Module-7

Introduction to Information Theory

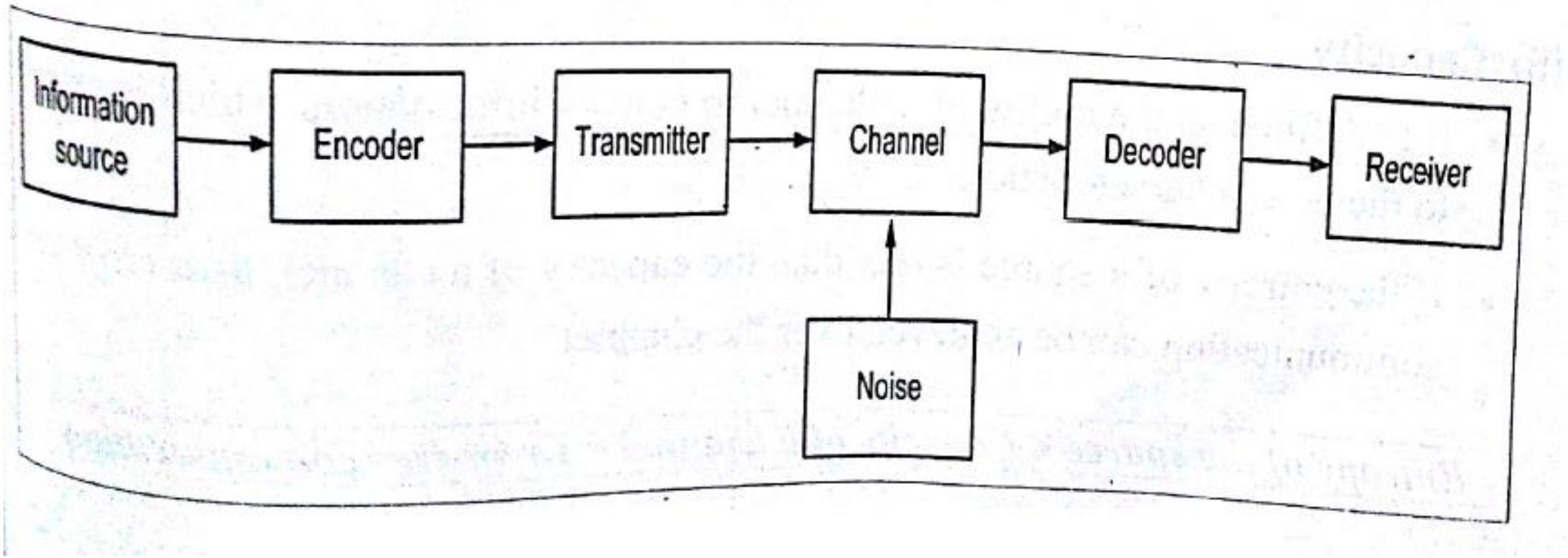
Topics to be discussed

- Entropy
- Mutual information and channel capacity theorem
- Error Correction Codes

Fundamentals of information theory

- Information theory is a branch of probability, which can be applied to the study of communication systems.
- In general, communication information is statistical in nature and the main aim of information theory is to study the simple ideal statistical communication models.
- Information theory deals with "***mathematical modeling*** and ***analysis*** of a communication system rather than with physical sources and physical channels".

Block diagram of an information system



Condition for Error free communication

(i) Entropy

- It is defined in terms of a ***"probabilistic behaviors"*** of a source of information.

(ii) Capacity

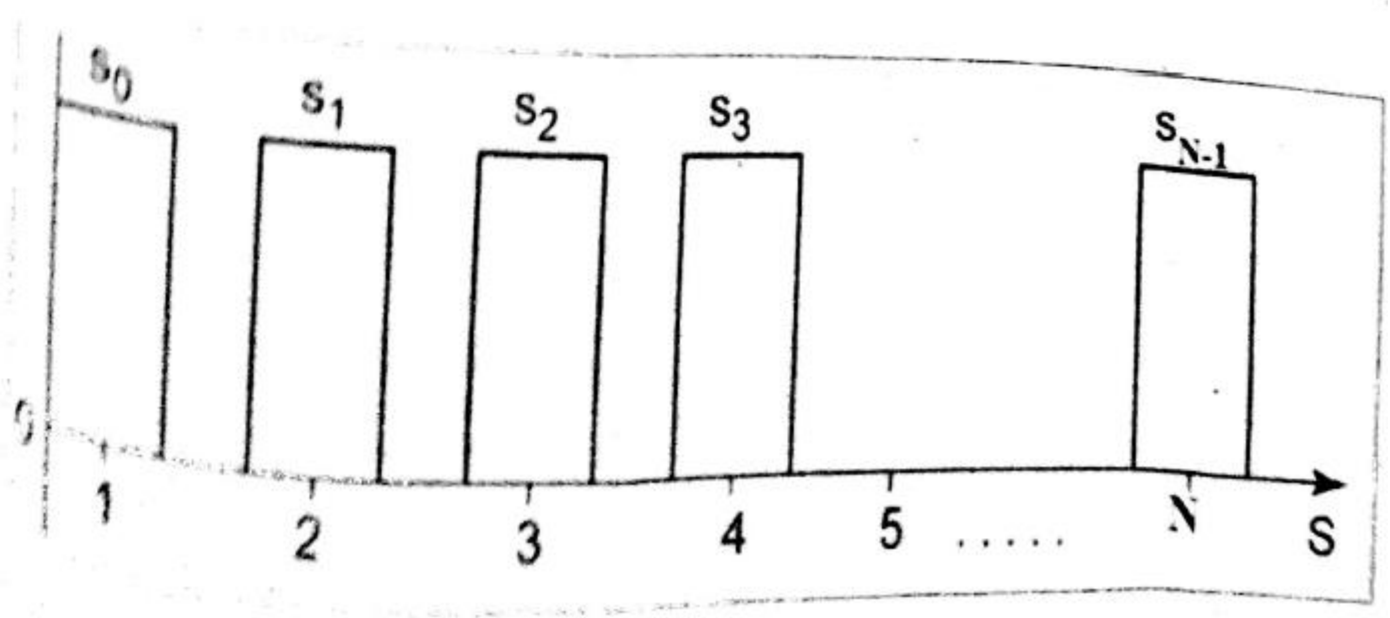
- It is defined as the ability of a channel to convey information, which is related to the noise characteristics.
- If the entropy of a source is ***less*** than the capacity of a channel, then error-free communication can be achieved over the channel.

Entropy of the source < Capacity of a channel = Error-free communication

Discrete message

- The output emitted by a source during every unit of time is called discrete messages

$$S = \{s_0, s_1, s_2, \ldots\ldots s_{N-1}\}$$



Amount of information

- The amount of information or messages transmitted over a channel is defined in statistical terms such as probability of occurrence.
- If the ***probability of occurrence*** of an event is ***more***, there will be a ***very less amount of information***; otherwise, if the ***probability of occurrence*** of an event is ***less***, then there will be ***more amount of information***.

$$I(s_i) = \log\left(\frac{1}{p_i}\right) \quad \text{for } i = 0, 1, 2; \ldots N-1$$

Properties of information

(i) *If we are absolutely certain on the outcome of an event, even before it occurs, there is* ***no information gained****.*

$$I(s_i) = 0 \qquad \textit{for } p_i = 1$$

Proof:

$$I(s_i) = \log_2 \frac{1}{p_i}$$

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

$$= \log_2 1 = 0$$

Properties of information

(ii) ***Non negative quantity, that is, $I(s_i) \geq 0$ for $0 \leq p_i \leq 1$***

This condition provides some or no information, but never brings about a loss of information.

(iii) $I(x_i) > I(y_i)$ *or* $p(x_i) < p(y_i)$

The less probable an event has the more information.

(iv) $I(x_i, y_i) = I(x_i) + I(y_i)$

If x_i and y_i are statistically independent.

Entropy

- An average information per individual message or symbol is called Entropy

*The **entropy** of a source is defined as the source which produces **average information** per **individual message** or symbol in a particular interval. It is also called as comentropy.*

$$H(S) = E[I(s_i)]$$

$$= \sum_{i=0}^{N-1} p_i \, I(s_i)$$

$$\text{Entropy } H(S) = \sum_{i=0}^{N-1} p_i \log_2\left(\frac{1}{p_i}\right)$$

Properties of Entropy

The entropy of a discrete memoryless source is bounded as,

$$0 \leq H \leq \log_2 N$$

where, N is the number of symbols of the alphabet S of the source.

Properties of Entropy

(1) ***Entropy is zero, if the event is sure or it is impossible. This lower bound on entropy corresponds to no uncertainty.***

i.e., $H(S) = 0$ if $P_i = 0$ (or) $P_i = 1$

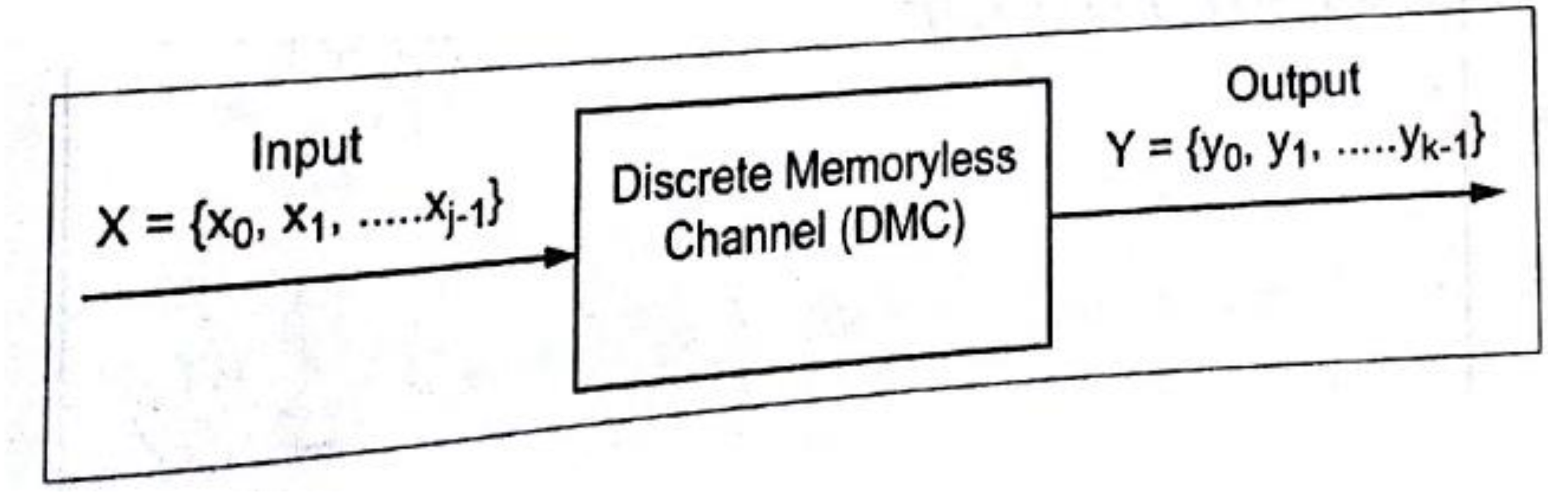
(2) ***Entropy $H(S) = \log_2 N$, when all the N symbols are equally likely in the alphabet S,*** that is, $P_i = \frac{1}{N}$. This upper bound on entropy corresponds to maximum uncertainty.

Rate of Information

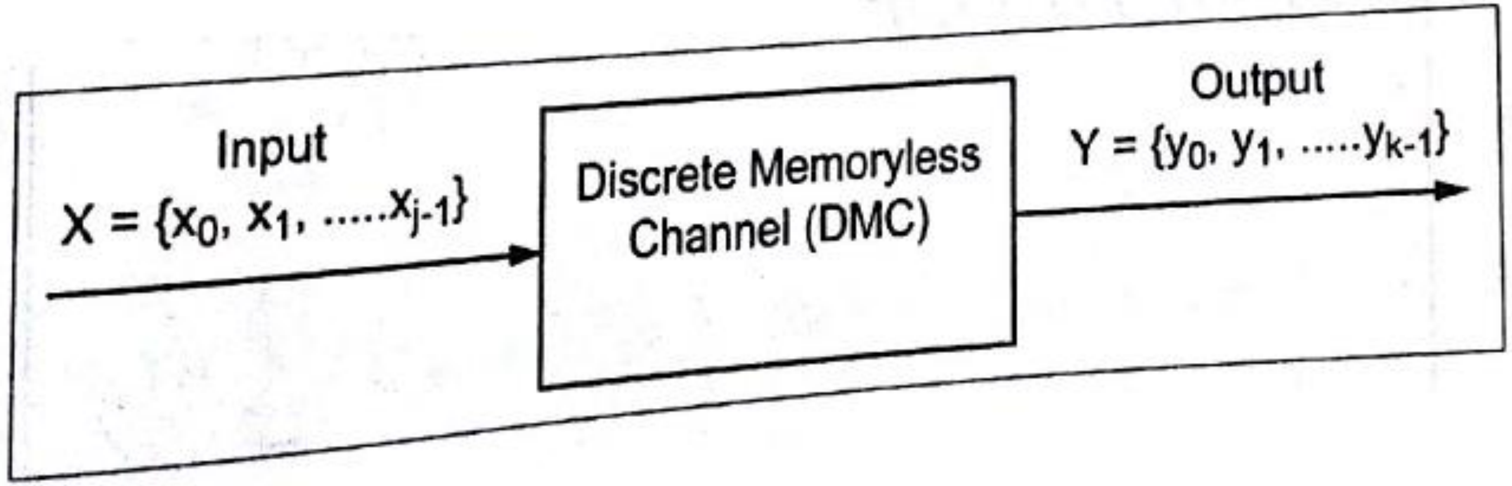
The rate of information (R) is defined as "the average number of bits of information per second"

$$R = r\,H(S)$$

Discrete Memoryless Channel



Discrete Memoryless Channel



- $P(y_k/x_j)$ is the conditional probability of obtaining output y_k given that the input is x_j and is called as a ***channel transition probability.***
- If $k = j$ then $P(y_k/x_j)$ represents a conditional probability of *correct reception.* If $k \neq j$ then $P(y_k/x_j)$ represents a conditional probability of error.

The Channel Matrix

$$
\mathbf{P} = \begin{bmatrix}
p(y_0/x_0) & p(y_1/x_0) & p(y_2/x_0) & \cdots\cdots & p(y_{K-1}/x_0) \\
p(y_0/x_1) & p(y_1/x_1) & p(y_2/x_1) & \cdots\cdots & p(y_{K-1}/x_1) \\
p(y_0/x_2) & p(y_1/x_2) & p(y_2/x_2) & \cdots\cdots & p(y_{K-1}/x_2) \\
\vdots & \vdots & & & \\
\vdots & \vdots & & & \\
p(y_0/x_{J-1}) & p(y_1/x_{J-1}) & p(y_2/x_{J-1}) & \cdots\cdots & p(y_{K-1}/x_{J-1})
\end{bmatrix}
$$

The Channel Matrix

From the probability theory,

$$P(XY) = P(Y/X)\,P(X)$$

That is, the Joint Probability of X and Y is given as,

$$P(x_j, y_k) = P(X = x_j, Y = y_k)$$

$$= P(Y = y_k / X = x_j)\,P(X = x_j)$$

$$P(x_j, y_k) = P(y_k / x_j)\,P(x_j)$$

The Channel Matrix

- Hence, the marginal probability distribution of output random variable Y is given as,

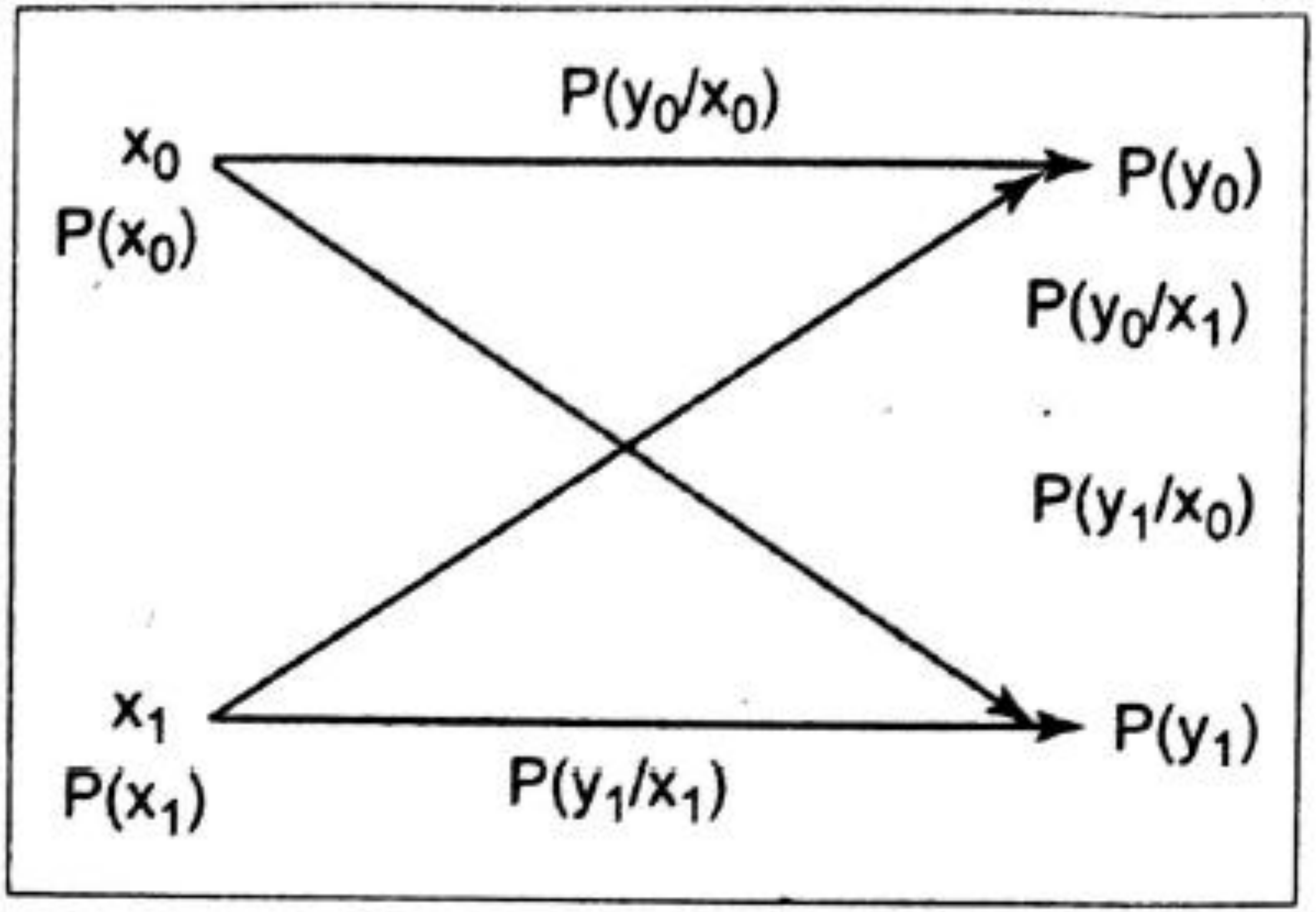
$$P(y_k) = P(Y = y_k)$$

$$= \sum_{j=0}^{J-1} P(Y = y_k / X = x_j) P(X = x_j)$$

$$P(y_k) = \sum_{j=0}^{J-1} P(y_k / x_j) P(x_j) \quad \text{for } k = 0, 1, 2 \ldots\ldots K-1$$

Dr. T.Ilavarasan, Associate Professor,SENSE, VIT

Binary communication channel



Binary communication channel

- The probabilities of y_0 and y_1 can be written as,

$$P(y_0) = P(y_0/x_0)\,P(x_0) + P(y_0/x_1)\,P(x_1) \quad \text{and,}$$

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

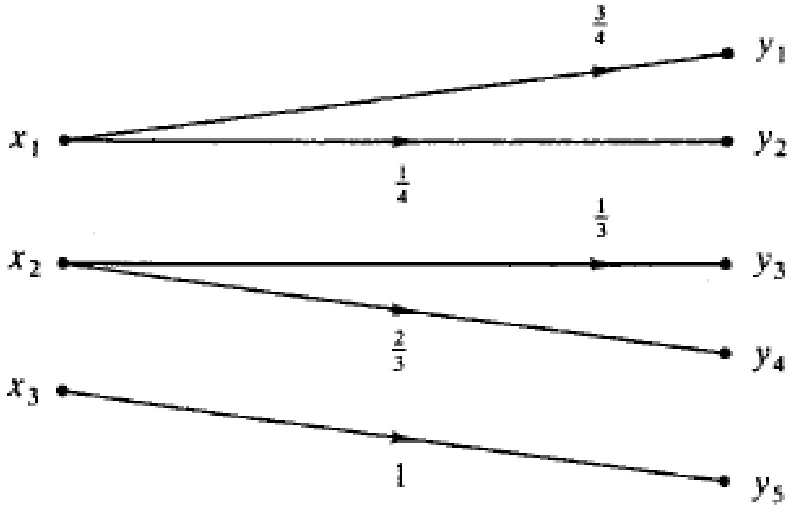
- It can be represented using a matrix as,

$$\begin{bmatrix} P(y_0) \\ P(y_1) \end{bmatrix} = P(x_0)P(x_1)\begin{bmatrix} P(y_0/x_0) & P(y_1/x_0) \\ P(y_0/x_1) & P(y_1/x_1) \end{bmatrix}$$

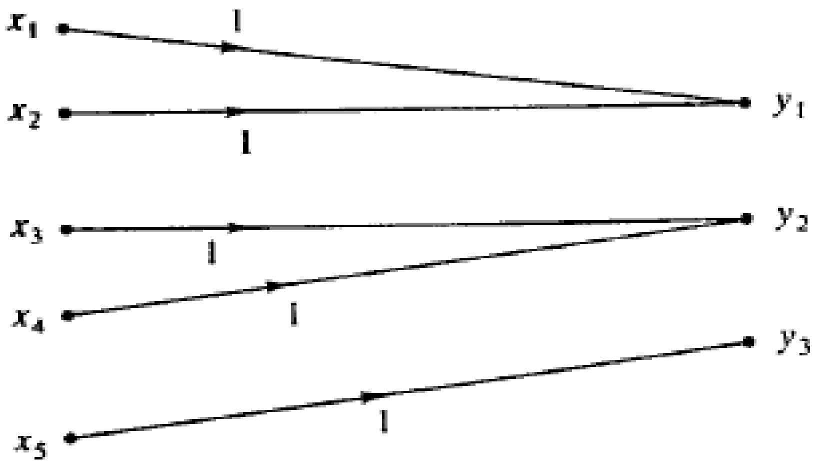
Lossless channel

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

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



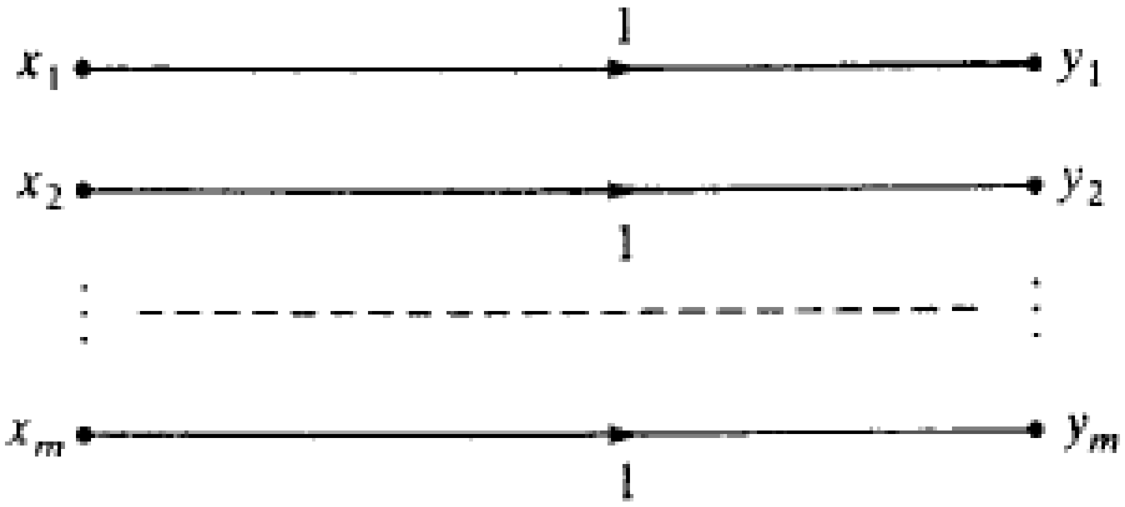
Deterministic channel



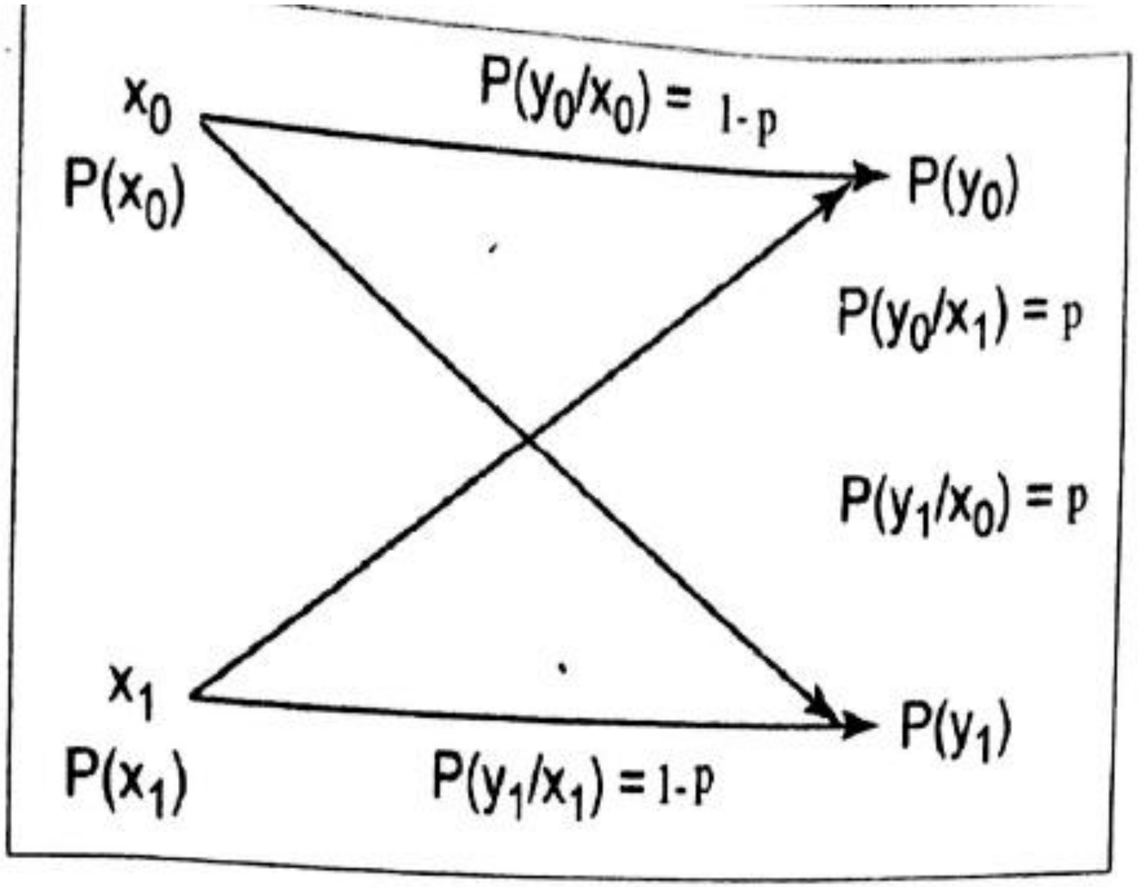
$$[P(Y|X)] = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Noiseless channel

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



Binary symmetric channel



Binary symmetric channel

- Binary Communication Channel is said to be symmetric if,

$$P(y_0/x_0) = P(y_1/x_1) = 1-p$$

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

Mutual Information

- Mutual Information is simply defined as "the difference between the two values H(X) – H(X/Y) - represents our uncertainty about the channel input that is resolved by observing the channel output".

$$I(X;Y) = \text{Initial uncertainty} - \text{Final uncertainty}$$
$$= H(X) - H(X/Y)$$

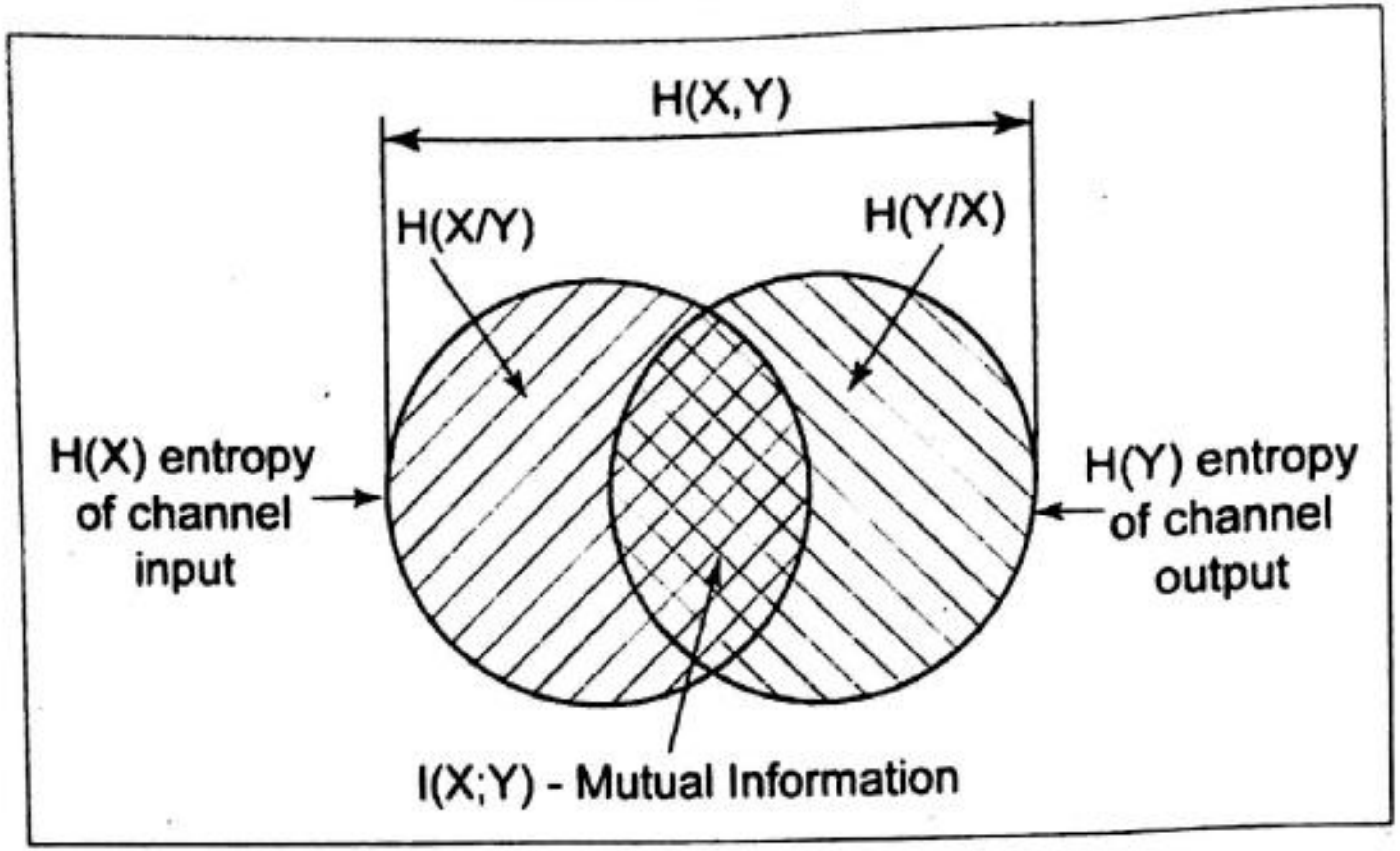
- The quantity H(X/Y) is called ***conditional entropy***. It represents the amount of uncertainty about the channel input X after the channel output Y and has been observed (known).

$$H(X/Y) = \sum_{j=0}^{J} \sum_{k=0}^{K} P(x_j, y_k) \log_2 \left(\frac{1}{P(x_j / y_k)} \right)$$

Joint Probability of X and Y is given as,

$$P(x_j, y_k) = P(x_j / y_k) \, P(y_k)$$

Mutual Information



Mutual Information

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

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

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

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

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

Properties of Mutual Information

Properties of $I(X;Y)$:

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

2. $$I(X;Y) \geqslant 0$$

3. $$I(X;Y) = H(Y) - H(Y|X)$$

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

Channel capacity

- Consider a discrete memoryless channel (DMC) with an input alphabet X, an output alphabet Y, and transition probabilities $P(y_k / x_j)$, where $j = 0, 1, 2 \ldots.. J - 1$ and $k = 0, 1, 2 \ldots\ldots K - 1$.

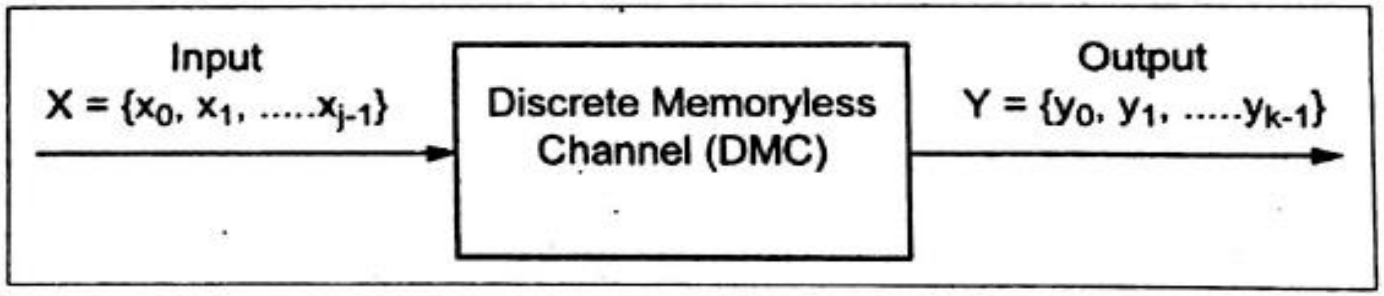


Figure 21.5. Discrete Memoryless Channel

- The mutual information of the channel is expressed as

$$I(X;Y) = \sum_{k=o}^{K-1} \sum_{j=0}^{J-1} P(x_j, y_k) \log_2 \left[\frac{P(x_j / y_k)}{P(x_j)} \right]$$

Channel capacity

- The mutual information I(X; Y) indicates *a measure of the average information per symbol transmitted in the system.* Shannon has introduced a significant concept of channel capacity (C), this defined as the ***maximum of mutual information***.

$$C = \max_{\{P(x_j)\}} I(X;Y) = \max\,[H(X) - H(X/Y)]$$

- The channel capacity C is measured in ***bits per channel use*** or ***bits per transmission***.

Channel Efficiency

- The transmission efficiency or channel efficiency is defined *as*,

$$\eta = \frac{\text{Mutual information}}{\text{Maximum mutual information}}$$

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

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

Redundancy

- The redundancy of the channel is defined as,

$$R = 1 - \eta$$

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

Noise free channel

- The Mutual Information for a noise free channel is given as,

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

- Therefore, the channel capacity here is calculated as,

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

- We know that, from property of Entropy, max $H(X) = \log_2 N$ bits / message, where, N – total number of messages. Hence, the Channel Capacity for a noise – free channel is,

$$\mathbf{C = \log_2 N \text{ bits / Symbol}}$$