Student Number: [redacted] Name: Bryan Hoang

3. (15 points)

(a) **Answer:**

Claim. The statement is $\boxed{\text{true}}$.

Proof. Let a_1, a_2, and a_3 be source symbols we want to encode, with codewords c_1, c_2, and c_3 of corresponding lengths l_1, l_2 and l_3. We have 3 cases of binary prefix codes to consider.

Case 1. WLOG, suppose that a_1 is decoded first. That means $c_1 = 00$. Then suppose that a_2 is decoded next, which means that $c_2 = 10$. The next part of the sequence, 110, can't be decoded by a_1 or a_2, so then a_3 is decoded next. Then $c_3 = 110$.

Continuing to decode the rest of the sequence yields $a_1, a_2, 010$. 010 can't be decoded using the available codewords used so far, so we must conclude that sequence can't be a concatenation of the codewords in this case.

Case 2. WLOG, suppose that a_1 is decoded first. That means $c_1 = 00$. Then suppose that a_3 is decoded next, which means that $c_3 = 101$. The next part of the sequence, 110, can't be decoded by a_1 or a_3, so then a_2 is decoded next. Then $c_2 = 10$.

Continuing to decode the rest of the sequence yields $a_1, a_2, 010$. 010 can't be decoded using the available codewords used so far, so we must conclude that sequence can't be a concatenation of the codewords in this case.

Case 3. WLOG, suppose that a_3 is decoded first. That means $c_3 = 001$. Then suppose that a_1 is decoded next, which means that $c_1 = 01$. The next part of the sequence, 10, can't be decoded by a_3 or a_1, so then a_2 is decoded next. Then $c_2 = 10$.

Continuing to decode the rest of the sequence yields $a_3, 00$. 00 can't be decoded using the available codewords used so far, so we must conclude that sequence can't be a concatenation of the codewords in this case.

In all possible cases of decode the sequence of bits with a binary prefix code, the sequence can't be a concatenation of the codewords in this case.

$\therefore$ the statement is **true**. $\square$

(b) **Answer:**

Claim. The statement is $\boxed{\text{false}}$.

Proof. Since the source is uniformly distributed, we have that

$$p_X(a_i) = p_i = \frac{1}{9}, \quad \forall i \in \{1, \dots, 9\}.$$

Then applying Huffman's algorithm to create code $\mathcal{C}$ yields

Student Number: [redacted] Name: Bryan Hoang

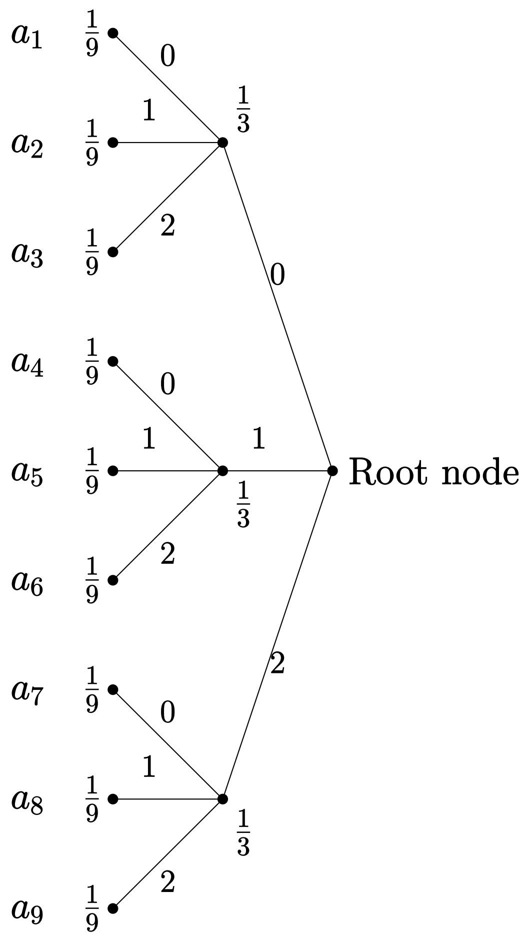

Figure 1: A Huffman tree used to design the Huffman code

From Figure 1, we can see that the Huffman code is

$$
\begin{aligned}
f : \mathcal{X} &\to \{0, 1, 2\}^* \\
a_1 &\to 00 \\
a_2 &\to 01 \\
a_3 &\to 02 \\
a_4 &\to 10 \\
a_5 &\to 11 \\
a_6 &\to 12 \\
a_7 &\to 20 \\
a_8 &\to 21 \\
a_9 &\to 22.
\end{aligned}
$$

Student Number: [redacted] Name: Bryan Hoang

Hence, the length variance of the first-order Huffman code $\mathcal{C}$ is

$$\begin{aligned}\mathrm{Var}(l(\mathcal{C})) &= \mathrm{E}[l^2(\mathcal{C})] - \mathrm{E}[l^2(\mathcal{C})]^2 \\ &= \sum_{i=1}^{9} p_i l_i^2 - \left(\sum_{i=1}^{9} p_i l_i\right)^2 \\ &= \sum_{i=1}^{9} \frac{1}{9}(2)^2 - \left(\sum_{i=1}^{9} \frac{1}{9}(2)\right)^2 \\ &= 4 - 2^2 \\ &= 0.\end{aligned}$$

$\therefore$ the statement is **false**. □

(c) **Answer:**

Claim. The statement is $\boxed{\text{true}}$.

Proof. Since the source distribution is 2-adic, the binary first-order Shannon code is optimal. If one were to compare its average code rate to the average code rate of a binary first-order Huffman code, which is optimal, they would be the same.

$\therefore$ the statement is **true**. □