Unit-V

Information Theory

Since its inception, the main role of Information Theory has been to provide the engineering and scientific communities with a mathematical framework for the theory of communication by establishing the fundamental limits on the performance of various communication systems.

5.1 Information source

The system which produces message or information is called **information source.**
The message may be an electrical message, speech message or picture message
Information source can be classified as
1. Analog source
2. Discrete source

5.1.1 Discrete message:

The output emitted by a source at unit time interval is known as **discrete messages.**
It can be represented using discrete random variable which takes symbols from fixed finite alphabet S.

$S = \{s_0, s_1, s_2 \dots . s_{k-1}\}$

It is shown in Fig. 5.1.

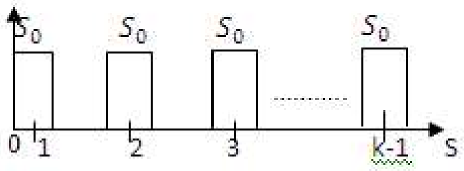


Fig.5.1 .Representation of Discrete Messages

5.1.2 Information content

The information emitted from source is called **information content**
Total information content = entropy + redundant information content

5.1.3 Redundant Information

The bits which we used to check errors is known as **redundant information**

5.2 Concept of amount of information

The amount of information or messages transmitted over a channel is represented in statistical terms such as probability of occurrence.

Principle:

If the probability of occurrence is more, the amount of information is less. Similarly If the probability of occurrence is less, the amount of information is more.

Assume x_j as an event, so that $p(x_j)$ is the occurance of an event

Then the information due toe the event is

$I(x_j) = log_2 \left[\frac{1}{p(x_j)}\right]$ 5.1

Assme x_j and y_k as two independent events

Therefore $I(x_j, y_k) = \log_2 \left[\frac{1}{p(x_j y_k)}\right]$ 5.2

$= log_2 \left[\frac{1}{p(x_j)}\right] + log_2 \left[\frac{1}{p(y_k)}\right]$ 5.3

$= I(x_j) + I(y_k)$ 5.4

Where $I(x_j) = log_2 \left[\frac{1}{p(x_j)}\right]$ 5.5

The above expression indicates that the amount of information is related to the logarithmic of the inverse of the probability of occurane of an event $p(x_j)$

5.2.1 Units of Information

Different units of information can be defined fro different bases of algorithms

Base '2'=the unit is bit

Base '10'= the unit is decit

5.3 Average Informtaion or Entropy

Definition

It is defined as the process of producing average information per individual message in a particular interval

Let L be the total message, $m_1, m_2 \dots . m_k$ represents discreate messages

and $p_1, p_2 \dots . p_k$ represents the probability of discreate messages.

The amount of information in m_1 is

$I_1 = log_2 \left[\frac{1}{p_1}\right]$ 5.6

The total amount of information due to m_1 message is

$$I_{t1} = p_1 L log_2 \left[\frac{1}{p_1}\right]$$ 5.7

The total information due to L message is

$$I_t = I_{t1} + I_{t2} + \cdots \ldots . I_{tk}$$ 5.8

$$I_t = p_1 L log_2 \left[\frac{1}{p_1}\right] + p_2 L log_2 \left[\frac{1}{p_2}\right] + \cdots \ldots . p_k L log_2 \left[\frac{1}{p_k}\right]$$ 5.9

The average information per message will be

$$Average\ information = \frac{Total\ information}{Total\ Messages}$$ 5.10

$$Average\ information = \frac{I_t}{L}$$ 5.11

The average information per message is also called as entropy

$$H(S) = \frac{I_t}{L}$$ 5.12

$$H(S) = \frac{p_1 L log_2\left[\frac{1}{p_1}\right] + p_2 L log_2\left[\frac{1}{p_2}\right] + \cdots \ldots p_k L log_2\left[\frac{1}{p_k}\right]}{L}$$ 5.13

$$H(S) = L\left[\frac{p_1 log_2\left[\frac{1}{p_1}\right] + p_2 log_2\left[\frac{1}{p_2}\right] + \cdots \ldots p_k log_2\left[\frac{1}{p_k}\right]}{L}\right]$$ 5.14

$$H(S) = p_1 L log_2 \left[\frac{1}{p_1}\right] + p_2 L log_2 \left[\frac{1}{p_2}\right] + \cdots \ldots . p_k L log_2 \left[\frac{1}{p_k}\right]$$ 5.15

$$H(S) = \sum_{k=1}^{K} p_k L log_2 \left[\frac{1}{p_k}\right]$$ 5.16

The above expression gives the entropy of a discrete memory less source

5.3.1 Extension of a Discreate Memoryless Source

The entropy of extended discrete memory less source is

$$H(S^n) = nH(S)$$ 5.17

Where $H(S^n)$ is entropy of extended source and $H(S)$ is entropy of original source

5.3.2 Rate of Information (R)
The rate of information is defined as the average number of bits of information per second. It is given as

$R = rH$ 5.18

Where R is the rate at which messages generated from the source .
H is the average number of bits of information per message (ie) Entropy
and r is the average message generated from the source.

5.4 Discreate Memory Less Channel
A Discreate Memory Less Channel is a Statistical Model With an Input X and an Output Y.

5.4.1 Discrete Channel
A channel is said to be **discrete** when both input and output are having finite sizes

5.4.2 Memoryless channel
In **memory less channel** the output at any instant depends on present input

5.4.3 Transition Probability
Let $x_1, x_2 \ldots. x_j, y_1, y_2 \ldots. y_k$ represents the input levels and output levels respectively.

The transition probability $P(\frac{y_k}{x_j})$ for input and output is given as

$P\left(\frac{y_k}{x_j}\right) = P(Y = y_k \,/\, X = x_j)$ 5.19

5.4.4 Joint Probability
From the probability theory

We have $P(XY) = P\left(\frac{Y}{X}\right) P(X)$ 5.20

Therefore the joint probability of X and Y is given as

$P(x_j, y_k) = P\left(\frac{y_k}{x_j}\right) P(x_j)$ 5.21

5.4.5 Marginal Probability
The marginal probability distribution of output Y is given as

$P(y_k) = \sum_{k=0}^{K-1} P(x_j, y_k)$ 5.22

Where $P\left(\frac{y_k}{x_j}\right)$ is the transition probability and $P(x_j)$ is the probability of an input signal.

5.4.6 BINARY COMMUNICATION CHANNEL
If two symbols are transmitted over a discrete memory less channel then the channel is said to be a binary communication channel.
The probabilities of y_o and y_1 can be written as

$$P(y_0) = P\left(\frac{y_0}{x_0}\right)P(x_0) + P\left(\frac{y_0}{x_1}\right)P(x_1) \qquad 5.23$$

$$P(y_1) = P\left(\frac{y_1}{x_1}\right)P(x_1) + P\left(\frac{y_1}{x_1}\right)P(x_0) \qquad 5.24$$

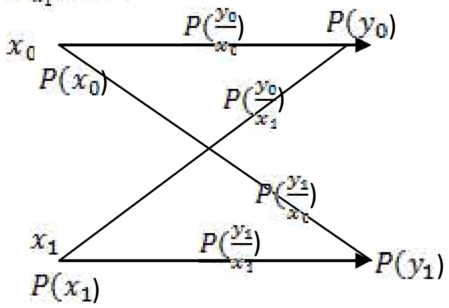


Fig.5.2 .Representation of Binary Communication Channel

It can be represented using a matrix as

$$\begin{bmatrix} P(y_0) \\ P(y_1) \end{bmatrix} = P(x_0)P(x_1)\begin{bmatrix} P(\frac{y_0}{x_0}) & P(\frac{y_1}{x_0}) \\ P(\frac{y_0}{x_1}) & P(\frac{y_1}{x_1}) \end{bmatrix} \qquad 5.25$$

5.4.7 Binary Symmetric Channel

Binary communication channel is said to be symmetric if

$$P\left(\frac{y_0}{x_0}\right) = P\left(\frac{y_1}{x_1}\right) = p \qquad 5.26$$

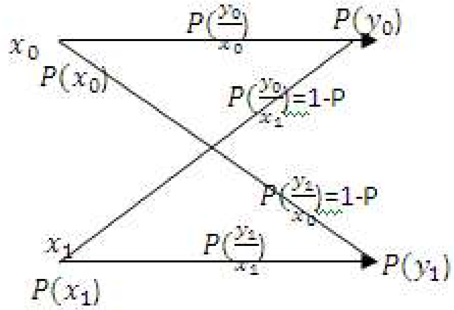


Fig. 5.3 .Representation of Binary Symmetric

The matrix representation is

$$\begin{bmatrix} P(y_0) \\ P(y_1) \end{bmatrix} = P(x_0)P(x_1)\begin{bmatrix} P & 1-P \\ 1-P & P \end{bmatrix} \qquad 5.27$$

5.4.8 Equivocation (Conditional Entropy)

The conditional entropy $H(\frac{X}{Y})$ is called equivocation .

It is defined as

$$H\left(\frac{X}{Y}\right) = \sum_{j=1}^{J}\sum_{k=1}^{K} P(x_j, y_k) log_2 (\frac{1}{P(x_j, y_k)}) \quad 5.28$$

$H\left(\frac{X}{Y}\right)$ represents uncertainty of X, on average , when Y is known.

$$H\left(\frac{Y}{X}\right) = \sum_{j=1}^{J}\sum_{k=1}^{K} P(x_j, y_k) log_2 (\frac{1}{P(y_k, x_j)}) \quad 5.29$$

Joint Entropy $H\left(\frac{X}{Y}\right)$ is given as

$$H(X,Y) = \sum_{j=1}^{J}\sum_{k=1}^{K} P(x_j, y_k) log_2 (\frac{1}{P(x_j, y_k)}) \quad 5.30$$

5.4.9 Rate of Transmission Over A Discrete Memoryless Channel

We know that the rate of information

$$R = rH \ bits/sec \quad 5.31$$

Therefore the rate of transmission into the channel is

$$D_{in} = rH(X) bits/sec$$

Where r is the rate at which Symbols generated from the source. When the information transmitted over the channel, errors may introduce. If there is loss of data it is represented as $H\left(\frac{X}{Y}\right)$.

Therefore the average rate of information across the channel is

$$D_t = r\left(H(X) - H\left(\frac{X}{Y}\right)\right) bits \ / \ sec. \quad 5.32$$

If no error occur during transmission, then

$$H\left(\frac{X}{Y}\right) = 0 \quad 5.33$$

Therefore

D_t=rH(X) bits/sec 5.34

Hence

$$D_t = D_{in} \quad 5.35$$

`

5.5 MUTUAL INFORMATION

It is defined as the difference between $H(X)$ and $H(\frac{X}{Y})$

Where $H(X)$ is entropy of channel input

and $H(\frac{X}{Y})$ represents uncertainity about the channel input after observing the channel output.

It is denoted as $I(X;Y)$, which is given by

$I(X;Y) = H(X) - H(\frac{X}{Y})$ 5.36

Mean of entropy $H(\frac{X}{Y})$ over the input is given as

$H\left(\frac{X}{Y}\right) = \sum_{k=1}^{K} H(\frac{X}{Y} = y_k) p(y_k)$ 5.37

$H\left(\frac{X}{Y}\right) = \sum_{j=1}^{J} \sum_{k=1}^{K} log_2 \frac{1}{p(x_j/y_k)} p\left(x_j/y_k\right) p(y_k)$ 5.38

Therefore

$I(X;Y) = H(X) - \sum_{j=1}^{J} \sum_{k=1}^{K} log_2 \frac{1}{p(x_j/y_k)} p\left(x_j/y_k\right) p(y_k)$ 5.39

5.6 Channel Capacity

It is defined as the maximum mutual information $I(X;Y)$ in any single use of the channel, where the maximization is over all possible input probability distributions $\{p\left(x_j\right)\}$ on X.

Where C is measured in bits per channel use or bits per transmission

$C = \max_{\{p(x_j)\}} I(X;Y)$ 5.40

The calculation of C involves maximization of mutual information $I(X;Y)$ over j variables $p(x_0), p(x_0), p(x_0), \ldots . p(x_0)$. It must satisfy two conditions

(i) $p\left(x_j\right) \geq 0 \; for \; all \; j$

(ii) $\sum_{j=1}^{J} p\left(x_j\right) = 1$

5.6.1 Channel efficiency

The transmission efficiency or channel efficiency is defined as

$\eta = \frac{Actual\, transmission}{Maximum \; transmission}$ 5.41

$\eta = \frac{I(X;Y)}{Max I(X;Y)}$ 5.42

$\eta = \frac{I(X;Y)}{C}$ 5.43

5.6.2 Chananel Redundancy

The redundancy of the channel is defined as

$R = 1 - \eta$ 5.44

$R = 1 - \frac{I(X;Y)}{C}$ 5.45

5.6.3 Noise free channel

We know that mutual information

$I(X;Y) = H(X) - H(\frac{X}{Y})$ 5.46

The mutual information for a noise free channel is given as

$I(X;Y) = H(X)$ 5.47

The channel capacity is $C = maxI(X;Y)$

$C = maxH(X)$ 5.48

$C = log_2 k\ bits/message$ 5.49

5.6.4 Symmetric Channel

A symmetric channel is defined as one, which has two properties

(i) $H\left(\frac{Y}{x_j}\right)$ *is independent of k*

(ii) $\sum_{j=1}^{J} p(\frac{y_k}{x_j})$ is independent of j

The mutual information for a symmetric channel is

$I(X;Y) = H(Y) - H(\frac{Y}{X})$ 5.50

$I(X;Y) = H(Y) - \sum_{j=1}^{J} H(Y/x_j)p(x_j)$ 5.51

$I(X;Y) = H(Y) - A\sum_{j=1}^{J} p(x_j)$ 5.52

Where

$A = H(\frac{Y}{x_j})$ 5.53

We know that

$\sum_{j=1}^{J} p(x_j) = 1$ 5.54

Therefore

I $(X;Y) = H(Y) - A$ 5.55

The channel capacity is

$C = max\ I(X;Y)$ 5.56

$C = max\ (H(Y) - A)$ 5.57

$C = max\ (H(Y)) - A$ 5.58

$C = log_2 K - A)$bits/message 5.59

5.6.5 Binary Symmetric Channel(BSC)

The channael capacity for binary symmetric chnanle is

$C = log_2 K - A)$bits/message 5.60

$C = log_2 K - H(\frac{Y}{x_j}))$ 5.61

$C = log_2 K - H(\frac{Y}{x_j}))$ 5.62

$C = log_2 K - \sum_{k-1}^{2} p(y_k/x_j) log_2 \frac{1}{p(y_k/x_j)}$ 5.63

5.7 SHANNONS THEOREM

Shannon's information theory for a discrete memory less sources and channels involves three theorems. They are

1. Shannon's first theorem or source coding theorem
2. Shannon's second theorem or channel coding theorem
3. Shannon's third theorem or information capacity theorem or Shannon's Hartley theorem

5.7.1. Shannon's first theorem or source coding theorem

This theorem include Shannon fano coding and Huffman coding to remove redundancy and to improve efficiency

5.7.1.1Source coding to improve average information per bit

An important problem in communication system is the efficient representation of data generated by a source.Source encoding or source coding is the process by which the representation is achieved.The device which performs source coding or encoding is called source encoder

Variable length code

It is the encoding process where two codeword's are used.

1. Short codeword is used to represent frequently occurring messages or symbols
2. Longer code word is used to represent rarely occurring symbols

eg) E is more frequently used than Q

Therefore E is represented as a shorter code word and Q is represented as longer code word

Requirement

The efficient source encoder should satisfy the following requirements

1. The code words generated by the encoder should be binary in nature
2. Every code word should represent a single message .

Consider a system as

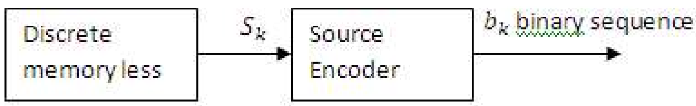


Fig.5.4 .Transmitter section (Diccrete)

The discrete memory less source produces an output S_k which is converted by the source encoder into a block of 0's and 1's denoted by b_k.The binary codeword for the symbol S_k have length l_k is measured in bits.

The average codeword length

$$\overline{L} = \sum_{k=0}^{K-1} p_k\, l_k \qquad 5.64$$

Where $\bar{L}$ is the average number of bits per source symbol

Let L_{min} denotes the minimum value of $\bar{L}$.Therefore the coding efficiency of the source encoder is

$\eta = \frac{L_{min}}{\bar{L}}$ 5.65

The source encoder is said to be efficient when coding efficiency approaches' unity

Since $\bar{L} \geq L_{min}$ η always less than one.The L_{min} value can be determined by the Shannon's first theorem called source coding theorem.

Statement

Shannon's first theorem is stated as for any distortion less source, a discrete memory less source of entropy ?H and average codeword length $\bar{L}$ is bounded as

$\bar{L} \geq H$ 5.66

If $L_{min} = H$ then

$\eta = \frac{H}{\bar{L}}$ 5.67

5.7.1.2 Code Redundancy

It is defined as

$\gamma = 1 - \eta$ 5.68

5.7.1.3Code Variance(σ^2)

It is defined as

$\sigma^2 = \sum_{k=0}^{K-1} p_k \ (n_k - \bar{L})$ 5.69

Where p_k is the probability of k_{th} symbol

n_k is the number of bits assigned to k_{th} symbol

$\bar{L}$ is the average code word length

5.7.2.Shannon's second theorem or channel coding theorem or channel coding theorem

Need for channel coding

The presence of noise in a channel causes discrepancies between the output and input data sequences of a digital communication system.

For example in wireless communication channel the error probability may be 10^{-1}. But the required error probability is 10^{-6}. In order to achieve this channel coding is used. The aim of channel coding is to increase the resistance of a digital communication system to a channel noise.

Channel coding consists of converting the incoming data sequence into a channel input sequence and inverse mapping the output sequence into output data sequence in such a way that the overall effect of channel noise on the system is minimized.

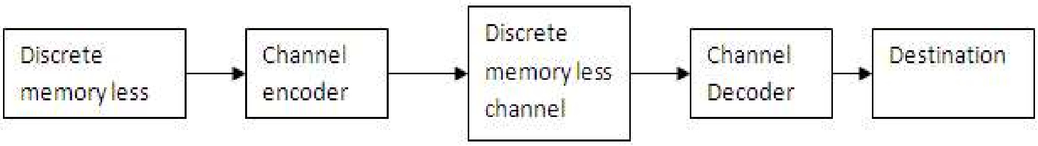


Fig.5.5 . Block Diagram of Discrete Memoryless channel

The step taken is to introduce redundancy in the channel encoder so as to reconstruct the original source sequence as accurately as possible.
Consider block codes in which the message sequence is divided into sequential blocks of K bits. Then the K bits are mapped into N bit block. So that the number of redundant bits added is N-K bits. The code rate is given as
Code rate $r=K/N<< 1$ 5.70

Only with less error probability accurate reconstruction is possible.Channel coding is efficient when the code rate is high, which can be defined using channel capacity
Consider a discrete memory less source S, entropy H(S) bits per source symbol. Assume that source emits and delivers symbols for every T_s seconds.
Channel capacity per units = C/T_c bits/sec. This gives the maximum rate of information over the channel.
Statement
The theorem states that the channel capacity has a fundamental limits on the rate at which the transmission of error free messages can takes place over a discrete memory less channel.
Consider a discrete memory less source S, entropy H(S) bits per source symbol . Assume the source emits and delivers symbols for every T_s seconds

Let DMC has capacity C and used every T_c seconds.
Then if $H(S)/T_s \leq C/T_c$ 5.71
There exists a coding scheme for which source output can be transmitted over the channel and can be reconstructed with small error probability.
Else if $H(S)/T_s=C/T_c$ 5.72
The system is said to be signaling at the critical rate
Else if $H(S)/T_s > C/ T_c$ 5.73
No possibility of transmission and reconstruction with small probability of error.
In the channel encoder the encoded symbol transmission rate is $R=1/T_c$ symbols per sec
The channel coding theorem states that $1/T_s \leq C/T_c$. Also $r \leq C$. Where r is the code rate.

5.7.3 Shannon's third theorem or Information capacity theorem or Shannon Hartley theorem

Here idea of mutual information is used to find the information capacity of band limited Gaussian channels. Consider a zero mean stationary process $X_k(t)$ which is band limited to B hertz.
Let X_k be a constant random variable which is uniformly sampled at the nyquist rate of 2B samples/sec. These samples are transmitted in T sec over a noisy channel. Therefore the number of samples K=2BT.
The channel output is disturbed by additive white Gaussian noise of zero mean and power spectral density $N_o/2$.
Output $Y_k=X_k+N_k$ 5.74

The power transmitted from the input is

$E[X_k^2]=P$ 5.75

Definition

It is defined as the maximum of mutual information between the channel input X_k and channel output Y_k over all possible distributions on the input X_k that satisfy the power constraint $E[X_k^2]=P$

Let $I(X_k;Y_k)$ be the mutual information. Mutual information can be expressed as

$I(X_k;Y_k)=H(X_k)-H(X_k/Y_k)$ 5.76

We have

$H(N_k)= H(X_k/Y_k)= H(Y_k/X_k)$ 5.77

Where $H(Y_k/X_k)$ is the conditional entropy

$H(N_k)$ is the differential entropy of N_k

Therefor e

$I(X_k;Y_k)=H(Y_k)- H(N_k)$ 5.78

The information capacity can be evaluated using the following points

1. The variance of sample Y_k of the received signals is $P+\sigma^2$. Hence differential entropy of Y_k is

$H(Y_k)=1/2 \log_2[2\pi e(P+\sigma^2)]$ 5.79

2. The variance of sample $N_k= \sigma^2$. Hence differential entropy of N_k is
3. $H(N_k)=1/2 \log_2 [2\pi e(\sigma^2)]$

Therefore

$I(X_k;Y_k)=H(Y_k)- H(N_k)$ 5.80

$I(X_k;Y_k)= 1/2 \log_2[2\pi e(P+\sigma^2)]- 1/2 \log_2 [2\pi e(\sigma^2)]$ 5.81

$I(Xk; Yk) = \log_2[\frac{2\pi e(P+\sigma^2)}{2\pi e(\sigma^2)}]$ 5.82

$I(Xk; Yk) = \frac{1}{2}\log_2[1+\frac{P}{\sigma^2}]$ bits/ transmission 5.83

Information capacity can be expressed as

$C = Blog_2\left[1+\frac{P}{N_o B}\right]$bits/sec 5.84

$C = Blog_2\left[1+\frac{S}{N}\right]$bits/sec 5.85

It ha s three important parameters

(i) Channel bandwidth

(ii) Average transmitted power

(iii) Noise power spectral density at the output

5.7.3.1 BANDWIDTH S/N TRADE OFF

Channel capacity of a gaussian channel is given as

$C = Blog_2[1+\frac{S}{N}]bits/sec.$ 5.86

Above equation indicates that channel capacity depends on two factors

1. Bandwidth of the channel
2. Signal to noise ratio

5.7.3.2 NOISELESS CHANNEL

Noiseless channel has infinite capacity. If there is no noise in the channel then N=0.Hence S/N=∞. Such channel is called as noiseless channel. Then the capacity of the channel will be

$C = Blog_2[1+\infty]bits/sec.$ 5.87

$C = \infty\ bits/sec.$ 5.88

Thus noiseless channel has infinite capacity. Infinite bandwidth channel has limited capacity.

Consider $C = Blog_2[1+\frac{S}{N}]bits/sec.$ 5.88

If bandwidth is infinite, then the channel capacity is limited. Because if bandwidth increases then noise power also increases.

5.8 Source coding techniques

There exist practical methods to design efficient codes. The most used algorithms are:
Shannon-Fano coding,
Huffman coding,
Lempel-Ziv algorithm,

Applied in: fax systems, commands "pack", "compress", "gzip" in UNIX,...

These algorithms can of course be used to code the k-th order extension of the source (then 1source message = 1 word of *k* letters), what is moreover necessary to improve efficiency, but at the price of an increased complexity.

5.8.1 Shannon-Fano codes

It is based on the following procedure, which can be represented using a tree:

1) Arrange the source messages such as the probabilities are in the descending order
2) Divide the list of messages into two (*Q*) subsets as balanced as possible, in the sense of the sum of elementary
probabilities messages.
3) Assign respectively the symbol “0” and “1”, (up to ... *Q-1*) to the first and second(*up to Q-1*) subsets (rootdivided into Q = 2 branches)
4) Repeat the steps 2) 3) with each subset (nodes divided into 2 (or Q) new branches) until that the operation becomes impossible (then each message has become a corresponding code-word left on the tree).

Example: A Discreate memoryless source has 6 symbols s1,s2,s3,s4,s5 and s6 with probabilities 0.4,0.1,0.2,0.1,0.1 and 0.1 respectively. Construct a Shannon Fano Code and also calculate its efficiency.

Solution
Given

Symbols	**Probabilities**
S1	0.4
S2	0.1
S3	0.2
S4	0.1
S5	0.1
S6	0.1

Step 1:
Arrange the given probabilities in decreasing order

Symbols	**Probabilities**
S1	0.4
S3	0.2
S2	0.1
S4	0.1
S5	0.1
S6	0.1

Step 2:
Partitioning
[x1]=[s1,s2]; [x2]=[s2,s4,s5,s6]
Or
[x1]=[s1]; [x2]=[s3,s2,s4,s5,s6]

Considering the first partitioning
We have

Symbols	Probability		Stage 1		Stage 2		Stage 3	Code Word	Length
S1	0.4	0.6	0	0.4	0			00	2
S3	0.2		0	0.2	1			01	2
S2	0.1	0.4	1	0.2	0	0.1	0	100	3
S4	0.1		1		0	0.1	1	101	3
S5	0.1		1	0.2	1	0.1	0	110	3
S6	0.1		1		1	0.1	1	111	3

$$\bar{L} = \sum_{k=1}^{K} P_k L_k$$

$$\bar{L} = \sum_{k=1}^{6} P_k L_k$$

$\bar{L} = P_1L_1 + P_2L_2 + P_3L_3 + P_4L_4 + P_5L_5 + P_6L_6$

$\bar{L} = (0.4x2) + (0.2x3) + (0.1x3) + (0.1x3) + (0.1x3) + (0.1x3)$

$\bar{L} = 0.8 + 0.4 + 0.3 + 0.3 + 0.3 + 0.3$

$\bar{L} = 2.4\ bits/symbol$

Step 3:
Entropy

$$H(S) = \sum_{j=1}^{J} P(x_j) log_2(\frac{1}{P(x_j)})$$

On substituting the values we get
H(S)=2.32 information bits/symbol

Step 4:
Efficiency

$$\eta = \frac{H}{L}$$

$$\eta = \frac{2.32}{2.4}$$

$$\eta = 96.7\%$$

5.8.2 Huffman codes

It uses the same principle as Shannon Fano coding i.e assigning different number of binary digits to the messages according to their probability of occurrence
This coding method leads to the lowest possible value of $\bar{L}$ for a given message sequence M, resulting in a maximum efficiency.
Hence it is also known as the minimum redundancy code or optimum code.

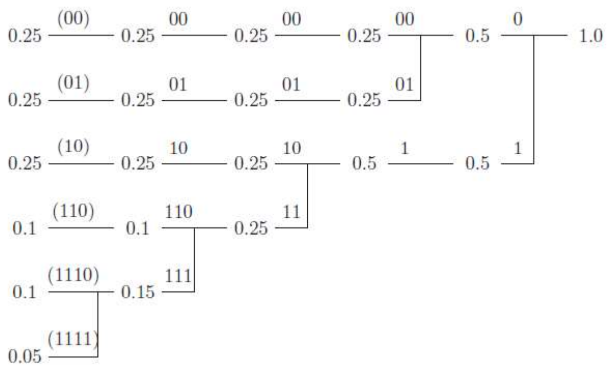
The Huffman code is ***optimum*** in the sense that no other instantaneous code for the same alphabet (and same probabilities distribution) can have a better efficiency (i.e. a lower expected length). In particular, the code efficiency obtained by Huffman algorithm is greater than or equal to the code efficiency obtained by Shannon-Fano algorithm.
But this efficiency is 1 (optimum code is not necessarily *absolutely optimal* code !).It can be proved that an instantaneous *optimum* code must satisfies the following properties:

- the shorter codewords are assigned to source messages with higher probabilities,24
- lengths of the two (Q) longest (*i.e.* less likely) codewords are equal
- the two (Q) longest codewords differ only in the last symbol (and correspond to the two least likely source messages).

Algorithm is presented in the particular case of a binary code (*Q = 2*) but is easily generalized for Q >2.
It is based on the following procedure, which can be represented by using a tree:

1) Arrange the "outputs" (source messages in the initial iteration) in decreasing order of their probabilities.
2) Combine the two (Q) least probable messages together into a single new "output" (node of the tree) that replaces the two (Q) previous ones, and whose probability is the sum of the corresponding probabilities.
3) If the number of remaining "outputs" is 1 (the remaining node is the root of the tree), then go to the next step;otherwise go to step 1 (and increment the number of iterations), with a new list to be arranged, with a number of "outputs" reduced.
For encoding (assignment of the Q-ary symbols to the different nodes, we have to proceed backward i.e. from the root of the tree (last iteration node) to the different terminal nodes (including first iteration nodes)
4) Assign arbitrary "0" and "1" (up to "Q-1") as first symbol of the 2 (Q) words (nodes) corresponding to the 2 (Q) remaining outputs (last iteration node with sum = 1).
5) If an output is the result of the merger of 2 (Q) outputs in a preceding iteration, append the current word (node) with a "0" and "1" (up to "Q-1") to obtain the word (node) for the preceding outputs and repeat 5). If no output is preceded by another output in a preceding iteration, then stop (first iteration).

Example 5.8.1 Consider a source with alphabet {1, 2, 3, 4, 5, 6} and symbol probabilities 0.25, 0.25, 0.25, 0.1, 0.1 and 0.05, respectively. Obtain the Huffman code
Solution. By following the Huffman encoding procedure, the Huffman code is obtained as 00, 01, 10, 110, 1110, 1111.



5.8.3 Lempelziv Coding
Drawback of Huffman code is that
(i) It requires knowledge of the probabilistic model of the source

(ii) With modeling text, the storage requirements prevent the Huffman coding from capturing the higher-order relationship between words

Principle:

Encoding is done by " Parsing the source data stream into segments that are shortest subsequences not encountered previously.

Example:

Find the encoded blocks for an input binary sequence 1111100000 using Lempel Ziv Code

Encoder:

Step 1:

Assume that the binary symbols 0 and 1 are already stored in that order in the code book. It can be written as,

Subsequences stored: 0, 1

Data to be parsed:11111000000

Step 2:

Assume the encoding process begins at the left, the shortest subsequences from the left other than 0 & 1 is 11. It is stored as

Subsequences stored: 0, 1,11

Data to be parsed:11100000

Step 3:

The shortest subsequences other than 0, 1,11 is 111. It is shown as

Subsequences stored: 0, 1,11,111

Data to be parsed:00000

Step 4:

The shortest subsequences which is not encountered previously is

Subsequences stored: 0, 1,11,111,00

Data to be parsed:000

Step 5:

The shortest subsequences found according to Lempel Ziv Principle is

Subsequences stored: 0, 1,11,111,00,000

Data to be parsed:…..

Since there is no data to be parsed, the code of binary sequences is shown as

Code Book: 0,1,11,111,00,000

The encoded block has been generated by the following row arrangement.

Numerical positions	1	2	3	4	5	6
Code Book Subsequences	0	1	1 1	11 1	0 0	00 0
Numerical Representation			2 2	3 2	1 1	5 1
Binary Encoded Blocks			010 1	011 1	001 0	101 0

5.9 Problems

1. Calculate amount of information if $P_k = 1/2$

Solution:

Given:

$P_k = 1/2$

Amount of Information $\boldsymbol{I_k = log_2(\frac{1}{P_k})}$

$$I_k = log_2(\frac{1}{1/2})$$

$$I_k = log_2 2$$

$$\boldsymbol{I_k = 1\ bit}$$

2.Calculate the total amount of information for binary symbols '0' and '1' which is transmitted with probabilities ¾ and ¼ respectively, also compare its amount of information.

Solution:

$P_1 = 3/4$

$P_2 = 1/4$

$$\boldsymbol{I_1 = log_2(\frac{1}{P_1})}$$

$$I_1 = log_2(\frac{1}{3/4})$$

$$I_1 = log_2(\frac{4}{3})$$

$$I_1 = \frac{log_{10}(\frac{4}{3})}{log_{10}(2)}$$

$$\boldsymbol{I_1 = 0.4150\ bit}$$

$I_2 = log_2(\frac{1}{P_2})$

$I_2 = log_2(\frac{1}{1/4})$

$I_2 = log_2(4)$

$I_2 = \frac{log_{10}4}{log_{10}2}$

$\boldsymbol{I_2 = 2\ bits}$

3..A source s emits symbolsms1,s2,s3 with probabilities 0.25,0.5 and 0.25. Find entropy of a source S

Solution:

$$\boldsymbol{H(S) = \sum_{k=1}^{K} P(x_j) log_2(\frac{1}{P(x_j)})}$$

$$H(S) = P_1 log_2\left(\frac{1}{P_1}\right) + P_2 log_2\left(\frac{1}{P_2}\right) + P_3 log_2\left(\frac{1}{P_3}\right)$$

$$H(S) = 0.25 log_2\left(\frac{1}{0.25}\right) + 0.5 log_2\left(\frac{1}{0.5}\right) + 0.25 log_2\left(\frac{1}{0.25}\right)$$

$$H(S) = 0.25\frac{log_{10}\left(\frac{1}{0.25}\right)}{log_{10}2} + 0.5\frac{log_{10}\left(\frac{1}{0.5}\right)}{log_{10}2} + 0.25\frac{log_{10}\left(\frac{1}{0.25}\right)}{log_{10}2}$$

$$H(S) = 0.25(2) + (0.5)(1) + (0.25)(2)$$

$$\boldsymbol{H(S) = 1.5\ bits/symbol}$$

4.An analog signal is band limited to 250 Hz and sampled at Nyquist rate. The samples are quantized into 4 levels. The probabilities of occurrence of these levels are p_1=p_3=1/8 and p_2=p_4=3/8. Find out the information rate of a source.

Solution:

Information rate $\boldsymbol{R = rH}$

$r = 2B$

$r = 2x250$

$r = 500\ levels/sec$

$$\boldsymbol{H(S) = \sum_{k=1}^{K} P(x_j) log_2(\frac{1}{P(x_j)})}$$

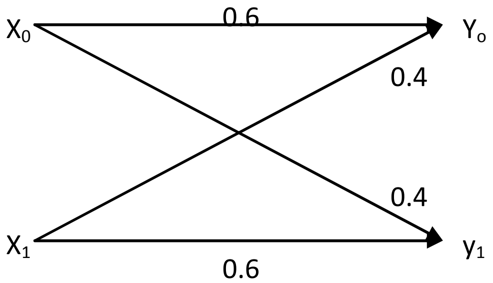
$$H(S) = P_1 log_2\left(\frac{1}{P_1}\right) + P_2 log_2\left(\frac{1}{P_2}\right) + P_3 log_2\left(\frac{1}{P_3}\right) + P_4 log_2\left(\frac{1}{P_4}\right)$$

$$H(S) = \frac{1}{8}log_2(8) + \left(\frac{3}{8}\right)log_2\left(\frac{8}{3}\right) + \left(\frac{1}{8}\right)log_2(8) + (\frac{3}{8})log_2\left(\frac{8}{3}\right)$$

$$H(S) = 0.375 + 0.5306 + 0.375 + 0.5306$$

$$\boldsymbol{H(S) = 1.8113\ bits/symbol}$$

5.A binary symmetric channel is shown in fig. Find the channel matrix of the resultant channel.Find P(y_o) and P(y_1) if P(x_0)=0.3,P(x_1)=0.7



Solution:

Channel Matrix

$$P\left(\frac{Y}{X}\right) = \begin{bmatrix} P(\frac{y_o}{x_o}) & P(\frac{y_1}{x_o}) \\ P(\frac{y_o}{x_1}) & P(\frac{y_1}{x_1}) \end{bmatrix}$$

Channel Matrix $P\left(\frac{Y}{X}\right) = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}$

$$P(Y) = P\left(\frac{Y}{X}\right)P(X)$$

$$P(Y) = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}[0.3 \quad 0.7]$$

$$P(Y) = [0.18 + 0.28 \qquad 0.120.42]$$

$$P(Y) = [0.46 \quad 0.54]$$

We know that $P(Y) = [P(y_o)\,P(y_1)]$

$$P(y_o) = 0.46 \quad P(y_1) = 0.54$$