Cédric Richard

Discrete source coding

Exercise 1

Indicate for each of the following codes whether it is regular, decodable, instantaneous and complete: $\mathcal{C}_1 = \{00, 01, 10, 11\}$, $\mathcal{C}_2 = \{0, 01, 11\}$, $\mathcal{C}_3 = \{0, 10, 11\}$, $\mathcal{C}_4 = \{0, 11, 111\}$.

Exercise 2

We consider a source S that can emit 5 symbols, each of which has a probability p_i in the table below. This table also provides two possible binary codings $\mathcal{C}_1$ and $\mathcal{C}_2$ of S. Indicate whether these codes are decodable and instantaneous. Calculate the average $\overline{n}_1$ and $\overline{n}_2$ lengths of their codewords. Compare to the minimum average codewords length $\overline{n}_{\min}$ required for S.

s_i	s_1	s_2	s_3	s_4	s_5
p_i	0.50	0.18	0.14	0.12	0.06
$\mathcal{C}_1$	0	10	11	101	1001
$\mathcal{C}_2$	00	10	11	010	011

Exercise 3

We consider a random variable X that can take n values distributed according to the following distribution: $P(X = x_i) = (1/2)^i$ for $1, 2, \ldots, n-1$, and $P(X = x_n) = (1/2)^{n-1}$. Determine the minimum average length $\overline{n}_{\min}(X)$. Propose a binary code using Huffman's method. Calculate the average length of its codewords. Discuss.

Exercise 4

A printer uses the following commands:

Raise the stylus (RS)
Press the stylus (PS)
move the stylus left (–X)
move the stylus right (+X)
move the stylus up (+Y)
move the stylus down (–Y).

Calculate the minimum average number of bits required for this set of commands if their probabilities are given by:

$$P_{\mathrm{RS}} = P_{\mathrm{PS}} = P_{\text{-X}} = 0.1 \qquad P_{\text{+X}} = 0.3 \qquad P_{\text{+Y}} = P_{\text{-Y}} = 0.2$$

Build a Shannon's binary code. Build a Huffman's binary code. Compare the two solutions.

Exercise 5

A high school has to communicate a list of undergraduate results for 2500 students. These results are as follows: 250 A, 375 B, 1125 C, 625 failed, 125 absent. Build a binary Huffman's code to compress the corresponding file. Calculate the average length of the codewords. Calculate the file size if the information are encoded using a fixed-length code with 8 bits. Evaluate the gain in file size achieved by using the Huffman's code.

Problem 1

We consider a code consisting of two words of length 2, two words of length 3 and one word of length 4.

1. Show that it exists a decodable binary code respecting these codeword lengths. Draw a possible code tree. Modify this tree in order the reduce the average codeword length.
2. We assign the following probabilities $\{0.50, 0.18, 0.14, 0.12, 0.06\}$ to the 5 states of the source. Associate these probabilities with the codewords proposed previously so as to minimize the average length of the codewords. Calculate the average codewords length and show that there exist binary codes with better performance.
3. Propose a binary code using Huffman's method. Compare the average length of its codewords to the one obtained in the previous question.

Problem 2

Consider a Markov source where $p = \frac{1}{10}$ and $q = \frac{2}{10}$.

$$P(S_n = 0|S_{n-1} = 1) = p$$
$$P(S_n = 1|S_{n-1} = 1) = 1 - p$$
$$P(S_n = 1|S_{n-1} = 0) = q$$
$$P(S_n = 0|S_{n-1} = 0) = 1 - q.$$

1. Determine the stationary distribution of the source. Calculate the entropy of the source without taking the dependency of the states into account. Calculate in this case the minimum average length of binary codewords required to encode this source.
2. Calculate the entropy of the Markov source, which assumes that the dependency of successive states is taken into account. Calculate in this case the minimum average length of binary codewords to encode this Markov source.
3. Consider the extension of order 2 for S. Calculate its entropy. Calculate the minimum average length of binary codewords to encode this source. Propose a Huffman's binary code and calculate the average length of its codewords.