DC

UNIT-IV

8:47 65%

Information theory

Importance of information theory :-

1. it is used for mathematical analysis and data analysis of digital communication system.

2. this theory is used for calculating the amount of informat to represent a message (or) carrier symbol

3. the amount of information can be calculated by the probability of message (or) symbol i.e., $P(x_i)$

Information theory is necessary for analysis of digital data and calculation of amount of information to repre a message (or) symbol.

Information (or) amount of information :-

the amount of information can be calculated interms of bits the information is always indirectly proportional to probability of particular Event.

(or)

it can be defined as calculation of total no of bits required to represent a message (or) Event

let us consider the communication system transmit messages $m_1, m_2, m_3 \ldots m_K$ with probability occurence of $P_1, P_2, P_3 \ldots P_K$ the amount of information theory transmitted to the messages m_K with the probability of P_K is given as

$$I_K = \log_2 \frac{1}{P_K} \text{ Bits}$$

Here we can represent amount of infamation in 3 measurments.

$$I_K = \log_2 \frac{1}{P_K} \text{ Bits (or) Binits}$$

$$I_K = \log_e \frac{1}{P_K} \text{ Nats}$$

$$I_K = \log_{10} \frac{1}{P_K} \text{ Decits}$$

In our information theory we preferred first notation to calculate amount of information carried by message which is mathematically convinient method to represent the information interms of bits.

Properties of information :-

1. The information is always indirectly proportional to probability when $p=0$ the amount of information is "infinity" bits and when $p=1$ the amount of information is '1' bits

2. if there are 'm' equally independent symbols then the amount of information is carried by each symbol is

$I_K =$ 'N' bits (or) 'n' bits

where $M = 2^n$

3. if 'M' no of symbols are possible then the ~~possible~~ probability of each symbol is '$\frac{1}{M}$'. The amount of information is expressed as $I_K = \log_2 \frac{1}{P_K}$

$$I_K = \log_2 \frac{1}{\frac{1}{m}}$$

$$\boxed{I_K = \log_2 M}$$

4. If the receiver knows the information theory being transmitted (or) if the Event is sure then the information is carried by message is '0'.

we know that the amount of information is Carrie by a message is given by $I_K = \log_2 \frac{1}{P_K}$

According to the statement the Event is sure $P_K =$

sub $P_K = 1$ $I_K = \log_2 \frac{1}{1} = \log_2 1 = 0$

$$\boxed{I_K = 0}$$

If information 'I_1' Carried by a message 'm_1' and 'I_2' information is carried by a message 'm_2'. the overall information Carried by both messages. is given as

$$\boxed{I(m_1 m_2) = I(m_1) + I(m_2)}$$

proof:

$I(m_1)$ is the information Carried by a message 'm_1'

then $I(m_1) = \log_2 \frac{1}{P(m_1)}$

lly $I(m_2) = \log_2 \frac{1}{P(m_2)}$

Here m_1, m_2 are independent messages

therefore the overall (message) information carried by m_1 and m_2 can be return as

$$I(m_1, m_2) = \log_2 \frac{1}{P(m_1) P(m_2)}$$

$$= \log_2 \frac{1}{P(m_1)} + \log_2 \frac{1}{P(m_2)}$$

$$I(m_1 m_2) = I(m_1) + I(m_2)$$

$I(m_1 m_2) = I(m_1) + I(m_2)$

5. The amount of information and the probability values are always a positive integer value.

Average Information (or) Entropy :-

It is defined as the ratio of between total amount of information to total no. of messages the average information rate is necessary for all types of digital Electronics to calculate used space and free space from the total storage capacity.

this calculation is important to store the digital data in the memory devices.

$$\text{Entropy} = \frac{\text{total amount of information}}{\text{total no. of messages}}$$

it is expressed as Bits/symbol.

Expression for avg information (or) Entropy :-

Consider a source Emits 'm' different messages the messages are $m_1, m_2, m_3 \ldots m_m$ with the probabilities of $P_1, P_2, P_3 \ldots P_m$. If the messages are produced from in a single source and which are available in sequence, that sequence is indicated with 'L' then the probabilities of Each messages is given as

$m_1 = P_1 L$

$m_2 = P_2 L$

$m_3 = P_3 L$

$\vdots \quad \vdots$

$m_m = P_m L$

we know that amount of information in the first message is given as $I_1 = \log_2 \frac{1}{P_1}$

there are $P_1 L$ no. of messages of 'mi' message are transmitted then the total information carried by 'mi' messages is given as $I_1(\text{total}) = P_1 L \log_2 \frac{1}{P_1}$

$$\text{lly} \quad I_2(\text{total}) = P_2 L \log_2 \frac{1}{P_2}$$

$$I_3(\text{total}) = P_3 L \log_2 \frac{1}{P_3}$$

$$I_m(\text{total}) = P_m L \log_2 \frac{1}{P_m}$$

$$I(\text{total}) = P_1 L \log_2 \frac{1}{P_1} + P_2 L \log_2 \frac{1}{P_2} + \dots + P_m L \log_2 \frac{1}{P_m}$$

w.k.T $H(x) = \frac{I(\text{total})}{L}$

$$H(x) = \frac{\cancel{L}\left(P_1 \log_2 \frac{1}{P_1} + P_2 \log_2 \frac{1}{P_2} + \dots + P_m \log_2 \frac{1}{P_m}\right)}{\cancel{L}}$$

$$H(x) = P_1 \log_2 \frac{1}{P_1} + P_2 \log_2 \frac{1}{P_2} + \dots + P_m \log_2 \frac{1}{P_m}$$

$$H(x) = \sum_{K=1}^{m} P_K \log_2 \frac{1}{P_K} \quad \text{Bits/symbols}$$

(or)

$$\boxed{H(x) = -\sum_{K=1}^{m} P_K \log_2 P_K} \quad \text{Bits/symbols}$$

properties of entropy

1. The entropy is always a positive integer value

2. The avg information is always less than to total amount of information which is expressed as $H(x) < I(total)$

3. If the event is sure (or) the event is impossible then the entropy of such event is zero

$P_K = 1 \rightarrow$ Event is sure

$P_K = 0 \rightarrow$ Event is impossible

proof w.K.T $H = P_K \log_2 \frac{1}{P_K}$

when $\boxed{P_K = 1}$

$= 1 \log_2 1$

$H = 0$

$\boxed{P_K = 0}$

$H = P_K \log_2 \frac{1}{P_K}$

$H = 0 \log_2 \frac{1}{0}$

$\boxed{H = 0}$

4. Consider a source emits 'm' no. of symbols (or) messages if all 'm' messages are equally probable and that have same probability $\frac{1}{m}$, then the entropy of such event is $\boxed{H = \log_2 m}$

Proof w.K.T $H = \sum_{K=1}^{M=m} P_K \log_2 \frac{1}{P_K}$ $\quad P = \frac{1}{m}$

$H = P_1 \log_2 \frac{1}{P_1} + P_2 \log_2 \frac{1}{P_2} + P_3 \log_2 \frac{1}{P_3} + \dots + P_m \log_2 \frac{1}{P_m}$

$P_1 = P_2 = P_m = \frac{1}{m}$

$H = \frac{1}{m} \log_2 m + \frac{1}{m} \cdot \log_2 m + \frac{1}{m} \log_2 m + \dots + \frac{1}{m} \log_2 m$

$H = \frac{1}{\cancel{m}} \cancel{m} \log_2 m$ $\quad \boxed{H = \log_2 m}$ is proved.

2m)

ormation rate: (R).

The no.of bits transmitted per second is usually called as information rate which is calculated by taking the product of symbol rate and Entropy.

$$\boxed{R = \text{symbol rate}(r) \times \text{Entropy } (H)}$$

symbol rate = No.of symbols transmitted per second

$$R = \text{symbols/sec} \times \text{Bits/symbols}$$

$$\boxed{R = \text{Bits/sec}}$$

the unit for information rate is Bits/sec

Entropy

P) * prove that $0 \leq H_{max} \leq \log_2^m$

Proof: To prove the above property we will use the followi property of natural logarithm $\ln x \leq x-1$ for $x>0$

let us consider any two probability distributions $\{p_1, p_2 \ldots p_m\}$ and $\{q_1, q_2 \ldots q_m\}$ on the alphabet $X = \{x_1, x_2 \ldots x_m\}$ of the discrete memoryless source. then let us consider the term $\sum_{K=1}^{m} p_K \log_2 \left[\frac{q_K}{p_K}\right]$

this term can also written as $\sum_{K=1}^{m} p_K \frac{\log_{10}\left[\frac{q_K}{p_K}\right]}{\log_{10}2}$

multiply the R.H.S by $\log_{10}^e$ and rearrange terms as follows.

$$\sum_{K=1}^{m} P_K \log_2\left[\frac{q_K}{P_K}\right] = P_K \sum_{K=1}^{m} \frac{\log_{10}^{e}}{\log_{10}^{2}} \cdot \frac{\log_{10}\left[\frac{q_K}{P_K}\right]}{\log_{10}^{e}}$$

$$= \sum_{K=1}^{m} P_K \log_2^{e} \quad \log_e\left[\frac{q_K}{P_K}\right]$$

Here $\log_e\left[\frac{q_K}{P_K}\right] = \ln\left[\frac{q_K}{P_K}\right]$

Hence equ becomes

$$\sum_{K=1}^{m} P_K \log_2\left[\frac{q_K}{P_K}\right] = \log_2^{e} \sum_{K=1} P_K \ln\left[\frac{q_K}{P_K}\right]$$. from equ we write

$$\ln\left[\frac{q_K}{P_K}\right] \leq \left[\frac{q_K}{P_K} - 1\right]$$

Hence above Equation becomes

$$\sum_{K=1}^{m} P_K \log_2\left[\frac{q_K}{P_K}\right] \leq \log_2^{e} \sum_{K=1}^{m} P_K \left[\frac{q_K}{P_K} - 1\right]$$

$$\sum_{K=1}^{m} P_K \log_2\left[\frac{q_K}{P_K}\right] \leq \log_2^{e} \sum_{K=1}^{m} \left[q_K - P_K\right]$$

$$\sum_{K=1}^{m} P_K \log_2\left[\frac{q_K}{P_K}\right] \leq \log_2^{e}\left[\sum_{K=1}^{m} q_K - \sum_{K=1}^{m} P_K\right]$$

Here note that $\sum_{K=1}^{m} q_K = 1$ as well as $\sum_{K=1}^{m} P_K = 1$

Hence above equation becomes $\sum_{K=1}^{m} P_K \log_2\left[\frac{q_K}{P_K}\right] \leq 0$

Now let us consider that $q_K = \frac{1}{m}$ for all k.

that is all symbols in the alphabet are Equally Likely

in the above equation becomes $\log_2 q_k - \log P_k$

$$\sum_{K=1}^{m} P_K\left[\log_2 q_K + \log_2 \frac{1}{P_K}\right] \leq 0$$

$$\sum_{K=1}^{m} P_K \log_2 q_K + \sum_{K=1}^{m} P_K \log_2 \frac{1}{P_K} \leq 0$$

$$\sum_{K=1}^{m} P_K \log_2 \frac{1}{P_K} \leq - \sum_{K=1}^{m} P_K \log_2 q_K$$

$$\sum_{K=1}^{m} P_K \log_2 \frac{1}{P_K} \leq \sum_{K=1}^{m} P_K \log_2 \frac{1}{q_K} \qquad \therefore q_K = \frac{1}{m}$$

$$\sum_{K=1}^{m} P_K \log_2 \frac{1}{P_K} \leq \sum_{K=1}^{m} P_K \log_2 m$$

$$\sum_{K=1}^{m} P_K \log_2 \frac{1}{P_K} \leq \log_2 m \sum_{K=1}^{m} P_K$$

$\sum_{K=1}^{m} P_K = 1$ then the above eqn becomes

$$\sum_{K=1}^{m} P_K \log_2 \frac{1}{P_K} \leq \log_2 m$$

The L.H.S of above Equation is Entropy $H(x)$ with arbitary probability distribution $H(x) \leq \log_2 m$ this is the proof of upperbound on entropy and max Value of entropy is $H(x)_{max} = \log_2 m$

we know that $H(x)$ is non-negative. Hence lower boun on $H(x)$ will be '0'. in it is proved that upper boun on $H(x)$ is $\log_2 m$.

$$\boxed{0 \leq H(x) \leq \log_2 m}$$

Above Equation gives lower & upper bound on $H(x)$.

Here 'm' is the no. of message Emitted by the source

Channel matrix Representation :-

it is used to define the properties of output symbols and channel capacity and this calculation is very important to know the o/p symbol probability.

there are two information sources

1. Discrete memory source

Discrete mssg 2. Discrete memoryless Source.

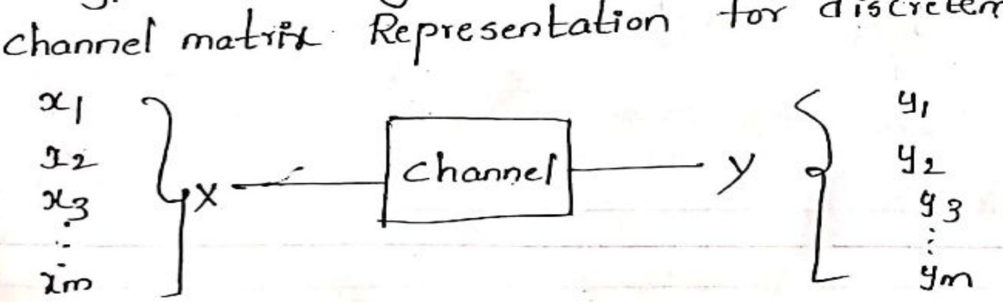
Discrete memory Source:

These source generates the o/p, which depends upon past output and present i/p to store the past output value there is a need of memory device which source is called "Discrete memory source".

Discrete memoryless source :

In this operation the source o/p is purely depends on present input of the source, that there is no need of any memory device. this source generated messages are called "Discrete messages" which is mainly prefable all types of digital communications.

the channel matrix Representation for discrete memoryless sourc



the i/p 'x' consists of $x_1, x_2, x_3 \ldots x_m$. The probability of i/p symbols are assumed $p(x_i)$ the o/p 'y' consists of y_1, y_2, y_3 y_m with the probability of $P(y_j)$, then the transition probability of channel is denoted by $P\left(\frac{y_j}{x_i}\right)$

The channel transition probability matrix is given as

$$P\left[\frac{Y}{X}\right] = \begin{bmatrix} P\left(\frac{y_1}{x_1}\right) & P\left(\frac{y_2}{x_1}\right) & P\left(\frac{y_3}{x_1}\right) & \ldots & P\left(\frac{y_m}{x_1}\right) \\ P\left(\frac{y_1}{x_2}\right) & P\left(\frac{y_2}{x_2}\right) & P\left(\frac{y_3}{x_2}\right) & \ldots & P\left(\frac{y_m}{x_2}\right) \\ P\left(\frac{y_1}{x_m}\right) & P\left(\frac{y_2}{x_m}\right) & P\left(\frac{y_3}{x_m}\right) & \ldots & P\left(\frac{y_m}{x_m}\right) \end{bmatrix}$$

the matrix $P\left(\frac{y}{x}\right)$ is also called as "channel matrix".
Here each input to the channel results in some output in each row of channel matrix is '1'.
the i/p symbol probability $p(x)$ is represented by a row matrix which is given as.

$$[P(\bar{x})] = [\, p(x_1) \quad p(x_2) \quad p(x_3) \; \ldots \; P(x_m)\,]$$

the o/p symbols probability is represented by a row matrix which is given as

$$[P(y)] = [\, p(y_1) \quad P(y_2) \quad P(y_3) \; \ldots \; P(y_m)\,]$$

<u>Conditional probability (or) Transition probability</u>:-
the i/p symbol probability $p(x)$ and the o/p symbol probability $p(y)$ are related to conditional probability.
According to probability theory $p(A)$, $P(B)$

$$P(B) = P(A)\, P\left(\frac{B}{A}\right)$$

$$P(A) = P(B)\, P\left(\frac{A}{B}\right)$$

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

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

Joint Probability $[P(x,y)]$

$$[P(x,y)] = [P(x)]_d \left[P\left(\frac{y}{x}\right)\right]$$

where
$$[P(x)]_d = \begin{bmatrix} P(x_1) & 0 & 0 & 0 \\ 0 & P(x_2) & 0 & 0 \\ 0 & 0 & P(x_3) & 0 \\ 0 & 0 & 0 & P(x_m) \end{bmatrix}$$

types of channel

there are 4 types of channel

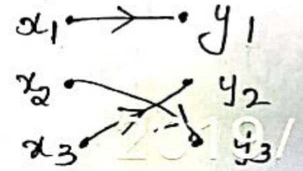
1. Lossless channel
2. determinstic channel
3. Noiseless channel
4. Binary symmetric channel

Lossless channel :-

In this channel matrix each column should have one non-Zero Element then that particular channel is called Lossless channel

$$P\left[\frac{Y}{X}\right] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

Transition diagram



Deterministic channel :-

In this channel matrix Each row should contain on non-Zero Element then that particular transition channel is called Deterministic channel.

$x_1 \to y_1$
$x_2 \to y_1$
$x_3 \to y_2$
y_3

$$P\left[\frac{y}{x}\right] = \begin{array}{c} \\ x_1 \\ x_2 \\ x_3 \end{array}\begin{array}{c} \begin{array}{ccc} y_1 & y_2 & y_3 \end{array} \\ \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \end{array}$$

Transition diagram

$x_1 \to y_1$
$x_2 \to y_1$
$x_3 \to y_2$
y_3

Noiseless channel :

this channel should be lossless and deterministic channel that means Every row and column should contain one non-Zero Elements then the transition channel is called "Noiseless channel"

$$P\left(\frac{y}{x}\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Transition diagram

$x_1 \to y_1$

$x_2 \to y_2$

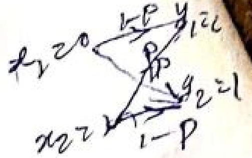
$x_3 \to y_3$

Binary symmetric channel :-

it contains two input symbols and two output symbols

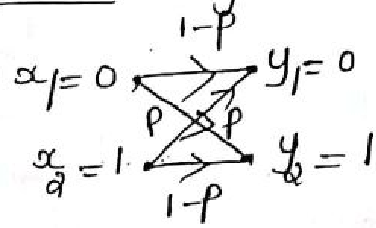
A binary symmetric channels $x_1 = 0$ $x_2 = 1$ and two outputs $y_1 = 0$ and $y_2 = 1$ this channel is symmetric.

In the second case if the i/p probability is '0' but the output probability will be '1' and if the input probability '1' and output probability is '0' this channel called "Binary symmetric channel"



$$P\left[\frac{Y}{X}\right] = \begin{array}{c c} & \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \end{array}$$

Transition diagram



Conditional Entropy:-

Conditional Entropy is also called "equivocation conditional entropy".

$H\left(\frac{X}{Y}\right)$ represents uncertainity of x, on average 'y' is known.

Ily $H\left(\frac{Y}{X}\right)$ represents uncertainity of y, on avg 'x' is known. (transmitted)

Conditional Entropy indicates the information lost across the noisy channel.

mathematically Equation $H\left(\frac{Y}{X}\right)$ and $H\left(\frac{X}{Y}\right)$ are as follows:

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

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

Marginal and conditional Entropies and Redundancy

the conditional entropy and joint entropy are used to calculate average information rate in multiple i/p source to multiple o/p source.

Here the i/p symbol (X), probability is $P(X_i)$ and the o/p symbol (Y), probability is $P(Y_j)$

the joint probability of source is $P(x_i, y_j)$

the conditional probability of source $P\left(\frac{x_i}{y_j}\right), P\left(\frac{y_j}{x_i}\right)$

if we add all joint probabilities for fixed y_j we get

$$P(y_j) = \sum_{i=1}^{m} P(x_i, y_j)$$

Ily if we add all joint probabilities for fixed x_i we get

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

Joint Entropy:

The joint entropy represents "entropy of Joint occurance of two or more Events".

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

$H(X,Y)$ represents entropy of joint occurance of 'x' and 'y'.

Marginal Entropy:-

when the entropy of individual (Elements) Event is evaluat from joint probabilities entropy of the Events

if the joint occurance of x_i and y_j Equation

$P(y_j) = \sum_{i=1}^{m} P(x_i, y_j)$ and equ $P(x_i) = \sum_{j=1}^{m} P(x_i, y_j)$

the probabilities of occurance $P(x_i)$ and $P(y_j)$ the entropies of x_i and y_j

i.e $H(x) = \sum_{i=1}^{m} P(x_i)$

Putting for $P(x_i)$ from equation

$$H(X) = \sum_{i=1}^{m} \left(\sum_{j=1}^{m} P(x_i, y_j) \right)$$

Ily $H(Y) = \sum_{j=1}^{m} P(x_i, y_j)$

putting for $P(y_j) = \sum_{i=1} P(x_i, y_j)$

$$H(Y) = \sum_{j=1}^{m} \left(\sum_{i=1}^{m} P(x_i, y_j) \right)$$

Entropies $H(X)$ and $H(Y)$ are called "Marginal Entropy"

P) Prove that $H(X,Y) = H\left(\frac{X}{Y}\right) + H(Y)$

(or)

$$H(X,Y) = H\left(\frac{Y}{X}\right) + H(X)$$

Proof: Consider equation

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

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

from probability theory we get

$$P(AB) = P\left(\frac{A}{B}\right) P(B)$$

$$P(x_i, y_j) = P\left(\frac{x_i}{y_j}\right) P(y_j)$$

putting this result in $\log_2$ term eqn

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

w.k.t $\log_2 P\left(\frac{x_i}{y_j}\right) P(y_j) = \log_2 P\left(\frac{x_i}{y_j}\right) + \log_2 P(y_j)$

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

$$H(X,Y) = \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m} p(x_i,y_j)\log_2 \frac{1}{P(\frac{x_i}{y_j})}}_{\text{first term}} - \sum_{j=1}^{m} \underbrace{\left\{\sum_{i=1}^{m} p(x_i,y_j)\right\}}_{\text{Second term}} \log_2 P(y_j)$$

Here first term represents $H(\frac{X}{Y}) = \sum_{i=1}^{m}\sum_{j=1}^{m} p(x_i,y_j)\log_2 \frac{1}{P(\frac{x_i}{y_j})}$

second term represents $P(y_j) = \sum_{i=1}^{m} P(x_i,y_j)$

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

$$H(X,Y) = H(\frac{X}{Y}) + \sum_{j=1}^{m} P(y_j)\log_2 \frac{1}{P(y_j)}$$

$$H(X,Y) = H(\frac{X}{Y}) + H(Y)$$

As per defination of Entropy second term in above Equis

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

$$\boxed{H(X,Y) = H(\frac{X}{Y}) + H(Y)}$$ Hence proved

P $H(X,Y) = H(\frac{Y}{X}) + H(X)$

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

w.r.t probability theory $P(AB) = P(\frac{B}{A}) \cdot P(A)$

$$P(x_i,y_j) = P(\frac{y_j}{x_i})\,P(x_i)$$

putting this result in $\log_2$ term Equation

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

$$= -\sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i,y_j)\log_2\left[P(\frac{y_j}{x_i})\,P(x_i)\right]$$

$$= -\sum^{m}\sum^{m} p(x_i,y_j)\log_2 P(\frac{y_j}{x_i}) - \sum_{i=1}^{m}\sum_{j=1}^{m} P(x_i,y_j)\log_2 P$$

$$H(X,Y) = \underbrace{\sum_{i=1}^{m}\sum_{j=1}^{m} p(x_i,y_j)\log_2 \frac{1}{P(\frac{y_j}{x_i})}}_{\text{first term}} - \sum_{i=1}^{m}\underbrace{\left\{\sum_{j=1}^{m} p(x_i,y_j)\right\}}_{\text{Second term}}\log_2 P$$

As per equ the first term of above equ is

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

Second term in the above eq is $p(x_i) = \sum_{j=1}^{m} p(x_i,y_j)$

$$H(X,Y) = H\left(\frac{Y}{X}\right) - \sum_{i=1}^{m} p(x_i)\log_2 p(x_i)$$

$$H(X,Y) = H\left(\frac{Y}{X}\right) + \sum_{i=1}^{m} p(x_i)\log_2 \frac{1}{p(x_i)}$$

As per defination of Entropy second term in above equ is $H(X) = \sum_{i=1}^{m} p(x_i)\log_2 \frac{1}{P(x_i)}$

$$\boxed{H(X,Y) = H\left(\frac{Y}{X}\right) + H(X)}$$

Hence proved

Channel capacity of a symmetric (or) uniform :-

for a symmetric (or) uniform channel, there are 'M' no.of output messages and 'N' no.of i/p messages

$$C = [\log_2 M - h]\cdot r \text{ bits/sec}$$

where $h = \sum_{j=1}^{m} P\left(\frac{y_j}{x_i}\right)\log_2 \frac{1}{P(\frac{y_j}{x_i})}$ for any 'i'

Channel capacity using Muraga's method:

let us consider the binary channel having channel matrix

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

let us form a new equation as follows

$$P\begin{bmatrix} \left(\frac{y_1}{x_1}\right) & \left(\frac{y_2}{x_1}\right) \\ \left(\frac{y_1}{x_2}\right) & \left(\frac{y_2}{x_2}\right) \end{bmatrix}\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} P\left(\frac{y_1}{x_1}\right)\log_2 P\left(\frac{y_1}{x_1}\right) + P\left(\frac{y_2}{x_1}\right)\log_2 P\left(\frac{y_2}{x_1}\right) \\ P\left(\frac{y_1}{x_2}\right)\log_2 P\left(\frac{y_1}{x_2}\right) + P\left(\frac{y_2}{x_2}\right)\log_2 P\left(\frac{y_2}{x_2}\right) \end{bmatrix}$$

Here Q_1 and Q_2 are two unknown Variables

Above equ is solved for Q_1 and Q_2, then the capacity is given as $C = \log_2 \{2^{Q_1} + 2^{Q_2}\}$ bits/sec

Mutual information :-

it is defined as the amount of information transferred when x_i is transmitted 'y_j' is received. it is represented by $I(x_i, y_j)$ and given as

$$I(x_i, y_j) = \log \frac{P\left(\frac{x_i}{y_j}\right)}{P(x_i)} \text{ bits}$$

Average mutual information : $[I(X, Y)]$

it is defined as the amount of information being gained by receiver if x_i transmitted and y_j received by the y_j what amount of information is gained by receiver can be calculated by avg. mutual information

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

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log \frac{P\left(\frac{x_i}{y_j}\right)}{P(x_i)}$$

Properties:

1) The mutual information of the channel is symmetric

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

2) it can be expressed interms of entropies of channel i/p&o/p conditionentropies.

3)
$$I(X;Y) = H(X) - H\left(\frac{X}{Y}\right)$$

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

Here $H\left(\frac{X}{Y}\right)$ and $H\left(\frac{Y}{X}\right)$ are conditional entropies

4) Mutual information is always positive $I(X;Y) \geq 0$

5) mutual information is related to joint entropy $H(X,Y)$ by following relation

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

Proofs

Prove mutual information of channel is symmetric $I(X;Y) = I(Y;X)$

let us consider some standard relationships from probability theory

$$P(x_i, y_j) = P\left(\frac{x_i}{y_j}\right) P(y_j)$$

$$P(AB) = P\left(\frac{A}{B}\right) P(B) \rightarrow ①$$

$$P(x_i y_j) = P\left(\frac{y_j}{x_i}\right) P(x_i)$$

$$P(AB) = P\left(\frac{B}{A}\right) P(A) \rightarrow ②$$

from eq ① and eq ②

$$P\left(\frac{x_i}{y_j}\right) P(y_j) = P\left(\frac{y_j}{x_i}\right) P(x_i)$$

$$\frac{P\left(\frac{x_i}{y_j}\right)}{P(x_i)} = \frac{P\left(\frac{y_j}{x_i}\right)}{P(y_j)} \rightarrow ③$$

The mutual information is given as

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

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

Hence we can write $I(Y;X)$ as follows

$$I(Y;X) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P\left(\frac{y_j}{x_i}\right)}{P(y_j)} \rightarrow ④$$

from eq ③ the above Eq can be written as

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

$$= I(X;Y)$$

$$\boxed{I(X;Y) = I(Y;X)}$$ Hence proved.

2) Prove that mutual information is always positive $I(X;Y) \geq 0$

Proof: $$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P\left(\frac{x_i}{y_j}\right)}{P(x_i)} \rightarrow ①$$

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

Putting above value of $P\left(\frac{x_i}{y_j}\right)$ in eq ①

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

we know that $\log_2\left(\frac{x}{y}\right)$ can be written as $-\log_2\left(\frac{y}{x}\right)$

Hence

$$I(X;Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log_2 \frac{P(x_i) P(y_j)}{P(x_i, y_j)} \rightarrow ②$$

Earlier we have desired one result given by equ it states that

$$\sum_{K=1}^{m} p_K \log_2 \left[\frac{q_K}{p_K}\right] \le 0$$

this result applied to Eq ② we can consider p_K be $p(x_i, y_j)$ then q_K be $P(x_i) P(y_j)$

Both p_K and q_K are two probability distributions on same alphabet this eq ② becomes

$$-I(X;Y) \le 0,$$

$$\boxed{I(X;Y) \ge 0}$$

③ prove that $I(X;Y) = H(X) - H\left(\frac{X}{Y}\right)$

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

proof:-

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

Here $\left(\frac{X}{Y}\right)$ is the information (or) uncertainity in 'x' after 'y' is received. In other words $H\left(\frac{X}{Y}\right)$ is the information lost in the noisy channel.

it is the avg conditional self information

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_j) \log_2 \frac{P\left(\frac{x_i}{y_j}\right)}{p(x_i)} \rightarrow ②$$

let us write the above Equ

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

from eq ① above eq can be written as

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

$P(x_i)$

the conditional entropy $H\left(\frac{X}{Y}\right)$ is $\sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P\left(\frac{x_i}{y_j}\right)}$.

$$I(X;Y) = \underbrace{\sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)}} - H\left(\frac{X}{Y}\right)$$

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

from equation represents entropy $H(X) = \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)}$

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

$I(X;Y)$ is the avg mutual information transferred per symbol across the channel it is equal to source entropy minus information lost in the noisy channel is given as

similarly consider the avg mutual information is

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

$$= \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P(y_j)} - \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{1}{P\left(\frac{y_j}{x_i}\right)}$$

the conditional entropy $H\left(\frac{Y}{X}\right)$ is given as

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

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

$$P(y_j) = \sum_{i=1}^{n} P(x_i, y_j)$$

$$= \sum_{j=1}^{m} P(y_j) \log_2 \frac{1}{P(y_j)} - H\left(\frac{Y}{X}\right)$$

from equ represents entropy $H(Y) = \sum_{j=1}^{m} P(y_j) \log_2 \frac{1}{P(y_j)}$

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

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

w.k.t $\quad H(X;Y) = H\left(\frac{X}{Y}\right) + H(Y)$

$$H\left(\frac{X}{Y}\right) = H(X,Y) - H(Y)$$

w.k.t

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

$$I(X;Y) = H(X) - \left[H(X,Y) - H(Y)\right]$$

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

$$\boxed{I(X;Y) = H(X) + H(Y) - H(X,Y)}$$

Hence proved