Information theory and coding: Introduction

Information theory is theoretical part of communication developed by American mathematician **Shannon**.

It deals with mathematical modelling and analysis of a communication system rather than dealing with physical sources (camera, keyboard, microphone etc.) and physical channels (wires, cables, fibre, satellite, radio etc).

Note: Lot of mathematics is involved in manufacturing of things we use in daily-life… like, manufacturing of pen, paper, flash memory, mobiles, antenna design, garment manufacturing ete etc etc

Information theory addresses and answers below fundamental questions of communication theory:

- Data compression
- Data transmission
- Data storage
- Error detection and correction codes
- Ultimate Data rate over a noisy channel (for reliable communication)

Every morning one reads newspaper to receive information. A message is said to convey information, if two key elements are present in it.

- Change in knowledge or meaning
- Uncertainty (unpredictability)

Amount of information contained in a message is inter-related to its **probability of occurrence**. Information is always about something (occurrence of an event etc.). It may be a true or lie.

Consider the following messages:

1. The sun rises in the east
2. Gopal will get Nobel prize in physics
3. Scattered rain
4. Cyclone storm
5. Son is born to his wife who is a mother of two daughters
6. COVID-19 pandemic disease

By intuition, understand that the above 6 messages carry different information.

S.No.	Message	Remark
1	The sun rises in the east	There is little information in this statement because everybody can expect this event.
2	Gopal will get Nobel prize in physics	This message will contain lot of information. Because it is not frequent. It is a rare event, so lot of information.
3	Scattered rain	Rare event, so lot of information
4	Cyclone storm	Rare event, so lot of information
5	Son is born to his wife who is a mother of two daughters	Rare event, so lot of information
6	COVID-19 pandemic disease	Rare event, so lot of information

- Message with lower probability contains higher information content and vice-versa
- Frequent events have less information
- Rare events have more information
- Note that greater the uncertainty (unpredictable), higher the value of information
- Certain events (sure events) carry least information
- Uncertain events (unsure events) carry least information

The messages in the table above have different probabilities of occurrence and hence contain different value of information.

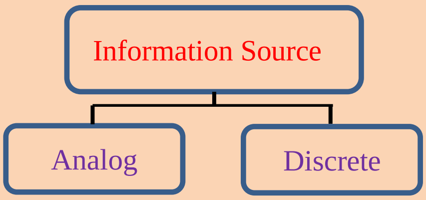
Information theory talks about:

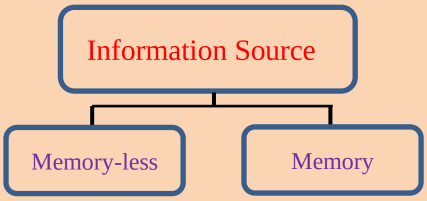
- How to measure the amount of information?
- How to measure the correctness of information?
- What to do if information gets corrupted by errors?
- How much memory does it require to store information?

DMS (Discrete Memory-less Source)

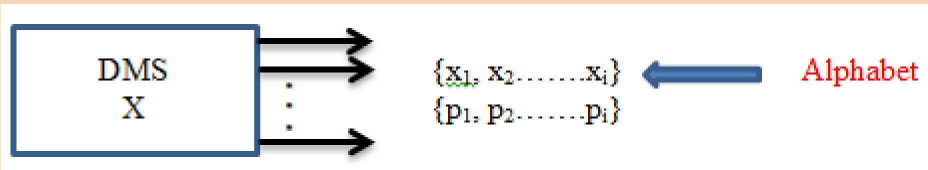
Information sources can be classified as either memory or memory-less. A memory-less source is one for which each symbol produced is independent of previous symbols. A memory source is one for which a current symbol depends on the previous symbols.

An information source is an object that produces an event. A practical source in a communication system is a device that produces messages. It can be either alalog or digital (discrete). Analog sources can be transformed into discrete sources through the use of sampling and quantization techniques.





A discrete information source is a source that has only a finite set of symbols as possible outputs.



Let X having alphabets $\{x_1, x_2, \ldots x_m\}$. Note that set of source symbols is called source alphabet.

A DMS is described by the list of symbols, probability assignment to these symbols and the specification of rate of generating these symbols by source. A discrete information source consists of a discrete set of letters or symbols. In general, any message emitted by the source consists of a string or sequence of symbols.

Example:

A source is one, which generates sequence of symbols called '**alphabet**'. Below table shows various information sources and the type of messages.

Source	Message	Alphabet
Camera	Picture	RGB values
Type writer	Text	{a, b,….z}
Microphone	Voice	{amplitudes}

Information Source

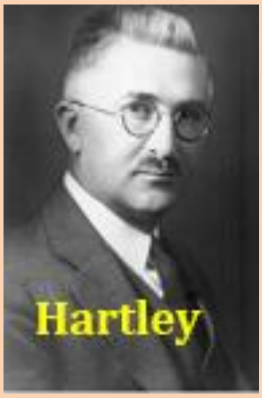
Information source generates sequence of symbols called '**alphabet**'. Below table shows various information sources and the type of messages.

Source	Message	Alphabet
Camera	Picture 	RGB values = Red, Green, Blue each ranging from 0 to 255
Computer Keyboard	Text 	{A, B, …. C, a, b,….z, 0,1,2….9, ?, $, %, #, etc}
Microphone	Voice 	{amplitudes} = {0.5v, 1.3v, 4.7v, ……. etc}

Measure of information

The amount of information in a message depends only upon the uncertainty of the event. Amount of information received from the knowledge of occurrence of an event may be related to the probability of occurrence of that event. Few messages received from different sources will contain more information than others. Information should be proportional to the uncertainty (doubtfulness) of an outcome.

According to **Hartley**, information in a message is a logarithmic function. Let $\mathbf{m_1}$ is any message, then information content of this message is given as:



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

$I(m_1)$ = Information in message m_1
$p(x_1)$ = probability of occurrence of message m_1

- Unit of information = **bits** (if base = 2)
- Unit of information = **Hartley** or **decit** (if base = 10)
- Unit of information = **nat** (natural unit) (if base = e)

- Information must be inversely proportional to p

Message: Sun will rise in the east
Here probability (p) = 1, so information (I) = 0
Note that maximum value of probability = 1

- Information must be a non-negative quantity since each message contains some information.

Let source, X having alphabets--- $\{x_1, x_2, \ldots x_m\}$

Information of symbol $x_1 = I(x_1)$
Information of symbol $x_2 = I(x_2)$
Information of symbol $x_3 = I(x_3)$
Information of symbol $x_i = I(x_i)$

$$I(x_1) = \log_2 \frac{1}{P(x_1)}$$
$$I(x_2) = \log_2 \frac{1}{P(x_2)}$$
$$I(x_3) = \log_2 \frac{1}{P(x_3)}$$

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

$p(x_1)$ = probability of occurrence of symbol x_1
$p(x_2)$ = probability of occurrence of symbol x_2
$p(x_3)$ = probability of occurrence of symbol x_3
$p(x_i)$ = probability of occurrence of symbol x_i

Properties of $I(x_i)$

i. $I(x_1) = 0$ if $p(x_1) = 1$
ii. $I(x_1) \geq 0$
iii. $I(x_1, x_2) = I(x_1) + I(x_1)$ if x_1 and x_1 are independent
iv. $I(x_1) > I(x_2)$ if $p(x_1) < p(x_2)$

Note: we can easily generalize these properties to x_i

Q: A source generates one of the 5 possible messages during each message interval. The probabilities of these messages are:
$p_1 = \frac{1}{2}, p_2 = \frac{1}{16}, p_3 = \frac{1}{8}, p_4 = \frac{1}{4}, p_5 = \frac{1}{16}$. Find the information content of each message.

$I(m_1) = \log_2\left(\frac{1}{p_1}\right) = \log_2 2 = 1\ bit$ ⟶ Information in message m_1

$I(m_2) = \log_2\left(\frac{1}{p_2}\right) = \log_2 16 = 4\ bits$ ⟶ Information in message m_2

$I(m_3) = \log_2\left(\frac{1}{p_3}\right) = \log_2 8 = 3\ bit\ s$ ⟶ Information in message m_3

$I(m_4) = \log_2\left(\frac{1}{p_4}\right) = \log_2 4 = 2\ bit\ s$ ⟶ Information in message m_4

$I(m_5) = \log_2\left(\frac{1}{p_5}\right) = \log_2 16 = 4\ bit\ s$ ⟶ Information in message m_5

Q: In a binary PCM if 0 occurs with probability ¼ and 1 occurs with probability ¾, then calculate the amount of information carried by each bit.

$$I(bit\ 0) = \log_2 \frac{1}{P(x_1)} = \log_2 \frac{1}{\frac{1}{4}} = \log_2 4 = \log_2 2^2 = 2\ \log_2 2 = 2 * 1 = 2\ bits$$
$$I(bit\ 1) = \log_2 \frac{1}{P(x_2)} = \log_2 \frac{1}{\frac{3}{4}} = \log_2 \frac{4}{3} = \log_2 1.33 = \frac{\log_{10} 1.33}{\log_{10} 2} = \frac{0.125}{0.301} = 0.415\ bits$$

- Bit 0 has probability ¼ and it has 2 bits of information
- Bit 1 has probability ¾ and it has 0.415 bits of information

Q: If $I(x_1)$ is the information carried by symbol x_1 and $I(x_2)$ is the information carried by message x_2 then prove that the amount of information carried compositely due to x_1 and x_2 is $I(x_1, x_2) = I(x_1) + I(x_2)$

Symbol = message

$$I(x_1) = \log_2 \frac{1}{p(x_1)}$$
$$I(x_2) = \log_2 \frac{1}{p(x_2)}$$

$p(x_1)$ = probability of occurrence of message (or symbol x_1)
$p(x_2)$ = probability of occurrence of message (or symbol x_2)

Note that messages x_1 and x_2 are independent, therefore
Joint probability $p(x_1, x_2) = p(x_1) * p(x_2)$
Information carried compositely due to x_1 and x_2 is

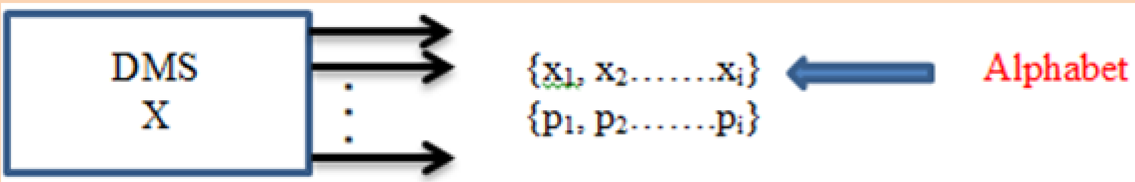
$$I(x_1, x_2) = \log_2 \frac{1}{P(x_1,x_2)} = \log_2 \frac{1}{p(x_1).p(x_2)} = \log_2 \left[\frac{1}{p(x_1)} . \frac{1}{p(x_2)}\right] = \log_2 \frac{1}{P(x_1)} + \log_2 \frac{1}{P(x_2)} = I(x_1) + I(x_2)$$

$[since \ \log_2 ab = \log_2 a + \log_2 b]$

Thus proved.

Information contained in independent outcomes should add. So, Information given by two independent messages is the sum of the information contained in individual messages.

ENTROPY



Entropy H(X) = Average information of a source X

- Here average information rather than information of a single symbol.
- This average information is called entropy.
- Also note that Entropy defines ultimate data compression of a source. Data compression provides a practical means for the efficient storage and transmission of data (text, audio, video etc.)
- Larger entropy represents larger average information.

Entropy is defined as H(X) and is given by:

$$H(X) = \sum_1^l p(x_i) \, I(x_i) = \sum_1^l p(x_i) \log_2 \frac{1}{p(x_i)}$$

Where X = source and l = size of alphabet

Q: A sample space of 5 messages with probabilities are given as: P(s) = {0.25, 0.25, 0.25, 0.125, 0.125}. Find Entropy of the source.

$$H(X) = \sum_1^5 P(x_i) \log_2 \frac{1}{P(x_i)} = P(x_1) \log_2 \frac{1}{P(x_1)} + P(x_2) \log_2 \frac{1}{P(x_2)} + P(x_3) \log_2 \frac{1}{P(x_3)} + P(x_4) \log_2 \frac{1}{P(x_4)} + P(x_5) \log_2 \frac{1}{P(x_5)}$$
$$= 0.25\log_2 \frac{1}{0.25} + 0.25\log_2 \frac{1}{0.25} + 0.25\log_2 \frac{1}{0.25} + 0.125\log_2 \frac{1}{0.125} + 0.125\log_2 \frac{1}{0.125}$$

$= 3 \times 0.25\log_2 \frac{1}{0.25} + 2 \times 0.125\log_2 \frac{1}{0.125}$
$= 3 \times 0.25 \times \log_2 4 + 2 \times 0.125 \times \log_2 8$
$= 3 \times 0.25 \times 2 + 2 \times 0.125 \times 4$
$= 2.25$ bits/ symbol

Q: An unfair dice with four faces and p(1) = 1/2, p(2) = 1/4, p(3) = p(4) = 1/8. Find entropy H (answer= 7/4).

Q: A sample space p(s) = {0.5, 0.125, 0.125, 0.125, 0.125}. Find Entropy of the source.

Q: A sample space p(s) = {0.75, 0.0625, 0.0625, 0.0625, 0.0626}. Find Entropy of the source.

Q: Find the entropy of a binary source

Q: A Discrete Memory-less Source (DMS) X has four symbols x_1, x_2, x_3, x_4 with probabilities $p(x_1) = 0.4$, $p(x_2) = 0.3$, $p(x_3) = 0.2$ and $p(x_4) = 0.1$. Calculate H(X)

$$H(X) = \sum_{1}^{4} P(x_i)\log_2 \frac{1}{P(x_i)} = P(x_1)\log_2 \frac{1}{P(x_1)} + P(x_2)\log_2 \frac{1}{P(x_2)} + P(x_3)\log_2 \frac{1}{P(x_3)} + P(x_4)\log_2 \frac{1}{P(x_4)}$$

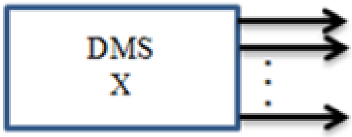
$$= 0.4\log_2 \frac{1}{0.4} + 0.3\log_2 \frac{1}{0.3} + 0.2\log_2 \frac{1}{0.2} + 0.1\log_2 \frac{1}{0.1}$$

$$= 0.4\log_2 \frac{1}{0.4} + 0.3\log_2 \frac{1}{0.3} + 0.2\log_2 \frac{1}{0.2} + 0.1\log_2 \frac{1}{0.1} = 1.85\ bits/symbol$$

Q: A source produces one of the 4 possible symbols during each interval having probabilities: $p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = p_4 = \frac{1}{8}$. Obtain information content of each of these symbols.

Q: A source emits independent sequences of symbols from a source alphabet containing five symbols with probabilities 0.4, 0.2, 0.2, 0.1 and 0.1. Compute the entropy of the source.

Information rate (R)



R = r H(X)
Where R = information rate (bps = bits per sec)
r = baud rate = symbols/sec
H(X) = entropy or average information (bits/symbol)

Q: An analog signal is band limited to 1000 Hz and sampled at Nyquist rate. The samples are quantized into 4 levels. Each level represents 1 symbol. Probabilities of the symbols are: $p(x_1) = p(x_4) = 1/8$ and $p(x_2) = p(x_3) = 3/8$. Obtain information rate of the source.

$$Entropy, H(X) = \sum_{1}^{4} P(x_i)\log_2\frac{1}{P(x_i)} = P(x_1)\log_2\frac{1}{P(x_1)} + P(x_2)\log_2\frac{1}{P(x_2)} + P(x_3)\log_2\frac{1}{P(x_3)} + P(x_4)\log_2\frac{1}{P(x_4)}$$

$$= \frac{1}{8}\log_2 8 + \frac{3}{8}\log_2\frac{8}{3} + \frac{1}{8}\log_2 8 + \frac{3}{8}\log_2\frac{8}{3} = 1.5\ bits/symbol$$

$f_m = 1000$ Hz (given)

Nyquist rate $f_s = 2f_m = 2000$ Hz = 2000 symbols/sec. This is nothing but number of symbols generated by source = **r**

R = r H(X)
Where R = information rate (bps = bits per sec)
r = baud rate = symbols/sec
H(X) = entropy or average information (bits/symbol)

We know that **R = r H(X)**

$$= 2000\ \frac{symbols}{sec} \times 1.5\ \frac{bits}{symbol} = 3000\ \frac{bits}{sec} = 3000\ bps$$

Mutual information

Mutual information about two messages measures the amount of information that one conveys about the other. It is defined as:

$$I(X;Y) = \sum_{x,y} p(x,y)\log_2\frac{p(x,y)}{p(x).p(y)}$$

X = information source
Y = another information source
x= message generated by source X
y = message generated by source Y
p(x) = probability of occurrence of x
p(y) = probability of occurrence of y

Mutual information means message x says about y and vice versa.

If X and Y are independent sources, then joint probability p(x, y) = p(x) . p(y)

Now mutual information between x and y is given as:

$$I(X;Y) = \sum_{x,y} p(x,y)\log_2\frac{p(x,y)}{p(x).p(y)} = \sum_{x,y} p(x,y)\log_2\frac{p(x).p(y)}{p(x).p(y)} = \sum_{x,y} p(x,y)\log_2 1 = 0$$

Digital communication system (Shannon model of communication)

- The basic goal of a communication system is to transmit some information from source to the destination.
- Message = Information
- Information consists of letters, digits, symbols, sequence of letters, digits, symbols etc.
- Information theory gives an idea about what can be achieved or what cannot be achieved -in a communication system.

Shannon gave ideas on:

- Signal processing operations such as compressing data
- Storing and communicating data on a noisy channel (channel with errors)

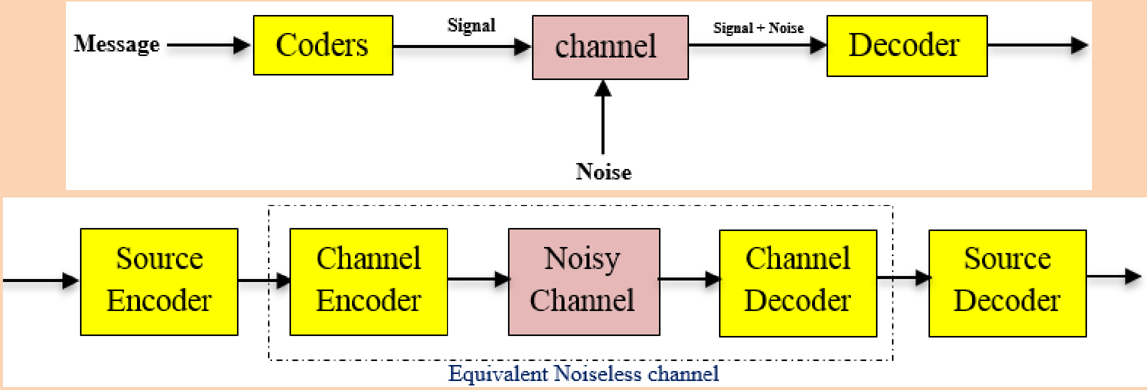


Fig: Digital communication system (Shannon model of communication)

A practical source in a communication system is a device that produces messages. It can be either alalog or digital (discrete). So, there is a source and there is a destination. Messages are transferred from one point to another.

We deal mainly with discrete sources since analog sources can be transformed to discrete sources through the use of sampling and quantization techniques.

Coding theory

Coding is the most important application of information theory. DMS output is converted into binary codes. Device that performs this conversion is called the source encoder.

- Source encoder takes care of data compression
- Channel encoder do reliable data transmission
- Decoding is exactly the reverse process of encoding
- Note that communication channel is at the heart of the communication problem. Channel noise corrupts the transmitted signal, causing unavoidable decoding errors at the receiver.

Scratch on the CD is also an example of noise. Noise is unwanted signal.

Source coding = Entropy Coding.

- Job: Data compression. Compression algorithms are of great importance when processing and transmitting audio, images and video
- Note that during compression no significant information is lost
- Entropy defines the minimum amount of necessary information
- Source coding is done at transmitter

Various source coding techniques are Huffman coding, Shannon-Fano, Lempel Ziv coding, PCM, DPCM, DM and adaptive DM (ADPCM).

Channel coding = Error-Control Coding.

Encoder adds additional bits to actual information in order to detect and correct transmission errors.
Various channel coding techniques are: Hamming codes, cyclic codes, BCH codes, block coding, convolutional coding, turbo coding etc.

Assume that 1-bit of information is transmitted from source to destination. If bit – 0 is transmitted, bit – 0 must be received.

Transmitter	Receiver	Remark
0	0	Good
0	1	Error
1	0	Error
1	1	Good

- Error correcting code adds just the right kind of redundancy as possible (i.e., error correction) needed to transmit the data efficiently across noisy channel.

Source coding: Also known as entropy coding. Here the information generated by the source is compressed. Note that during compression no significant information is lost. Entropy defines the minimum amount of necessary information. Source coding is done at transmitter.

Source encoder transforms information from source into different information bits, while implementing data compression. Various source coding techniques are Huffman coding, Lempel Ziv coding, PCM, DPCM, DM and adaptive DM (ADPCM).

Channel coding: Also known as error-control coding. Encoder adds additional bits to actual information in order to detect and correct transmission errors. Channel coding is done at receiver. Various channel coding techniques are: block coding, convolutional coding, turbo coding etc.

Q: In a binary PCM if 0 occurs with probability ¼ and 1 occurs with probability ¾, then calculate the amount of information carried by each bit.

$$I(bit\ 0) = \log_2 \frac{1}{P(x_1)} = \log_2 \frac{1}{\frac{1}{4}} = \log_2 4 = \log_2 2^2 \quad = 2\ \log_2 2 = 2 * 1 = 2\ bits$$

$$I(bit\ 1) = \log_2 \frac{1}{P(x_2)} = \log_2 \frac{1}{\frac{3}{4}} = \log_2 \frac{4}{3} = \log_2 1.33 \quad = \frac{\log_{10} 1.33}{\log_{10} 2} = \frac{0.125}{0.301} = 0.415\ bits$$

- Bit 0 has probability ¼ and it has 2 bits of information
- Bit 1 has probability ¾ and it has 0.415 bits of information

Q: A Discrete Memory-less Source (DMS) X has four symbols x_1, x_2, x_3, x_4 with probabilities $p(x_1) = 0.4$, $p(x_2) = 0.3$, $p(x_3) = 0.2$ and $p(x_4) = 0.1$. Calculate H(X)

$$H(X) = \sum_{1}^{4} P(x_i)\log_2\frac{1}{P(x_i)} = P(x_1)\log_2\frac{1}{P(x_1)} + P(x_2)\log_2\frac{1}{P(x_2)} + P(x_3)\log_2\frac{1}{P(x_3)} + P(x_4)\log_2\frac{1}{P(x_4)}$$

$$= 0.4\log_2\frac{1}{0.4} + 0.3\log_2\frac{1}{0.3} + 0.2\log_2\frac{1}{0.2} + 0.1\log_2\frac{1}{0.1}$$

$$= 0.4\log_2\frac{1}{0.4} + 0.3\log_2\frac{1}{0.3} + 0.2\log_2\frac{1}{0.2} + 0.1\log_2\frac{1}{0.1} \quad = 1.85\ bits/symbol$$

Q: A source produces one of the 4 possible symbols during each interval having probabilities: $p_1 = \frac{1}{2}, p_2 = \frac{1}{4}, p_3 = p_4 = \frac{1}{8}$. Obtain information content of each of these symbols.

Q: A source emits independent sequences of symbols from a source alphabet containing five symbols with probabilities 0.4, 0.2, 0.2, 0.1 and 0.1. Compute the entropy of the source.

Solution: Source alphabet = (s1, s2, s3, s4, s5)
Probs. of symbols = p1, p2, p3, p4, p5 = 0.4, 0.2, 0.2, 0.1, 0.1

(i) Entropy of the source = H = $-\sum p_i \log p_i$ bits / symbol

Substituting we get,
$H = -[p_1 \log p_1 + p_2 \log p_2 + p_3 \log p_3 + p_4 \log p_4 + p_5 \log p_5]$
$= -[0.4 \log 0.4 + 0.2 \log 0.2 + 0.2 \log 0.2 + 0.1 \log 0.1 + 0.1 \log 0.1]$
H = 2.12 bits/symbol

Q: An analog signal is band limited to 1000 Hz and sampled at Nyquist rate. The samples are quantized into 4 levels. Each level represents 1 symbol. Thus, there are 4 levels (symbols) are:
$p(x_1) = p(x_4) = 1/8$ and $p(x_2) = p(x_3) = 3/8$. Obtain information rate of the source.

$$Entropy, H(X) = \sum_{1}^{4} P(x_i)\log_2\frac{1}{P(x_i)} = P(x_1)\log_2\frac{1}{P(x_1)} + P(x_2)\log_2\frac{1}{P(x_2)} + P(x_3)\log_2\frac{1}{P(x_3)} + P(x_4)\log_2\frac{1}{P(x_4)}$$

$$= \frac{1}{8}\log_2 8 + \frac{3}{8}\log_2\frac{8}{3} + \frac{1}{8}\log_2 8 + \frac{3}{8}\log_2\frac{8}{3} = 1.5\ bits/symbol$$

Information rate (R)

R = r H(X)
Where R = information rate (bps) = rate at which symbols are generated.
r = Number of symbols generated by source (symbols/sec)
H(X) = entropy or average information (bits/symbol)

f_m = 1000 Hz (given)
Nyquist rate $f_s = 2f_m$ = 2000 Hz = 2000 symbols/sec. This is nothing but number of symbols generated by source, which is **r**

We know that **R = r H(X)**
$= 2000\ \frac{symbols}{sec}$ x $1.5\ \frac{bits}{symbol} = 3000\ \frac{bits}{sec} = 3000\ bps$

Shannon Channel Capacity Theorem

The Shannon channel capacity theorem defines the theoretical maximum bit rate (number of bits per second) for a noisy channel

Capacity $C = B\ x\ log_2(1 + \frac{S}{N})$

Every communication channel has a speed limit, measured in bps. This famous Shannon limit and the formula for capacity of communication channel is given as:

Where, C = Channel capacity (bps)
B = Bandwidth of signal
S = Signal power (Watt)
N = Noise power (Watt)

$$SNR = \frac{S}{N} = \frac{Signal\ Power}{Noise\ Power}\ (no\ unit)$$

Generally, SNR is expressed in decibel (dB) units. [dB = 10 $\log_{10}$ (S/N)]

- Channel capacity is the ultimate transmission rate of a communication system.
- Channel capacity is defined as the ability of a channel to convey information.

Bad news:

It is mathematically impossible to get error free communication above the limit.

- No matter how sophisticated error correction scheme we use
- No matter how much we compress the data

We can't make the channel go faster than the limit without losing some information

Good news:

Below the Shannon limit, it is possible to transmit the information with zero error. Shannon mathematically proved that use of encoding techniques allow us to reach maximum limit of channel capacity without any errors regardless of amount of noise.

Capacity $C = B\ x\ log_2(1 + \frac{S}{N})$

Q: Calculate the bit rate for a noisy channel with SNR 300 and bandwidth of 3000Hz

Solution: The bit rate for a noisy channel according to Shannon theorem can be calculated as follows:

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

$$\log_a x = \frac{\log_{10} x}{\log_{10} a}$$

$= 3000 * log_2(1 + 300)$
$= 3000\ log_2(301)$
= 3000 x 8.23
= 24,690 bps

Q: Calculate the capacity of a Gaussian channel with a BW B= 1 MHz and SNR of 30 dB.

Every communication channel has a speed limit, measured in bps. This famous Shannon limit and the formula for capacity of communication channel is given as:

Bad news:

It is mathematically impossible to get error free communication above the limit.

- No matter how sophisticated error correction scheme we use
- No matter how much we compress the data

We can't make the channel go faster than the limit without losing some information

Good news:

Below the Shannon limit, it is possible to transmit the information with zero error. Shannon mathematically proved that use of encoding techniques allow us to reach maximum limit of channel capacity without any errors regardless of amount of noise.