UNIT- 1 : INFORMATION THEORY.

- The information theory is used for mathematical modeling and analysis of the communication systems.
- It is related to concepts of statistical properties of messages/sources, channels, noise interference, etc.

DISCRETE MEMORYLESS SOURCE.

- Consider the source, which emits discrete symbols randomly from the set of fixed alphabet (i.e) $X = \{x_0, x_1, x_2, \dots x_{K-1}\}$.
- The various symbols in 'X' have probabilities of p_0, p_1, p_2, etc which can be written as,

$$P(X = x_k) = p_k, \quad k = 0, 1, 2, \dots K-1.$$

- This set of probabilities satisfy the following condition, $\sum_{k=0}^{K-1} p_k = 1$.
- Such an information source is called discrete information source.
- The idea of information is related to 'uncertainty' or 'surprise'.
- Consider the emission of symbol $X = x_k$ from the source.
- If the probability of x_k is $p_k = 0$, then such a symbol is impossible.
- Similarly when probability $p_k = 1$, then such a symbol is sure.

— In both the cases, there is no 'surprise' and hence no information is produced when symbol x_k is emitted.

— If the probability p_k is low, there is more surprise or uncertainty.

— Before the event $X = x_k$ is emitted, there is an amount of uncertainty.

— When the symbol $X = x_k$ occurs, there is an amount of surprise.

— After the occurance of the symbol $X = x_k$, there is gain in amount of information.

INFORMATION (MEASURE OF INFORMATION)

— Let us consider the communication system which transmits messages $m_1, m_2, m_3, \ldots$ with probabilities of occurance $p_1, p_2, p_3, \ldots$ The amount of information transmitted through the message m_k with probability p_k is given as,

$$\left.\begin{array}{r}\text{Amount of}\\ \text{Information}\end{array}\right\} I_k = \log_2\left(\frac{1}{p_k}\right) \text{ bits.}$$

PROPERTIES OF INFORMATION.

i) If there is more uncertainty about the message, information carried is also more.

ii) If receiver knows the message being transmitte the amount of information carried is zero.

iii) If I_1 is the information carried by message m_1 and I_2 is the information carried by

m_2, then amount of information carried due to m_1 and m_2 is $I_1 + I_2$.

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

Proof for properties of Information.

Property (ii) If the receiver knows the message being transmitted, the amount of information carried is zero.

Proof: Rx knows the message.

$\Rightarrow$ probability of occurance, $P_K = 1$.

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

$$= \frac{\log_{10}(1)}{\log_{10}(2)} = 0 \text{ bits.}$$

Property (iii) message m_1 carries I_1 Information and message m_2 carries I_2 Information. Then amount of information carried compositely due to m_1 & m_2 is $I_1 + I_2$

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

and $$I_2 = \log_2\left(\frac{1}{P_2}\right).$$

m_1 & m_2 are independent $\Rightarrow$ The probability of the composite message is $P_1 P_2$.

Information carried compositely due to m_1 and m_2 is, $I_{1,2} = \log_2\left(\frac{1}{P_1 P_2}\right)$

$$= \log_2\left(\frac{1}{P_2} \cdot \frac{1}{P_2}\right) = \log_2\left(\frac{1}{P_1}\right) + \log_2\left(\frac{1}{P_2}\right)$$

$$I_{1,2} = I_1 + I_2 \qquad (\because \log AB = \log A + \log B).$$

Property (iv): If there are M equally likely and independent messages, then $(M=2^N)$ the amount of information carried by each message will be N bits.

$\Rightarrow$ Probability of each message, $p_K = \frac{1}{M}$.

$$I_K = \log_2\left(\frac{1}{p_K}\right) = \log_2(M)$$

and $M = 2^N$

$$\Rightarrow I_K = \log_2(2^N) = N \log_2(2)$$

$$= N \frac{\log_{10}(2)}{\log_{10}(2)} = N \text{ bits}.$$

ENTROPHY (AVERAGE INFORMATION).

- Consider that we have M-different messages. Let these messages be $m_1, m_2, m_3, \ldots m_M$ and they have probabilities of occurrance as $p_1, p_2, \ldots, p_m$.
- Suppose that a sequence of L messages are transmitted.

→ If L is very large, then we may say that
$p_1 L$ number of m_1 messages are transmitted
$p_2 L$ messages of m_2 are transmitted.
$p_3 L$ messages of m_3 are transmitted.
⋮
$p_M L$ messages of m_M are transmitted.

– Information due to message m_1 will be,

$$I_1 = \log_2 (1/p_1) .$$

– Since there are $p_1 L$ number of messages of m_1, the total information due to all message of m_1 will be,

$$I_{1(total)} = p_1 L \log_2 (1/p_1) .$$

– Similarly the total information carried due to the sequence of L messages will be,

$$I_{(total)} = I_{1(total)} + I_{2(total)} + \cdots + I_{M(total)}$$

$$\therefore I_{(total)} = p_1 L \log_2 \left(\frac{1}{p_1}\right) + p_2 L \log_2 \left(\frac{1}{p_2}\right) + \cdots + p_M L \log_2 \left(\frac{1}{p_M}\right) .$$

$$I_{(total)} = L \left[p_1 \log_2 \left(\frac{1}{p_1}\right) + p_2 \log_2 \left(\frac{1}{p_2}\right) + \cdots + p_M \log_2 \left(\frac{1}{p_M}\right) \right] \quad \text{— (1)}$$

— The average information per message will be,

$$\text{Avg. Info} = \frac{\text{Total information}}{\text{No. of messages.}}$$

$$= \frac{I_{(total)}}{L}$$

— Average information is represented by Entrophy. It is represented by H.

$$\Rightarrow \text{Entrophy } (H) = \frac{I_{(total)}}{L}.$$

Sub. from eqn ①

$$H = P_1 \log_2 (1/P_1) + P_2 \log_2 (1/P_2) + \cdots + P_M \log_2 (1/P_M).$$

$$\Rightarrow \boxed{H = \sum_{k=1}^{M} P_k \log_2 (1/P_k).}$$

PROPERTIES OF ENTROPHY.

i) Entrophy is zero if the event is sure or it is impossible.

(i.e) $H = 0$ if $P_k = 0$ or 1

Proof: $H = \sum_{k=1}^{M} P_k \log_2 (1/P_k)$

Sub $P_k = 0$

$$\Rightarrow H = \sum_{k=1}^{M} (0) \log_2 (1/0) = 0.$$

Sub $p_k = 1$ in $H = \sum_{k=1}^{M} p_k \log_2 (1/p_k)$.

$$\Rightarrow H = \sum_{k=1}^{M} (1) \log_2 (1) = \sum_{k=1}^{M} \frac{\log_{10}(1)}{\log_{10}(2)} = 0.$$

ii) When $p_k = 1/M$ for all the 'M' symbols, then the symbols are equally likely. For such source entropy is given as $H = \log_2 (M)$.

Proof: $p_k = 1/M$.

(i.e) $p_1 = p_2 = \cdots p_M = 1/M$.

$$H = \sum_{k=1}^{M} p_k \log_2 (1/p_k)$$

$$H = p_1 \log_2 (1/p_1) + p_2 \log_2 (1/p_2) + \cdots + p_M \log_2 (1/p_M)$$

$$H = \frac{1}{M} \log_2 (M) + \frac{1}{M} \log_2 (M) + \cdots + \frac{1}{M} \log_2 (M)$$

(Add 'M' number of terms).

$$\Rightarrow H = M \left(\frac{1}{M} \log_2 (M) \right)$$

$$H = \log_2 (M).$$

iii) Upper bound on entropy is given as,

$$H_{max} = \log_2 (M).$$

Proof: To prove the above property, we will use the property of natural logarithm.

$$\ln x \leq x - 1 \quad \text{for } x \geq 0. \quad \text{--- (1)}$$

- Let us consider any two probability distribution $\{p_1, p_2, \dots, p_M\}$ and $\{q_1, q_2, \dots, q_M\}$ on the alphabet $X = \{x_1, x_2, \dots, x_M\}$ of the discrete memoryless source.
- consider this term $\sum_{k=1}^{M} p_k \log_2\left(\frac{q_k}{p_k}\right)$.

$$\Rightarrow \sum_{k=1}^{M} p_k \log_2\left(\frac{q_k}{p_k}\right) = \sum_{k=1}^{M} p_k \frac{\log_{10}(q_k/p_k)}{\log_{10}(2)}.$$

$\times$ & $\div$ RHS by $\log_{10}(e)$ and rearrange the terms.

$$\Rightarrow \sum_{k=1}^{M} p_k \log_2\left(\frac{q_k}{p_k}\right) = \sum_{k=1}^{M} p_k \frac{\log_{10}(e)}{\log_{10}(2)} \cdot \frac{\log_{10}(q_k/p_k)}{\log_{10}(e)}.$$

$$= \sum_{k=1}^{M} p_k \log_2(e) \cdot \log_e\left(q_k/p_k\right).$$

Here, $\log_e(q_k/p_k) = \ln(q_k/p_k)$.

$$\Rightarrow \sum_{k=1}^{M} p_k \log_2(q_k/p_k) = \sum_{k=1}^{M} p_k \log_2(e) \ln(q_k/p_k)$$

$$= \log_2(e) \sum_{k=1}^{M} p_k \ln(q_k/p_k).$$

from eqn ① we can write, $\ln(q_k/p_k) \leq \left(\frac{q_k}{p_k} - 1\right)$

$$\Rightarrow \sum_{k=1}^{M} p_k \log_2\left(\frac{q_k}{p_k}\right) \leq \log_2(e) \cdot \sum_{k=1}^{M} p_k \left(\frac{q_k}{p_k} - 1\right).$$

$$\leq \log_2(e) \sum_{k=1}^{M} q_k - p_k.$$

$$\leq \log_2(e) \left[\sum_{k=1}^{M} q_k - \sum_{k=1}^{M} p_k\right]$$

– We know that $\sum_{k=1}^{M} p_k = 1$ and $\sum_{k=1}^{M} q_k = 1$.

$$\Rightarrow \sum_{k=1}^{M} p_k \log_2\left(\frac{q_k}{p_k}\right) \leq 0. \quad \text{— (2)}$$

– Now let us consider that $q_k = 1/M$ for all k.
(i.e) all the symbols of the alphabet are equally likely.

$$\text{eqn (2)} \Rightarrow \sum_{k=1}^{M} p_k \left[\log_2 q_k + \log_2 1/p_k\right] \leq 0.$$

$$\sum_{k=1}^{M} p_k \log_2(q_k) + \sum_{k=1}^{M} p_k \log_2(1/p_k) \leq 0.$$

$$\sum_{k=1}^{M} p_k \log_2(1/p_k) \leq - \sum_{k=1}^{M} p_k \log_2(q_k).$$

$$\leq \sum_{k=1}^{M} p_k \log_2\left(\frac{1}{q_k}\right).$$

Sub $q_k = 1/M$.

$$\Rightarrow \sum_{k=1}^{M} p_k \log_2(1/p_k) \leq \sum_{k=1}^{M} p_k \log_2(M).$$

$$\leq \log_2(M) \underbrace{\sum_{k=1}^{M} p_k}_{=1}.$$

$$\Rightarrow \sum_{k=1}^{M} p_k \log_2(1/p_k) \leq \log_2(M).$$

The LHS of the above equation is entropy $H(x)$ with arbitrary probability distribution.

(i.e) $\boxed{H(X) \leq \log_2(M)}$

- This is the proof of upper bound on entrophy.
- The maximum value of entrophy is

$$H_{max}(X) = \log_2(M)$$

INFORMATION RATE.

- The information rate is represented by R and it is given as,

$$R = r H.$$

Here, $r \rightarrow$ rate at which messages are generated.

$R \rightarrow$ Information rate.

$H \rightarrow$ Entrophy or average information.

- Information rate (R) is represented in average number of bits of information per second.
- It is calculated as,

$$R = \left(r \text{ in } \frac{\text{messages}}{\text{second}}\right) \times \left(H \text{ in } \frac{\text{Information bits}}{\text{message}}\right)$$

$$R = \text{Information bits / second}.$$

SOURCE CODING THEOREM (SHANNON'S FIRST THEOREM)

— The efficient source coder should satisfy following requirement.

i) The codewords generated by the encoder should be binary in nature.

ii) The source code should be unique in nature. (i.e) every codeword should represent unique message.

— Let there be 'L' number of messages emitted by the source.

— The probability of the k^{th} message is p_k and the number of bits assigned to this message be n_k.

— Then the average number of bits ($\bar{N}$) in the codeword of the message is given as,

$$\bar{N} = \sum_{k=0}^{L-1} p_k n_k .$$

— Let N_{min} be the minimum value of $\bar{N}$.

— Then the coding efficiency of the source encoder is defined as,

$$\eta = \frac{N_{min}}{\bar{N}} .$$

— The source encoder is called efficient if coding efficiency (η) approaches unity.

— In other words $N_{min} \leq \bar{N}$ and coding efficiency is maximum when $N_{min} \simeq \bar{N}$.

- The value of N_{min} can be determined using Shannon's first theorem, called source coding theorem.
- It is stated as follows,

"Given a discrete memoryless source of entropy H, the average codeword length $\bar{N}$ for any distortionless source encoding is bounded as, $\bar{N} \geq H$ "

- This limit says that the average number of bits per symbol cannot be made smaller than entropy H. Hence, $N_{min} = H$.
- The efficiency of source encoder can be written as, $\eta = \frac{H}{\bar{N}}$

CODE REDUNDANCY. (γ)

- It is the measure of redundancy of bits in the encoded message sequence.
- It is given as, $\gamma = 1 - \eta$.

  $\eta \rightarrow$ code efficiency.
- Redundancy should be as minimum as possible.

CODE VARIANCE (σ^2)

- Variance of the code is given as,

$$\sigma^2 = \sum_{K=0}^{M-1} P_K (n_K - \bar{N})$$

$$\sigma^2 = \sum_{k=0}^{M-1} p_k (n_k - \bar{N}).$$

Here, $M \rightarrow$ No. of symbols.

$p_k \rightarrow$ probability of k^{th} symbol.

$n_k \rightarrow$ No. of bits assigned to k^{th} symbol.

$\bar{N} \rightarrow$ Average codeword length.

— Variance is the measure of variability in codeword lengths.

— Variance should be as minimum as possible.

SHANNON – FANO ALGORITHM.

— It is used to encode the messages depending upon their probabilities.

— This algorithm allots less number of bits for highly probable messages and more number of bits for rarely occuring messages.

Steps:-

— The messages are listed in a column and the probabilities of occurrance of those messages are listed correspondingly in the next column.

— Partitioning of probabilities is made such that sum of probabilities in both the partitions are almost equal.

- The messages in the upper partition are assigned bit '0' and lower partition are assigned bit '1'.
- Those partitions are further subdivided into new partitions following the same rule.
- The partitioning is stopped when there is only one message in partition.
- The codeword is obtained by reading the bits of a particular message row wise through all the columns.
- In Shannon-fano algorithm, average number of binary digits per message are reduced and maximum information is conveyed by every binary digit (binit).

HUFFMAN CODING.

- This algorithm assigns different number of binary digits to the messages according to their probabilities of occurance.
- This type of coding makes average number of binary digits per message nearly equal to Entropy.

Steps :-

i) The messages are arranged according to their decreasing order of probabilities.

ii) The two messages of lowest probabilities are assigned binary '0' and '1'.

iii) The lowest two probabilities are added.

iv) The sum of probabilities in Stage-I ~~are~~ is placed in stage-II such that the probabilities are in descending order.

v) Steps ii), iii) & iv) are repeated again. until only two probability value being assigned 0 and 1 respectively are obtained.

vi) Codeword is obtained by tracing from Stage-I to the last stage. Thus the traced sequence is 1100 (for example) and thus the codeword is 0011.

EXAMPLE FOR SHANNON - FANO & HUFFMAN CODES

A discrete memoryless source has five symbols x_1, x_2, x_3, x_4 and x_5 with probabilities 0.4, 0.19, 0.16, 0.15 and 0.15 respectively.

i) construct a Shannon-Fano code for the source and calculate code efficiency η.

ii) Repeat i) for Huffman's code. compare the two techniques of source coding.

Solution:

i) To obtain Shannon - Fano code.

MESSAGE	PROBABILITY OF MESSAGE	STAGES I	STAGES II	STAGES III	CODE WORD FOR MESSAGE	NO. OF BITS PER MESSAGE (i.e) n_K.
x_1	0.4	0			0	1
x_2	0.19	1	0	0	100	3
x_3	0.16	1	0	1	101	3
x_4	0.15	1	1	0	110	3
x_5	0.15	1	1	1	111	3

— The entrophy (H) is given by,

$$H = \sum_{K=1}^{M} P_K \log_2 (1/P_K).$$

$$H = \sum_{K=1}^{5} P_K \log_2 (1/P_K)$$

$$H = 0.4 \log_2 (1/0.4) + 0.19 \log_2 (1/0.19) + 0.16 \log_2 (1/0.16) + 0.15 \log_2 (1/0.15) + 0.15 \log_2 (1/0.15)$$

$$H = 2.281 \text{ bits/message.}$$

— The average number of bits per message,

$$\bar{N} = \sum_{K=0}^{L-1} P_K n_K$$

$$\bar{N} = 0.4(1) + 0.19(3) + 0.16(3) + 0.15(3) + 0.15(3)$$

$\bar{N} = 2.35.$

– The code efficiency is given by,

$$\eta = \frac{H}{\bar{N}} = \frac{2.2281}{2.35} = 0.948.$$

ii) To obtain Huffman code.

MESSAGE	STAGE-I	STAGE-II	STAGE-III	
x_1	0.4	0.4	0.4	0.65 (0)
x_2	0.19	0.3	0.35 (0)	0.4 (1)
x_3	0.16	0.19 (0)	0.3 (1)	
x_4	0.15 (0)	0.16 (1)		
x_5	0.15 (1)			

– Obtaining codewords by tracing along the path.

MESSAGE	PROBABILITY	DIGITS OBTAINED BY TRACING $b_0\ b_1\ b_2$	CODEWORD $b_2\ b_1\ b_0$	No. OF DIGITS n_k.
x_1	0.4	1	1	1
x_2	0.19	000	000	3
x_3	0.16	100	001	3
x_4	0.15	010	010	3
x_5	0.15	110	011	3

- Average no. of bits per message,

$$\bar{N} = \sum_{k=0}^{L-1} P_k n_k$$

$$\bar{N} = 0.4(1) + 0.19(3) + 0.16(3) + 0.15(3) + 0.15(3)$$

$$\bar{N} = 2.35.$$

- code efficiency,

$$\eta = \frac{H}{\bar{N}} = \frac{2.2281}{2.35} = 0.948.$$

- Thus the code efficiency of Shannon-fano code and Huffman code is same in this example.

Example(2) Problem.

- compare the Huffman coding and Shannon-fano algorithms for data compression. for a discrete memoryless source 'x' with six symbols $x_0, x_1, \ldots x_6$. find a compact code if the probability distribution is $P(x_1) = 0.3$, $P(x_2) = 0.25$, $P(x_3) = 0.2$, $P(x_4) = 0.12$, $P(x_5) = 0.08$ and $P(x_6) = 0.05$.
calculate entrophy of the source, average length of the code, efficiency and redundancy of the code.

Discrete Memoryless Channels.

→ The discrete memoryless channel has input x and output y.

→ Both x & y are random variables.

→ The channel is discrete when both x and y are discrete.

→ The channel is called memoryless (Zero memory) when current O/p depends only on current I/p.

→ The channel is decribed in terms of I/p alphabet O/p alphabet and the set of transition probabilities.

→ The transition probability $P(y_j/x_i)$ is the conditional probability of y_j is received, given that x_i was transmitted.

- If $i=j$, then $P(y_j/x_i)$ represents conditional probability of correct reception.
- If $i \neq j$, then $P(y_j/x_i)$ represents conditional probability of error.

→ The conditional probability matrix of the channel is given by,

$$\underset{n \times m}{P} = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) & \cdots & P(y_m/x_1) \\ P(y_1/x_2) & P(y_2/x_2) & \cdots & P(y_m/x_2) \\ \vdots & & & \\ P(y_1/x_n) & P(y_2/x_n) & \cdots & P(y_m/x_n) \end{bmatrix} \quad \text{—(1)}$$

→ It is called the channel matrix or probability transition matrix.

→ Each row of the matrix represents fixed I/P and each column of the matrix represents fixed O/P.

→ The summation of all transition probabilities along the row is equal to 1

(i.e) $P(y_1/x_1) + P(y_2/x_1) + \cdots + P(y_m/x_1) = 1$

This is applicable to other rows also.

$$\Rightarrow \sum_{j=1}^{m} P(y_j/x_i) = 1$$

(i.e) for the fixed input x_i, the O/P can be any of $y_1, y_2, y_3, \ldots, y_m$. The summation of all these possibilities is equal to 1.

→ From the probability theory, we know that

$$P(AB) = P(B/A)\,P(A).$$

↓

Joint probability of A and B.

→ Here if $A = x_i$ and $B = y_j$ then,

$$P(x_i y_j) = P(y_j/x_i) \cdot P(x_i). \quad \text{—— (2)}$$

↓

Joint probability of x_i and y_j.

→ If we add all the joint probabilities for fixed y_j then we get $P(y_j)$.

(i.e) $\displaystyle\sum_{i=1}^{n} P(x_i, y_j) = P(y_j).$ —— (3)

→ The above equation gives the probability of getting symbol y_j.

→ From eqns (2) & (3)

$$P(y_j) = \sum_{i=1}^{n} P(y_j/x_i)\, P(x_i) \quad \text{--- (4)}.$$

Here $j = 1, 2, 3, \ldots, m$

→ If we are given the probabilities of I/P symbols and transition probabilities, it is possible to calculate the probabilities of O/P symbols.

→ Error occurs when i^{th} symbol is transmitted but j^{th} symbol is received. Hence, the error probability P_e can be obtained as,

$$P_e = \sum_{\substack{j=1 \\ j \neq i}}^{m} P(y_j). \quad \text{--- (5)}$$

Sub eqn. (4) in eqn (5)

$$\Rightarrow P_e = \sum_{\substack{j=1 \\ j \neq i}}^{m} \sum_{i=1}^{n} P(y_j/x_i)\, P(x_i) \quad \text{--- (6)}$$

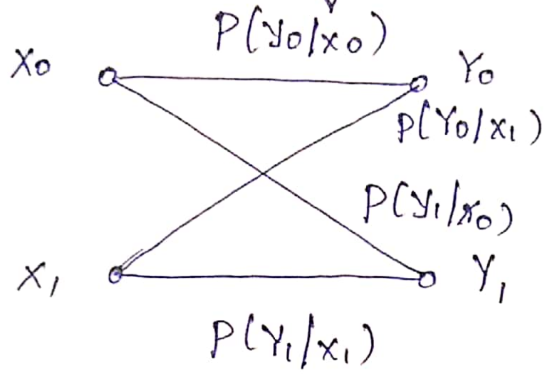
→ The probability of correct reception will be,

$$P_c = 1 - P_e \quad \text{--- (7)}.$$

→ Thus all probabilities will contribute to error except $i=j$. This is because in case of $i=j$, correct symbol is received.

BINARY COMMUNICATION CHANNEL.

Consider the case of the discrete channel where there are only two symbols transmitted.



/ Binary communication channel /

We can write the equations for probabilities of y_0 and y_1 as,

$$P(y_0) = P(y_0/x_0)\,P(x_0) + P(y_0/x_1)\,P(x_1).$$

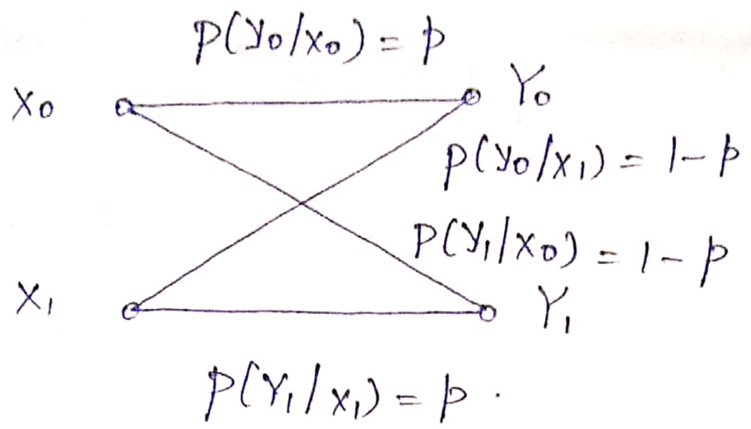
and $$P(y_1) = P(y_1/x_1)\,P(x_1) + P(y_1/x_0)\,P(x_0).$$

Above equations can be written in the matrix form as,

$$\begin{bmatrix} P(y_0) \\ P(y_1) \end{bmatrix} = \underbrace{\begin{bmatrix} P(y_0/x_0) & P(y_0/x_1) \\ P(y_1/x_0) & P(y_1/x_1) \end{bmatrix}}_{\text{probability transition matrix.}} \begin{bmatrix} P(x_0) \\ P(x_1) \end{bmatrix}$$

BINARY SYMMETRIC CHANNEL.

- The binary communication channel is said to be symmetric if $P(y_0/x_0) = P(y_1/x_1) = p$



/ Binary Symmetric channel /.

For the above channel,

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

EQUIVOCATION (CONDITIONAL ENTROPHY)

- The conditional entrophy $H(X/Y)$ is called equivocation. It is defined as,

$$H(X/Y) = \sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i, y_j) \log_2 \frac{1}{p(x_i/y_j)}$$

and the joint entrophy $H(XY)$ is given as,

$$H(XY) = \sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i, y_j) \log_2 \frac{1}{p(x_i, y_j)}$$

- The conditional entrophy.

$H(X/Y) \rightarrow$ uncertainty of X, on average, when Y is known.

$H(Y/X) \rightarrow$ uncertainty of Y, on average, when X is transmitted.

— H(Y/x) can be given as,

$$H(Y/x) = \sum_{i=1}^{M} \sum_{j=1}^{N} p(x_i, y_j) \log_2 \frac{1}{p(y_j/x_i)}$$

— The conditional entrophy H(X/Y) is an average measure of uncertainty in X after Y is received. In other words, H (X/Y) represents the information lost in the noisy channel.

RATE OF INFORMATION TRANSMISSION OVER A DISCRETE CHANNEL.

— The entrophy of the symbol gives average amount of information going into the channel. (i.e) $H(X) = \sum_{i=1}^{M} p_i \log_2 \left(\frac{1}{p_i}\right)$.

— Let the symbols be generated at the rate of 'r' symbols per second. Then the average rate of information going into the channel is given as,

$$D_{in} = r\,H(X) \text{ bits/sec.}$$

— Error are introduced in the data during the transmission.

— Because of these errors some information is lost in the channel.

— The conditional entrophy H(X/Y) is the measure of information lost in the channel.

- Hence, the information transmitted over the channel will be,

$$\text{Tx. Information} = H(x) - H(x/y).$$

- Hence, the average rate of information transmission D_t across the channel will be,

$$D_t = [H(x) - H(x/y)].r \quad \text{bits/sec.}$$

- When the noise becomes very large, then x & y become statistically independent.

$\Rightarrow$ $H(x/y) = H(x)$ and hence no information is transmitted over the channel.

- In case of errorless transmission $H(x/y) = 0$, hence, $D_{in} = D_t$.

(i.e) the I/p information rate is the same as information rate across the channel.

- No information is lost, when $H(x/y) = 0$.

CAPACITY OF A DISCRETE MEMORYLESS CHANNEL.

- The channel capacity is denoted as C.

It is given as,

$$C = \max_{p(x)} \{D_t\}.$$

Sub $D_t = [H(x) - H(x/y)].r$

$$\Rightarrow C = \max_{p(x)} \{H(x) - H(x/y)\}.r$$

- The maximum is taken with respect to probability of random variable X.
- The channel capacity can be defined as the maximum possible rate of information transmission across the channel.

MUTUAL INFORMATION.

- The mutual information is defined as the amount of information transferred when x_i is transmitted and y_j is received.
- It is represented by $I(x_i, y_j)$ and given as,

$$I(x_i, y_j) = \log_2 \left[\frac{P(x_i/y_j)}{P(x_i)} \right] \text{ bits.}$$

Here, $I(x_i, y_j) \rightarrow$ Mutual information.

$P(x_i/y_j) \rightarrow$ conditional probability that x_i was transmitted and y_j is received.

$P(x_i) \rightarrow$ probability of symbol x_i for transmission.

- The average mutual information is represented by $I(X;Y)$. It is calculated in bits / Symbol.
- The average mutual information is defined as the amount of source information gained per received symbol.

– It is given as,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j)\, I(x_i, y_j).$$

– Thus, $I(x_i, y_j)$ is weighted by joint probabilities $p(x_i, y_j)$ over all the joint events possible.

– Sub the value of $I(x_i; y_j)$,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 \frac{p(x_i/y_j)}{p(x_i)}.$$

PROPERTIES OF MUTUAL INFORMATION.

i) The mutual information of the channel is symmetric. (i.e) $I(X;Y) = I(Y;X)$.

ii) The mutual information can be expressed in terms of entropies of channel I/p or O/p and conditional entropies.

(i.e) $I(X;Y) = H(X) - H(X/Y)$

$= H(Y) - H(Y/X)$.

Here, $H(X/Y)$ & $H(Y/X) \rightarrow$ conditional entropies.

iii) The mutual information is always positive.

(i.e) $I(X;Y) \geq 0$.

iv) The mutual information is related to the joint entrophy $H(X,Y)$ by the following relation:

$$I(X;Y) = H(X) + H(Y) - H(X,Y).$$

CHANNEL CAPACITY.

— channel capacity can be expressed in terms of mutual information.

— Mutual information is given by,

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 \frac{p(x_i/y_j)}{p(x_i)},$$

where,

$$p(x_i; y_j) = p(y_j/x_i)\, p(x_i).$$

We know that,

$$\frac{p(x_i/y_j)}{p(x_i)} = \frac{p(y_j/x_i)}{p(y_j)}.$$

$$\Rightarrow I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(y_j/x_i)\, p(x_i) \log_2 \left(\frac{p(y_j/x_i)}{p(y_j)} \right).$$

and $p(y_j)$ can be represented as,

$$p(y_j) = \sum_{i=1}^{n} p(y_j/x_i)\, p(x_i).$$

$$\Rightarrow I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(y_j|x_i)\, p(x_i) \log_2 \left[\frac{p(y_j|x_i)}{\sum_{i=1}^{n} p(y_j|x_i)\, p(x_i)} \right]$$

— Mutual information is obtained from transition probabilities $p(y_j/x_i)$ and $p(x_i)$.

— Transition probabilities $p(y_j|x_i)$ are the characteristic of the channel but $p(x_i)$ are independent of the channel.

- The channel capacity of the discrete memoryless channel is given as maximum average mutual information.
- The maximization is taken with respect to input probabilities $P(x_i)$.

(i.e) $$C = \max_{P(x_i)} I(X;Y).$$

DIFFERENTIAL ENTROPHY AND MUTUAL INFORMATION FOR CONTINUOUS ENSEMBLES.

→ DIFFERENTIAL ENTROPHY.

- consider, a continuous random variable X having probability density function $f_X(x)$.

Then, $$h(X) = \int_{-\infty}^{\infty} f_X(x) \log_2 \left[\frac{1}{f_X(x)}\right] dx.$$

Here, $h(X) \rightarrow$ Differential entrophy of X.

→ MUTUAL INFORMATION.

- Mutual information for the continuous variables can be defined as,

$$I(X;Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y) \log_2 \left[\frac{f_X(x|y)}{f_X(x)}\right] dx\,dy$$

Here, $f_{XY}(x,y) \rightarrow$ Joint pdf of X and Y.

$f_X(x|y) \rightarrow$ Conditional pdf of X given Y.

— The mutual information has similar properties to those of continuous random variables.

i) $I(X;Y) = I(Y;X)$.

ii) $I(X;Y) \geq 0$.

iii) $I(X;Y) = h(X) - h(X/Y)$.

iv) $I(X;Y) = h(Y) - h(Y/X)$.

CHANNEL CAPACITY THEOREM.

— The channel considered is a bandlimited, powerlimited white gaussian noise channel.

Step-1 : Assume a zero mean stationary process $X(t)$ is bandlimited to B Hz. Let X_k, $k=1,2,\ldots,n$ indicate the continuous random variables obtained by sampling $X(t)$.

— Let Y_k, $k=1,2,\ldots,n$ denote the samples of received signal. They are related to X_k as,

$$Y_k = X_k + N_k \quad , \quad k=1,2,\ldots,n.$$

Here N_k are samples of white Gaussian noise of zero mean and variance of σ^2.

Step-2 : The channel capacity for the channel described is,

$$C = \max_{f_{X_k}(x)} \left\{ I(X_k;Y_k) : X_k \text{ Gaussian} \right\}$$

Here $I(X_K ; Y_K)$ is average mutual information and $f_{X_K}(x)$ is the pdf of X_K.

Step-3: The average mutual information $I(X_K ; Y_K)$ is given as,

$$I(X_K ; Y_K) = h(Y_K) - h\left(\frac{Y_K}{X_K}\right).$$

and we know that $h\left(\frac{Y_K}{X_K}\right) = h(N_K)$.

$$\Rightarrow I(X_K ; Y_K) = h(Y_K) - h(N_K).$$

Step-4: The variance of Y_K equals $S+\sigma^2$. Here S is the average transmitted power.

$$\Rightarrow h(Y_K) = \frac{1}{2}\log_2\left[2\pi e(S+\sigma^2)\right]$$

($\because$ The differential entropy of random variable 'x' having gaussian distribution,

(i.e) $h(x) = \frac{1}{2}\log_2(2\pi e\sigma^2)$.

Here variance of $Y_K = (S+\sigma^2)$.

$$\Rightarrow h(Y_K) = \frac{1}{2}\log_2(2\pi e(S+\sigma^2))).$$

— The variance of noise is equal to σ^2.

$$\Rightarrow h(N_K) = \frac{1}{2}\log_2(2\pi e\sigma^2)$$

Step-5: Substituting the values in $I(X_K ; Y_K)$

$$\Rightarrow I(X_K ; Y_K) = \frac{1}{2}\log_2\left[2\pi e(S+\sigma^2)\right] - \frac{1}{2}\log_2(2\pi e\sigma^2)$$

$$\Rightarrow I(X_K; Y_K) = \frac{1}{2}\log_2\left[\frac{2\pi e\,(S+\sigma^2)}{2\pi e\,\sigma^2}\right]$$

$$= \frac{1}{2}\log_2\left(\frac{S+\sigma^2}{\sigma^2}\right)$$

$$= \frac{1}{2}\log_2\left(1+\frac{S}{\sigma^2}\right).$$

$$\left(\because \log A - \log B = \log\left(\frac{A}{B}\right)\right).$$

$$\Rightarrow I(X_K; Y_K) = \frac{1}{2}\log_2\left(1+\frac{S}{\sigma^2}\right).$$

This is nothing but the channel capacity of Gaussian channel. (i.e) $C = \frac{1}{2}\log_2\left(1+\frac{S}{\sigma^2}\right)$.

— When this channel is used over the bandwidth of 'B' Hz, then the above equation becomes,

$$C = B\log_2\left(1+\frac{S}{N}\right) \text{ bits/sec.}$$

Shannon's Theorems On Channel Capacity.

→ Channel coding theorem (Shannon's second theorem).

Statement :

Given a source of M equally likely messages, with $M \gg 1$, which is generating information at a Rate 'R'. Given channel with channel capacity C. Then if, $R \leq C$, there exists a coding technique such that the output of the source may be transmitted over the channel with a probability of error in the received message which may be made arbitrarily small.

Explanation:

— This theorem says that it is possible to transmit information without any error even if noise is present. Coding techniques are used to detect and correct the errors.

→ Shannon Hartley theorem for Gaussian channel (continuous channel).

— Channel capacity theorem.

→ When Shannon's theorem of channel capacity is applied specifically to a channel in which the noise is gaussian is known as Shannon – Hartley theorem.

→ It is also called Information capacity theorem.

→ The channel capacity of a white band limited gaussian channel is,

$$C = B \log_2 \left(1 + \frac{S}{N}\right) \text{ bits/sec.}$$

Where, $B \rightarrow$ channel Bandwidth.

$S \rightarrow$ Signal power.

$N \rightarrow$ total noise power within the channel BW.

→ We know that signal power is given as,

$$P = \int_{-B}^{B} \text{power spectral density.}$$

→ B is the Bandwidth, and the power spectral density of white noise is $\frac{No}{2}$.
→ Hence, Noise power becomes,

$$N = \int_{-B}^{B} \frac{No}{2} \cdot df.$$

$$N = No.B.$$

TRADEOFF BETWEEN BANDWIDTH & SIGNAL TO NOISE RATIO.

— The channel capacity of the Gaussian channel is given as,

$$C = B \log_2 \left(1 + \frac{S}{N}\right).$$

— The above equation shows that the channel capacity depends on two factors.

i) B → Bandwidth of the channel.

ii) S/N → Signal to Noise ratio.

— Noiseless channel has infinite capacity.

→ No noise in the channel, then N=0.

$$\Rightarrow \frac{S}{N} = \infty.$$

Such a channel is called noiseless channel.

→ Capacity of such channel,

$$C = B \log_2 (1+\infty) = \infty.$$

→ Thus noiseless channel has infinite capacity.

– Infinite bandwidth has limited capacity.

→ If BW = ∞, the channel capacity is limited.

→ Because, as BW increases, noise power also increases. Noise power is given by $N = N_0 B$.

→ Due to this increase in noise power, S/N ratio decreases.

→ Hence, even if BW – 'B' approaches infinity, capacity doesnot approach infinity.

→ As $B \to \infty$, capacity approaches an upper limit. This upper limit is given by,
$$C_\infty = \lim_{B \to \infty} C = 1.44 \frac{S}{N_0}.$$

PROOF FOR $C_\infty = \lim_{B \to \infty} C = 1.44 \frac{S}{N_0}$.

Noise power is given as, $N = N_0 B$.

⇒ WKT: $C = B \log_2 \left(1 + \frac{S}{N}\right)$ — (1) /channel capacity/

Sub. $N = N_0 B$ in eqn (1)

$$\Rightarrow C = B \log_2 \left(1 + \frac{S}{N_0 B}\right) \quad \text{— (2)}$$

Rearranging the above equation,

$$C = \frac{S}{N_0} \cdot \frac{N_0 B}{S} \log_2 \left(1 + \frac{S}{N_0 B}\right) \quad \text{— (3)}$$

$$C = \frac{S}{N_o} \log_2 \left(1+\frac{S}{N_oB}\right)^{\frac{N_oB}{S}}.$$

$$C = \frac{S}{N_o} \log_2 \left(1+\frac{S}{N_oB}\right)^{1/\left(\frac{S}{N_oB}\right)} \quad \text{— (4)}$$

— Let us apply the limits as $B \to \infty$.

$$\Rightarrow C_\infty = \lim_{B\to\infty} C = \lim_{B\to\infty} \frac{S}{N_o} \cdot \log_2 \left(1+\frac{S}{N_oB}\right)^{1/\left(\frac{S}{N_oB}\right)}$$

— In the above equation put $x = \frac{S}{N_oB}$, as $B \to \infty$, $x \to 0$.

(i.e) $$C_\infty = \frac{S}{N_o} \lim_{x=0} \log_2 (1+x)^{1/x}.$$

— The standard relation, $\lim_{x=0} (1+x)^{1/x} = e$, then the above equation becomes,

$$C_\infty = \frac{S}{N_o} \log_2 (e) = \frac{S}{N_o} \cdot \frac{\log_{10}(e)}{\log_{10}(2)}.$$

$$C_\infty = 1.44 \frac{S}{N_o}.$$

— This gives the upper limit on channel capacity as bandwidth 'B' approaches infinity.

RATE / BANDWIDTH & SIGINAL TO NOISE RATIO, (E_b/N_o) TRADEOFF.

- Let the system is transmitting at a rate R_b, which is equal to channel capacity C.
- Then average transmitted signal power will be,

$$S = E_b \cdot C$$

Here, $E_b \rightarrow$ Transmitted energy per bit.

- The channel capacity is bits/sec. Hence, the product of E_b and C gives average signal power.
- Capacity of continuous channel is given by,

$$C = B \log_2 \left(1 + \frac{S}{N}\right).$$

- As defined above sub. $S = E_b \cdot C$ and $N = N_o B$. in the above equation.

$$\Rightarrow C = B \log_2 \left(1 + \frac{E_b \cdot C}{N_o \cdot B}\right).$$

$$\Rightarrow \frac{C}{B} = \log_2 \left(1 + \frac{E_b}{N_o} \cdot \frac{C}{B}\right)$$

$$\frac{C}{B} = \frac{\log_e \left(1 + \frac{E_b}{N_o} \frac{C}{B}\right)}{\log_e (2)}.$$

(Changing to base 'e' from base '2')

Scanned by CamScanner

$$\Rightarrow \frac{C}{B} \log_e (2) = \log_e \left(1 + \frac{E_b}{N_o} \frac{C}{B}\right)$$

$$\log_e (2)^{C/B} = \log_e \left(1 + \frac{E_b}{N_o} \frac{C}{B}\right).$$

$$\Rightarrow (2)^{C/B} = \left(1 + \frac{E_b}{N_o} \cdot \frac{C}{B}\right).$$

$$1 + \frac{E_b}{N_o} \cdot \frac{C}{B} = (2)^{C/B}$$

$$\frac{E_b}{N_o} \cdot \frac{C}{B} = (2)^{C/B} - 1$$

$$\frac{E_b}{N_o} = \frac{(2)^{C/B} - 1}{C/B}$$

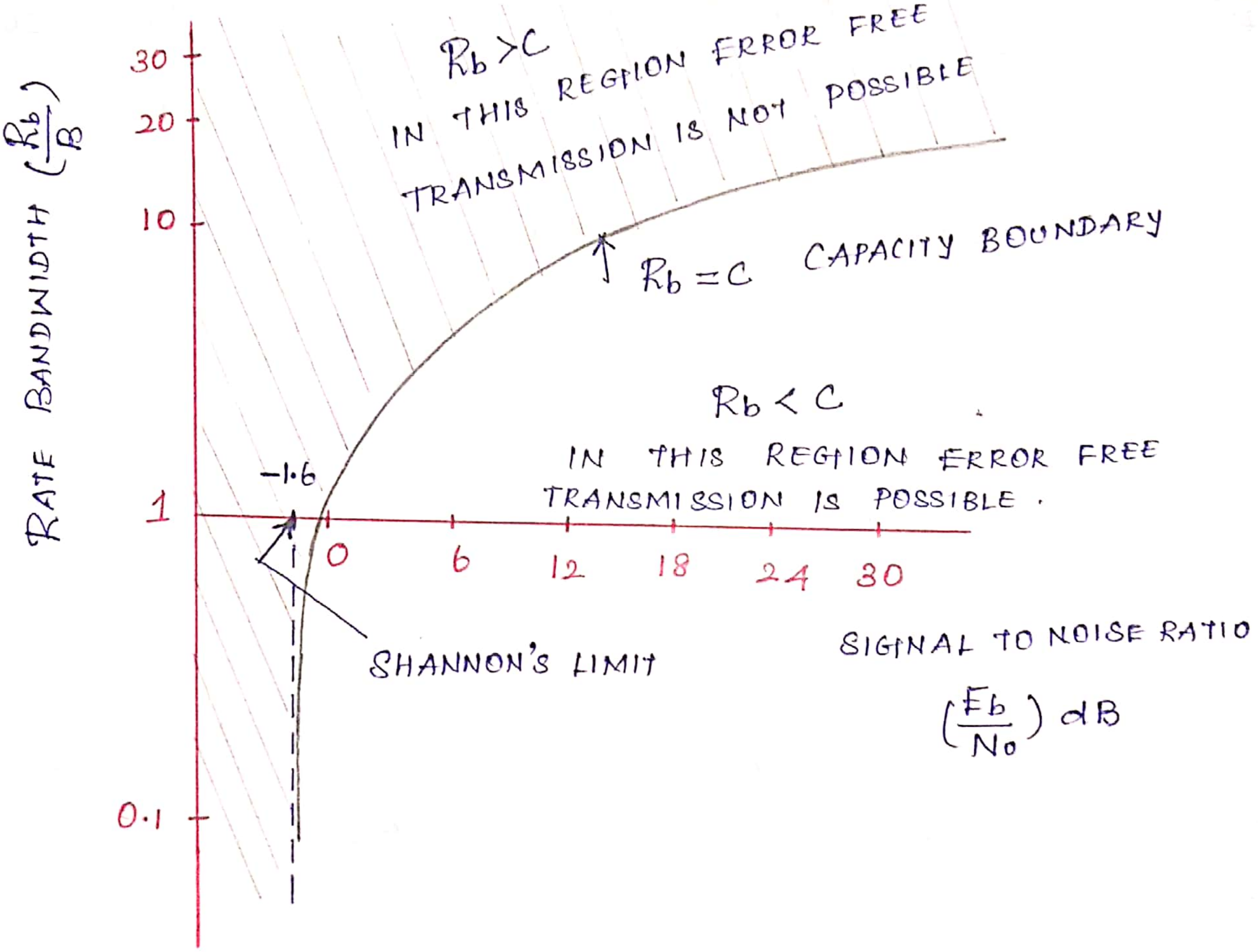
— In this equation, $\frac{E_b}{N_o}$ is the energy /bit to noise spectral density ratio.

— And C/B is called the bandwidth efficiency.

— When $C = R_b$, then the above equation becomes, $\frac{E_b}{N_o} = \frac{2^{(R_b/B)} - 1}{(R_b/B)}$.

Here $(R_b/B) \rightarrow$ rate bandwidth.

(Rb/B) Vs (S/N) RATIO DIAGRAM.



CONCLUSIONS:

i) for infinite bandwidth, $C_\infty = 1.44 \frac{S}{N_o}$.

Here, $S = E_b C = E_b C_\infty$.

$$\Rightarrow C_\infty = 1.44 . \frac{E_b C_\infty}{N_o}$$

$$\Rightarrow \frac{E_b}{N_o} = \frac{1}{1.44} = 0.693.$$

$$\left(\frac{E_b}{N_o}\right)_{dB} = 10 \log \left(\frac{E_b}{N_o}\right) = 10 \log (0.693)$$

$$= -1.6 \text{ dB}.$$

— Thus $\left(\frac{E_b}{N_0}\right) = -1.6\,dB$ for $B \to \infty$.

This value of E_b/N_0 is called Shannon's Limit. and the capacity at Shannon's Limit is, $C_\infty = 1.44 \frac{S}{N_0}$.

— $R_b = C \to$ Capacity boundary.

$R_b > C \to$ Error-free transmission is not possible.

$R_b < C \to$ Error-free transmission is possible.

— Probability of error for an ideal systema

$$P_e = \begin{cases} 1 & \text{for } R_b \geq C \\ 0 & \text{for } R_b < C. \end{cases}$$

PROBLEMS

1. A binary channel matrix is given as,

$$\begin{array}{c|cc} & y_1 & y_2 \leftarrow \text{O/p's} \\ \hline \text{I/p's} \begin{cases} x_1 \\ x_2 \end{cases} & \begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix} \end{array}$$

Determine $H(X)$, $H(X/Y)$, $H(Y/X)$ and mutual information $I(X;Y)$.

Solution:-

Source probabilities are not given.

$\Rightarrow$ Assume $p(x_1) = 1/3$ and $p(x_2) = 2/3$ (say).

Channel matrix,

$$P = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix}$$

To obtain $H(X)$:

$$H(X) = p(x_1) \log_2 \frac{1}{p(x_1)} + p(x_2) \log_2 \frac{1}{p(x_2)}$$

$$= 1/3 \log_2 (3) + 2/3 \log_2 (3/2)$$

$$= 0.9182 \text{ bits / Symbol.}$$

To obtain $H(Y)$:

$$\begin{bmatrix} P(y_1) \\ P(y_2) \end{bmatrix} = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix} \begin{bmatrix} P(x_1) & P(x_2) \end{bmatrix}$$

$$\begin{bmatrix} p(y_1) \\ p(y_2) \end{bmatrix} = \begin{bmatrix} 2/3 & 1/3 \\ 1/10 & 9/10 \end{bmatrix} \begin{bmatrix} 1/3 & 2/3 \end{bmatrix}$$

$$= \begin{bmatrix} 0.2889 \\ 0.7111 \end{bmatrix}$$

$$p(y_1) = 0.2889$$

$$p(y_2) = 0.7111$$

$$H(Y) = p(y_1) \log_2 \frac{1}{p(y_1)} + p(y_2) \log_2 \frac{1}{p(y_2)}$$

$$= 0.2889 \log_2 \frac{1}{0.2889} + 0.7111 \log_2 \frac{1}{0.7111}$$

$$= 0.8672 \text{ bits/symbol.}$$

To calculate joint entrophy $H(X, Y)$.

$$p(x_i y_j) = p(y_j / x_i) \cdot p(x_i)$$

$$p(x_1 y_1) = p(y_1 / x_1) p(x_1) = \frac{2}{3} \times \frac{1}{3} = 2/9$$

$$p(x_1 y_2) = p(y_2 / x_1) p(x_1) = \frac{1}{3} \times \frac{1}{3} = 1/9$$

$$p(x_2 y_1) = p(y_1 / x_2) p(x_2) = \frac{1}{10} \times \frac{2}{3} = 2/30$$

$$p(x_2 y_2) = p(y_2 / x_2) p(x_2) = \frac{9}{10} \times \frac{2}{3} = 18/30$$

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

$$= p(x_1y_1)\log_2\frac{1}{p(x_1y_1)} + p(x_1y_2)\log_2\frac{1}{p(x_1y_2)} + p(x_2y_1)\log_2\frac{1}{p(x_2y_1)} + p(x_2y_2)\log_2\frac{1}{p(x_2y_2)}.$$

$$= \frac{2}{9}\log_2(9/2) + \frac{1}{9}\log_2(9) + \frac{2}{30}\log_2(30/2) + \frac{18}{30}\log_2(30/18).$$

$= 1.5365$ bits/symbol.

<u>To obtain conditional entropies $H(x/y)$ and $H(y/x)$.</u>

$$H(x/y) = H(x,y) - H(y)$$
$$= 1.5365 - 0.8672$$
$$= 0.6692 \text{ bits/symbol.}$$

$$H(y/x) = H(x,y) - H(x)$$
$$= 1.5365 - 0.9182$$
$$= 0.6183 \text{ bits/symbol.}$$

<u>To Obtain mutual information $I(x;y)$</u>

$$I(x;y) = H(x) - H(x/y)$$
$$= 0.9182 - 0.6692$$
$$= 0.249 \text{ bits/symbol.}$$

$$I(X;Y) = H(Y) - H(Y/X)$$

$$= 0.8672 - 0.6183$$

$$= 0.249 \text{ bits/Symbol.}$$

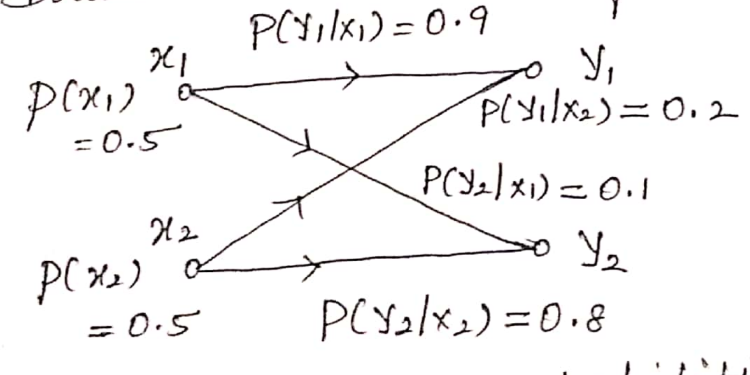
2. The channel transition matrix is given by, $\begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$. Draw the channel diagram and determine the probabilities associated with outputs assuming equiprobable inputs. Also find the mutual information $I(X;Y)$ for the channel.

Solution: The data given is,

$$P = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

I/p's are equiprobable, $P(x_1) = P(x_2) = 0.5$.

i) Draw the channel diagram.



ii) To obtain O/p probabilities:

$$\begin{bmatrix} P(y_1) \\ P(y_2) \end{bmatrix} = \begin{bmatrix} P(x_1) & P(x_2) \end{bmatrix} \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix}$$

Sub values,

$$\begin{bmatrix} p(y_1) \\ p(y_2) \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

$$= \begin{bmatrix} (0.5\times0.9)+(0.5\times0.2) \\ (0.5\times0.1)+(0.5\times0.8) \end{bmatrix} = \begin{bmatrix} 0.55 \\ 0.45 \end{bmatrix}$$

$p(y_1) = 0.55$ and $p(y_2) = 0.45$.

iii) To obtain mutual information.

$$I(x;y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i y_j) \log_2 \frac{P(x_i/y_j)}{p(x_i)}.$$

WKT: $p(x_i y_j) = p(x_i/y_j) \cdot p(y_j)$.

$$p(x_i y_j) = p(y_j/x_i) \cdot p(x_i)$$

$$P(x_i/y_j)\, p(y_j) = p(y_j/x_i) \cdot p(x_i)$$

$$\frac{P(x_i/y_j)}{p(x_i)} = \frac{p(y_j/x_i)}{p(y_j)}.$$

$$\Rightarrow I(x;y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(y_j/x_i)\, p(x_i) \log_2 \frac{p(y_j/x_i)}{p(y_j)},$$

$$= p(y_1/x_1)p(x_1)\log_2 \frac{P(y_1/x_1)}{p(y_1)} + P(y_1/x_2)p(x_2)\log_2 \frac{P(y_1/x_2)}{p(y_1)} +$$

$$P(y_2/x_1)\, P(x_1) \log_2 \frac{P(y_2/x_1)}{p(y_2)} + P(y_2/x_2)p(x_2)\log_2 \frac{P(y_2/x_2)}{p(y_2)}$$

$$I(x;y) = (0.9)(0.5)\log_2\left(\frac{0.9}{0.55}\right) + (0.2)(0.5)\log_2\left(\frac{0.2}{0.55}\right) + (0.1)(0.5)\log_2\left(\frac{0.1}{0.45}\right) + (0.8)(0.5)\log_2\left(\frac{0.8}{0.45}\right).$$

$$= 0.3197 - 0.1459 - 0.1085 + 0.332$$

$I(x;y) = 0.3973$ bits/symbol.

3. An analog signal having 4KHz bandwidth is sampled at 1.25 times the nyquist rate and each sample is quantized into one of 256 equally likely levels. Assuming the samples to be statistically independent.

i) What is information rate of this source?

ii) can the O/p of this source transmitted without error over an AWGN channel with a BW of 10KHz and (S/N) ratio of 20dB?

iii) find (S/N) ratio required for error-free transmission of part ii)

iv) find the bandwidth required for an AWGN channel for an errorfree transmission of the O/p of this source if (S/N) ratio is 20 dB.

Solution: The BW of the signal is

$B = 4$ KHz.

Nyquist rate $= 2 \times B = 8$ KHz.

$\Rightarrow$ Sampling rate $= r =$ 1.25 × Nyquist rate

$r = 1.25 \times 8000 = 10{,}000$ messages/s.

Since, the samples are quantized into 256 equally likely levels, $M = 256$ samples.

$\Rightarrow P_K = 1/256$. (Probability of occurance)

$\Rightarrow H = \log_2 (M)$ (Entrophy)

$H = \log_2 (256) = 8$ bits/sample.

i) Information Rate (R)

$R = rH = 10{,}000 \times 8$

$R = 80{,}000$ bits/second.

ii) $B = 10KHz$ & $(S/N) = 20dB$.

$(S/N)_{dB} = 10 \log_{10} (S/N)$

$20 = 10 \log_{10} (S/N)$

$(S/N) = 100$.

Capacity of AWGN channel is given by,

$C = B \log_2 (1 + S/N)$.

$C = 10\,000 \log_2 (1+100)$

$C = 10000 \dfrac{\log_{10} (101)}{\log_{10} (2)} = 66.582\,k$ bits/s

— If $R \le C$, Error free transmission is possible.

— Here, $R = 80{,}000$ bits/s & $C = 66582$ bits/s

$\Rightarrow R > C$

$\therefore$ Error free transmission is not possible

iii) (S/N) ratio for error-free Tx.

— from ii) we have $R = 80{,}000$ bits/s & $B = 10{,}000$ Hz.

— for errorfree Tx : $R \le C$.

$$\Rightarrow R \le C = B \log_2 (1 + S/N).$$

$$80000 \le 10000 \log_2 (1 + S/N).$$

$$8 \le \log_2 (1 + S/N).$$

$$8 \le \frac{\log_{10} (1 + S/N)}{\log_{10} (2)}.$$

$$2.40824 \le \log_{10} (1 + S/N)$$

$$S/N \geq 255$$

$$(S/N)_{dB} = 10 \log_{10} (255) = 24\, dB.$$

iv) To determine BW.

Given : $(S/N) = 20 dB$.

$$\Rightarrow (S/N)_{dB} = 10 \log_{10} (S/N) = 20$$

Scanned by CamScanner

$$20 = 10 \log (S/N)$$

$$(S/N) = 100$$

WKT: for error-free Tx, $R \leq C$.

$$\Rightarrow \quad R \leq C = B \log_2 (1 + S/N)$$

$$R \leq B \log_2 (1 + S/N).$$

Here, $R = 80000$ bits/s.

$$S/N = 100$$

$$\Rightarrow \quad 80000 \leq B \log_2 (1 + 100)$$

$$B \geq \frac{80000}{\log_2 (101)}$$

$$\therefore \quad B \geq 11.974 \text{ KHz}$$

This is the bandwidth required for error-free transmission.

4. consider a telegraph source having two symbols dot and dash. The dot duration is 0.2 sec: and the dash duration is 3 times of the dot duration. The probability of the dot's occuring is twice that of dash and time b/w symbols is 0.2 sec. calculate

information rate of the telegraph source.

Solution:

i) To calculate probabilities of dot & dash.

- Let probability of dash be 'p'. Then probability of dot will be '2p'. and

$$p + 2p = 1 \Rightarrow p = 1/3$$

$\therefore$ $p(\text{dash}) = 1/3$ and $p(\text{dot}) = 2/3$.

ii) To calculate entrophy of the source.

$$H = \sum_{K=1}^{M} P_K \log_2 (1/P_K).$$

$$H = 1/3 \log_2 (3) + 2/3 \log_2 (3/2)$$

$$H = 0.9183 \text{ bits /symbol.}$$

iii) To calculate average symbol rate.

dot duration $= T_{dot} = 0.2\,s$.

dash duration $= T_{dash} = 3 \times 0.2 = 0.6\,s$.

Duration b/w symbols $= T_{sym} = 0.2\,s$.

/ consider 1200 symbols are Tx /.

No. of dots $= 1200 \times p(\text{dot}) = 800$

No. of dash $= 1200 \times p(\text{dash}) = 400$

Total time for this string.

T = dots duration + dash duration + $(1200 \times T_{sym})$

$T = (800 \times 0.2) + (400 \times 0.6) + (1200 \times 0.2)$

$T = 640\, s.$

— Hence, average symbol rate will be,

$$r = \frac{1200}{T} = 1.875 \text{ symb/sec.}$$

iv) To calculate Information rate.

Info Rate : $r = rH$

$r = 1.875 \times 0.9183$

$r = 1.7218$ bits/sec.

— Thus the average information rate of the telegraph source will be 1.7218 bits/sec.