1. \[ I(X^n; L^n) = H(X^n) - H(X^n | L^n) = m - \sum_{i=1}^{n} H(X_i | X^{i-1}, L^n) \]

and \[ H(X_1 | L^n) = H(X_1) \] since \( L^n \) and \( X_1 \) are independent

\[ H(X_i | X^{i-1}, L^n) = 0 \] for \( i = 2, 3, \ldots, n \)

since \( X_i \) is a function of \( X_1 \) and \( X_2 \) for \( i = 2, 3, \ldots, n \)

2. a. Conditioning on \( X = x \in X \), \( \hat{M} = n - 1 \) is a geometric variable

with probability \( p(x) \), i.e.,

\[ P(M^\prime | X = x) = p(x) (1 - p(x))^{m-1}, \quad m \geq 1 \]

\[ \Rightarrow E[\hat{M} | X = x] = \frac{1}{p(x)} \] (longer waiting time for lower probability)

\[ \Rightarrow E[\hat{M}] = \sum_{x \in X} p(x) E[\hat{M} | X = x] = 1/x_1 \]

b. \[ E[\log \hat{M}] = \sum_{x \in X} p(x) E[\log \hat{M} | X = x] \]

\[ \leq \sum_{x \in X} p(x) \log E[\hat{M} | X = x] \]

\[ = \sum_{x \in X} p(x) \log \frac{1}{p(x)} = H(X) \]
3. $M_n = X_1 X_2 \ldots X_n$

a. $\phi_{M_n}(m) = \begin{cases} \frac{1}{2^n} & \text{for } m = 2^n \\ \left(1 - \frac{1}{2^n}\right) & \text{for } m = 0 \end{cases}$

b. $E[M_n] = 2^n \frac{1}{2^n} + 0 \left(1 - \frac{1}{2^n}\right) = 1$

$$H(M_n) = -\frac{1}{2^n} \log_2 \frac{1}{2^n} - \left(1 - \frac{1}{2^n}\right) \log_2 \left(1 - \frac{1}{2^n}\right) \to 0 \quad \text{as } n \to \infty$$

c. Shannon code:

$$l_0 = \left\lceil -\log_2 \left(1 - \frac{1}{2^n}\right) \right\rceil = 1 \quad \text{if } n = 4$$

$$l_{2^n} = \left\lceil -\log_2 \frac{1}{2^n} \right\rceil = n = 4$$

$$R = \frac{1}{16} \cdot 4 + \left(1 - \frac{1}{16}\right) \cdot 1 = 1.1875$$

$$H(X) = 0.3393$$

$$R - H(X) = 0.8502 < 1$$
4. We need to identify $X_1, X_2, X_3$ using a binary code. Therefore the average length of the binary code (= number of binary questions) is bounded as

$$H(X_1X_2X_3) \leq \mathbb{E}[\# \text{questions}] \leq H(X_1X_2X_3) + 1$$

where

$$H(X_1X_2X_3) = H(p) + H(p_2) + H(p_3)$$

5. $$- \frac{1}{5} \log_2 \binom{0,0,1,0,0} = - \frac{1}{5} \log_2 (0.1 (1-0.1)^4) = 0.7860$$

$$H(X) = H(0.1) = 0.469$$

$$\Rightarrow x^5 \notin A_{0.1}^{(5)}$$

The smallest value of $\epsilon$ is

$$0.7860 - 0.469 = 0.3170$$
e. No. For instance, the average
\[ E[\mathcal{S}_n] = E[\sum_{i=1}^{n} X_i] = n \cdot E[X] = n \left( \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 2 \right) = \frac{3}{2} n \]
defends on time \( n \).

b. \( H(S_1, \ldots, S_n) = H(X_1, \ldots, X_n) = n \cdot H(\frac{1}{2}) = n \)

c. Even though it is not stationary, we can write \( H(X) = \lim_{n \to \infty} \frac{H}{n} = 1 \)

7. a. \( \mu_0 = \frac{1}{2} = P_{Y}(0) \)
\( \mu_1 = \frac{1}{2} = P_{Y}(1) \)

b. \( H(X) = \lim_{n \to \infty} \frac{1}{n} H(X(Y)X_n) = \lim_{n \to \infty} \frac{1}{n} H(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \)
\( \approx \) is a function of \( X^n \)
\[ = \lim_{n \to \infty} \frac{1}{n} H(Y^n) + \lim_{n \to \infty} \frac{1}{n} H(X^n | Y^n) \]
\[ = H(Y) + 2 = H(Y_2 | Y_1) + 2 \]
\[ = \frac{1}{2} H(\frac{1}{3}) + \frac{1}{2} H(\frac{1}{3}) + 2 \]
\[ = H(\frac{1}{3}) + 2 \]
a. In order to show that the code is prefix-free, one must show that the intervals corresponding to the given code do not overlap.

\[ \frac{1}{2^{2i}} \leq \frac{1}{2 \log P_i} \leq P_i \leq P_{i-1} \checkmark \]

b. Easy!

c. Easy!