Source Coding and Kraft Inequality

Tutorial : 20/03

Q1.)

Given is a stationary discrete random process $\boldsymbol{S} = \{S_n\}$ with independent and identically distributed random variables S_n. The alphabet $\mathcal{A}_S$ for the random variables S_n includes six letters and is given by $\mathcal{A}_S = \{\mathrm{a}, \mathrm{b}, \mathrm{c}, \mathrm{d}, \mathrm{e}, \mathrm{f}\}$.

(a) The following table shows three sets of codeword lengths for the given alphabet $\mathcal{A}_S$. The codeword length assignments are denoted as “set A”, “set B” and “set C”. An assignment $\ell(a_i) = M$ describes that, for a selected codeword set, a codeword of M bits is assigned to the alphabet letter a_i.

a_i	assignment of codeword lengths $\ell(a_i)$		
	set A	set B	set C
a	2	2	1
b	2	2	3
c	2	2	3
d	4	3	3
e	4	3	4
f	4	4	4

Determine for each of the given sets, using the Kraft inequality, whether it is possible to construct a uniquely decodable code with the corresponding codeword lengths. [3 points]

(b) Develop a prefix code for the set of codeword lengths given in the table below. The codeword that is assigned to a letter a_i shall have $\ell(a_i)$ bits. Write the codewords of the developed code directly into the table. [2 points]

(c) Consider the prefix code developed in (b). Is it possible to find a pmf $p(a_i)$ for which the developed code yields an average codeword length $\bar{\ell}$ that is equal to the entropy (S_n) ? Briefly explain your answer. If such a pmf exists, write down a table with the corresponding probability masses. [3 points]

<u>Note:</u> The codeword lengths for the prefix code developed in (b) are equal to the codeword lengths for “set C”, which has been investigated in (a).

Q2.)

Consider a source emitting 7 symbols, $a_1, a_2, \ldots, a_7$, with probabilities 0.49, 0.26, 0.12, 0.04, 0.04, 0.03 and 0.02, respectively. Find a binary Huffman code for the source.

Kraft Inequality:

Theorem (*Kraft inequality*) For any instantaneous code (prefix code) over an alphabet of size D, the codeword lengths $l_1, l_2, \ldots, l_m$ must satisfy the inequality

$$\sum_i D^{-l_i} \leq 1$$

Conversely, given a set of codeword lengths that satisfy this inequality, there exists an instantaneous code with these word lengths.

Proof : https://mortada.net/simple-proof-for-krafts-inequality.html