UNIT_IV

INFORMATION THEORY

Information theory is a branch of probability theory which may be applied to the study of the communication systems.

In communication system, information theory deals with mathematical modelling and analysis of a communication system rather than with physical sources and physical channels.

Information

Few messages produced by an information source contain more information than others.

Let us consider an example, Pune University declares large number of results. Suppose that you are received the following messages –

1. Result is declared today
2. Result is declared today
3. Result is declared today
4. Electronics and Communication branch results are declared today.

First 3 messages are very common. Since large number of examinations are conducted, every day results are declared by university. Therefore such message give very less information to you.

But, when you receive 4^{th} message, you will forget everything and go to collect the result. Thus, the amount of information received by 4^{th} message is very large.

This shows that probability of occurrence of 4^{th} message is very small but amount of information received is great.

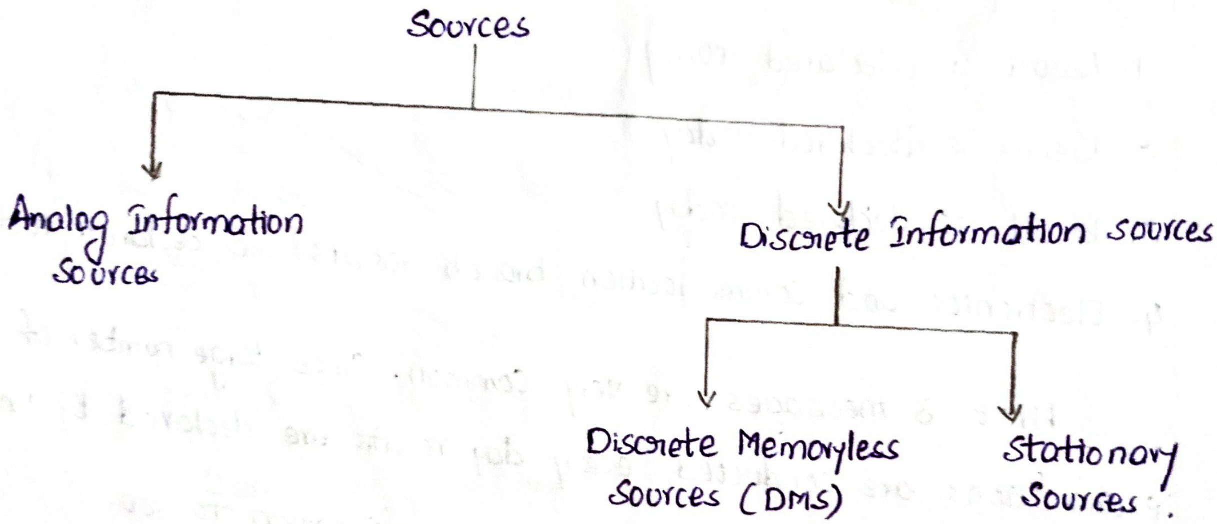
→ Information Sources:

A practical source in a communication system is a device which produces messages and it can be either analog or discrete.

Analog Sources: Analog sources can be transformed to discrete sources through the use of sampling and quantization techniques.
Eg: Radio and TV Broadcasting

Discrete Information Source: It is a source which has only a finite set of symbols as possible outputs. The set of source symbols is called the source alphabet and the elements of the set are called symbols or letters. Eg: Digital computer or a storage device such as magnetic tape or opticaldisk.

→ Classification of Information sources



(i) Discrete Memoryless sources (DMS)

If the current output letter (or symbol) is statistically independent from all the past and future outputs, then the source is called as "Discrete Memoryless source" (DMS)

Eg: If a binary source generates a random sequence 1011011.... such that each binary digit is statistically independent of all the other digits then the source is called as a "Binary Memoryless source".

(ii) Stationary Sources

If the output of a discrete source is statistically dependent on its past or future outputs, then it is called as the discrete stationary source.

Eg: Generating the English text.

Mathematical Model for the information sources:

Consider the source which emits the discrete symbols randomly from the set of fixed alphabet that is

$$X = [x_0, x_1, x_2 \ldots\ldots x_m]$$

The various symbols in 'x' have probabilities of $P_0, P_1, P_2 \ldots P_{k-1}$ which can be written as

$$P[X = x_k] = P_k \quad \text{where } k = 0, 1, 2, \ldots k-1$$

The set of probabilities is given as

$$\sum_{k=0}^{k-1} P_k = 1$$

The probability of x_k is P_k

If $P_k = 0$ then symbol is impossible

If $P_k = 1$ then symbol is possible.

Information Content of a discrete Memoryless source (DMS)

The amount of information contained in an event is closely related to its "uncertainty".

The words "uncertainty, surprise and information" are all related to each other.

Before an event occurs, there is an uncertainty, when the event occurs there is an amount of surprise and after the occurrence of an event there is a gain of information.

Messages containing high probability of occurrence convey relatively little information.

We know that if an event is certain (that is, the event occurs with probability 1), it conveys zero information.

Information Content of a symbol (i.e Logarithmic Measure of Information)

Consider a Discrete Memoryless source (DMS) denoted by 'X' and having alphabet $\{x_1, x_2 \ldots x_m\}$.

The information content of a symbol x_i denoted by $I(x_i)$ is defined by

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

where $P(x_i)$ is the probability of occurrence of symbol x_i.

This is the defintion of information.

$$I(x_i) = \log_2 \frac{1}{P_{x_i}} = \frac{\log_{10}(1/P(x_i))}{\log_{10} 2}$$

Properties of Information $I(x_i)$

The information content of a symbol x_i, denoted by $I(x_i)$ satisfies the following properties:

1. The information content $I(x_i)$ must be a non-negative quantity since each message contains some information. In the worst case, $I(x_i)$ can be equal to zero.

   i.e $I(x_i) \geq 0$

2. The amount of information of a symbol x_i is denoted by $I(x_i)$, must be inversely related to $P(x_i)$.

$I(x_i)$ must approach '0' as $P(x_i)$ approaches unity.

For example, let us consider the message "sun will rise in the east". This message does not contain any information since the sun will rise in the east with probability '1'.

$$I(x_i) = 0 \text{ for } P(x_i) = 1$$

3. The information content of a message having higher probability of occurrence is less than the information content of a message having lower probability.

$$I(x_i) > I(x_j) \quad \text{if}$$

$$P(x_i) < P(x_j)$$

4. If I_1 is the information carried by message x_1 and I_2 is the information carried by x_2, the amount of information carried componently if x_1 and x_2 are independent then

$$I(x_1, x_2) = I(x_1) + I(x_2)$$

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

→ Calculate the amount of information if it is given that $P(x_i) = \frac{1}{4}$

Sol. We know that amount of information $I(x_i)$ of a discrete symbol x_i is given by

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

$$= \frac{\log_{10}(1/P_{x_i})}{\log_{10} 2} = \frac{\log_{10} 4}{\log_{10} 2} = \frac{\log_{10} 2^2}{\log_{10} 2} = 2 \text{ bits}$$

→ A source produces one of four possible symbols during each interval having probabilities $P(x_1) = \frac{1}{2}$, $P(x_2) = \frac{1}{4}$, $P(x_3) = P(x_4) = \frac{1}{8}$. Obtain the information content of each of these symbols.

Sol. Given $P(x_1) = \frac{1}{2}$, $P(x_2) = \frac{1}{4}$, $P(x_3) = \frac{1}{8}$, $P(x_4) = \frac{1}{8}$

We know that the information content $I(x_i)$ of a symbol x_i is given by

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

We can write

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

$$= \frac{\log_{10}\left(\frac{1}{P(x_1)}\right)}{\log_{10} 2}$$

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

$$= 1 \text{ bit}$$

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

$$= \frac{\log_{10} 4}{\log_{10} 2} = 2 \text{ bits}.$$

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

$$= \frac{\log_{10} 8}{\log_{10} 2} = \frac{\log_{10} 2^3}{\log_{10} 2}$$

$$= 3 \text{ bits}.$$

$$I_4 = \log_2 (1/P(x_i))$$

$$= \frac{\log_{10} 8}{\log_{10} 2} = 3 \text{ bits}.$$

→ Calculate the amount of information if binary digits (binits) occur with equal likelihood in a binary PCM system.

Sol. We know that in binary PCM, there are only two binary levels i.e 1 or 0. Since these two binary levels occur with equal likelihood, their probability of occurrence will be

$$P(x_1) \text{ for '0' level} = P(x_2) \text{ for '1' level} = \frac{1}{2}$$

Therefore, the amount of information content will be given as

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

$$= \frac{\log_{10} 2}{\log_{10} 2} \quad ; \quad = \frac{\log_{10} 2}{\log_{10} 2}$$

$$= 1 \text{ bit} \quad ; \quad = 1 \text{ bit}$$

Hence, the correct identification of binary digit (binit) in binary PCM carries 1 bit of information.

→ In a binary PCM if '0' occur with probability $\frac{1}{4}$ and '1' occur with probability equal to $\frac{3}{4}$, then calculate the amount of information carried by each binit.

Sol Here, given that binit '0' has $P(x_1) = \frac{1}{4}$ &

binit '1' has $P(x_2) = \frac{3}{4}$

The amount of information is given as

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

$$= \frac{\log_{10} 4}{\log_{10} 2} \qquad = \frac{\log_{10}\left(\frac{4}{3}\right)}{\log_{10} 2}$$

$$= 2 \text{ bits.} \qquad = 0.415 \text{ bits}$$

Hence, we observed that binit '0' has probability $\frac{1}{4}$ and it carries 2 bits of information.

Whereas binit '1' has probability $\frac{3}{4}$ and it carries 0.415 bits of information.

Thus, it reveals that the fact, if probability of occurrence is less, then the information carried is more and vice-versa.

→ Prove the statement stated as under

" If a receiver knows the message being transmitted, the amount of information carried will be zero"

sol. Here it is stated that receiver knows the message. This means that only one message is transmitted. The probability of occurrence of this message will be $P(x_i) = 1$

The amount of information carried by this type of message will be

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

$$= \frac{\log_{10} 1}{\log_{10} 2} \quad ; \quad [\because P(x_i) = 1]$$

$$= 0 \text{ bits}$$

This proves the statement that if the receiver knows message, the amount of information carried will be zero.

→ If $I(x_1)$ is the information carried by symbol x_1 and $I(x_2)$ is the information carried by symbol x_2 then prove that the amount of information carried compositely due to x_1 and x_2 is

$$I(x_1, x_2) = I(x_1) + I(x_2)$$

sol The individual amounts carried by symbols x_1 and x_2 are,

$$I(x_1) = \log_2 \frac{1}{P(x_1)} \text{ — } (1) \quad ; \quad I(x_2) = \log_2 \frac{1}{P(x_2)} \text{ — } (2)$$

where $P(x_1)$ is probability of symbol x_1

$P(x_2)$ is probability of symbol x_2

Since messages x_1 and x_2 are independent, the probability of composite message is $P(x_1) P(x_2)$.

Therefore, information carried compositely due to symbols x_1 and x_2 will be

$$I(x_1, x_2) = \log_2 \frac{1}{[P(x_1)\,P(x_2)]}$$

$$= \log_2 \left[\left(\frac{1}{P(x_1)}\right)\left(\frac{1}{P(x_2)}\right)\right]$$

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

$$I(x_1, x_2) = I(x_1) + I(x_2) \quad [\because \text{From eq } ① \,\&\, ②]$$

→ If there are M equally likely and independent symbols, then prove that amount of information carried by each symbol will be

$I(x_i) = N$ bits. where $M = 2^N$ and N is an integer

Sol) Given that all the M symbols are equally likely and independent, Therefore the probability of occurrence of each symbol must be $\frac{1}{M}$.

The amount of information $I(x_i)$ of a discrete symbol x_i is given by

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

The probability of each message is $P(x_i) = \frac{1}{M}$ — ②

Substitute ② in ①, We get

$$I(x_i) = \log_2 M$$

$$= \log_2 2^N \quad [\text{Given } M = 2^N]$$

$$= N \cdot \frac{\log_{10} 2}{\log_{10} 2}$$

$$= N \text{ bits}$$

Hence proved

Entropy: (Average Information)

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

The flow of information in a system can fluctuate widely because of randomness involved into the selection of the symbols.

Thus, Entropy is defined as the "average information per message".

Mathematical Expression

The Mean value of $I(x_i)$ over the alphabet of source 'x' with m different symbols is represented by $H(X)$ and given by

$$H(X) = E[I(X_i)] = E[I(x_i)]$$

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

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

The above equation shows that the entropy of a source is dependent only on the probabilities of the symbols in an alphabet of the source.

→ Properties of Entropy

1. Entropy is zero if the event is sure or it is impossible i.e

$$H = 0 \quad \text{if} \quad P_k = 0 \text{ or } 1$$

2. When $P_k = \frac{1}{M}$ for all the 'M' symbols, then the symbols are equally likely. For such source entropy is given as

$$H = \log_2 M$$

3. Upper bound on entropy is given as

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

→ Calculate entropy when $P_k = 0$ and when $P_k = 1$

We know Entropy $H = \sum_{k=1}^{M} P_k \log_2 \frac{1}{P_k}$ — ① .

If $P_k = 1$ then eq ① will be

$$H = \sum_{k=1}^{M} 1 \cdot \log \frac{1}{1} = \sum_{k=1}^{M} \frac{\log_{10} 1}{\log_{10} 2} = 0 \qquad [\because \log_{10} 1 = 0]$$

If $P_k = 0$, Instead of putting $P_k = 0$ directly Let us consider the limiting case i.e

$$H = \sum_{k=1}^{M} \underset{P_k \to 0}{Lt} P_k \log\left(\frac{1}{P_k}\right)$$

$$= 0.$$

Thus, Entropy is zero for both certain and most rare message.

→ Show that if there are 'M' number of equally likely messages, then entropy of the source is $\log_2 M$

Sol We know that for 'M' number of equally likely messages, probability is

$$P = \frac{1}{M}$$

The probability is same for all 'M' messages i.e.

$$P_1 = P_2 \cdots\cdots P_M = \frac{1}{M}$$

$$H = \sum_{k=1}^{M} P_k \log_2\left(\frac{1}{P_k}\right)$$

$$= P_1 \log_2 \frac{1}{P_1} + P_2 \log_2 \frac{1}{P_2} + \cdots\cdots + P_M \log_2 \frac{1}{P_M}$$

$$= \frac{1}{M}\log_2 M + \frac{1}{M}\log_2 M + \cdots\cdots + \frac{1}{M}\log_2 M$$

$$= \frac{1}{M} \cdot M \cdot \log_2 M$$

$$= \log_2 M$$

→ Verify the following expression $0 \leq H(X) \leq \log_2 M$ where M is the size of the alphabet of X.

Sol. Proof of the lower bound

$$0 \leq P(x_i) \leq 1,$$

$$\frac{1}{P(x_i)} \geq 1 \quad \text{and} \quad \frac{1}{P(x_i)} \geq 0$$

Then it follows that $P(x_i) \log_2 \frac{1}{P(x_i)} \geq 0$.

Thus, $H(X) = \sum_{i=1}^{m} P(x_i) \log_2 \frac{1}{P(x_i)} \geq 0$ if and only if $P(x_i) = 0$ or 1.

Since $\sum_{i=1}^{m} P(x_i) = 1$ when $P(x_i) = 1$ then $P(x_j) = 0$ for $i \neq j$

Thus, only in this case $H(X) = 0$.

Proof of the upper bound

Let us consider two probability distributions $[P(x_i) = P_k]$ and $[Q(x_i) = Q_k]$ on the alphabet $\{x_i\}$, $i = 1, 2 \ldots M$

We know that $\sum_{k=1}^{M} P_k \log_2\left(\frac{q_k}{P_k}\right) = \sum_{k=1}^{M} P_k \frac{\log_{10}\left(\frac{q_k}{P_k}\right)}{\log_{10} 2}$ — ①

$$= \sum_{k=1}^{M} P_k \cdot \frac{\log_{10} e}{\log_{10} 2} \cdot \frac{\log_{10}\left(\frac{q_k}{P_k}\right)}{\log_{10} e}$$

$$= \sum_{k=1}^{M} P_k \cdot \log_2 e \cdot \log_e \left(\frac{q_k}{P_k}\right)$$

$$= \log_2 e \sum_{k=1}^{M} P_k \log_e\left(\frac{q_k}{P_k}\right)$$

$$= \log_2 e \sum_{k=1}^{M} P_k \cdot \ln\left(\frac{q_k}{P_k}\right) \quad \text{— ②}$$

Using the inequality

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

so, $\ln\left(\frac{q_k}{P_k}\right) \leq \frac{q_k}{P_k} - 1$

∴ Eq ② can be write as

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

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

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

$$\leq \log_2 e \, [1 - 1]$$

$$\therefore \sum_{k=1}^{M} P_k \log_2\left(\frac{q_k}{P_k}\right) \leq 0$$

Now, Let us consider that $q_k = \frac{1}{M}$ for all 'k'. That is all symbols in the alphabet are equally likely. Then above equation becomes

$$\sum_{k=1}^{M} P_k \left[\log_2 q_k + \log \frac{1}{P_k} \right] \le 0$$

$$\sum_{k=1}^{M} P_k \log_2 q_k + \sum_{k=1}^{M} P_k \log \frac{1}{P_k} \le 0$$

$$\cancel{\sum_{k=1}^{M} P_k \log_2 q_k + \sum_{k=1}^{M} P_k \log \frac{1}{P_k}} \quad \sum_{k=1}^{M} P_k \log \frac{1}{P_k} \le -\sum_{k=1}^{M} P_k \log_2 q_k$$

$$\sum_{k=1}^{M} P_k \log \frac{1}{P_k} \le \sum_{k=1}^{M} P_k \log \left(\frac{1}{q_k}\right)$$

Putting $q_k = \frac{1}{M}$ in the above equation

$$\therefore \sum_{k=1}^{M} P_k \log \frac{1}{P_k} \le \sum_{k=1}^{M} P_k \log_2 M$$

$$\le \log_2 M \sum_{k=1}^{M} P_k$$

$$\sum_{k=1}^{M} P_k \log \frac{1}{P_k} \le \log_2 M \qquad \left[\because \sum_{k=1}^{M} P_k = 1\right]$$

$$\boxed{H(X) \le \log_2 M}$$

The Maximum value of Entropy is

$$H_{max}(X) \le \log_2 M.$$

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

(i) Calculate $H(X)$

(ii) Find the amount of information contained in the messages $x_1 x_2 x_1 x_3$ and $x_4 x_3 x_3 x_2$ and compare with the $H(X)$ obtained in part(i).

Sol: We know that entropy is given by

$$H(X) = -\sum_{i=1}^{4} P(x_i) \log [P(x_i)]$$

(i) $H(X) = -[P(x_1)\log P(x_1) + P(x_2)\log(P(x_2)) + P(x_3)\log(P(x_3)) + P(x_4)\log P(x_4)]$

$= -[0.4\log(0.4) + 0.3\log(0.3) + 0.2\log(0.2) + 0.1\log(0.1)]$

$= 1.85$ b/symbol.

(ii) Now, we have $P(x_1 x_2 x_1 x_3) = P(x_1)P(x_2)P(x_1)P(x_3)$

$= (0.4)(0.3)(0.4)(0.2)$

$= 0.0096$

$\therefore I(x_1 x_2 x_1 x_3) = -[\log_2 P(x_1 x_2 x_1 x_3)]$

$= -[\log_2 0.0096] = \dfrac{\log_{10}\left(\frac{1}{0.0096}\right)}{\log_{10} 2} = \dfrac{2.0177}{0.301}$

$= 6.70$ bits/symbol. $\Rightarrow I(x_1 x_2 x_1 x_3) < 7.4\ [= 4H(X)]$

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

$= 0.0012$

$\therefore I(x_4 x_3 x_3 x_2) = -\log_2 0.0012 = \dfrac{\log\left(\frac{1}{0.0012}\right)}{\log 2} = \dfrac{2.92}{0.301}$

$= 9.70$ bits/symbol.

We conclude, $I(x_4 x_3 x_3 x_2) > 7.4\ [= 4H(X)]$ bits/symbol.

→ Given a binary memoryless source 'X' with two symbols x_1 and x_2. Prove that $H(X)$ is maximum when both x_1 and x_2 equiprobable.

Sol. Here, Let us assume that

$$P(x_1) = \alpha \quad \text{so that} \quad P(x_2) = 1-\alpha$$

We know that entropy is given by

$$H(X) = -\sum_{i=1}^{m} P(x_i)\log P(x_i) \text{ bits/symbol}$$

$$= P(x_1)\log\frac{1}{P(x_1)} + P(x_2)\log\frac{1}{P(x_2)}$$

$$= \alpha\log\frac{1}{\alpha} + (1-\alpha)\log\frac{1}{(1-\alpha)}$$

The above equation can also write as

$$H(X) = -\alpha\log\alpha - (1-\alpha)\log(1-\alpha)$$

Differentiating above equation with respect to α, we get

$$\frac{d}{d\alpha}H(X) = \frac{d}{d\alpha}\left[-\alpha\log\alpha - (1-\alpha)\log(1-\alpha)\right]$$

$$= \frac{d}{d\alpha}\left[-\alpha\log\alpha\right] - \frac{d}{d\alpha}\left[(1-\alpha)\log(1-\alpha)\right]$$

$$= -\left[\alpha\cdot\frac{1}{\alpha} + \log\alpha\cdot 1\right] - \frac{d}{d\alpha}\left[\log(1-\alpha)\right] + \frac{d}{d\alpha}\left[\alpha\log(1-\alpha)\right]$$

$$= -\left[1+\log\alpha\right] - \frac{1}{(1-\alpha)}(0-1) + \alpha\cdot\frac{1}{(1-\alpha)}(0-1) + \log(1-\alpha)$$

$$= -1-\log\alpha + \frac{1}{1-\alpha} - \frac{\alpha}{1-\alpha} + \log(1-\alpha)$$

$$= -1-\log\alpha + \frac{1-\alpha}{1-\alpha} + \log(1-\alpha)$$

$$= -1-\log\alpha + 1 + \log(1-\alpha)$$

$$= \log(1-\alpha) - \log\alpha \quad = \log_2\left[\frac{1-\alpha}{\alpha}\right]$$

The maximum value of $H(X)$ requires that

$$\frac{d}{d\alpha} H(X) = 0$$

$$\frac{1-\alpha}{\alpha} = 0$$

$$1-\alpha = \alpha \Rightarrow 2\alpha = 1$$

$$\alpha = \frac{1}{2}$$

$H(X) = 0$ when $\alpha = 0$ or 1.

When $P(x_1) = P(x_2) = \frac{1}{2}$, $H(X)$ is maximum

$$H(X) = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2$$

$$= 1 \text{ bits/symbol}$$

→ For a source transmitting two independent messages m_1 and m_2 having probabilities of p and $(1-p)$ respectively. Prove that the entropy is maximum when both the messages are equally likely. Also, plot the variation of entropy (H) as a function of probability (p) of one of the messages.

Sol. It is given that,

Messages m_1 has a probability of p and

Message m_2 has a probability of $(1-p)$

Therefore, the average information per message is given as

$$H = P\log_2\left(\frac{1}{P}\right) + (1-P)\log_2 \frac{1}{(1-P)}$$

1. Entropy when both the messages are equally likely

$$P = 1-P$$

$$2P = 1$$

$$P = 1/2$$

$$\therefore H = \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2$$

$$= \frac{1}{2} + \frac{1}{2} = 1 \text{ bit/message}.$$

2. Variation of H with probability p

Following table shows values of H for different values of (P).

P	0	0.2	0.4	0.5	0.6	0.8	1
1-P	1	0.8	0.6	0.5	0.4	0.2	0
H	0	0.72	0.97	1	0.97	0.72	0

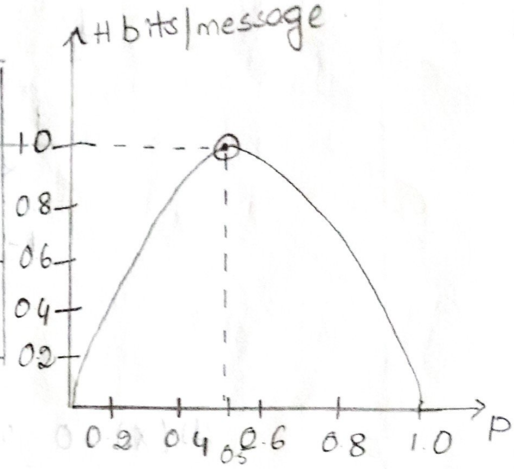


Fig: Average information H for two messages plotted as a function of probability p of one of the messages.

→ Information Rate :

Def: If the time rate at which source 'X' emits symbols is r (symbols/s), the information rate 'R' of the source is given by

$$R = rH(X) \text{ b/sec}$$

where R is information rate

H(X) is entropy or average information

r is rate at which symbols are generated.

Information rate 'R' is represented in average no. of bits of information per second. It is calculated as under.

$$R = \left[r \text{ in } \frac{\text{Symbols}}{\text{Second}} \right] \times \left[H(X) \text{ in } \frac{\text{Information bits}}{\text{Symbol}} \right]$$

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

→ Given a telegraph source having two symbols, dot and dash. The dot duration is 0.2 sec. The dash duration is 3 times the dot duration. The probability of the dot's occurring is twice that of the dash, and the time between symbols is 0.2 sec. Calculate the information rate of the telegraph source.

Sol. Given that

$P(dot) = 2P(dash)$

$P(dot) + P(dash) = 1$

$2P(dash) + P(dash) = 1$

$P(dash) = \frac{1}{3}$

$P(dot) = \frac{2}{3}$

Further, we know that the entropy is given by

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

$$= \sum_{i} P(dot) \log \frac{1}{P(dot)} + P(dash) \log \frac{1}{P(dash)}$$

$$= \frac{2}{3} \log_2 \frac{3}{2} + \frac{1}{3} \cdot \log_2 3$$

$$= 0.667(0.585) + 0.333(1.585)$$

$$= 0.92 \text{ bits/symbol}.$$

Also, given that

$t_{dot} = 0.2s$, $t_{dash} = 0.6s$, $t_{space} = 0.2 sec$.

Average time per symbol will be

$$T_s = P(dot)t_{dot} + P(dash)t_{dash} + t_{space}$$

$$= \frac{2}{3} \cdot (0.2) + \frac{1}{3} \cdot (0.6) + 0.2$$

$$= 0.533 \text{ sec/symbol}.$$

Average symbol rate

$$r = \frac{1}{T_s} = 1.875 \text{ symbols/sec}.$$

Average information rate

$$R = rH(X)$$

$$= 1.875(0.92)$$

$$= 1.725 \text{ b/sec}.$$

→ A discrete source emits one of five symbols once every millisecond with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}$ and $\frac{1}{16}$ respectively. Determine the source entropy and information rate.

Sol: We know that source entropy is given as

$$H(X) = \sum_{i=1}^{m} P(x_i) \log_2 \frac{1}{P(x_i)} = \sum_{i=1}^{5} P(x_i) \log_2 \frac{1}{P(x_i)} \text{ bits/symbol}$$

$$= \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + \frac{1}{16}\log_2 16 + \frac{1}{16}\log_2 16$$

$$= \frac{1}{2} + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{16}\cdot 4 + \frac{1}{16}\cdot 4$$

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

The symbol rate $r = \frac{1}{T_b}$

$$= \frac{1}{10^{-3}} = 1000 \text{ symbols/sec}$$

The information rate is expressed as

$$R = rH(X)$$

$$= 1000 \times 1.875 = 1875 \text{ bits/}\cancel{\text{symbol}}\text{ sec}$$

→ An analog signal bandlimited to 10 KHz is quantized in 8 levels of a PCM system with probabilities of $\frac{1}{4}, \frac{1}{5}, \frac{1}{5}, \frac{1}{10}, \frac{1}{10}, \frac{1}{20}, \frac{1}{20}$ & $\frac{1}{20}$ respectively. Find the entropy and the rate of information.

Sol: According to the sampling theorem,

$$f_s = 2f_m = 2\times 10 = 20 \text{ KHz}.$$

The Entropy is

$$H(X) = \sum_{i=1}^{8} P(x_i) \log \frac{1}{P(x_i)}$$

$$= \frac{1}{4}\log_2 4 + \frac{1}{5}\log_2 5 + \frac{1}{5}\log_2 5 + \frac{1}{10}\log_2 10 + \frac{1}{10}\log_2 10 + \frac{1}{20}\log_2 20$$

$$+ \frac{1}{20}\log_2 20 + \frac{1}{20}\log_2 20$$

$$H = \frac{1}{4}\log_2 4 + \frac{2}{5}\log_2 5 + \frac{2}{10}\log_2 10 + \frac{3}{20}\log_2 20 = 2.84 \text{ bits/sym}$$

$r = 20000$ message/sec

$R = rH(X) = 20000 \times 2.84 = 56800$ bits/sec.

Discrete Memoryless channels (DMC)

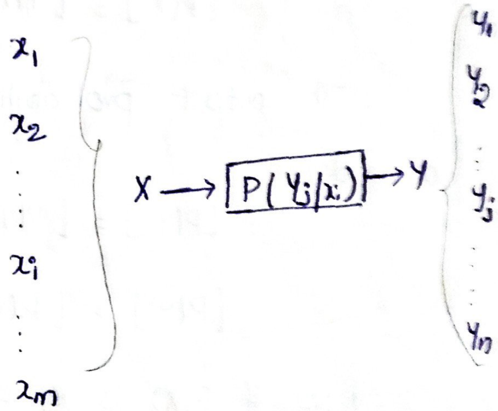
→ Channel Representation

A Communication channel may be defined as the path or medium through which the symbols flow to the receiver end.

A Discrete Memoryless channel (DMC) is a statistical Model with an input X and an output 'Y' as shown in Fig.

During each unit of the time, the channel accepts an input symbol from X, and it generates an output symbol from 'Y'. The channel is said to be "discrete" when the alphabets of X and Y are both finite. Also, it is said to be "Memoryless" when the current output depends on only the current input and not on any of the previous inputs.

A diagram of a DMC with m inputs and n outputs has been illustrated in Fig. The input X consists of input symbols $x_1, x_2 \ldots x_m$. The probabilities of these Source symbols $P(x_i)$ are assumed to be known. The outputs Y consists of output symbols $y_1, y_2 \ldots y_n$. Each possible input-to-output path is indicated along with a conditional probability $P(y_j/x_i)$, where $P(y_j/x_i)$ is the conditional probability of obtaining output y_j given that the input is x_i and is called a channel transition probability.



→ The Channel Matrix

A channel is Completely specified by the complete set of transition probabilities. Accordingly, the channel in the above figure is often specified by the matrix of transition probabilities $P[Y/X]$. This matrix is given by

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

This matrix $[P(Y|X)]$ is called the channel Matrix.

Each input to the channel results in some output, each row of the channel matrix must sum to unity. This means that

$$\sum_{j=1}^{n} P(y_j/x_i) = 1 \text{ for all } i$$

Now, if the input probabilities $P(X)$ are represented by the row matrix, then we have

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

The output probabilities $P(Y)$ are represented by the row matrix as under:

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

$$[P(Y)] = [P(X)]\, P[(Y|X)]$$

Now, if $P(X)$ is represented as a diagonal matrix, then we have

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

then $[P(X,Y)] = [P(X)]_d\, [P(Y|X)]$

where the (i,j) element of matrix $[P(X,Y)]$ has the form $P(x_i\, y_j)$

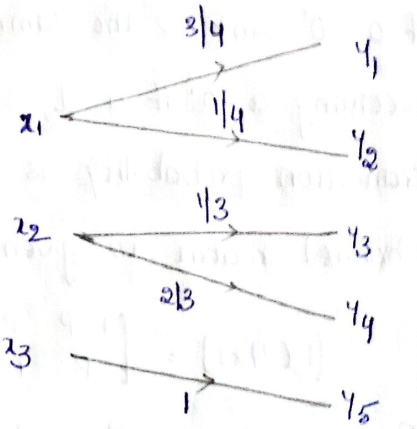
The matrix $[P(X,Y)]$ is known as the joint probability matrix, and the element $P(x_i, y_j)$ is the joint probability of transmitting x_i and receiving y_j.

→ Types of channels:

Other than continous & discrete channels, there are some special type of channels with their own channel matrices.

1. Lossless channel

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

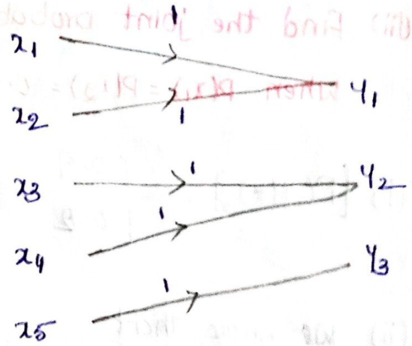


$$P[(Y/X)] = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

In lossless transmission channel, no source information is lost in transmission.

Deterministic channel

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



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

Since each row has only one-zero element, therefore, this element must be unity. Thus, when a given source symbol is sent in the deterministic channel, it is clear which output symbol will be received.

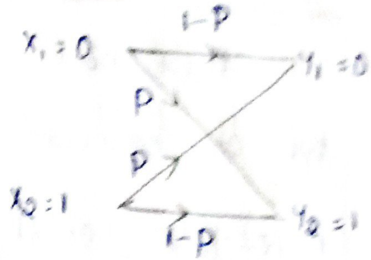
Noiseless channel

A channel is called noiseless if it is both ~~noise~~loss less & deterministic. The input and output alphabets are of the same size, that is $m = n$ for the Noiseless channel.

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

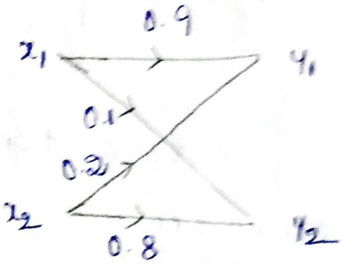
Binary Symmetric Channel (BSC)

A BSC channel has two inputs ($x_1 = 0$, $x_2 = 1$) and two outputs ($y_1 = 0$, $y_2 = 1$). This channel is symmetric because the probability of receiving a 1 if a '0' sent is the same as the probability of receiving a '0' if a 1 is sent. This common transition probability is denoted by 'p'. The Channel matrix is given by



$$[P(Y|X)] = \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix}$$

→ Given a Binary channel shown in given figure
(i) Find the channel matrix of the channel
(ii) Find $P(y_1)$ & $P(y_2)$ when $P(x_1) = P(x_2) = 0.5$
(iii) Find the joint probabilities $P(x_1, y_2)$ & $P(x_2, y_1)$ when $P(x_1) = P(x_2) = 0.5$



sol

(i) $$[P(Y|X)] = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

(ii) We know that

$$[P(Y)] = [P(X)][P(Y|X)]$$

$$= \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}_{1\times 2} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}_{2\times 2} = \begin{bmatrix} 0.45+0.1 & 0.05+0.4 \end{bmatrix}_{1\times 2}$$

$$= \begin{bmatrix} 0.55 & 0.45 \end{bmatrix}$$

$\therefore P(y_1) = 0.55$, $P(y_2) = 0.45$

(iii) $$[P(X, Y)] = [P(X)]_d [P(Y|X)]$$

$$= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} = \begin{bmatrix} 0.45 & 0.05 \\ 0.1 & 0.4 \end{bmatrix}$$

Now let us compare with standard matrix form as under

$$[P(X,Y)] = \begin{bmatrix} P(x_1, y_1) & P(x_1, y_2) \\ P(x_2, y_1) & P(x_2, y_2) \end{bmatrix} = \begin{bmatrix} 0.45 & 0.05 \\ 0.1 & 0.4 \end{bmatrix}$$

$P(x_1, y_2) = 0.05$, $P(x_2, y_1) = 0.1$

→ Two binary channels are connected in cascade as shown in Fig

(i) Find the overall channel matrix of the resultant channel, and draw the resultant equivalent channel diagram

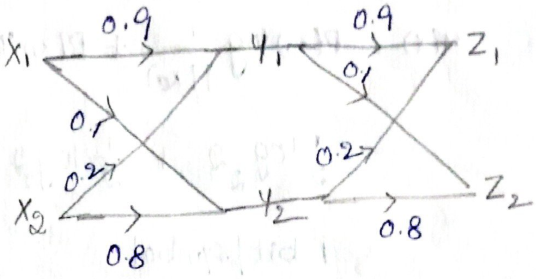


Fig: Two binary channels in cascade

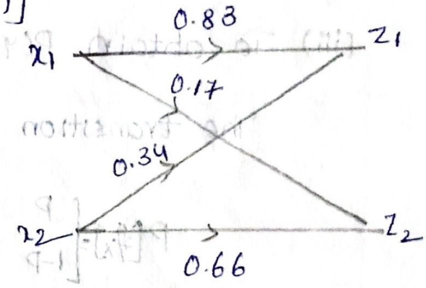
(ii) Find $P(z_1)$ & $P(z_2)$ when $P(x_1) = P(x_2) = 0.5$

Sol (i) We know that

$$[P(Y)] = [P(X)][P(Y/X)] = [P(Y)][P(Z/Y)]$$

$$[P(Z)] = [P(X)][P(Y/X)][P(Z/Y)] = [P(X)][P(Z/X)]$$

Thus from Fig, we have



$$P(Z/X) = [P(Y/X)][P(Z/Y)]$$

$$= \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix} \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}$$

$$= \begin{bmatrix} 0.81 + 0.02 & 0.09 + 0.08 \\ 0.18 + 0.16 & 0.02 + 0.64 \end{bmatrix} = \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix}$$

(ii) $P(Z) = [P(X)][P(Z/X)]$

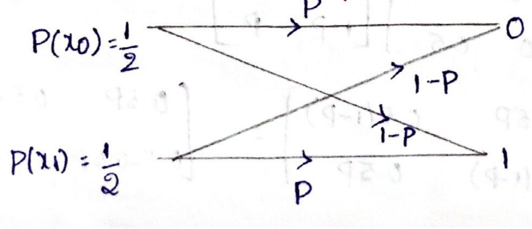
$$= [0.5 \quad 0.5] \begin{bmatrix} 0.83 & 0.17 \\ 0.34 & 0.66 \end{bmatrix}$$

$$= [0.585 \quad 0.415]$$

Compare with standard expression $[P(z_1) \quad P(z_2)]$

$\therefore P(z_1) = 0.585, \quad P(z_2) = 0.415$

→ Binary symmetric channel shown in the given figure. Find the rate of information transmission across this channel for $p = 0.8$ and 0.6. The symbols are generated at the rate of 1000 per second. $P(x_0) = P(x_1) = \frac{1}{2}$. Also determine channel input information rate.



sol (i) To obtain entropy of the Source :

$$H(X) = P(x_0)\log\frac{1}{P(x_0)} + P(x_1)\log\frac{1}{P(x_1)}$$

$$= \frac{1}{2}\log_2 2 + \frac{1}{2}\log_2 2$$

$= 1$ bit/symbol

(ii) To obtain input information rate

$R = rH(X)$

Given $r = 1000$ symbols per sec

$R = 1000 \times 1 = 1000$ bits/sec

(iii) To obtain $P(Y_0)$ & $P(Y_1)$

The transition matrix for the given BSC channel is

$$P[Y/X] = \begin{bmatrix} P & 1-P \\ 1-P & P \end{bmatrix}$$

Then

$$[P(Y)] = P(X)\,[P(Y|X)]$$

$$= [0.5 \quad 0.5]\begin{bmatrix} P & 1-P \\ 1-P & P \end{bmatrix}$$

$$= [0.5P + 0.5(1-P) \qquad 0.5(1-P) + 0.5P]$$

$$= [0.5P + 0.5 - 0.5P \qquad 0.5 - 0.5P + 0.5P]$$

$$[P(Y)] = [0.5 \quad 0.5]$$

$\therefore P(Y_1) = 0.5 \qquad P(Y_2) = 0.5$

(iv) To obtain $P(x_i, y_j)$ and $P(x_i/y_j)$

We know that Mutual Probability

$$P(X,Y) = [P(X)]_d\,[P(Y|X)]$$

$$= \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}\begin{bmatrix} P & 1-P \\ 1-P & P \end{bmatrix}$$

$$= \begin{bmatrix} 0.5P & 0.5(1-P) \\ 0.5(1-P) & 0.5P \end{bmatrix} = \begin{bmatrix} 0.5P & 0.5-0.5P \\ 0.5-0.5P & 0.5P \end{bmatrix}$$

$$\therefore P(X,Y) = \begin{bmatrix} 0.5P & 0.5-0.5P \\ 0.5-0.5P & 0.5P \end{bmatrix} = \begin{bmatrix} P(X_1 Y_1) & P(X_1 Y_2) \\ P(X_2 Y_1) & P(X_2 Y_2) \end{bmatrix}$$

Thus $P(X_1 Y_1) = 0.5P$ $\quad P(X_1 Y_2) = 0.5-0.5P$

$P(X_2 Y_1) = 0.5-0.5P$ $\quad P(X_2 Y_2) = 0.5P$

Then the conditional probability $P(X|Y)$ can be obtain by

$$P(X|Y)\, P(Y) = P(Y|X)\, P(X)$$

$$P(X|Y) = \frac{P(Y|X) \cdot P(X)}{P(Y)}$$

$$= \frac{\begin{bmatrix} P & 1-P \\ 1-P & P \end{bmatrix} \begin{bmatrix} 0.5 & 0.5 \end{bmatrix}}{\begin{bmatrix} 0.5 & 0.5 \end{bmatrix}}$$

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

(v) To obtain information rate across the channel:

The conditional entropy $H(X|Y)$ is given by equation

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

$$= P(x_1 y_1) \log \frac{1}{P(x_1|y_1)} + P(x_1 y_2) \log \frac{1}{P(x_1|y_2)} + P(x_2 y_1) \log \frac{1}{P(x_2|y_1)}$$

$$+ P(x_2 y_2) \log \frac{1}{P(x_2|y_2)}$$

$$= 0.5P \log \frac{1}{(0.5P)} + 0.5(1-P) \log \frac{1}{(1-P)0.5} + 0.5(1-P) \log \frac{1}{0.5(1-P)}$$

$$+ 0.5P \log \frac{1}{0.5P}$$

$$H(X/Y) = \frac{1}{2} P \log \frac{2}{P} + \frac{1}{2}(1-P)\log \frac{2}{(1-P)} + \frac{1}{2}(1-P)\log \frac{2}{(1-P)} + \frac{1}{2} P \log \frac{2}{P}$$

$$= P \log_2 \frac{2}{P} + (1-P)\log_2 \frac{2}{(1-P)}$$

For $P = 0.8$ then $H(X/Y)$ becomes

$$H(X/Y) = 0.8 \log_2 \left(\frac{2}{0.8}\right) + (0.2)\log_2\left(\frac{2}{0.2}\right)$$

$$= 0.7219 \text{ bits/symbol}$$

$$R = [H(X) - H(X/Y)]r$$

$$= [1 - 0.7219]\,1000$$

$$= 278 \text{ bits/sec}$$

For $P = 0.6$, $H(X/Y)$ becomes

$$H(X/Y) = 0.6 \log_2 \frac{2}{0.6} + (0.4)\log \frac{2}{0.4}$$

$$= 0.97 \text{ bits/symbol}$$

Hence, the average rate of information transmission 'R' across the channel will be

$$R = [H(X) - H(X/Y)]r$$

$$= [1 - 0.97]\,1000$$

$$= 29 \text{ bits/sec.}$$

Thus, the above results indicate that the information transmission rate across the channel decreases rapidly as 'p' approaches $\frac{1}{2}$.

→ 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

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

where $I(x_i, y_j)$ is the mutual Information

$P(x_i/y_j)$ is the conditional probability that x_i was transmitted and y_j is received

$P(x_i)$ is the 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.

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

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

→ **Properties of Mutual Information**

1. The mutual information of the channel is symmetric i.e

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

2. The mutual information can be expressed in terms of entropies of channel input or output and conditional entropies i.e,

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

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

where $H(X/Y)$ & $H(Y/X)$ are conditional entropies

3. The Mutual information is always positive i.e

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

4. The Mutual information is related to the joint entropy $H(X,Y)$ by following relation

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

→ Prove that the Mutual information of the channel is symmetric i.e

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

Sol) We know that from probability theory.

$$P(x_i\, y_j) = P(x_i / y_j)\, P(y_j) \quad \text{—} ①$$

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

$P(x_i\ y_j)$ is the joint probability that x_i is transmitted & y_j is received

$P(x_i/y_j)$ is the conditional probability of that x_i is transmitted and y_j is received.

$P(y_j/x_i)$ is the conditional probability that y_j is received & x_i is transmitted.

$P(x_i)$ is the probability of symbol x_i for transmission

$P(y_j)$ is the probability of symbol y_j is received.

From eq ① & ② we can write

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

$$\frac{P(x_i/y_j)}{P(x_i)} = \frac{P(y_j/x_i)}{P(y_j)} \quad -③$$

We know, the average mutual information is given by

$$I(X;Y) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i\ y_j)\log_2 \frac{P(x_i/y_j)}{P(x_i)} \quad -④$$

$$I(Y;X) = \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i\ y_j)\log \frac{P(y_j/x_i)}{P(y_j)} \quad -⑤$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{m} P(x_i\ y_j)\log \frac{P(x_i/y_j)}{P(x_i)} \qquad \left[\because \frac{P(x_i\ y_j)}{P(x_i)} = \frac{P(y_j/x_i)}{P(y_j)}\right]$$

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

Thus, the mutual information of the discrete memoryless channel is symmetric.

2. Prove the following relationships

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

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

Sol) $H(X/Y)$ is the conditional entropy and it is given as

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

$H(X/Y)$ is the conditional entropy and $H(X/Y)$ is the information or uncertainty in X after Y is received

In otherwords $H(X/Y)$ is the information lost in the noisy channel

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

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

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

Let us use the standard probability relation

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

∴ The above equation will be

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

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

The $I(X;Y)$ is the average information transferred per symbol across the channel. It is equal to source entropy minus information lost in the noisy channel

Similarly consider the average mutual information

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

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

The conditional entropy $H(Y/X)$ is given as

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

Substitute in the above equation, then

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

The standard probability equation

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

$$\therefore I(Y;X) = \sum_{j=1}^{n} P(y_j) \log_2 \frac{1}{P(y_j)} - H(Y/X)$$

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

→ Prove that the mutual Information is always positive i.e

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

sol We know Mutual Information

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

By Considering from probability theory,

$$P(x_i/y_j) = \frac{P(x_i, y_j)}{P(y_j)} \quad -(2)$$

Substitute eq (2) in eq (1), We get

$$I(X;Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} P(x_i, y_j) \log_2 \frac{P(x_i, y_j)}{P(x_i) P(y_j)} \quad -(3)$$

We know that $\log_2 \frac{x}{y} = -\log_2 \frac{y}{x}$

Hence, the above equation becomes

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

We know that $\sum_{k=1}^{m} P_k \log_2 \left(\frac{q_k}{P_k}\right) \leq 0 \quad -(5)$

By Comparing eq (4) & (5) We get,

$P_k = P(x_i, y_j)$; $q_k = P(x_i) P(y_j)$.

Both P_k & q_k are two probability distributions on same alphabet

$-I(X;Y) \leq 0 \implies$ i.e $\boxed{I(X;Y) \geq 0}$

→ Verify the following expression

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

Sol. We know that

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

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

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

$$= -\sum_{j=1}^{m} \sum_{i=1}^{m} P(x_i, y_j) \log [P(x_i/y_j) \cdot P(y_j)]$$

$$= -\sum_{j=1}^{n} \sum_{i=1}^{m} P(x_i, y_j) \log [P(x_i/y_j)] - \sum_{j=1}^{n} \sum_{i=1}^{m} P(x_i, y_j) \log P(y_j)$$

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

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

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

Hence proved.

→ Prove the following

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

Sol. We know that

$$H(X,Y) = H(X/Y) + H(Y) \quad \text{— (1)}$$

then $H(X/Y) = H(X,Y) - H(Y)$ — (2)

We know the mutual information

$$I(X;Y) = H(X) - H(X/Y) \quad \text{— (3)}$$

Substitute eq (2) in eq (3), we get

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

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

Hence proved.

The Conditional & Joint Entropies

The input probabilities $P(x_i)$, output probabilities $P(y_j)$, transition probabilities $P(y_j/x_i)$ and Joint probabilities $P(x_i, y_j)$. Let us define the following various entropy functions for a channel with 'm' inputs and 'n' outputs

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

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

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

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

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

where $H(X)$ is the average uncertainity of the channel input

$H(Y)$ is the average uncertainity of the channel output

$H(X/Y)$ is the measure of the average uncertainity remaining about the channel input after the channel output has been observed.

$H(Y/X)$ is the average uncertainity of the channel output given that X was transmitted

$H(X, Y)$ is the joint entropy. It gives average uncertainity of the communication channel as a whole.

Two useful relationships among the various entropies as under

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

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

SOURCE CODING

Definition of source coding: A conversion of the output of a discrete memoryless source (DMS) into a sequence of binary symbols (i.e binary code word) is called source coding.

The device that performs this conversion is called the "Source encoder"

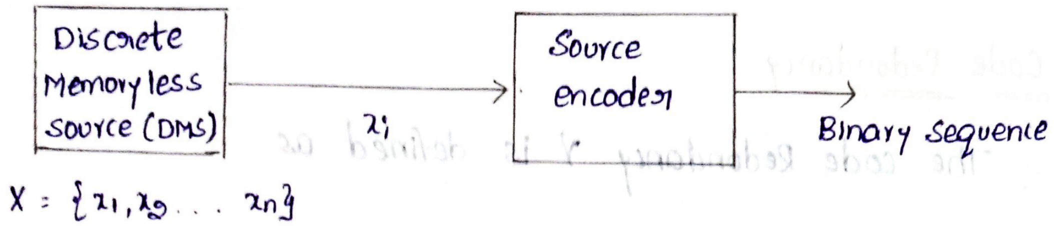


Fig: Block diagram for source coding.

→ **Objective of source coding:**

An objective of source coding is to minimize the average bit rate required for representation of the source by reducing the redundancy of the information source.

→ **Codeword length:**

Let 'X' be a DMS with finite entropy $H(X)$ and an alphabet $\{x_1, \dots x_m\}$ with corresponding probabilities of occurrence $P(x_i)$ $(i = 1, 2 \dots m)$

Let the binary codeword assigned to symbol x_i by the encoder have length 'n_i' measured in bits.

The Length of a codeword is the no. of binary digits in the codeword.

→ **Average codeword length:**

The average codeword length 'L' per source symbol is given by

$$\bar{N} \text{ (or) } L = \sum_{i=1}^{m} P(x_i) n_i$$

where $\bar{N}$ or L represents average no. of bits per source symbol.

n_i length of codeword.

→ Code efficiency

The code efficiency η is defined as under

$$\eta = \frac{H(X)}{L} \quad (\text{or}) \quad \frac{L_{min}}{L} \quad (\text{or}) \quad \frac{N_{min}}{\overline{N}}$$

where L_{min} (or) N_{min} is the minimum possible value of L (or) $\overline{N}$.

When η approaches unity, the code is said to be "efficient"

→ Code Redundancy

The code Redundancy γ is defined as

$$\gamma = 1 - \eta$$

→ Code Variance :

Variance of the code is given as

$$\sigma^2 = \sum_{k=0}^{L-1} P_k (n_k - \overline{N})^2$$

where $\overline{N}$ → average codeword length

n_k is the no. of bits assigned to k^{th} symbol

P_k is the probability of k^{th} symbol

$L-1$ (or) M is the no. of symbols.

Variance is the measure of variability in codeword lengths.

Variance should be as small as possible.

→ The source coding Theorem

The source coding theorem states that for a DMS X, with entropy $H(x)$, the average codeword length 'L' per symbol is bounded as

$$L \geq H(X)$$

'L' can be made as close to $H(x)$ as desired. Thus $L_{min} = H(X)$, the code efficiency can be written as

$$\eta = \frac{H(X)}{L}$$

Entropy Coding:

The design of a variable-length code such that its average codeword length approaches the entropy of DMS is often reffered to as "Entropy coding"

The two examples of entropy coding as

1. Shannon-Fano coding
2. Huffman coding

→ **Shannon-Fano coding**

The procedure of Shannon-Fano algorithm as follows:

1. List the source symbols in order of decreasing probability
2. Partition the set into two sets that are as close to equiprobable as possible, and assign '0' to the upper set and '1' to the lower set
3. Continue this process, each time partitioning the sets with as nearly equal probabilities as possible until further partitioning is not possible.

→ A discrete memoryless source has an alphabet of seven symbols with probability for its output as described here:

Symbol	S_0	S_1	S_2	S_3	S_4	S_5	S_6
probability	0.25	0.25	0.125	0.125	0.125	0.0625	0.0625

(i) Use the Shannon-Fano algorithm to develop an efficient code

(ii) For that code, calculate the average no. of bits/message

x_i	$P(x_i)$	Step 1	Step 2	Step 3	Step 4	Code	n_i
S_0	0.25	0	0			00	2
S_1	0.25	0	1			01	2
S_2	0.125	1	0	0		100	3
S_3	0.125	1	0	1		101	3
S_4	0.125	1	1	0		110	3
S_5	0.0625	1	1	1	0	1110	4
S_6	0.0625	1	1	1	1	1111	4

Average codeword length

$$\bar{N} = \sum_{i=0}^{6} P_i n_i$$

$= 0.25 \times 2 + 0.25 \times 2 + 0.125 \times 3 + 0.125 \times 3 + 0.125 \times 3 + 0.0625 \times 4 + 0.0625 \times 4$

$= 0.5 + 0.5 + 0.375 + 0.375 + 0.375 + 0.25 + 0.25$

$= 1 + 1.125 + 0.5$

$= 2.625$ bits/symbol

$$H = \sum_{i=0}^{6} P_i \log \frac{1}{P_i}$$

$= P_0 \log \frac{1}{P_0} + P_1 \log \frac{1}{P_1} + P_2 \log \frac{1}{P_2} + P_3 \log \frac{1}{P_3} + P_4 \log \frac{1}{P_4} + P_5 \log \frac{1}{P_5} + P_6 \log \frac{1}{P_6}$

$= 0.25 \log \frac{1}{0.25} + 0.25 \log \frac{1}{0.25} + 0.125 \log \frac{1}{0.125} + 0.125 \log \frac{1}{0.125} + 0.125 \log \frac{1}{0.125} + 0.0625 \log \frac{1}{0.0625} + 0.0625 \log \frac{1}{0.0625}$

$= 2.625$ bits/symbol

$\eta = \frac{H}{\bar{N}} = \frac{2.625}{2.625} = 1 = 100\%$

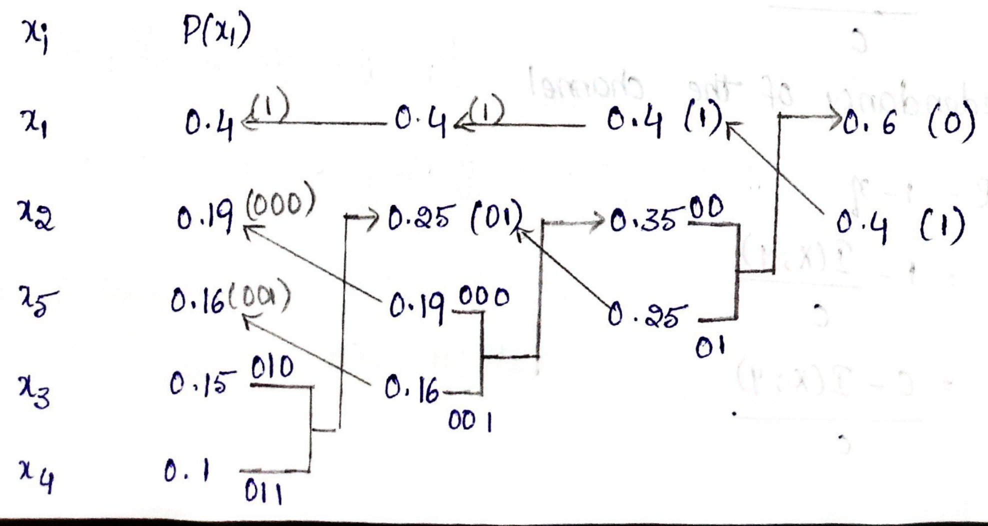
→ Huffman Coding

Huffman encoding results in an optimum code. It is the code that has the highest efficiency. The Huffman encoding procedure procedure is as follows :

1. List the source symbols in order of decreasing probability.
2. Combine the probabilities of the two symbols having the lowest probabilities and reorder the resultant probabilities, this step is called "Reduction 1". The same procedure is repeated until there are two ordered probabilities remaining.
3. Start encoding with the last reduction, which consist of exactly two ordered probabilities. Assign '0' as the first digit in the codewords for all the source symbols associated with the first probability; assign 1 to the second probability.
4. Now go back and assign '0' and '1' to the second digit for the two probabilities that were combined in the previous reduction step, retaining all assignments made in step 3.
5. Keep regressing this way until the first column is reached.

→ A DMS 'X' has five symbols x_1 to x_5 with $P(x_1) = 0.4$, $P(x_2) = 0.19$, $P(x_3) = 0.15$, $P(x_4) = 0.1$, $P(x_5) = 0.16$. Construct a Huffman code and find out the efficiency & Redundancy.

Sol



x_i	$P(x_i)$	code	Length (n_i)
x_1	0.4	1	1
x_2	0.19	000	3
x_3	0.15	010	3
x_4	0.1	011	3
x_5	0.16	001	3

$$H = -\sum_{k=1}^{5} P_k \log_2 P_k$$

$= 2.15$ bit / symbol

$$\bar{N} = \sum_{i=1}^{5} P_i n_i$$

$= 2.2$ bits / symbol

$$\eta = \frac{H}{\bar{N}} = \frac{2.15}{2.2}$$

$= 97.7\%$

→ Channel Capacity

It is defined as the maximum of Mutual Information

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

$$= \max [H(X) - H(X/Y)]$$

A suitable measure for efficiency of transmission of information introduced by comparing the actual rate and the upper bound of the rate of information transmission for a given channel

channel efficiency is defined as

$$\eta = \frac{\text{Actual information transmission}}{\text{Maximum information transmission}}$$

$$= \frac{I(X;Y)}{\max I(X;Y)}$$

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

The Redundancy of the channel

$$R = 1 - \eta$$

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

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

→ **Noise - free channel**

For a Noise-free channel, the Mutual information is

We know that,

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

$$I(X;Y) = H(X) \quad [\because H(X/Y) = 0]$$

Hence, the channel capacity of a Noise-free channel is

$$C = \max[I(X;Y)]$$

$$= \max[H(X)]$$

$$C = \log M \text{ bits/message}$$

where M is no. of messages.

→ **Symmetric channel**

A Symmetric channel is defined as

(i) $H(Y/x_j)$ is independent of j ; i.e the entropy corresponding to each row of $[P(Y/X)]$ is the same.

(ii) $\sum_{j=1}^{m} P(y_k/x_j)$ is independent of k, i.e the sum of all the columns of $[P(Y/X)]$ is same.

If $D = P(Y/X)$ is a square matrix, then for a symmetric channel, the rows and columns are identical, except for permutations.

$$D = P[(Y/X)] = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix}$$

For a symmetric channel,

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

$$= H(Y) - \sum_{j=1}^{m} H(Y/x_j)\, P(x_j)$$

$$= H(Y) - A \sum_{j=1}^{m} P(x_j) \qquad [A = H(Y/x_j)] \text{ is independent of } j$$

& hence is taken out of the summation sign.

$$I(X;Y) = H(Y) - A \qquad \left[\because \sum_{j=1}^{m} P(x_j) = 1\right]$$

The channel capacity of a symmetric channel is

$$C = \max [I(X;Y)]$$

$$= \max [H(Y) - A]$$

$$= \max [H(Y)] - A$$

$$C = \log_2 n - A$$

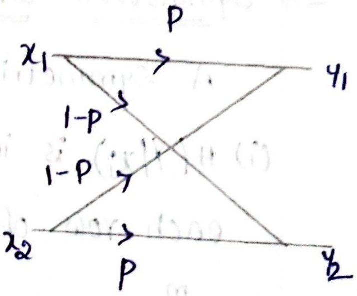
where 'n' is the total no. of received symbols.

→ <u>Binary Symmetric Channel</u> (BSC)

In BSC case $m = n = 2$.

The channel matrix $D = [P(Y/X)] = \begin{bmatrix} P & 1-P \\ 1-P & P \end{bmatrix}$

$$= \begin{bmatrix} P & q \\ q & P \end{bmatrix}$$



For the BSC shown in the above figure, the channel capacity for (i) $P = 0.9$; (ii) $P = 0.6$

Sol) We know the channel capacity of a symmetric channel is

$$C = \log n - A$$

$$= \log n - H(Y/x_j)$$

$$= \log 2 - \left[-\sum_{j=1}^{2} P(y_k/x_j) \log P(y_k/x_j)\right]$$

$$= \log 2 + P(y_1/x_1) \log P(y_1/x_1) + P(y_1/x_2) \log P(y_1/x_2)$$

$$= \log 2 + P\log P + (1-P)\log(1-P)$$

$$= \log 2 + P\log P + q\log q$$

$$= 1 - H(P)$$

$$= 1 - H(q)$$

(i) For $P = 0.9$

$C = 1 + 0.9 \log 0.9 + 0.1 \log 0.1$

$= 0.531$ bit/message

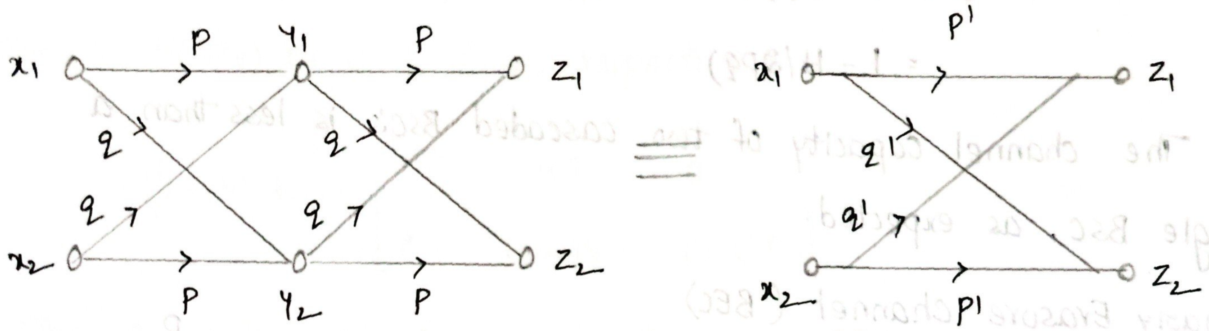
(ii) For $P = 0.6$

$C = 1 + 0.6 \log 0.6 + 0.4 \log 0.4$

$= 0.029$ bit/message

→ Cascaded Channels

Sometimes channels are to be cascaded for some reasons. Let us consider the case of two cascaded identical Binary symmetric channels as shown in below figure:



The message from X_1, reaches Z_1 in two ways:

1. $X_1 \xrightarrow{P} Y_1 \xrightarrow{P} Z_1 \Rightarrow P \cdot P = P^2$

2. $X_1 \xrightarrow{q} Y_2 \xrightarrow{q} Z_1 \Rightarrow q \cdot q = q^2$

Then $P' = P^2 + q^2$

$= (P+q)^2 - 2pq = 1 - 2pq$

Similarly, the message from x_1 reaches z_2 in two ways

1. $x_1 \xrightarrow{P} y_1 \xrightarrow{q} z_2 \Rightarrow P \cdot q = Pq$

2. $x_1 \xrightarrow{q} y_2 \xrightarrow{P} z_2 \Rightarrow q \cdot P = Pq$

Then $q' = Pq + Pq$

$q' = 2pq$

$\therefore P' + q' = 1 - 2pq + 2pq$

$= 1$

The channel matrix of the cascaded channel is

$$P(Z/X) = \begin{bmatrix} 1-2pq & 2pq \\ 2pq & 1-2pq \end{bmatrix}$$

$$= \begin{bmatrix} p' & q' \\ q' & p' \end{bmatrix}$$

The channel capacity of a BSC is given by

$$C = 1 - H(q)$$

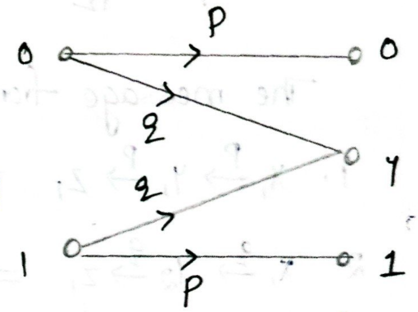
The channel capacity of cascaded channel is

$$C = 1 - H(q')$$

$$= 1 - H(2pq)$$

The channel capacity of two cascaded BSC's is less than a single BSC, as expected.

→ Binary Erasure channel (BEC)



A Binary Erasure channel (BEC) has two inputs (0,1) and three outputs (0, y, 1) as shown in Fig. Here '0' & '1' are transmitted and they are received 0, y & 1.

The symbol 'y' indicates that, due to noise, no deterministic decision can be made as to whether the received symbol is as '0' or '1'.

In otherwords, the symbol 'y' indicates that the output is erased. Hence the name Binary Erasure channel (BEC).

In practice, whenever decision is in favour of 'y', i.e whenever the deterministic decision in favour of '0' or '1' is not possible, the receiver requests the transmitter for re-transmission till the decision is taken either in favour of '0' or in favour of '1'.

For BEC, the channel matrix is

$$D = [P(Y/X)] = \begin{bmatrix} p & q & 0 \\ 0 & q & p \end{bmatrix}$$

Let us assume that $P(0) = \alpha$ & $P(1) = 1-\alpha$ at the transmitter.

$$H(X) = P(x_1)\log\frac{1}{P(x_1)} + P(x_2)\log\frac{1}{P(x_2)}$$

$$= P(0)\log\frac{1}{P(0)} + P(1)\log\frac{1}{P(1)}$$

$$= \alpha\log\frac{1}{\alpha} + (1-\alpha)\log\frac{1}{(1-\alpha)}$$

The joint probability $P(X,Y)$ can be found by multiplying the rows of $P(Y/X)$ by α & $(1-\alpha)$ respectively.

$$P(X,Y) = \begin{bmatrix} \alpha p & \alpha q & 0 \\ 0 & (1-\alpha)q & (1-\alpha)p \end{bmatrix}$$

The summation of columns give

$P(y_1) = \alpha p$; $P(y_2) = \alpha q + (1-\alpha)q$

$P(y_3) = (1-\alpha)p$; $P(y_2) = q$

The conditional probability matrix $P(X/Y)$ can be found by dividing the columns of $P(X,Y)$ by $P(y_1)$, $P(y_2)$ & $P(y_3)$ respectively.

$$P(X/Y) = \begin{bmatrix} \frac{\alpha p}{\alpha p} & \frac{\alpha q}{q} & \frac{0}{(1-\alpha)p} \\ \frac{0}{\alpha p} & \frac{(1-\alpha)q}{q} & \frac{(1-\alpha)p}{(1-\alpha)p} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \alpha & 0 \\ 0 & 1-\alpha & 1 \end{bmatrix}$$

$$H(X/Y) = -\sum_{j=1}^{2}\sum_{i=1}^{3} P(x_j\, y_k)\log P(x_j/y_k)$$

$$= -[\alpha p\log 1 + \alpha q\log\alpha + (1-\alpha)q\log(1-\alpha) + (1-\alpha)p\log 1]$$

$$= -q\left[\alpha \log \alpha + (1-\alpha)\log(1-\alpha)\right]$$

$$= qH(X)$$

$$H(X|Y) = (1-p)\,H(X)$$

The mutual information $I(X;Y) = H(X) - H(X|Y)$

$$= H(X) - (1-p)\,H(X)$$

$$= pH(X)$$

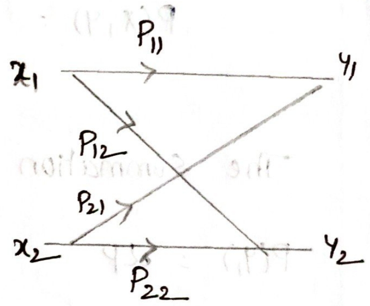
$\therefore$ The channel capacity $C = \max(I(X;Y))$

$$= \max(pH(X))$$

$$= p \cdot \max H(X)$$

$$C = p$$

→ Binary Channel



In practice, we come across binary channels with non-symmetric structures. A binary channel is shown in Figure. The channel matrix is

$$D = [P[Y/X]] = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$$

To find the channel capacity of a binary channel, the auxilliary variables Q_1 & Q_2 are defined by

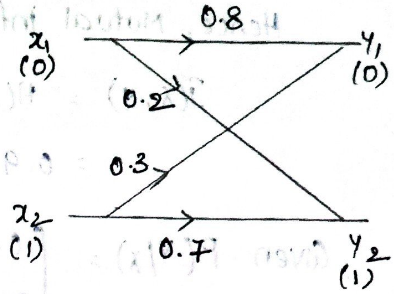
$$[P][Q] = -[H]$$

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}\begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} P_{11}\log P_{11} + P_{12}\log P_{12} \\ P_{21}\log P_{21} + P_{22}\log P_{22} \end{bmatrix}$$

The channel capacity is given by,

$$C = \log\left(2^{Q_1} + 2^{Q_2}\right)$$

→ Find the mutual information and channel capacity of the channel shown in given figure for $P(x_1) = 0.6$, $P(x_2) = 0.4$



sol The channel matrix $D = [P(Y/X)] = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix}$

The joint probability $P(X,Y) = [P(X)]_d \, [P(Y/X)]$

$$= \begin{bmatrix} 0.6 & 0 \\ 0 & 0.4 \end{bmatrix} \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} 0.48 & 0.12 \\ 0.12 & 0.28 \end{bmatrix}$$

The output probabilities can be obtained by

$$P(Y) = P(X) \cdot P(Y/X)$$

$$= [0.6 \quad 0.4] \cdot \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} = [0.48 + 0.12 \quad 0.12 + 0.28]$$

$$= [0.6 \quad 0.4]$$

The matrix $[P(x/y)]$ is obtained by dividing the columns of $[P(X,Y)]$ by $P(y_1)$ and $P(y_2)$ respectively.

$$P(X/Y) = \frac{P(X,Y)}{P(Y)} = \begin{bmatrix} \frac{0.48}{0.6} & \frac{0.12}{0.4} \\ \frac{0.12}{0.6} & \frac{0.28}{0.4} \end{bmatrix}$$

$$P(X/Y) = \begin{bmatrix} 0.8 & 0.3 \\ 0.2 & 0.7 \end{bmatrix}$$

The entropies $H(X)$ and $H(X/Y)$ can be found as

$$H(X) = -\sum_{i=1}^{2} P(x_i) \log P(x_i)$$

$$= -[0.6 \log_2 0.6 + 0.4 \log_2 0.4] \Rightarrow 0.971 \text{ bits/symbol}$$

$$H(X/Y) = -\sum_{j=1}^{2} \sum_{i=1}^{2} P(x_i, y_j) \log P(x_i/y_j)$$

$$= -[0.48 \log 0.8 + 0.12 \log 0.3 + 0.12 \log 0.2 + 0.28 \log 0.7]$$

$$= 0.786 \text{ bits/symbol}$$

Hence, Mutual Information

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

$$= 0.971 - 0.786 = 0.185 \text{ bits/symbol}$$

Given $P(Y/X) = \begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$

We know that the channel capacity of a binary channel is

$$[P][Q] = -[H]$$

$$\begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} P_{11} \log P_{11} + P_{12} \log P_{12} \\ P_{21} \log P_{21} + P_{22} \log P_{22} \end{bmatrix}$$

$$\therefore C = \log(2^{Q_1} + 2^{Q_2})$$

i.e $\begin{bmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{bmatrix} \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} = \begin{bmatrix} 0.8 \log 0.8 + 0.2 \log 0.2 \\ 0.3 \log 0.3 + 0.7 \log 0.7 \end{bmatrix}$

$$0.8Q_1 + 0.2Q_2 = 0.8 \log 0.8 + 0.2 \log 0.2 \quad -① $$

$$0.3Q_1 + 0.7Q_2 = 0.3 \log 0.3 + 0.7 \log 0.7 \quad -②$$

By solving the above equations, we get

$$Q_1 = -0.6568$$

$$Q_2 = -0.9764$$

Therefore the channel capacity of a given binary channel is

$$C = \log(2^{+Q_1} + 2^{Q_2})$$

$$= \log(2^{-0.6568} + 2^{-0.9764})$$

$$= \log(1.146)$$

$$C = 0.2 \text{ bits/message}$$

Capacity of a Gaussian channel (or) Shannon-Hartley Theorem

Statement: If the channel is disturbed by a white Gaussian noise one can transmit information at a rate of 'C' bits per second, where 'C' is the channel capacity and is expressed as

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

where B is channel Bandwidth in Hz
S is Signal power
N is Noise power

Proof: The received signal contain both signal and noise then received signal amplitude is $\sqrt{S+N}$

where S is Average signal power
N is Average Noise power

The root mean square of Noise voltage is $\sqrt{N}$

The no. of distinct levels that can be distinguished without error can be expressed as

$$M = \frac{\sqrt{S+N}}{\sqrt{N}}$$

$$M = \sqrt{1 + \frac{S}{N}}$$

The maximum amount of information carried by each pulse having $\sqrt{1+\frac{S}{N}}$ distinct levels is given by

$$I = \log_2 \sqrt{1 + \frac{S}{N}}$$

$$I = \frac{1}{2} \log_2 \left(1 + \frac{S}{N}\right) \text{ bits}$$

If a channel can transmit a maximum of 'k' pulses per second. Then channel capacity 'C' is given by

$$C = \frac{k}{2}\log_2\left(1+\frac{S}{N}\right) \text{ bits per second}$$

Since each pulse carry a maximum information of $\frac{1}{2}\log_2\left(1+\frac{S}{N}\right)$ bits then a system with Bandwidth 'B' Hz can transmit a maximum of 2B pulses per second then channel capacity is

$$C = 2B \cdot \frac{1}{2}\log_2\left(1+\frac{S}{N}\right)$$

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

To transmit the information at a given rate, we may reduce the signal power transmitted provided that the Bandwidth is increased correspondingly

Similarily the BW may be reduced if we have to increase signal power.

→ **Tradeoff between Bandwidth and Signal to Noise Ratio**

The channel capacity of the gaussian channel is given as

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

Above equation shows that the channel capacity depends on two factors :

1. Bandwidth (B) of the channel
2. Signal-to-Noise ratio $\left(\frac{S}{N}\right)$

Noiseless channel has infinite capacity

If there is no noise in the channel then $N=0$. Hence $\frac{S}{N}=\infty$ such channel is called noiseless channel. Then capacity of such channel will be

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

$$= \infty$$

Thus noiseless channel has infinite capacity. However practically N is always finite and therefore the channel capacity is finite.

Infinite Bandwidth channel has limited capacity

Now if Bandwidth 'B' is infinite, the channel capacity is limited. This is because bandwidth increases, noise power (N) also increases. Noise power is given by

$$N = N_0B \quad (\text{or}) \quad \eta B$$

Due to increase in noise power, (S/N) signal to noise ratio decreases. Hence even if 'B' approaches infinity, capacity does not approach infinity. As $B\to\infty$, capacity approaches an upper limit known as "shannons limit".

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

$$C = \frac{S}{\eta}\cdot\frac{\eta B}{S}\log\left(1+\frac{S}{\eta B}\right) \quad \text{— (1)}$$

If $x = \frac{S}{\eta B}$, then B approaches infinity, x approaches zero

we know $\underset{x\to 0}{Lt}\ \frac{1}{x}\log(1+x) = \log_2 e = 1.44$ — (2)

Comparing eq ③ with ① and substitute the value then we get

$$\lim_{B\to\infty} C = \frac{S}{\eta}\log_2 e$$

$$= 1.44\frac{S}{\eta}$$

Let us consider the trade off between the bandwidth and S/N ratio

Let $\frac{S}{N} = 15$ and $B = 5kHz$

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

$$= 5k\log_2(1+15) = 5000\log_2 16$$

$$= 5000 \times 4 = 20k \text{ bits/sec.}$$

Then, if the S/N ratio is increased to 31, the BW for the same channel capacity can be found from

$$C = 20K = B\log_2\left(1+\frac{S}{N}\right)$$

$$20k = B\log_2(1+31) = B\log_2 32$$

$$20k = B \times 5$$

$$B = 4kHz$$

Therefore 20% reduction in BW (5kHz to 4kHz) requires a 65% increase in the signal power.

Thus, to decrease the Bandwidth, the signal power has to be increased. Similarly, to decrease the signal power the Bandwidth must be increased.

$\frac{S_1}{N_1} = 15 \rightarrow S_1 = 15N_0B_1 = 75N_0 - 100$
$\frac{S_2}{N_2} = 31 \rightarrow S_2 = 31N_0B_2 = 124N_0 - 165$ $\}$ 65%

$5 \rightarrow 100$
$4 \rightarrow \frac{4\times 100}{5} = 80$ $\}$ $\rightarrow 20\%$

→ A Gaussian channel has 1 MHz Bandwidth. Calculate the channel capacity if the signal power to noise spectral density ratio (S/η) is 10^5 Hz. Also find the maximum information rate.

Sol Given: Bandwidth (B) = 1 MHz

$S/\eta = 10^5$ Hz.

$$C = B \log_2 \left(1 + \frac{S}{N}\right) = B \log_2 \left(1 + \frac{S}{N_0 B}\right)$$

$$= 10^6 \log_2 \left(1 + \frac{10^5}{10^6}\right)$$

$$= 13,800 \text{ bits/sec.}$$

The maximum information rate is $R_{max} = 1.44 \frac{S}{\eta}$

$$= 144 \times 10^5 = 1,44,000 \text{ bits/sec.}$$

→ For an AWGN channel with 4 KHz bandwidth and noise power spectral density $\frac{N_0}{2} = 10^{-12}$ W/Hz, the signal power required at the receiver is 0.1 mW. Calculate capacity of this channel.

Sol Given $B = 4 \times 10^3$ Hz

$S = 0.1$ mW

$\frac{N_0}{2} = 10^{-12}$ W/Hz.

$N = N_0 . B$

$= 10^{-12} \times 2 \times 4 \times 10^3$

$= 8 \times 10^{-9}$ W.

The channel capacity $C = B \log_2 \left(1 + \frac{S}{N_0 B}\right)$

$$= 4000 \log_2 \left(1 + \frac{0.1 \times 10^{-3}}{8 \times 10^{-9}}\right)$$

$$= 4000 \log_2 (12501)$$

$$= 54.44 \text{ kbits/sec.}$$

→ 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 output of this source transmitted without error over an AWGN channel with a bandwidth 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 error-free transmission of the output of this source if S/N ratio is 20dB.

Sol: Given: $B = 4KHz$

Nyquist rate $= 2B = 2 \times 4k = 8KHz$

$r = 1.25 \times$ Nyquist rate

$= 1.25 \times 8000 = 10,000$ messages/second.

Since the samples are quantized into 256 equally likely levels, there will be $M = 256$ samples. Each sample will have a probability of occurrence as $\frac{1}{256}$. The entropy for such messages is given i.e

$$H = \log_2 M$$

$$= \log_2 256 = 8 \text{ bits/symbol}$$

(i) To obtain the Information rate (R):

$$R = rH$$

$$= 10,000 \times 8 = 8000 \text{ bits/second.}$$

(ii) To check for error-free transmission of $B = 10K\,Hz$ & $\frac{S}{N} = 20dB$

Giventhat the signal to noise ratio is 20 dB

$$\left(\frac{S}{N}\right)_{dB} = 10 \log \frac{S}{N}$$

$$20 = 10 \log_{10} \frac{S}{N}$$

$$\frac{S}{N} = 10^2 = 100$$

The capacity of AWGN channel is $C = B \log(1 + \frac{S}{N})$

$$= 10 \times K \log(1 + 100)$$

$$= 10000 \frac{\log(101)}{\log 2}$$

$$= 66.582 \text{ kbits/sec}$$

If $R \leq C$, then error-free transmission is possible. Here $R = 80,000$ & $C = 66.582$ kbits/sec. Hence $R > C$. Therefore error-free transmission is not possible.

(iii) $\frac{S}{N}$ ratio for error-free transmission:

We have $R = 80000$ & $B = 10000$. For error-free transmission $R \leq C$

Hence
$$R \leq C = B \log(1 + \frac{S}{N})$$

$$R \leq B \log_2(1 + \frac{S}{N})$$

$$80000 \leq 10000 \log_2(1 + \frac{S}{N})$$

$$8 \leq \log_2(1 + \frac{S}{N}) \quad \text{(or)} \quad 1 + \frac{S}{N} = 2^8 = 256$$

$$\frac{S}{N} = 255$$

$$8 \leq \frac{\log(1 + \frac{S}{N})}{\log 2} \Rightarrow$$

$$2.40824 \leq \log(1 + \frac{S}{N}) \Rightarrow \frac{S}{N} \geq 255 \quad \text{(or)} \quad \left(\frac{S}{N}\right)_{dB} = 10 \log 255$$

$$= 24 \text{ dB}$$

(iv) To determine Bandwidth:

The given $\frac{S}{N}$ ratio is 20dB. Hence,

$$\left(\frac{S}{N}\right)_{dB} = 10 \log \frac{S}{N}.$$

$$20 = 10 \log \frac{S}{N}.$$

$$\frac{S}{N} = 100$$

For error-free transmission $R \leq C$. Hence

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

$$R \leq B \log_2 \left(1 + \frac{S}{N}\right)$$

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

$$\leq B \log_2 (101)$$

$$\frac{80\,000}{\log_2(101)} \leq B$$

$$\therefore B \geq 11.974 \text{ kHz}.$$

This is the Bandwidth required for error-free transmission.

→ Shannon's theorem on channel capacity

The information is transmitted through the channel with rate 'R' called information rate. Shannon's theorem says that it is possible to transmit information with an arbitrarily small probability of error provided that information rate 'R' is less than or equal to a rate 'C' called channel capacity.

Thus, channel capacity is the maximum information rate with which the error probability is within tolerable limits.

Statement of the theorem:

Given a source of M equally likely messages, with $M >> 1$, which is generating information at a Rate 'R' and with channel capacity 'C' then if

$$R \le C$$

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 arbitrarilly small.

Negative statement of channel coding theorem

Given a source of 'M' equally likely messages with $M >> 1$, which is generating information at a rate 'R'; then if

$$R > C$$

the probability of error is close to unity for every possible set of 'M' transmitter signals.

Thus, the negative statement of shannon's theorem says that if $R > C$, then every message will be in error.

→ Continous channel

A no. of communication systems use continous sources and thus use the channel continously. AM, FM, PM are examples of systems using continous channel.

If $p(x)$ is the probability density function associated with the signal $x(t)$, then the entropy of the source is

$$H(X) = E[-\log p(x)]$$

$$H(x) = -\int_{-\infty}^{\infty} p(x)\log p(x)\,dx.$$ "This is called differential entropy of X."

In a similar way, the different entropies associated with two-dimensional random variables with joint density $P(x,y)$, conditional densities $P(x/y)$, $P(y/x)$ and marginal densities $P_1(x)$ & $P_2(y)$ may be defined as

$$H(x) = E[-\log P_1(x)] = -\int_{-\infty}^{\infty} P_1(x)\log P_1(x)\,dx$$

$$H(y) = E[-\log P_2(y)] = -\int_{-\infty}^{\infty} P_2(y)\log P_2(y)\,dy$$

$$H(x,y) = E[-\log P(x,y)] = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)\log P(x,y)\,dx.dy$$

$$H(x/y) = E[-\log P(x/y)] = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)\log P(x/y)\,dx.dy$$

$$H(y/x) = E[-\log P(y/x)] = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)\log P(y/x)\,dx.dy$$

In continous case, $\int_{-\infty}^{\infty} P(x)\,dx = 1$

and $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} P(x,y)\,dx.dy = 1$

Hence, the density functions $p(x)$ and $p(x,y)$ need not be less than 1 for all values of the random variable. This may lead to a negative entropy.