Proakis

6.1-6.4-6.22-6.26

Question 1

A source has an alphabet $\{a_1, a_2, a_3, a_4, a_5, a_6\}$ with corresponding probabilities $\{0.1, 0.2, 0.3, 0.05, 0.15, 0.2\}$. Find the entropy of this source. Compare this entropy with the entropy of a uniformly distributed source with the same alphabet.

$$H(x) = \sum_{i=1}^{6} P_i \log_2 P_i$$

Question 2

Let X be a geometrically distributed random variable; i.e.,

$$P(X = k) = p(1 - p)^{k-1} \qquad k = 1, 2, 3, \ldots$$

1. Find the entropy of X.
2. Knowing that $X > K$, where K is a positive integer, what is the entropy of X?

$$\sum_{k=1}^{\infty} (1-p)^{k-1} \Rightarrow \frac{1}{1-(1-p)} \qquad \sum_{k=1}^{\infty} (1-p)^{k-1}(k-1) \Rightarrow \frac{(p-1)^2}{(1-p)p^2}$$

Question 3

A source has an alphabet $\{a_1, a_2, a_3, a_4\}$ with corresponding probabilities $\{0.1, 0.2, 0.3, 0.4\}$.

1. Find the entropy of the source.
2. What is the minimum required average code word length to represent this source for error-free reconstruction?
3. Design a Huffman code for the source and compare the average length of the Huffman code with the entropy of the source.

Question 4

Design a ternary Huffman code for a source with output alphabet probabilities given by $\{0.05, 0.1, 0.15, 0.17, 0.13, 0.4\}$. (Hint: You can add a dummy source output with zero probability.)

Question 1

$$H(X) = -\sum_{i=1}^{6} P_i \log_2 P_i = -(0.1 \log_2 0.1 + 2\times 0.2 \log_2 0.2$$
$$+ 0.3 \log_2 0.3 + 0.05 \log_2 0.05 + 0.15 \log_2 0.15)$$

$$= 2.4087 \text{ bits/symbol}$$

If the source symbols are equiprobable

then $P_i = \frac{1}{6}$

$$H_u(X) = -\sum_{i=1}^{6} P_i \log_2 P_i = -\log_2 \frac{1}{6} = \log_2 6$$

$$= 2.5850 \text{ bits/symbol}$$

As it is observed the entropy of the source is less than that of a uniformly distributed source

Question 2

(1) $$H(X) = -\sum_{k=1}^{\infty} p(1-p)^{k-1} \log_2 \left(p(1-p)^{k-1}\right)$$

$$= -p \log_2(p) \sum_{k=1}^{\infty} (1-p)^{k-1} - p \log_2(1-p) \sum_{k=1}^{\infty} (k-1)(1-p)^{k-1}$$

$$= -p \log_2(p) \frac{1}{1-(1-p)} - p \log_2(1-p) \frac{(1-p)}{(1-(1-p))^2}$$

$$= -\log_2(p) - \frac{(1-p)}{p} \log_2(1-p)$$

(2) Clearly $p(X=k \mid X>K) = 0$ for $k \leq K$

if $k > K$, then

$$p(X=k \mid X>K) = \frac{p(X=k, X>K)}{P(X>K)}$$

$$= \frac{p(1-p)^{k-1}}{P(X>K)}$$

$$P(X>K) = \sum_{k=K+1}^{\infty} p(1-p)^{k-1} = p\left(\sum_{k=1}^{\infty}(1-p)^{k-1} - \sum_{k=1}^{K}(1-p)^{k-1}\right)$$

$$= p\left(\frac{1}{1-(1-p)} - \frac{1-(1-p)^K}{1-(1-p)}\right) = (1-p)^K$$

$$p(X=k \mid X>K) = \frac{p(1-p)^{k-1}}{(1-p)^K}$$

If we let $k = K + \ell$ with $\ell = 1, 2, \ldots$

then, $$p(X=k \mid X>K) = \frac{p(1-p)^K (1-p)^{\ell-1}}{(1-p)^K}$$

$$= p(1-p)^{\ell-1}$$

that is $p(X=k \mid X>K)$ is the geometrically distributed. Hence using the results of the first part we obtain

$$H(X \mid X>K) = -\log_2(p) - \frac{1-p}{p}\log_2(1-p)$$

Question 3

1) The entropy of the source is

$$H(x) = -\sum_{i=1}^{4} p(a_i) \log_2 (p(a_i))$$

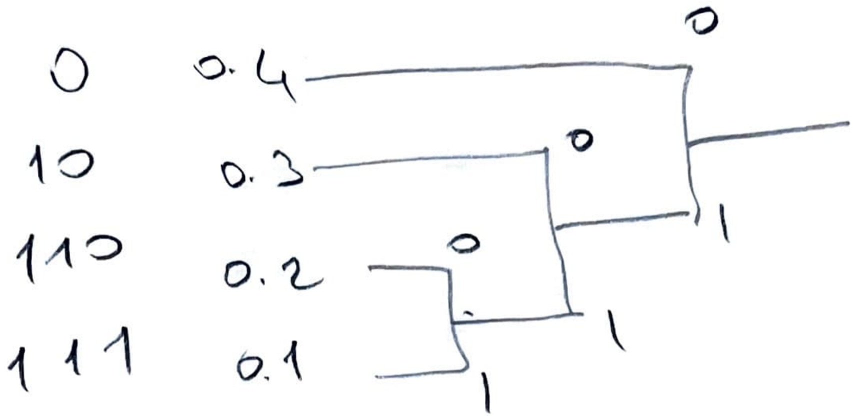
$= 1.8464$ bits/output

2) The av. codeword length is lower bounded by the entropy of the source. Hence for error free reconstruction the minimum possible average codeword length is $H(x) = 1.8464$

3) The following figure depicts the source Huffman coding scheme of the source

The average codeword length is

$$\bar{R}(x) = 3 \times (0.2 + 0.1) + 2 \times 0.3 + 0.4 = 1.9$$



Question 4

Add a dummy variable with length 0 probability to enable 3 by 3 grouping