UNIT - V
Information theory

Introduction.
The performance of the communication system is measured in terms of its error probability. An errorless transmission is possible when the probability of error at the receiver approaches zero

Uncertainty
Consider the source which emits the discrete symbols randomly from the set of fixed alphabet.

$$X = \{x_0, x_1, x_2, \dots x_{K-1}\}$$

the various symbols in 'x' have probabilities of $P_0, P_1, P_2, \dots$ etc. This set of probabilities satisfy the following condition.

$$\sum_{K=0}^{K-1} P_K = 1 \quad \text{(sum of probabilities is zero).}$$

Information
Consider the communication system which transmits messages $m_1, m_2, m_3, \dots$ with probabilities $P_1, P_2, P_3, \dots$
The amount of information transmitted through the message m_K with probability P_K is given as

$$I_K = \log_2\left(\frac{1}{P_K}\right) \text{ bits}$$

Properties of information

1) If there is more uncertainty about the message, information carried is more
2) If the receiver knows the message being transmitted the amount of information carried is zero

3) If I_1 is the information carried by message m_1, I_2 carried by m_2, the amount of information carried due to m_1 and m_2 is $I_1 + I_2$

4) If there are $M = 2^N$ equally likely messages, then amount of information carried by each message is N bits.

problem :- Calculate the amount of information if $P_K = \frac{1}{4}$.

sol:- $I_K = \log_2\left(\frac{1}{P_K}\right) = \log_2\left(\frac{1}{1/4}\right)$

$$= \log_2 4 = \log_2(2^2) = 2\log_2 2 = 2 \text{ bits}$$

Discrete memoryless Channels (DMC)

channel

A communication channel may be defined as the path or medium through which the symbols flow to the receiver. A DMC is a statistical model with an input X and an output Y. The channel is said to be discrete when the alphabets X and Y are finite. It is said to be memoryless when the current output depends on only the current input and not on the past inputs.

$$\left.\begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array}\right\} X \longrightarrow \boxed{P(y_j/x_i)} \longrightarrow Y \left\{\begin{array}{c} y_1 \\ y_2 \\ \vdots \\ y_n \end{array}\right.$$

Representation of a DMC

The Channel Matrix

A channel is completely described by the complete set of transition probabilities. The channel matrix is given by

$$\underset{\text{channel matrix}}{[P(Y|X)]} = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) & \cdots & P(y_n/x_1) \\ P(y_1/x_2) & P(y_2/x_2) & \cdots & P(y_n/x_2) \\ \vdots & & & \\ P(y_1/x_m) & P(y_2/x_m) & \cdots & P(y_n/x_m) \end{bmatrix}$$

Types of channels

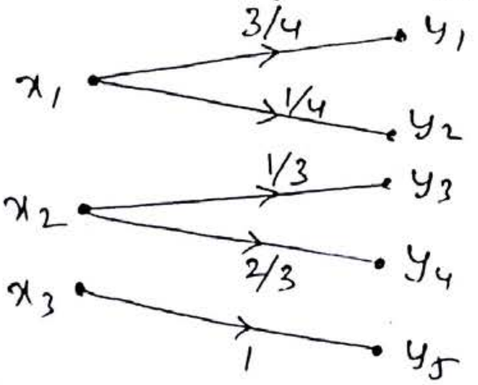
Other than continuous and discrete channels there are some special type of channels.

1) Lossless channel

A channel described by a channel matrix with only one non zero element in each column is called a lossless channel.

$$P(Y/X) = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

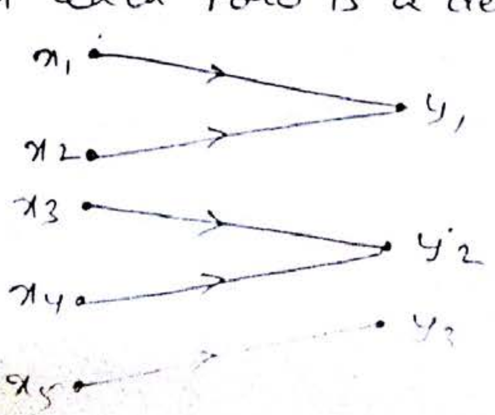
The channel diagram is as follows.



2) Deterministic channel

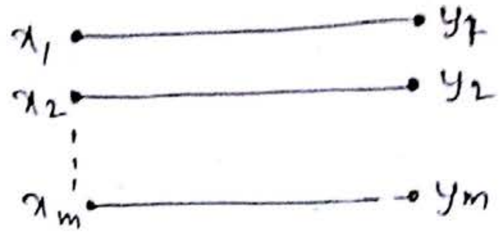
A channel described by a channel matrix with only one non zero element in each row is a deterministic channel

$$P(Y/X) = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$



3) Noiseless channel

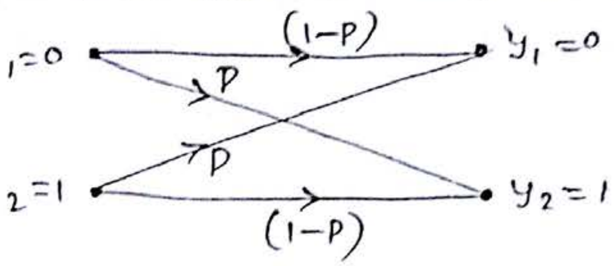
A channel is called noiseless if it is both lossless and deterministic.



$$P(Y/X) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

4) Binary symmetric channel (BSC)

The BSC is defined by the channel diagram shown.



$$P(Y/X) = \begin{bmatrix} (1-P) & P \\ P & (1-P) \end{bmatrix}$$

Probabilities

Input probabilities

$$P(X) = [P(x_1)\ P(x_2) \cdots P(x_m)]$$

output probabilities

$$P(Y) = [P(y_1)\ P(y_2) \cdots P(y_n)]$$

$P(X)$ as a diagonal matrix

$$P(X)_d = \begin{bmatrix} P(x_1) & 0 & 0 & \cdots & 0 \\ 0 & P(x_2) & 0 & \cdots & 0 \\ \vdots & \vdots & & & \\ 0 & 0 & 0 & \cdots & P(x_m) \end{bmatrix}$$

Joint probability matrix $P(X,Y) = [P(X)_d].[P(Y/X)]$

Conditional probability matrix $= P(Y/X)$

Entropy (Average information)

In a practical communication system, we usually transmit long sequences of symbols from an information source. Hence we are more interested in the average information that a source produces than the information content of a single symbol.

The mean value of $I(x_i)$ over the alphabet of source X with m different symbols is given by

$$H(X) = E[I(x_i)] = \sum_{i=1}^{m} P(x_i)\, I(x_i)$$

But $$I(x_i) = \sum_{i=1}^{m} \frac{1}{\log_2 P(x_i)}$$

$$\therefore\ H(X) = \sum_{i=1}^{m} P(x_i) \cdot \frac{1}{\log_2 [P(x_i)]}$$

$$= -\sum_{i=1}^{m} P(x_i) \log_2 P(x_i) \text{ bits/symbol.}$$

The quantity $H(X)$ is called the entropy of source X. It is a measure of the average information content per source symbol

Information rate

If the time rate at which source X emits symbols is r, the information rate R of the source is given by

$$R = r\, H(X) \text{ bits/second}$$

where R = information rate

$H(X)$ = entropy or average information

r = rate at which the symbols are generated

$\therefore R = r$ in $\frac{\text{symbols}}{\text{sec}} \times H(X)$ in $\frac{\text{bits}}{\text{symbol}}$

$= r\,H(X)$ bits/second.

problem

A Discrete memoryless source (DMS) X has four symbols x_1, x_2, x_3, x_4 with probabilities $P(x_1) = 0.4$, $P(x_2) = 0.3$, $P(x_3) = 0.2$, $P(x_4) = 0.1$

i) Calculate $H(X)$

ii) Find the amount of information contained in the messages $x_1 x_2 x_1 x_3$ and $x_4 x_3 x_3 x_2$.

sol:- Entropy $H(X) = -\sum_{i=1}^{4} P(x_i) \log_2 P(x_i)$

$= -0.4 \log_2(0.4) - 0.3 \log_2(0.3) - 0.2 \log_2(0.2) - 0.1 \log_2(0.1)$

$= 1.85$ bits/symbol.

ii) $P(x_1 x_2 x_1 x_3) = (0.4)(0.3)(0.4)(0.2) = 0.0096$

$I(x_1 x_2 x_1 x_3) = -\sum_{i=1}^{4} \log_2 P(x_1 x_2 x_1 x_3)$

$= -\log_2(0.0096) = 6.7$ bits/symbol

$P(x_4 x_3 x_3 x_2) = (0.1)(0.2)(0.2)(0.3) = 0.0012$

$I(x_4 x_3 x_3 x_2) = -\log_2(0.0012)$

$= 9.7$ bits/symbol

problem

The probabilities of the five possible outcomes of an experiment are $P(x_1) = \frac{1}{2}$, $P(x_2) = \frac{1}{4}$, $P(x_3) = \frac{1}{8}$, $P(x_4) = P(x_5) = \frac{1}{16}$ Determine the entropy and information rate if there are 16 outcomes per second.

sol:- Entropy $H(X) = \sum_{i=1}^{5} P(x_i) \log_2 \frac{1}{P(x_i)}$ bits/symbol.

$$H(X) = \frac{1}{2}\log_2\left(\frac{1}{1/2}\right) + \frac{1}{4}\log_2\left(\frac{1}{1/4}\right) + \frac{1}{8}\log_2\left(\frac{1}{1/8}\right)$$

$$+ \frac{1}{16}\log_2\left(\frac{1}{1/16}\right) + \frac{1}{16}\log_2\left(\frac{1}{1/16}\right)$$

$$= \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{4}{16} + \frac{4}{16} = \frac{15}{8} = 1.875 \text{ bits/outcome}$$

Information rate $R = rH(X)$

given $r = 16$

$\therefore R = 16 \times 1.875 = 30$ bits/sec.

Conditional and Joint Entropies

Using the input probabilities $P(x_i)$, the output probabilities $P(y_j)$, transition probabilities $P(y_j/x_i)$ and joint probabilities $P(x_i, y_j)$, the various entropy functions are defined as follows

m = no. of inputs, n = no. of outputs

$$H(X) = -\sum_{i=1}^{m} P(x_i)\log_2 P(x_i)$$

$$H(Y) = -\sum_{j=1}^{n} P(y_j)\log_2 P(y_j)$$

$$H(X/Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(x_i/y_j)$$

$$H(Y/X) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(y_j/x_i)$$

$$H(X,Y) = -\sum_{j=1}^{n}\sum_{i=1}^{m} P(x_i, y_j)\log_2 P(x_i, y_j)$$

$H(X)$ = average uncertainty of the channel input

$H(Y)$ = average uncertainty of the channel output

$H(X/Y)$ = conditional entropy which is a measure of the average uncertainty about the channel input after the channel o/p has been observed

$H(Y/X)$ = conditional entropy which is a measure of the average uncertainty of the channel o/p given that X is transmitted

$$H(X,Y) = H(X/Y) + H(Y)$$

$$H(X,Y) = H(Y/X) + H(X)$$

Mutual information

The mutual information denoted by $I(X;Y)$ of a channel is defined by

$$I(X;Y) = H(X) - H(X/Y) \text{ bits/symbol}$$

$I(X;Y)$ is the uncertainty about the channel input that is resolved by observing the channel output.

Properties of mutual information

1) $I(X;Y) = I(Y;X)$
2) $I(X;Y) \geq 0$
3) $I(X;Y) = H(Y) - H(Y/X)$
4) $I(X;Y) = H(X) + H(Y) - H(X,Y)$

which is known as Kraft inequality.

problem :- Consider a DMS X with symbols $x_i = 1, 2, 3, 4$

x_i	code A	code B	code C	code D
x_1	00	0	0	0
x_2	01	10	11	100
x_3	10	11	100	110
x_4	11	110	110	111

show that all codes except code B satisfy Kraft inequality

sol:- for code A, $n_1 = n_2 = n_3 = n_4 = 2$

$$K = \sum_{i=1}^{4} 2^{-n_i} = 2^{-2} + 2^{-2} + 2^{-2} + 2^{-2} = 1$$

For code B, $n_1 = 1, n_2 = n_3 = 2, n_4 = 3$

$$K = 2^{-1} + 2^{-2} + 2^{-2} + 2^{-3} = \frac{9}{8} > 1$$

For code C, $n_1 = 1, n_2 = 2, n_3 = 3, n_4 = 3$

$$K = 2^{-1} + 2^{-2} + 2^{-3} + 2^{-3} = 1$$

For code D, $n_1 = 1, n_2 = 3, n_3 = 3, n_4 = 3$

$$K = 2^{-1} + 2^{-3} + 2^{-3} + 2^{-3} = \frac{7}{8}$$

Shannon-Fano Coding

An efficient code can be obtained by the following procedure, known as Shannon-Fano coding.

Algorithm

1) List the source symbols in order of decreasing probability.
2) Partition the set into two sets that are as close to equiprobable as possible.
3) Assign a '0' to the first set and '1' to the second set

4) Continue this process, each time partitioning the sets with as nearly equal probabilities possible until further partitioning is not possible.

<u>problem</u> :- Find the Shannon - Fano code for the probabilities ~~as~~ shown.

message	probabilities												code word	length of code word
x_1	$\frac{16}{32}$	sum $= \frac{16}{32}$	0	only one message stop here									0	1
x_2	$\frac{4}{32}$		1	$= \frac{8}{32}$	0	$\frac{4}{32}$	0	only 1 message stop here					100	3
x_3	$\frac{4}{32}$	$= \frac{16}{32}$	1	$\frac{4}{32} + \frac{4}{32}$	0	$\frac{4}{32}$	1	only 1 message stop here					101	3
x_4	$\frac{2}{32}$	$\frac{4}{32} + \frac{4}{32} +$	1		1	$= \frac{4}{32}$	0	$\frac{2}{32}$	0	only 1 message stop here			1100	4
x_5	$\frac{2}{32}$	$\frac{2}{32} +$	1	$= \frac{8}{32}$	1	$\frac{2}{32} + \frac{2}{32}$	0	$\frac{2}{32}$	1	only 1 message stop here			1101	4
x_6	$\frac{2}{32}$	$\frac{2}{32} +$	1	$\frac{2}{32} + \frac{2}{32} +$	1	$= \frac{4}{32}$	1	$\frac{2}{32}$	0	only 1 message stop here			1110	4
x_7	$\frac{1}{32}$	$\frac{2}{32} + \frac{1}{32} +$	1	$\frac{2}{32} + \frac{1}{32} +$	1	$\frac{2}{32} + \frac{1}{32} +$	1	$= \frac{2}{32}$	1	$\frac{1}{32}$	0	only 1 message stop here	11110	5
x_8	$\frac{1}{32}$	$\frac{1}{32}$	1	$\frac{1}{32}$	1	$\frac{1}{32}$	1	$+ \frac{1}{32}$	1	$\frac{1}{32}$	1	only 1 message stop here	11111	5

Entropy (average info) $H(X) = \sum_{i=1}^{m} P(x_i) \log_2 \frac{1}{P(x_i)}$

$$= \left(\frac{16}{32}\right) \log_2 \frac{1}{16/32} + \frac{4}{32} \log_2 \frac{1}{4/32} + \frac{4}{32} \log_2 \frac{1}{4/32}$$

$$+ \frac{2}{32} \log_2 \frac{1}{2/32} + \frac{2}{32} \log_2 \frac{1}{2/32} + \frac{2}{32} \log_2 \frac{1}{2/32} + \frac{1}{32} \log_2 \frac{1}{1/32}$$

$$+ \frac{1}{32} \log_2 \frac{1}{1/32} = 2\frac{5}{16} \text{ bits/symbol}$$

Average length $L = \sum_{i=1}^{m} P(x_i) n_i$ $\qquad n_i = \text{length}$

$$= \left(\frac{16}{32}\right)(1) + \left(\frac{4}{32}\right)(3) + \frac{4}{32}(3) + \frac{2}{32}(4) + \frac{2}{32}(4) + \frac{2}{32}(4) + \frac{1}{32}(5) + \frac{1}{32}$$

$= 2\frac{5}{16}$ bits/symbol

Efficiency $\eta\% = \frac{H(X)}{L} = \frac{2\frac{5}{16}}{2\frac{5}{16}} \times 100 = 100\%$

Huffman Coding

Huffman encoding results in an optimum code. This code has highest efficiency. The algorithm is as follows.

Algorithm

1) List the source symbols in the order of decreasing probability.
2) Combine the probabilities of the two symbols having the lowest probabilities and record the resultant probabilities, this step is called reduction 1.
3) This procedure is repeated till there are two probabilities remaining
4) Start encoding with the last reduction. Assign '0' to the 1st and '1' to the second probability.
5) Repeat the procedure till the first column is reached

xample of Huffman coding.

ool	Probability	Step1	Step2			
1	0.25 10	0.25 10	0.25 10	0.25 10	0.5 0	0.5 0
2	0.25 11	0.25 11	0.25 11	0.25 11	0.25 10	0.5 1
3	0.125 010	0.125 010	0.25 00	0.25 00	0.25 11	
4	0.125 011	0.125 011	0.125 010	0.25 01		
5	0.125 000	0.125 000	0.125 011			
6	0.0625 0010	0.125 001				
7	0.0625 0011					

the above Huffman code is obtained by placing combined symbol as low as possible

Average length of the code $L = \sum_{i=1}^{m} p(x_i) n_i$

$= (0.25)(2) + (0.25)(2) + 0.125(3) + 0.125(3) + 0.125(3)$
$+ (0.0625)(4) + 0.0625(4) = 2.2$ bits/symbol

Huffman code can be implemented by placing the combined symbol as high as possible or as low as possible. The average length of the code word is the same in both cases.