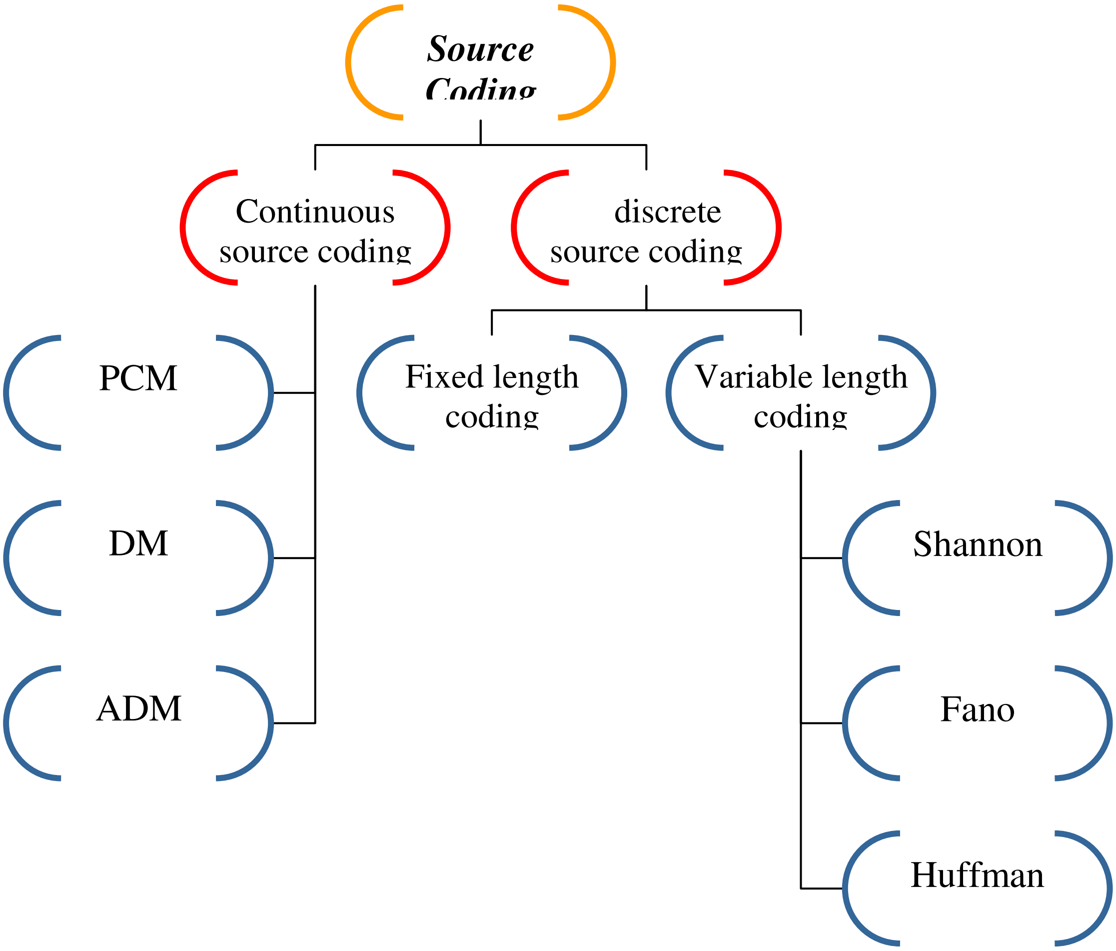
Source Coding



Source Coding of discrete Source:-

A discrete source is that source that produces finite set of messages X_1, X_2,……, X_n with prob. $P(X_1)$, $P(X_2)$ …., $P(X_n)$.

A source Coder : will transform each message into finite sequences of digits called "Codeword" of the message.
If binary digits are used in the code word then we obtain what is called "binary source coder".
Ternary Source coder is also possible.

The selection of codeword is done according to the following considerations:

1. The average code length L_c must be as minimum as possible. This L_c is given by $L_c = \sum_{i=1}^{n} L_i . P(X_i)$ bit/symbol.
2. This codeword at the receiver must be **Uniquely** decodable.

Ex: A code table for certain binary code is given as

X_i	$P(X_i)$	Codeword	L_i
X_1	0.2	0	1
X_2	0.1	10	2
X_3	0.4	110	3
X_4	0.3	111	3

Find:

a) Average code length

b) If a received sequence is 1011000111110, check if this sequence can be uniquely decoded or not.

Sol:

a) $L_c = \sum_{i=1}^{n} L_i . P(X_i)$ = [1*0.2+2*0.1+3*0.4+3*0.3 = 2.5 bit / symbol

b) Using previous code table

10	110	0	0	111	110
X_3	X_2	X_1	X_1	X_4	X_2

NOTE

For previous example, the code is not optimum in terms of L_c. we can reduce L_c by giving less L_i for X_i with higher probability. Hence the previous example can be modified.

X_i	$P(X_i)$	Codeword	L_i
X_3	0.4	0	1
X_4	0.3	10	1
X_1	0.2	110	3
X_2	0.1	111	3

$L_c = \sum_{i=1}^{n} L_i.P(X_i)$ = [1*0.4+2*0.3+3*0.2+1*0.3 = 1.9 bit / symbol

- You can notice that with this new arrangement of messages, the codeword is less than before.

Code Efficiency & Redundancy:

A code with average code length L_c as coding efficiency

$$\eta = \frac{H(X)}{L_c.\log_2 D} \qquad \text{D=2 binary , D=3 Ternary}$$

$$R = 1 - \eta$$

For binary source coder

$$\eta = \frac{H(X)}{L_c}$$

Fixed Length Code: This is used when the source produces almost equiprobable messages $P(X_1)$, $P(X_2)$ …., $P(X_n)$ then $L_1 = L_2 \dots L_n = L_c$.

For binary coding:

1. $L_c = \log_2(n)$ bit/symbol if $n = 2^r$
   $r = 1,2,3,4$
   $n = 2,4,8,16$

   gives η = 100 %
2. $L_c = \text{int}[\log_2 n]+1$ if $n \# 2^r$
   Which gives less efficiency

Ex: For ten equiprobable messages coded in fixed code length. Find : L_c, codeword, η

Sol:

n = 10
L_c = int [$\log_2$ 10]+1 = 4 bit/message

H(X)= $\log_2$ 10 = 3.3219 bit/message

$$\eta = \frac{3.3219}{4} = 83\%$$

X_i	Codeword
X_1	0000
X_2	0001
X_3	0010
X_4	0011
X_5	0100
X_6	0101
X_7	0110
X_8	0111
X_9	1000
X_{10}	1001

Ex: Find coding efficiency of a fixed code length used to encode messages obtained from a fair dice: a. Once b. Twice c. Three times

Sol:

For a fair dice, n=6 equiprobable

a. Once
   L_c = int [$\log_2$ 6]+1 = 3 bit/message

   H(X)= $\log_2$ 6 = 2.584 bit/message

$$\eta = \frac{2.584}{3} = 86.1\%$$

b. Twice

n=6*6=36

L_c = int [$\log_2 36$]+1 = 6 bit/message

H(X)= $\log_2 36 = 5.1699$

$$\eta = \frac{5.1699}{6} = 86.1\%$$

c. Three times

n=6*6*6= 216

L_c = int [$\log_2 216$]+1 = 8 bit/message

H(X)= $\log_2 216 = 7.75488$

$$\eta = \frac{7.75488}{8} = 96.936\%$$

Variable length Code: When the messages probability are not equal then we use variable length codes, these codes are some times (minimum redundancy codes).
They will be explained for binary coding (D=2) and ternary coding (D=3) will be easily modified.

1. Shannon Codes

For messages $X_1, X_2, \ldots, X_n$ with prob. $P(X_1), P(X_2) \ldots, P(X_n)$ then:-

$$L_i = \begin{cases} -\log_2 P(X_i) \ldots\ldots\ldots\ldots\ldots\ldots P(X_i) = \left(\frac{1}{2}\right)^r = \left[\frac{1}{2}, \frac{1}{4}, \frac{1}{8} \ldots\right] \\ \text{int}[-\log_2 P(X_i)] + 1 \ldots\ldots\ldots P(X_i) \neq \left(\frac{1}{2}\right)^r \end{cases}$$

Also we define $w_i = \sum_{k=1}^{i} P(X_k) \ldots\ldots\ldots\ldots 1 \geq w \geq 0$

Note :

In encoding, messages must be arranged in a decreasing order of prob.

Ex: develop the Shannon code for the following set of messages, then find:

a. Code length efficiency
b. P(0) & P(1) at the encoder output

$$P(X_i)=[0.4 \quad 0.25 \quad 0.15 \quad 0.1 \quad 0.07 \quad 0.03]$$

Sol:

X_i	$P(X_i)$	w_i	L_i	Codeword c_i	0_i	1_i
X_1	0.4	0	2	00	2	0
X_2	0.25	0.4	2	01	1	1
X_3	0.15	0.65	3	101	1	2
X_4	0.1	0.8	4	1100	2	2
X_5	0.07	0.9	4	1110	1	3
X_6	0.03	0.97	6	111110	1	5

$L_1 = \text{int}\,[-\log_2 0.4]+1 = 2$ bit/message

$L_2 = -\log_2 0.25= 2$ bit/message

$L_3 = \text{int}\,[-\log_2 0.15]+1 = 3$ bit/message

$L_4 = \text{int}\,[-\log_2 0.1]+1 = 4$ bit/message

$L_5 = \text{int}\,[-\log_2 0.07]+1 = 4$ bit/message

$L_6 = \text{int}\,[-\log_2 0.03]+1 = 6$ bit/message

$w_1 = 0$

$w_2 = P(X_1)= 0.4$

$w_3 = P(X_1)+ P(X_2)=0.65$

$w_4 = P(X_1)+ P(X_2)+ P(X_3)=0.8$

$w_5 = P(X_1)+ P(X_2)+ P(X_3)+ P(X_4)=0.9$

$w_6 = P(X_1)+ P(X_2)+ P(X_3)+ P(X_4)+ P(X_5)=0.97$

$c_1 = w_1*2=0*2=\underline{0}$, $0*2=\underline{0}$

$c_2 = w_2*2=0.4*2=\underline{0}.8$, $0.8*2=\underline{1}.6$

$c_3 = w_3*2=0.65*2=\underline{1}.3$, $0.3*2=\underline{0}.6$, $0.6*2=\underline{1}.2$

$c_4 = w_4*2=0.8*2=\underline{1}.6$, $0.6*2=\underline{1}.2$, $0.2*2=\underline{0}.4$, $0.4*2=\underline{0}.8$

$c_5 = w_5*2=0.9*2=\underline{1}.8$, $0.8*2=\underline{1}.6$, $0.6*2=\underline{1}.2$, $0.2*2=\underline{0}.4$

$c_6 = w_6*2=0.97*2=\underline{1}.94$, $0.94*2=\underline{1}.88$, $0.88*2=\underline{1}.76$, $0.76*2=\underline{1}.52$

$0.52*2=\mathbf{1}.04$, $0.04*2=\underline{0}.08$

a.

$$H(X) = -\sum_{i=1}^{6} P(X_i).\log_2 P(X_i) = 2.1918 \text{ bits/message}$$

$$L_c = \sum_{i=1}^{6} L_i.P(X_i) = 2.61 \text{ bit/message}$$

$$\eta = \frac{2.1918}{2.61} = 83.9\%$$

b.

$$P(0) = \frac{\sum 0_i.P(X_i)}{L_c} = \frac{[2*0.4+1*0.25+1*0.15+2*0.1+1*0.07+1*0.03]}{2.61} = 0.574$$

$$P(1) = \frac{\sum 1_i P(X_i)}{L_c} \quad \text{or } P(1) = 1 - P(0) = 0.426$$

H.W 1. Develop Ternary Shannon code for the following set of messages:

$P(X_i)$= [0.4 0.25 0.15 0.1 0.07 0.03] then find:

a. Code length efficiency
b. P(0) & P(1) at the encoder output

NOTES:

1. base of log in evaluating L_i will be 3.
2. the condition of L_i will be $\left(\frac{1}{3}\right)^r$.
3. w_i is changed into ternary word of length L_i.

2. Shannon-Fano Code (Fano Code)

Procedure for binary:

a. Arrange messages in decreasing order of probability.
b. Find out a point in that order in which the sum of probability upward is almost equal to the sum of probability downward. Assign all messages upward as "0" and all messages downward as "1".
c. Repeat the previous step many times on upward and downward until all messages are separated.

Ex: Develop Shannon-Fano Code for the following messages:
$P(X_i)$= [0.4 0.25 0.15 0.1 0.07 0.03] then find:

a. Code length efficiency
b. P(0) & P(1) at the encoder output

Sol:

X_i	$P(X_i)$	Codeword c_i	L_i	0_i	1_i
X_1	0.4	0	1	1	0
X_2	0.25	10	2	1	1
X_3	0.15	110	3	1	2
X_4	0.1	1110	4	1	3
X_5	0.07	11110	5	1	4
X_6	0.03	11111	5	0	5

a.

H(X) = 2.1918 bits/message
L_c = 2.25 bit/message

$$\eta = \frac{2.1918}{2.25} = 97.413\%$$

b.

$$P(0) = \frac{\sum 0_i . P(X_i)}{L_c} = 0.4311$$

$$P(1) = \frac{\sum 1_i P(X_i)}{L_c}$$ or P(1)= 1 – P(0) = 0.5689

H.W 2. Develop Ternary Shannon-Fano code for the following set of messages:$P(X_i)$= [0.4 0.25 0.15 0.1 0.07 0.03] then find:

a. Code length efficiency
b. P(0) & P(1) at the encoder output

Hint: Find out two lines in each step that split the sum of prob. Into almost three equal parts giving them as 0,1,2.

3. Huffman Code:

Procedure:

a. Arrange messages in decreasing order of probability.
b. The two lowest probability messages are summed (joined) assigning one of them "0" and the other by "1". This sum will replace these two messages when messages are rewritten in a decreasing order.
c. Repeat the previous step many times until all messages are joined (summed) up with total sum of "1" unity.
d. Codeword for each message are read from left to right and writing the codeword bits from right to left.

Ex : Develop Shannon-Fano Code for the following messages:
$P(X_i)$= [0.4 0.25 0.15 0.1 0.07 0.03] then find code length efficiency

Sol:

X_i	$P(X_i)$									
X_1	0.4		0.4		0.4		0.4		0.6	0
X_2	0.25		0.25		0.25		0.35	0	0.4	1
X_3	0.15		0.15		0.2	0	0.25	1		
X_4	0.1		0.1	0	0.15	1				
X_5	0.07	0	0.1	1						
X_6	0.03	1								

X_i	c_i	L_i
X_1	1	1
X_2	01	2
X_3	001	3
X_4	0000	4
X_5	00010	5
X_6	00011	5

$H(X) = 2.1918$ bits/message
$L_c = 2.25$ bit/message

$$\eta = \frac{2.1918}{2.25} = 97.4\%$$

H.W 3. Develop Ternary Shannon-Fano code for the following set of messages: $P(X_i)= [0.4 \quad 0.25 \quad 0.15 \quad 0.1 \quad 0.07 \quad 0.03]$ then find Code length efficiency