Information Theory

o **Source Coding Methods**

Source coding method means replacing the letters and symbols by using a binary code (0, 1) in order to reduce the number of required Bits for the transfer of information provided by the source. Source coding methods are:

A. **Simple Coding Method**

The simple coding method assumes that there is a set of characters with the probabilities of its appearance, and each character will symbolizes by consecutive set of the binaries. According to this method, there are no specific steps to choosing groups, but it will select based on the experience taking into account to achieve the desired qualities of encoding.

Example: Language consists of four symbols (%, $, #, @) and that the probability of each symbol is 0.3, 0.5,0.1 and 0.1 respectively. Needed is to give three encodings of this language by using the simple coding method, and the calculation of the efficiency of coding each.

1. @ → 0, # → 10, $ → 110, % → 11.
2. @ → 00, # → 01, $ → 10, % → 11.
3. @ → 0, # → 10, $ → 110, % → 1110.
4. @ → 0, # → 01, $ → 011, % → 0111.
5. @ → 0, # → 10, $ → 110, % → 111.

P@ = 0.3 P# = 0.5   P$ = 0.1   P% = 0.1

$E = H / L$

$H = \sum_{i=1}^{4} P_i \log_2(1/P_i)$
$= P_@ * \log2(1/P_@) + P_\# * \log2(1/P_\#) + P_\$ * \log2(1/P_\$) + P_\% * \log2(1/P_\%)$
$= 0.3*3.322*\log_{10}(3.333) + 0.5*3.322*\log_{10}(2) + 2(0.1*3.322*\log_{10}(10))$
$= 0.521 + 0.5 + 2 * 0.332$
$H = 1.685$ bits

$L = \sum_{i=1}^{4} P_i * L_i$
$= P_@ * L_@ + P_\# * L_\# + P_\$ * L_\$ + P_\% * L_\%$

1. $L = 0.3*1 + 0.5*2 + 0.1*3 + 0.1*2$ = 1.8 bits

∴ E = 1.685 / 1.8
= 0.936

2. $L = 0.3*2 + 0.5*2 + 0.1*2 + 0.1*2$ = 2 bits

∴ E = 1.685 / 2
= 0.843

3. $L = 0.3*1 + 0.5*2 + 0.1*3 + 0.1*4$ = 2 bits

∴ E = 1.685 / 2
= 0.843

4. $L = 0.3*1 + 0.5*2 + 0.1*3 + 0.1*4$ $= 2$ bits

$\therefore E = 1.685 / 2$
$= 0.843$

5. $L = 0.3*1 + 0.5*2 + 0.1*3 + 0.1*3$ $= 1.9$ bits

$\therefore E = 1.685 / 1.9$
$= 0.887$

B. Shannon-Fano Method

This method can be summarized by using the following algorithm:

1. Symbols and letters are placed in a table in descending order based on the values of possibilities

2. The table is divided horizontally by a straight line into two parts upper and lower, so that the **sum of probabilities in the upper part** $\geq \frac{1}{2}$ **sum of probabilities in the lower part**.

3. Put bit (0) into the upper part, and bit (1) into lower part.

4. Re steps 2 and 3 until the division end.

5. The encoding is read by compiling the binaries from the left to the right.

Example: Language has five letters A, B, C, D and E. The probability for each character is 0.5, 0.2, 0.1, 0.1 and 0.1 respectively. Needed is to give the encoding using Fano - Shannon method, then calculate the encoding efficiency.

A	0.5	0			
B	0.2	1	0		
C	0.1	1	1	0	
D	0.1	1	1	1	0
E	0.1	1	1	1	1

$\therefore$ $A = 0$ $B = 10$ $C = 110$ $D = 1110$ $E = 1111$

$H = \sum_{i=1}^{5} P_i \log_2(1/P_i)$

$= P_A * \log_2(1/P_A) + P_B * \log_2(1/P_B) + P_C * \log_2(1/P_C) + P_D * \log_2(1/P_D) + P_E * \log_2(1/P_E)$

$= 0.5*3.322*\log_{10}(2) + 0.2*3.322*\log_{10}(5) + 3(0.1*3.322*\log_{10}(10))$

$= 0.5 + 0.464 + 0.996$

$= 1.96$ bits

$$L = \sum_{i=1}^{5} P_i * L_i$$
$$= P_A*L_A + P_B*L_B + P_C*L_C + P_D*L_D + P_E*L_E$$
$$= 0.5 *1 + 0.2*2 + 0.1*3 + 2(0.1*4)$$
$$= 2$$
$$E = H / L$$
$$= 1.96 / 2$$
$$= 0.98$$

H.W: You have a sequence of symbols (X) and their probabilities (P).

(X, P) = {(x1, 0.3), (x2, 0.1), (x3, 0.05), (x4, 0.15), (x5, 0.15), (x6,0.1)}

Find :   1. Give the binary coding using Fano - Shannon method.

                2. Calculate the encoding efficiency.

C. Hoffman Method

This method can be summarized by using the following algorithm:

1. Symbols and letters are placed in a table in descending order based on the values of possibilities

2. Put bit (0) for the last letter or symbol that has less probability, and bit (1) for the letter or symbol that precede it.

3. Put the two characters or two symbols that have low probability with each other, To form a new symbol, where the new probability is the sum of the probability of two letters or symbols

4. Re steps 1 to 3 until last letter or symbol.

5. The encoding is read by compiling the binaries from the right to the left.

Example: Language has four letters A, B, C and D. The probability of each letters is 0.4, 0.35, 0.2, and 0.05 respectively. Needed is to give the encoding using Huffman method then calculate the encoding efficiency.

∴ A = 0   B = 11   C = 101   D = 100

Information Theory
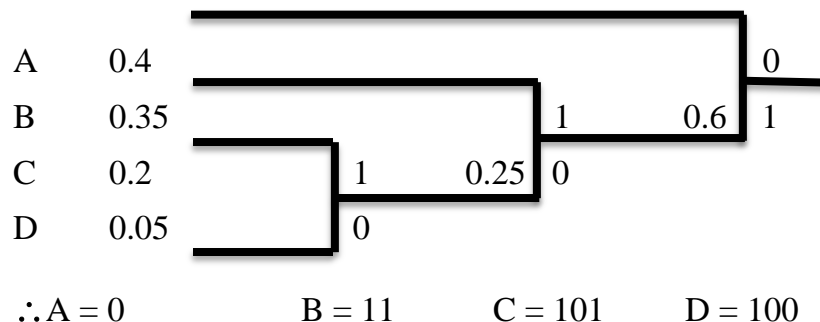
$H = \sum_{i=1}^{4}$ $P_i \log_2(1/P_i)$

$= P_A * \log_2(1/P_A) + P_B * \log_2(1/P_B) + P_C * \log_2(1/P_C) + P_D * \log_2(1/P_D)$

$= 0.4*3.322*\log_{10}(2.5) + 0.35*3.322*\log_{10}(2.857) + 0.2*3.322*\log_{10}(5)$

$\quad + 0.05*3.322*\log_{10}(20)$

$= 0.529 + 0.53 + 0.464 + 0.216$

$= 1.739$ bits

$L = \sum_{i=1}^{4}$ $P_i * L_i$

$= P_A*L_A + P_B*L_B + P_C*L_C + P_D*L_D$

$= 0.4 *1 + 0.35*2 + 0.2*3 + 0.05*3$

$= 1.85$

$E = H / L$

$= 1.7394 / 1.85$

$= 0.9402$

H.W: Language has five letters A, B, C, D and E. The probability of each character is 0.35, 0.17, 0.17, 0.15 and 0.16 respectively. Needed is to give the encoding using Huffman method then calculate the encoding efficiency.