that achieves $L_2$. Since $C^*$ is a UD code, then its lengths $\{l_i^*\}_{i=1}^m$ satisfy Kraft's inequality. The latter implies that there exists a prefix code $C' \in \mathcal{P}$ with the *same* lengths $\{l_i^*\}_{i=1}^m$. Hence $L(C') = \sum_{i=1}^m p_i (l_i^*)^{100} = L_2$. But by definition of $L_1$ we have that $L_1 \leq L(C')$. Thus $L_1 \leq L_2$.
Therefore

$$L_1 = L_2.$$

**Problem 5.3:** Recall that $\sum_{i=1}^m D^{-l_i} < 1$ iff there is an internal node in the code tree which has $< D$ children. The sequence which corresponds to the missing branch cannot be decoded.
For example, if $D = 3$ and

$$\begin{aligned} \alpha(x_1) &= 00 \\ \alpha(x_2) &= 01 \\ \alpha(x_3) &= 020 \\ \alpha(x_4) &= 021 \\ \alpha(x_5) &= 1 \\ \alpha(x_6) &= 2 \end{aligned}$$

then $\{l_i\} = \{1, 1, 2, 2, 3, 3\}$ and $\sum_{i=1}^6 3^{-l_i} < 1$. In this example, the sequence 022 can not be decoded.

**Problem 5.4:**

**(a)** Binary Huffman code:

$$\begin{aligned} p(x_1) &= 0.49; & \alpha(x_1) &= 0 \\ p(x_2) &= 0.26; & \alpha(x_2) &= 10 \\ p(x_3) &= 0.12; & \alpha(x_3) &= 110 \\ p(x_4) &= 0.04; & \alpha(x_4) &= 11100 \\ p(x_5) &= 0.04; & \alpha(x_5) &= 11101 \\ p(x_6) &= 0.03; & \alpha(x_6) &= 11110 \\ p(x_7) &= 0.02; & \alpha(x_7) &= 11111 \end{aligned}$$

**(b)** $L(\alpha) = 2.02$ bits/source sample (while the source entropy is $H(X) = 2.01$ bits/source sample).

**(c)** Ternary Huffman code:

$$\begin{aligned} p(x_1) &= 0.49; & \alpha(x_1) &= 0 \\ p(x_2) &= 0.26; & \alpha(x_2) &= 1 \\ p(x_3) &= 0.12; & \alpha(x_3) &= 20 \\ p(x_4) &= 0.04; & \alpha(x_4) &= 21 \\ p(x_5) &= 0.04; & \alpha(x_5) &= 220 \\ p(x_6) &= 0.03; & \alpha(x_6) &= 221 \\ p(x_7) &= 0.02; & \alpha(x_7) &= 222 \end{aligned}$$

In this case, $L(\alpha) = 1.34$ ternary digits/source sample; while $H_3(X) = 1.27$ ternary digits/source sample.

**Problem 5.8:**
Code $C(S_1)$ given that the state is $S_1$:

| Next State | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|
| $p(\cdot \vert S_1)$ | 1/2 | 1/4 | 1/4 |
| $C(S_1)$ | 0 | 10 | 11 |

So $\bar{l}_{C(S_1)} = (\frac{1}{2})(1) + (2)(\frac{1}{2}) = 1.5$ bits/source sample.

Code $C(S_2)$ given that the state is $S_2$:

| Next State | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|
| $p(\cdot \vert S_2)$ | 1/4 | 1/2 | 1/4 |
| $C(S_2)$ | 10 | 0 | 11 |

So $\bar{l}_{C(S_2)} = 2(\frac{1}{2}) + (1)(\frac{1}{2}) = 1.5$ bits/source sample.

Code $C(S_3)$ given that the state is $S_3$:

| Next State | $S_1$ | $S_2$ | $S_3$ |
|---|---|---|---|
| $p(\cdot \vert S_3)$ | 0 | 1/2 | 1/2 |
| $C(S_3)$ | – | 0 | 1 |

So $\bar{l}_{C(S_3)} = 1$ bit/source sample.

The average message length of the next symbol conditioned on the previous state $S = S_i$, is $\bar{l}_{C(S_i)}$ as computed above. Note that each of the above codes achieve the conditional entropy $H(U_n|U_{n-1} = S_i)$ lower bound.

To find the unconditional average # of bits per source symbol and the entropy rate $H(\mathcal{U})$, we need first to obtain the stationary distribution of the Markov chain, $\pi = (\pi_1, \pi_2, \pi_3)$ by solving

$$(\pi_1, \pi_2, \pi_3) = (\pi_1, \pi_2, \pi_3) \begin{bmatrix} 1/2 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{bmatrix}$$

This yields: $\pi_1 = 2/9$, $\pi_2 = 4/9$, $\pi_3 = 1/3$. Thus $\pi = (\pi_1, \pi_2, \pi_3) = (2/9, 4/9, 1/3)$. Hence, the unconditional average # of bits per source symbol is :

$$\begin{aligned} \bar{l} &= E[\bar{l}_{C(S_i)}] \\ &= \sum_{i=1}^{3} \pi_i \bar{l}_{C(S_i)} \\ &= (\frac{2}{9})(1.5) + (\frac{4}{9})(1.5) + (\frac{1}{3})(1) \\ &= \frac{4}{3} \text{ bits/source symbol,} \end{aligned}$$

while the entropy rate of the Markov chain is:

$$\begin{aligned} H(\mathcal{U}) &= H(U_2|U_1) \\ &= \sum_{i=1}^{3} \pi_i H(U_2|U_1 = S_i) \\ &= (\frac{2}{9})(1.5) + (\frac{4}{9})(1.5) + (\frac{1}{3})(1) \\ &= \frac{4}{3} \text{ bits/source symbol.} \end{aligned}$$

Hence: $\bar{l} = H(\mathcal{U})$; so in this case the entropy rate (which is the smallest theoretically achievable compression rate) is achieved.

**Problem 5.12:**

$\mathcal{X} = \{x_1, x_2, x_3, x_4\}$ with probability distribution $P = \{1/3, 1/3, 1/4, 1/12\}$.

**(a)**

$$\begin{aligned} p(x_1) = 4/12; \quad \alpha(x_1) &= \ 0 \\ p(x_1) = 4/12; \quad \alpha(x_2) &= \ 10 \\ p(x_1) = 3/12; \quad \alpha(x_3) &= \ 110 \\ p(x_1) = 1/12; \quad \alpha(x_4) &= \ 111 \end{aligned}$$

with $\bar{l} = 2$ bits/sample while $H(\mathcal{X}) = 1.855$ bits/sample.

**(b)** Both sets of lengths $(1, 2, 3, 3)$ and $(2, 2, 2, 2)$ satisfy the Kraft inequality and they both yield the same average codeword length of 2 bits per source symbol. Hence they are both optimal. For example, the second optimal Huffman code with length $(2, 2, 2, 2)$ is:

$$\begin{aligned} p(x_1) = 4/12; \quad \alpha(x_1) &= \ 00 \\ p(x_1) = 4/12; \quad \alpha(x_2) &= \ 01 \\ p(x_1) = 3/12; \quad \alpha(x_3) &= \ 10 \\ p(x_1) = 1/12; \quad \alpha(x_4) &= \ 11 \end{aligned}$$

**(c)** The Huffman code of part (a) assigns symbol $x_3$ a codeword of length 3. But the probability of $x_3$ is $1/4$, so its associated Shannon code length is $\log_2 4 = 2$. Hence, the Huffman codeword for a particular source symbol may be longer than the Shannon codeword for that symbol. However, on the average, the Huffman code yields the *smallest* expected codeword length.

**Problem 5.14:**

**(a)**

$$\begin{aligned} p(x_1) = 1/21; \quad \alpha(x_1) &= \ 0111 \\ p(x_2) = 2/21; \quad \alpha(x_2) &= \ 0110 \\ p(x_3) = 3/21; \quad \alpha(x_3) &= \ 010 \\ p(x_4) = 4/21; \quad \alpha(x_4) &= \ 11 \\ p(x_5) = 5/21; \quad \alpha(x_5) &= \ 10 \\ p(x_6) = 6/21; \quad \alpha(x_6) &= \ 00 \end{aligned}$$

$L(\alpha) = 51/21$ bits/sample.

**(b)**

$$\begin{aligned} p(x_1) = 1/21; \quad \alpha(x_1) &= \ 221 \\ p(x_2) = 2/21; \quad \alpha(x_2) &= \ 220 \\ p(x_3) = 3/21; \quad \alpha(x_3) &= \ 21 \\ p(x_4) = 4/21; \quad \alpha(x_4) &= \ 20 \\ p(x_5) = 5/21; \quad \alpha(x_5) &= \ 1 \\ p(x_6) = 6/21; \quad \alpha(x_6) &= \ 0 \end{aligned}$$

$L(\alpha) = 34/21$ ternary digits/sample. (Here, you should merge together only the two least likely symbols in the first step of the algorithm.)

**Problem B:**

**(i)** The code {0,1,11} is *not* UD since, for example, the sequence 111 can be decoded into the following different ways:

- 1,1,1
- 1,11
- 11,1

Furthermore, Kraft's inequality is violated.

**(ii)** The code {0,10,11} is UD since it is a prefix code.

**(iii)** The code {0,01,11} is also UD since it is a suffix code (by reversing the order of the codewords, the reversed codewords satisfy the prefix condition, hence we can uniquely decode them).

**(iv)** The code {0,01,10} is not UD.For example, the sequence 010 has 2 valid parsings.

**(v)** The code {0,01} is a suffix code, hence it is UD.

**(vi)** The code {00,01,10,11} is a prefix code, hence it is UD.

**Problem C:**
Symmetric Markov source: $\mathcal{X}= \{0,1\}$, $P(X_i = 0) = \frac{1}{2}$.

**(a)** First-order Huffman code:

$0 \longrightarrow 0$

$1 \longrightarrow 1$

$\bar{l}_1 = \overline{R}_1 = 1$ bit/source sample.

$H(\mathcal{X}) = h_b(0.8) = 0.7219$ bits/source sample.

**(b)** Second-order Huffman code ($n = 2$): $\mathcal{X}^2 = \{00, 01, 10, 11\}$.

$P(00) = 0.8(\frac{1}{2}) = 0.4 = P(11)$

$P(01) = 0.2(\frac{1}{2}) = 0.1 = P(10)$

$00 \longrightarrow 0$

$11 \longrightarrow 10$

$01 \longrightarrow 110$

$10 \longrightarrow 111$

$\bar{l}_2 = (1)(0.4) + (2)(0.4) + (3)(0.1) + (3)(0.1) = 1.8$ bits/source sample.
$\overline{R}_2 = \frac{1}{2}\bar{l}_2 = \frac{1.8}{2} = 0.9$ bits per source sample while $H(\mathcal{X}) = H(X_2|X_1) = h_b(0.8) = 0.7219$ bits/source sam
So this code is better than the first-order one.

**(c)** Third-order Huffman code ($n = 3$):

$$\mathcal{X}^3 = \{000, 001, 010, \ldots, 111\} \triangleq \{a_0, a_1, a_2, \ldots, a_7\}.$$

$$
\begin{aligned}
p(a_0) &= p(000) = \frac{1}{2}(0.8)^2 = 0.32 & p(a_4) &= p(100) = \frac{1}{2}(0.2)(0.8) = 0.08 \\
p(a_1) &= p(001) = \frac{1}{2}(0.8)(0.2) = 0.08 & p(a_5) &= p(101) = \frac{1}{2}(0.2)^2 = 0.02 \\
p(a_2) &= p(010) = \frac{1}{2}(0.2)(0.2) = 0.02 & p(a_6) &= p(110) = \frac{1}{2}(0.8)(0.2) = 0.08 \\
p(a_3) &= p(011) = \frac{1}{2}(0.2)(0.8) = 0.08 & p(a_7) &= p(111) = \frac{1}{2}(0.8)^2 = 0.32
\end{aligned}
$$

And the code is:

(0.32) $a_0 \longrightarrow 00$

(0.32) $a_7 \longrightarrow 01$

(0.08) $a_1 \longrightarrow 100$

(0.08) $a_3 \longrightarrow 101$

(0.08) $a_4 \longrightarrow 110$

(0.08) $a_6 \longrightarrow 1110$

(0.02) $a_2 \longrightarrow 11110$

(0.02) $a_5 \longrightarrow 11111$

$\bar{l}_3 = 2(0.64) + 3(0.24) + 4(0.08) + 5(0.04) = 2.52$ bits/source sample.

$\overline{R}_3 = \frac{\bar{l}_3}{3} = \frac{2.52}{3} = 0.84$ bits/source sample vs. $H(\mathcal{X}) = 0.7219$ bits/source sample. In general as $n \longrightarrow \infty$, $\overline{R} = \frac{\bar{l}}{n} \longrightarrow H(\mathcal{X})$.

**Please let me know if you see a mistake in this solution.**