4.1 A source produces a sequence of statistically independent binary digits with the probabilities $P(1) = 0.005$, $P(0) = 0.995$. These digits are taken 100 at a time and a binary code word is provided for every sequence of 100 digits containing 3 or fewer 1's.

(a) If the code words are all of the same length, find the minimum length required to provide the specified set of code words.

(b) Find the probability of getting a source sequence for which no code word has been provided.

(c) Use the Chebyshev inequality to bound the probability of getting a sequence for which no code word has been provided and compare with part (b).

4.2 A source produces statistically independent binary digits with the probabilities $P(0) = 3/4$, $P(1) = 1/4$. Consider sequence $\mathbf{u}$ of L successive digits and the associated inequality

$$\Pr\left[\left|\frac{I(\mathbf{u})}{L} - H(U)\right| \geq \delta\right] \leq \varepsilon \qquad \text{(i)}$$

where $H(U)$ is the entropy of the source.

(a) Find L_0 such that (i) holds for $L \geq L_0$ when $\delta = 0.05, \varepsilon = 1/10$. *Hint:* Use the Chebyshev inequality.

(b) Repeat for $\delta = 10^{-3}, \varepsilon = 10^{-6}$.

(c) Let A be the set of sequences $\mathbf{u}$ for which

$$\left|\frac{I(\mathbf{u})}{L} - H(U)\right| < \delta$$

Find upper and lower bounds for the number of sequences in A when $L = L_0$ for the cases in (a) and (b).

4.3 A source has an alphabet of 4 letters. The probabilities of the letters and two possible sets of binary code words for the source are given below:

For each code, answer the following questions (no proofs or numerical answers are required)

s_k	$P(s_k)$	$1^\#$	$2^\#$
s_1	0.4	1	1

s_2	0.3	01	10
s_3	0.2	001	100
s_4	0.1	000	1000

(a) Does the code satisfy the prefix condition?
(b) Is the code uniquely decodable?

4.4 A source has an alphabet of $K=6$, $P(s_k)$ are

$P(s_1)=1/2$, $P(s_2)=1/4$, $P(s_3)=1/8$, $P(s_4)=P(s_5)=1/20$, $p(s_6)=1/40$

Code the source using Huffman binary code, and what is its efficiency?

4.5 A source has an alphabet of $K=3$, $P(s_k)$ are 0.5，0.4，and 0.1, respectively

(a) Code the source using Huffman binary code, and what is its efficiency?

(b) Combine two symbols into a super-symbol, and do the Huffman binary coding again. what is its efficiency now?

4.6 A source has an alphabet of $K=8$, $P(s_k)$ are 0.2，0.15，0.15，0.1，0.1，0.1，0.1，0.1. Code the source using Huffman 3-ary code and compute its efficiency.

4.7 *How many fingers has a Martian?* Let

$$S=\begin{pmatrix} S_1, & \cdots, & S_m \\ p_1 & \cdots, & p_m \end{pmatrix}.$$

The *Si*'s are encoded into strings from a *D*-symbol output alphabet in a uniquely decodable manner. If m=6 and the codeword lengths are $(l_1,l_2,\ldots,l_6)$=(1,1,2,3,2,3), find a good lower bound on *D*. You may wish to explain the title of the problem.

4.8 *Huffman coding*. Consider the random variable

$$X=\begin{pmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{pmatrix}$$

(a) Find a binary Huffman code for X.
(b) Find the expected codelength and efficiency for this encoding (a).
(c) Find a ternary Huffman code for X.

(d) Find the expected codelength and efficiency for this encoding (b).

4.9 *More Huffman codes.* Find the binary Huffman code for the source with probabilities(1/3,1/5,1/5,2/15,2/15). Argue that this code is also optimal for the source with probabilities (1/5,1/5,1/5,1/5,1/5).

4.10 *Bad codes.* Which of these codes cannot be Huffman codes for any probability assignment? Please explain the reason.

(a) {0,10,11}.

(b) {00,01,10,110}.

(c) {01,10}.