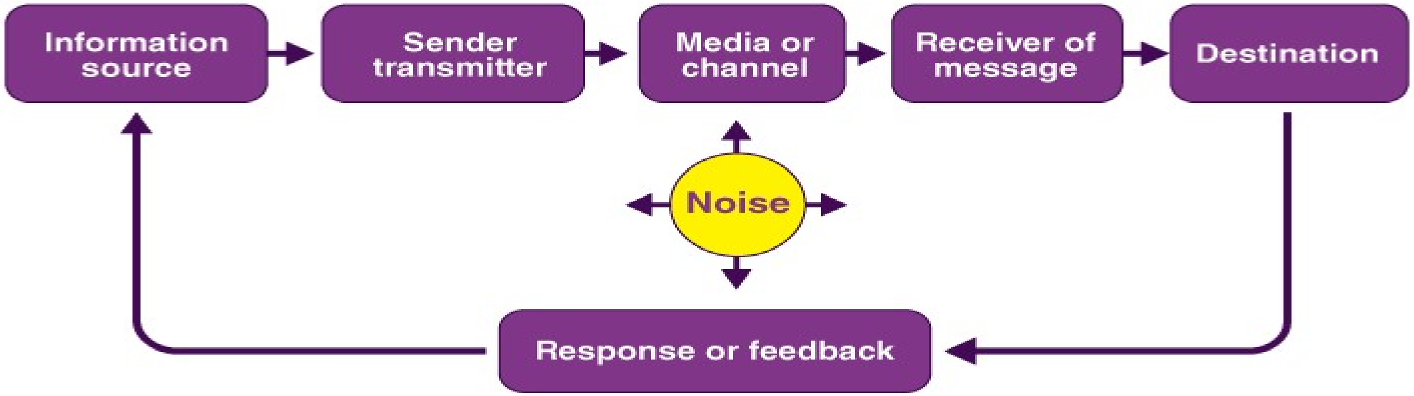
UNIT – IV

Information Theory and Source Coding: Discrete message and information content, Concept of amount of Information- Average Information, Entropy, Information Rate, Mutual Information and its properties; Source Coding to increase Average Information per bit-Source coding theorem, Shannon-Fano Coding, Huffman Coding; Channel Capacity of Gaussian Channel-Band width-S/N trade off.

1.Information is the source of a communication system, whether it is analog or digital.

2.Information theory is a mathematical approach to the study of coding of information along with the quantification, storage, and communication of information. Information theory examines the utilisation, processing, transmission and extraction of information. *The information communication over noisy channels, this theoretical concept was formalised by Claude Shannon (in his work called A Mathematical Theory of Communication) in 1948*. In his work, information is considered as a group of possible messages. The main goal is to transfer these messages over noisy channels and to have the receiving device redevelop the message with negligible error probability (regardless of the channel noise).



Information theory is a mathematical approach to the study of coding of information along with the quantification, storage, and communication of information.

Conditions of Occurrence of Events

- *If we consider an event, there are three conditions of occurrence.*
- *If the event has not occurred, there is a condition of uncertainty.*
- *If the event has just occurred, there is a condition of surprise.*
- *If the event has occurred, a time back, there is a condition of having some information.*
- *These three events occur at different times. The differences in these conditions help us gain knowledge on the probabilities of the occurrence of events*

The performance of the communication system is measured in terms of its error probability. An errorless transmission is possible when probability of error at the receiver approaches zero.

- Consider a discrete memoryless source with source alphabet $\mathcal{S} = \{s_0, s_1, s_2\}$, whose three distinct symbols have the following probabilities $\{ p_0=1/4, p_1=1/4, p_2=1/2\}$
- The entropy of the discrete random variable S representing the source as

$$H(S) = p_0 \log_2\left(\frac{1}{p_0}\right) + p_1 \log_2\left(\frac{1}{p_1}\right) + p_2 \log_2\left(\frac{1}{p_2}\right)$$

$$= \frac{1}{4} \log_2(4) + \frac{1}{4} \log_2(4) + \frac{1}{2} \log_2(2) \quad = \frac{3}{2} \text{ bits}$$

(3): Consider a Telegraph Source having 2 Symbols • (dot) & — (dash). (8)

The dot duration is 0.2 Sec & dash duration is 3 times of dot duration.

The Probability of dot is 2 times of dash Probability, and time b/n

Two Symbols is 0.2 Sec. Calculate R of The Telegraph Source.

Sol:-

$$p(\text{dash}) = P$$

$$p(\text{dot}) = 2\,p(\text{dash}) = 2P$$

$$p(\text{dot}) + p(\text{dash}) = 1 \Rightarrow P + 2P = 1 \Rightarrow P = 1/3$$

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

Entropy:- $$H = \sum_{k=1}^{2} P_k \log_2(1/P_k) = \frac{1}{3}\log_2 3 + \frac{2}{3}\log(3/2) = 0.9183 \text{ b/symbols}$$

$\therefore$ $T_{dot} = 0.2$ Sec, $T_{dash} = 3(T_{dot}) = 0.6$ Sec

Differance b/n dot & dash = 0.2 Sec. = T_{gaurd}

$\therefore$ Total Symbol / avg Symbol duration is

$$T_s = P_{dot} T_{dot} + P_{dash} T_{dash} + T_{guard}$$

$$T_s = \frac{1}{3} 0.2 + \frac{2}{3} 0.6 + 0.2 = 0.5333 \text{ Sec/symbol}$$

$\therefore$ Source Rate $r = \frac{1}{T_s} = \frac{1}{0.5333} = 1.875$ Symbols / sec

Information Rate $R = r.H = (1.875)(0.9183) = 1.725$ bps

Joint entropy

There is probability p(x), p(y) and joint probability p(x,y), we can define entropy H(X) and joint entropy H(X,Y).

Definition: Joint Entropy

If X and Y are discrete random variables and $p(x, y)$ is the value of their joint probability distribution at (x, y), then the joint entropy of X and Y is:

$$H(X, Y) = -\sum_{x \in X} \sum_{y \in Y} p(x,y) \log_2(1/P(x,y))$$

Conditional entropy

Suppose we have two random variables X and Y. Suppose that we know the outcome of X, and the question is, how much entropy is "left" in Y

$$P(x/y) = \frac{P(xy)}{P(y)} \quad (\text{or}) \quad P(y/x) = \frac{P(xy)}{P(x)}.$$

$$H(x/y_k) = -\sum_{j=1}^{m} \frac{P(x_j y_k)}{P(y_k)} \cdot \log_2 \frac{P(x_j, y_k)}{P(y_k)}$$

$H(x/y) =$ avg/mean value of $H(x/y_k)$ Then

$$H(x/y) = \sum_{k=1}^{n} H(x/y_k)\, P(y_k)$$

$$H(X/Y) = \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_j, y_k) \log_2 \frac{1}{P(x_j/y_k)}$$

$$H(Y/X) = \sum_{j=1}^{m} \sum_{k=1}^{n} P(x_j, y_k) \log_2 \frac{1}{P(y_k/x_j)}$$

Source Coding – eliminate redundancy in the data, send same information in fewer bits

- Goal is to find an efficient description of information sources
  - Reduce required bandwidth
  - Reduce memory to store
- *Memoryless* –If symbols from source are independent, one symbol does not depend on next
- *Memory* – elements of a sequence depend on one another, e.g. UNIVERSIT_?, 10-tuple contains less information since dependent

$$H(X)_{memory} < H(X)_{no\ memory}$$

- This means that it's more efficient to code information

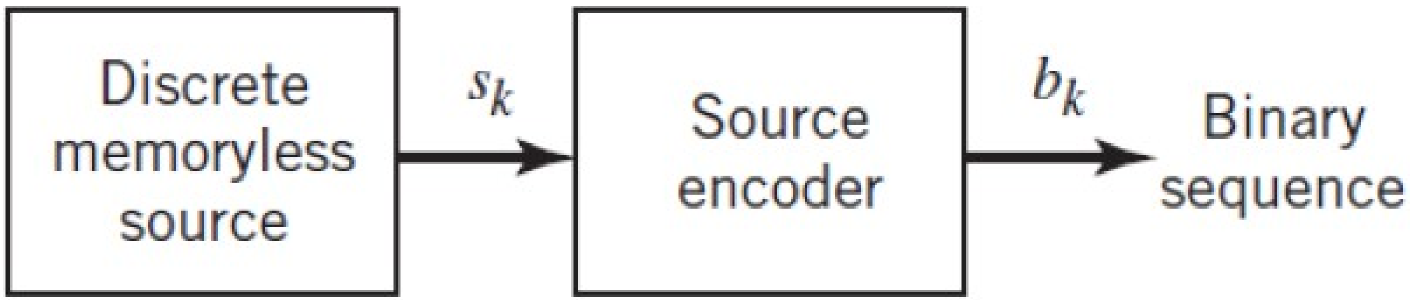
With memory as group of symbols

- 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.

- There are few terms related to source coding process:
  i. Code word length
  ii. Average code word length
  iii. Code efficiency
  iv. Code redundancy

Source-coding Theorem

- A discrete memoryless source whose output s_k is converted by the source encoder into a sequence of 0's and 1's, denoted by b_k. We assume that the source has an alphabet with K different symbols and that the kth symbol s_k occurs with probability p_k, $k = 0, 1, \ldots, K-1$.



- Let the binary codeword assigned to symbol s_k by the encoder have length l_k, measured in bits. We define the average codeword length L of the source encoder as represents the average number of bits per source symbol used in the source encoding process

$$\bar{L} = \sum_{k=0}^{K-1} p_k l_k$$

- We then define the coding efficiency of the source encoder as

$$\eta = \frac{L_{\min}}{\bar{L}}$$

- The entropy $H(S)$ represents a fundamental limit on the average number of bits per source symbol necessary to represent a discrete memoryless source, in that it can be made as small as but no smaller than the entropy $H(S)$.
- Thus, setting $L_{min} = H(S)$, defining the efficiency of a source encoder in terms of the entropy $H(S)$ as shown by

$$\eta = \frac{H(S)}{\bar{L}}$$

Code Word Length: Let X be a DMS with finite entropy H(X) and an alphabet $\{x_1, \ldots. x_m\}$ with corresponding probabilities of occurrence $P(x_i)$ $(i = 1,\ldots,m)$.

- Let the binary code word assigned to symbol x_i by the encoder have length n_i, measured in bits.
- The length of a code word is the number of binary digits in th code word.

Average Code Word Length: The average code word length L, per source symbol is given by

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

Code Efficiency: The code efficiency η is defined as under:

$$\eta = \frac{L_{min}}{L}$$

Code Redundancy: The code redundancy γ is defined as $\gamma = 1 - \eta$

SOURCE CODING THEOREM

- The source coding theorem states that for a DMS X, with entropy H(X), the average code word length L per symbol is bounded as $L \geq H(X)$

- Further, L can be made as close to H(X) as desired for some suitably chosen code.

- Thus, with $L_{min} = H(X)$, the code efficiency can be rewritten as $\eta = \frac{H(X)}{L}$

- **Classification of Codes:** Classification of codes is best illustrated by an example.
  1. Fixed-Length Codes
  2. Variable-Length Codes
  3. Distinct codes
  4. Prefix-Free Codes

Table-Binary codes

x_i	Code 1	Code 2	Code 3	Code 4	Code 5	Code 6
x_1	00	00	0	0	0	1
x_2	01	01	1	10	01	01
x_3	00	10	00	110	011	001
x_4	11	11	11	111	0111	0001

① Fixed length coding :-

(i). S_0, S_1, S_2, S_3 with $P_1 = 1/2$, $P_2 = 1/4$, $P_3 = 1/8$, $P_4 = 1/8$.

$$\therefore H(S) = \sum_{K=0}^{3} P_K \log_2\left(\frac{1}{P_K}\right) = \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \frac{1}{8}\log_2 8 + \log_2 8 \, \frac{1}{8}$$

$$\text{avg } H(S) = \frac{1}{2} + \frac{1}{4}(2) + \frac{1}{8}(3) + \frac{1}{8}(3) \Rightarrow 1 + \frac{3}{4} \Rightarrow 7/4 \text{ b/m}$$

$\therefore$

S_K ↓		P_K ↓	b_K ↓	l_K
S_0	⇒	1/2	00	= 2
S_1	⇒	1/4	01	= 2
S_2	⇒	1/8	10	= 2
S_3	⇒	1/8	11	= 2

$$\bar{L} = \sum_{K=0}^{3} P_K l_K = 2\left[\frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{8}\right]$$

$$\bar{L} = 2 \text{ b/m}$$

$$\eta = \frac{7/4}{2} = 87.5\%$$

$$\gamma \simeq 13\%$$

(2) Variable length coding :-

$H(S) = 7/4$ b/m

S_k	P_k	b_k	l_k
S_0 →	$1/2$	0	1
S_1 →	$1/4$	10	2
S_2 →	$1/8$	110	3
S_3 →	$1/8$	111	3

$$\bar{L} = \sum_{k=0}^{3} P_k l_k = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{8} + 3 \times \frac{1}{8} \Rightarrow 1 + 3/4 = 7/4 \text{ b/m}$$

$$\therefore \eta = \frac{7/4}{7/4} = 100\% \qquad \gamma = 0\%$$

Clearly observe That due to Variable length Source coder give high Coding efficiency & low redundancy.

Shannon Fano Algorithm is an entropy encoding technique for lossless data compression of multimedia.

1. The messages are first written in the order of decreasing probability.
2. Then divide the messages set into two most equiprobable subset X and Y.
3. The message of 1st set X is given bit 0 and message in the 2nd subset is given bit 1.
4. The procedure is now applied for each set seprately till end.
5. Finally we get the code word for respective symbol.
6. Calculation

Example – Find the code words occuring in the probability $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$ for symbols s_1, s_2, s_3 & s_4. Find efficiency and redundancy of code.

Symbol	Prob	Codeword	Length
s_1	1/2	0	1
s_2	1/4	10	2
s_3	1/8	110	3
s_4	1/8	111	3

where H = entropy $= \sum_{i=1} p_i \log_2(p_i)$

$$= \frac{1}{2}\log_2 2 + \frac{1}{4}\log_2 4 + \left(\frac{1}{8}\log_2 8\right) \times 2$$

$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{4}$$

$$= 1.75 \text{ bits/symbol}$$

(2) Variable length coding :-

$H(S) = 7/4$ b/m

s_k	P_k	b_k	l_k
$s_0 \rightarrow$	$1/2$	0	1
$s_1 \rightarrow$	$1/4$	10	2
$s_2 \rightarrow$	$1/8$	110	3
$s_3 \rightarrow$	$1/8$	111	3

$$\bar{L} = \sum_{k=0}^{3} P_k l_k = 1\times\frac{1}{2} + 2\times\frac{1}{4} + 3\times\frac{1}{8} + 3\times\frac{1}{8} \Rightarrow 1 + 3/4 = 7/4 \text{ b/m}$$

$$\therefore \eta = \frac{7/4}{7/4} = 100\% \qquad \gamma = 0\%$$

Clearly observe that due to variable length source codes give high coding efficiency & low redundancy.

Ex:- Construct the codewords for the messages $S_0, S_1, S_2, S_3, S_4, S_5, S_6, S_7$ with Probabilities $1/2, 1/8, 1/8, 1/16, 1/16, 1/16, 1/32, 1/32$ respectively using Shannon-Fano coding.

Sol:-

message S	Prob P_k	I	II	III	IV	V	No. of b/m l_k	Codeword b_k
S_0	$1/2$	0					1	0
S_1	$1/8$	1	0	0			3	100
S_2	$1/8$	1	0	1			3	101
S_3	$1/16$	1	1	0	0		4	1100
S_4	$1/16$	1	1	0	1		4	1101
S_5	$1/16$	1	1	1	0		4	1110
S_6	$1/32$	1	1	1	1	0	5	11110
S_7	$1/32$	1	1	1	1	1	5	11111

Sol:-

message S.	Prob P_K	I	II	III	IV	V	No. of b/m l_K	Codeword b_K
S_0	1/2	0					1	0
S_1	1/8	1	0	0			3	100
S_2	1/8	1	0	1			3	101
S_3	1/16	1	1	0	0		4	1100
S_4	1/16	1	1	0	1		4	1101
S_5	1/16	1	1	1	0		4	1110
S_6	1/32	1	1	1	1	0	5	11110
S_7	[illegible]	1	1	1	1	1	5	11111

Sum: 1/2; Sum 8/32; Sum 8/32; 4/32; 4/32

$$H(s) = \sum_{K=0}^{K-1} P_K \log_2\left(\frac{1}{P_K}\right)$$

$$H(s) = P_0 \log_2\left(\frac{1}{P_0}\right) + \cdots\ P_7 \log_2\left(\frac{1}{P_7}\right)$$

$$H(s) = \frac{1}{2}\log_2 2 + \frac{1}{8}\log_2 8 + \frac{1}{8}\log_2 8 + \frac{1}{16}\log_2 16 + \frac{1}{16}\log_2 16 + \frac{1}{16}\log_2 16$$

$$+ \frac{1}{32}\log_2 32 + \frac{1}{32}\log_2 32$$

$$H(s) = \frac{1}{2} + \left(\frac{1}{8}3\right)2 + 3\frac{1}{16}4 + 2\frac{1}{32}5$$

$$H(s) = \frac{1}{2} + \frac{3}{4} + \frac{3}{4} + \frac{5}{16} = \frac{37}{16}\ \text{b/m.}$$

$$\bar{L} = \sum_{K=0}^{K-1} l_K P_K = 1\left(\frac{1}{2}\right) + 3\left(\frac{1}{8}\right) + 3\left(\frac{1}{8}\right) + 4\left(\frac{1}{16}\right) + 4\left(\frac{1}{16}\right) + 4\left(\frac{1}{16}\right) + 5\left(\frac{1}{32}\right) + 5\left(\frac{1}{32}\right)$$

$$\bar{L} = \frac{37}{16} \text{ b/m.}$$

$\therefore$ coding efficiency $= \eta = \dfrac{H(S)}{\bar{L}} \times 100 = \dfrac{37/16}{37/16} \times 100 = 100\%$

Redundancy $\gamma = 1 - \eta = 1 - 1 = 0\%$

Huffman Coding

1. The source symbols are arranged in order of decreasing prob. Then the two of lowest probability are assigned bit 0 and 1.
2. Then combine last two symbols and move the combined Symbol as high as possible.
3. Repeat the above step untill end.
4. Code for each symbol is found by moving backward.
5. Calculation

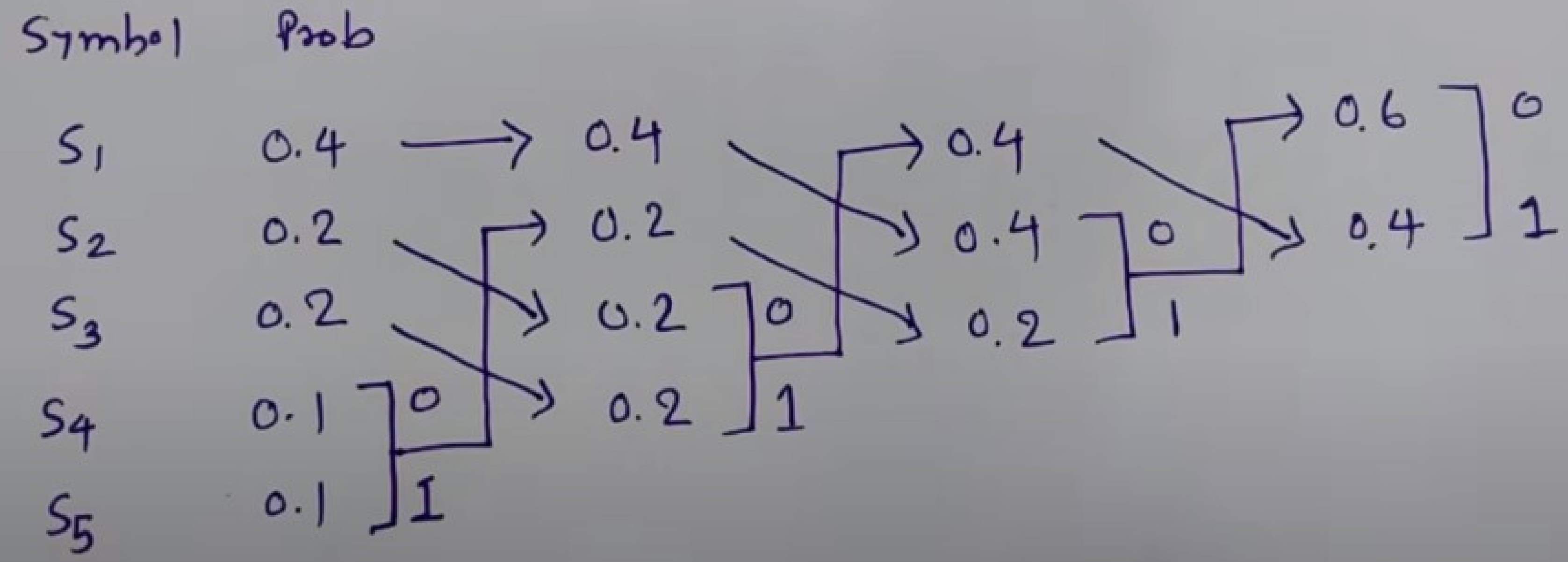
efficiency $\eta = \frac{H}{L \log_2 r}$

L = Avg Codeword $= \sum_{i=1}^{n} p_i n_i$

The Huffman coding is widely used in data compression techniques. It is used for both encoding and decoding.

- To be specific, the Huffman encoding algorithm proceeds as follows:
- 1. The source symbols are listed in order of decreasing probability. The two source symbols of lowest probability are assigned 0 and 1. This part of the step is referred to as the splitting stage.
- 2. These two source symbols are then combined into a new source symbol with probability equal to the sum of the two original probabilities. The probability of the new symbol is placed in the list in accordance with its value.
- 3. The procedure is repeated until we are left with a final list of source statistics (symbols) of only two for which the symbols 0 and 1 are assigned.

Example: Alphabet with prob = {0.4, 0.2, 0.2, 0.1, 0.1}
For symbols $\{s_1, s_2, \ldots, s_5\}$. Find Huffman codes and also Find efficiency & variance

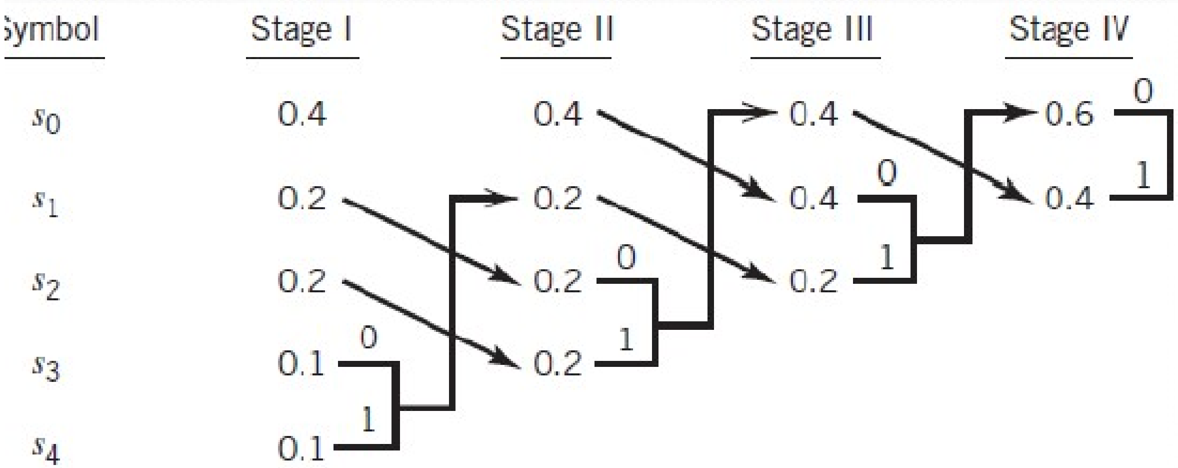


- The average codeword length is $\bar{L} = \sum_{k=0}^{K-1} p_k l_k$

$$\bar{L} = 0.4(2) + 0.2(2) + 0.2(2) + 0.1(3) + 0.1(3)$$
$$= 2.2 \text{ binary symbols}$$

- The entropy of the specified discrete memoryless source is

$$H(S) = 0.4 \log_2\left(\frac{1}{0.4}\right) + 0.2 \log_2\left(\frac{1}{0.2}\right) + 0.2 \log_2\left(\frac{1}{0.2}\right) + 0.1 \log_2\left(\frac{1}{0.1}\right) + 0.1 \log_2\left(\frac{1}{0.1}\right)$$
$$= 0.529 + 0.464 + 0.464 + 0.332 + 0.332 = 2.121 \text{ bits}$$



Symbol	Probability	Code word
s_0	0.4	00
s_1	0.2	10
s_2	0.2	11
s_3	0.1	010
s_4	0.1	011

Ex: Construct The Huffman code for The Symbols $S_0, S_1, \ldots S_7$ with Probabilities $1/2, 1/8, 1/8, 1/16, 1/16, 1/16, 1/32, 1/32$ respectively using Huffman coding.

Sol:

	I	II	III	IV	V	VI	VII
S_0	$1/2$ 0	$1/2$ 0	$1/2$ 0	$1/2$ 0	$1/2$ 0	$1/2$ 0	$1/2$ 0
S_1	$1/8$ 110	$1/8$ 110	$1/8$ 110	$1/8$ 110	$1/4$ 10	$1/4$ 10	$1/2$ 1
S_2	$1/8$ 111	$1/8$ 111	$1/8$ 111	$1/8$ 111	$1/8$ 110	$1/4$ 11	
S_3	$1/16$ 1010	$1/16$ 1010	$1/8$ 100	$1/8$ 100	$1/8$ 111		
S_4	$1/16$ 1011	$1/16$ 1011	$1/16$ 1010	$1/8$ 101			
S_5	$1/16$ 1000	$1/16$ 1000	$1/16$ 1011				
S_6	$1/32$ 10010	$1/16$ 1001					
S_7	$1/32$ 10011						

$$H(S) = \frac{37}{16} \text{ b/m}$$

$$\bar{L} = \frac{37}{16} \text{ b/m}$$

$S_0 - 0 \rightarrow 1$

$S_1 - 110 \rightarrow 3$

$S_2 - 111 \rightarrow 3$

$S_3 - 1010 \rightarrow 4$

$S_4 - 1011 \rightarrow 4$

$S_5 - 1000 \rightarrow 4$

$S_6 - 10010 - 5$

$S_7 - 10011 \rightarrow 5$

$\eta = 100\%$

$\gamma = 0\%$

Noisy Channel Coding Theorem

- Let a DMS with an alphabet X have entropy $H(X)$ and produce symbols every T_s seconds.
- Let a discrete memoryless channel have capacity C and be used once every T_c seconds.
- Then if $$\frac{H(X)}{T_s} \leq \frac{C}{T_c}$$

there exists a coding scheme for which the source output can be transmitted over the noisy channel and be reconstructed with an arbitrarily low probability of error

- Conversely if $$\frac{H(X)}{T_s} > \frac{C}{T_c}$$

It is not possible to transmit information over the channel and reconstruct it with an arbitrarily small probability of error.

- The parameter $\frac{C}{T_c}$ is called the **Critical Rate**.

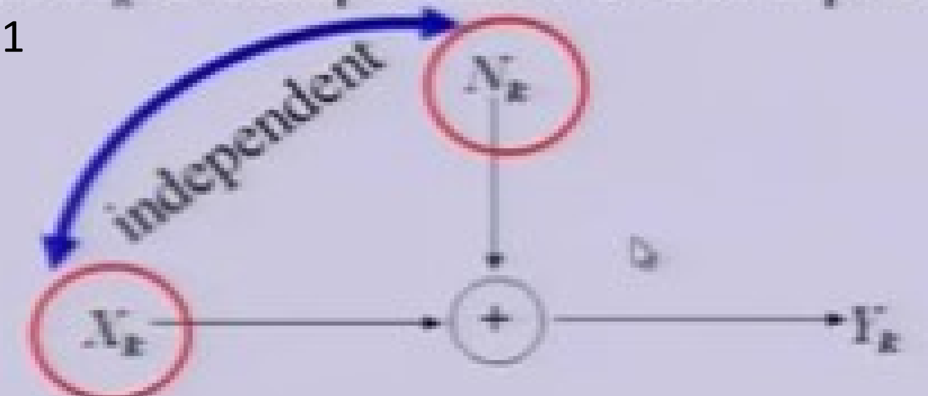
Gaussian Channel

- So far we have studied limits on the maximum rate at which information can be sent over a channel reliably in terms of the channel capacity.
- We next formulate the information capacity theorem for band-limited, power-limited Gaussian channels.

Gaussian Channel

- An important and useful channel is the **Gaussian Channel**
- This is a time discrete channel with output Y_k at time k.
- This output is the result of the sum of the input X_k and the noise Z_k. This noise is drawn from a Gaussian distribution with mean zero and variance σ^2.
- Thus, $Y_k = X_k + N_k$, where $N_k \sim N(0, \sigma^2)$.
- The noise N_k is independent of the input X_k.

Assumption1



Channel Capacity

$$C = \max_{P(x_j)} I(X;Y)$$

$$= \max_{P(x_j)} \sum_{j=0}^{q-1} \sum_{i=0}^{r-1} P(x_j) P(y_i \mid x_j) \log \frac{P(y_i \mid x_j)}{P(y_i)}$$

- $C \geq 0$, since $I(X;Y) \geq 0$.
- $C \leq \log|X|$, since $C = \max I(X;Y) \leq \max H(X) = \log|X|$.
- $C \leq \log|Y|$, since $C = \max I(X;Y) \leq \max H(Y) = \log|Y|$.

Capacity of Gaussian Channel

- Since the transmitter is usually power-limited, let us put a constraint on the *average* power in X_k:

$$E[X_k^2] = P \qquad k = 1, 2, \ldots, K.$$

- Thus, the information capacity of the channel (same as the channel capacity) is given by

$$C = \max_{f_{X_k}(x)} \{ I(X;Y) \mid E[X_k^2] = P \}$$

Capacity of Gaussian Channel

- Since the transmitter is usually power-limited, let us put a constraint on the *average* power in X_k:

$$E[X_k^2] = P \qquad k = 1, 2, \dots, K.$$

- Thus, the information capacity of the channel (same as the channel capacity) is given by

$$C = \max_{f_{X_k}(x)} \{ I(X;Y) \mid E[X_k^2] = P \}$$

probability density function of X_k.

Conditional differntial entropy

$$I(X_k;Y_k) = h(Y_k) - h(Y_k \mid X_k)$$

- X_k and N_k are independent random variables.
- Therefore, the conditional differential entropy of Y_k given X_k is equal to the differential entropy of N_k.
- Intuitively, this is because given X_k the uncertainty arising in Y_k is purely due to N_k. That is,

$$h(Y_k \mid X_k) = h(N_k)$$

Hence we can write

$$I(X_k;Y_k) = h(Y_k) - h(N_k)$$

$$I(X_k;Y_k) = h(Y_k) - h(Y_k \mid X_k)$$

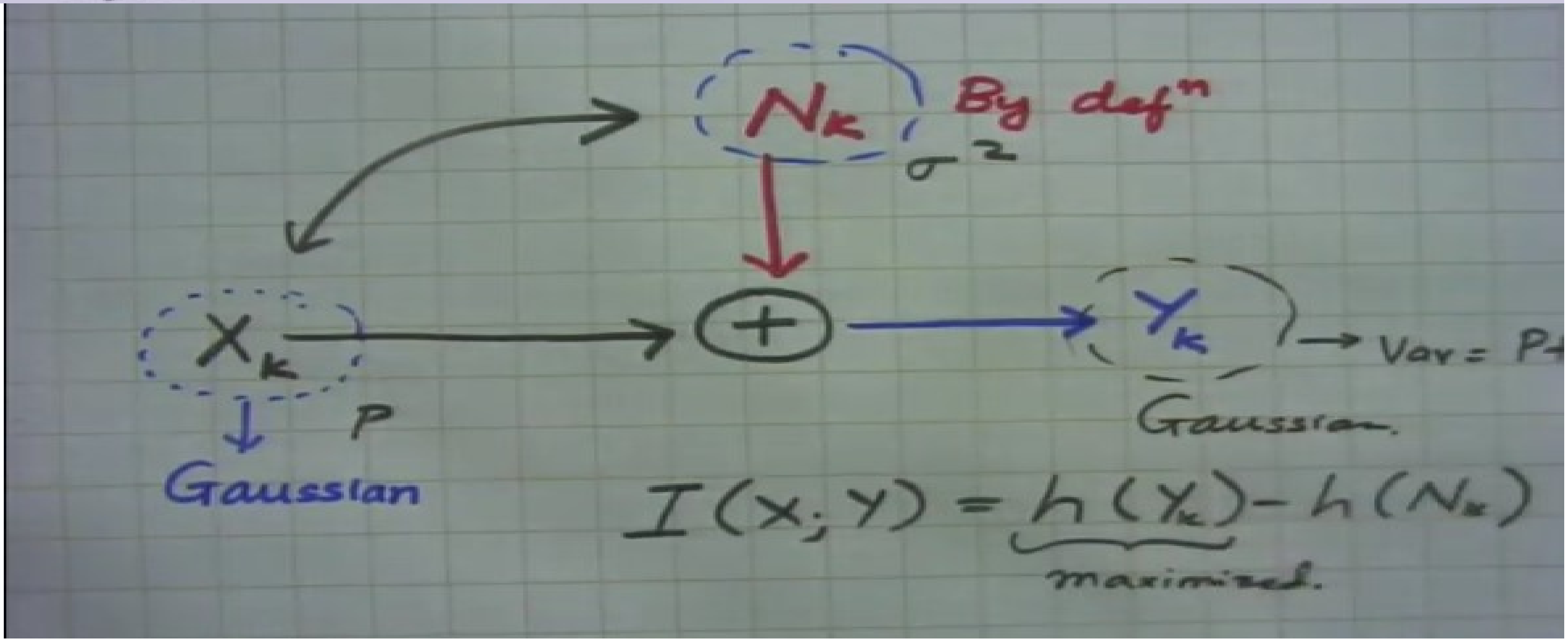
Capacity of Gaussian Channel

- If we assume Y_k to be Gaussian, and N_k is Gaussian by definition, then X_k is also Gaussian.
- This is because the sum (or difference) of two Gaussian random variables is also Gaussian.
- Thus, in order to maximize the mutual information between the channel input X_k and the channel output Y_k, the transmitted signal should be Gaussian.
- Therefore we can write

$$C = I(X;Y) \mid E[X_k^2] = P \text{ and } X_k \text{ is Gaussian}$$

1

- We know that if two independent Gaussian random variables are added, the variance of the resulting Gaussian random variable is the sum of the variances.
- Therefore, the variance of the received sample Y_k equals $P + N_0W$.



2

- We know that if two independent Gaussian random variables are added, the variance of the resulting Gaussian random variable is the sum of the variances.
- Therefore, the variance of the received sample Y_k equals $P + N_0W$.
- It can be shown that the differential entropy of a Gaussian random variable with variance σ^2 is

$$\frac{1}{2}\log_2(2\pi e\sigma^2)$$

Therefore, $h(Y_k) = \frac{1}{2}\log_2[2\pi e(P + N_0W)]$

Therefore, $h(Y_k) = \frac{1}{2}\log_2[2\pi e(P + N_0 W)]$

And, $h(N_k) = \frac{1}{2}\log_2[2\pi e(N_0 W)]$

$$C = h(Y_k) - h(N_k)$$

$$C = \frac{1}{2}\log_2[2\pi e(P + N_0 W)] - \frac{1}{2}\log_2[2\pi e(N_0 W)]$$

$$C = \frac{1}{2}\log_2\left(1 + \frac{P}{N_0 W}\right)$$

$$C = \frac{1}{2}\log_2\left(1+\frac{P}{N_0 W}\right) \text{ bits per channel use.}$$

- We are transmitting $2W$ samples per second, i.e., the channel is being used $2W$ times in one second.
- Therefore, the information capacity can be expressed as

$$C = W\log_2\left(1+\frac{P}{N_0 W}\right) \text{ bits per second}$$

- This basic formula for the capacity of the band-limited, AWGN waveform channel with a band-limited and average power-limited input was first derived by Shannon in 1948.

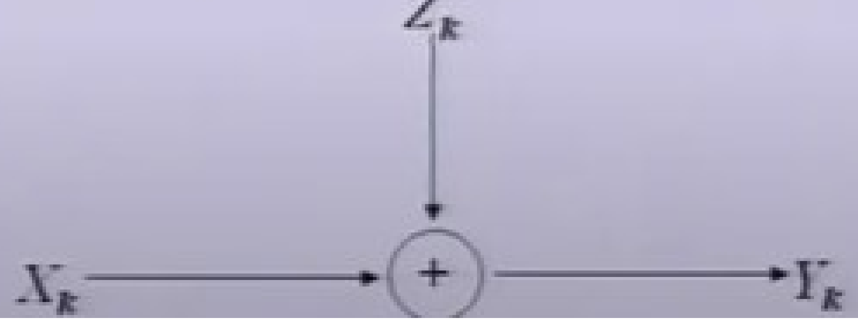
- It is known as the Shannon's third theorem, the **Information Capacity Theorem**.

Information Capacity Theorem

The information capacity of a band-limited, power-limited Gaussian channel can be expressed as

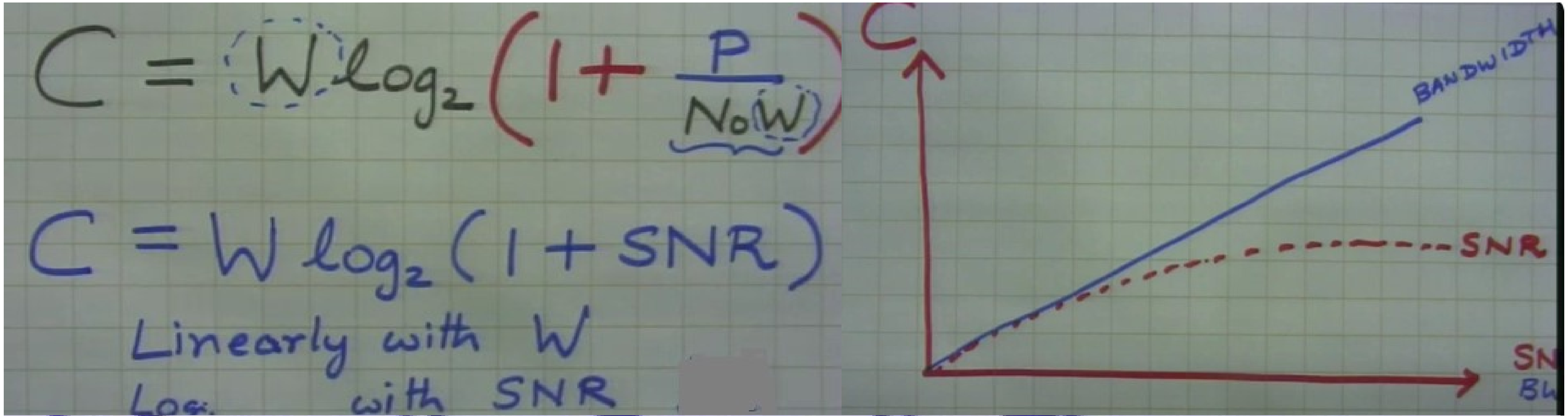
$$C = W \log_2\left(1 + \frac{P}{N_0 W}\right)$$

$N_0 W$ → Noise power



LAKIREDDY BALI REDDY COLLEGE OF ENGINEERING
(AUTONOMOUS)
Accredited by NAAC & NBA (CSE, IT, ECE, EEE & ME)
Approved by AICTE, New Delhi and Affiliated to JNTUK, Kakinada
L.B.Reddy Nagar, Mylavaram-521230, Krishna Dist, Andhra Pradesh, India

Channel Capacity of Gaussian Channel-Band width-S/N trade off



Information Capacity Theorem

- The **Information Capacity Theorem** is one of the important results in information theory.
- In a single formula one can see the trade off between the channel bandwidth, the average transmitted power and the noise power spectral density.
- Given the channel bandwidth and the SNR the channel capacity (bits/second) can be computed.
- This channel capacity is the fundamental limit on the rate of reliable communication for a power-limited, band-limited Gaussian channel.

Mutual Information

- the entropy H(X) is a measure of the prior uncertainty about X,
- How can we measure the uncertainty about X after observing Y? The conditional entropy of X selected from alphabet $\mathcal{X}$, given Y = y_k.
- We write $H(X|Y = y_0), \ldots, H(X|Y = y_{K-1})$ with probabilities $p(y_0), \ldots, p(y_{K-1})$, respectively. The expectation of entropy $H(X|Y = y_k)$ over the output alphabet $\mathcal{Y}$ is therefore given by

$$H(X|Y = y_k) = \sum_{j=0}^{J-1} p(x_j|y_k) \log_2\left(\frac{1}{p(x_j|y_k)}\right)$$

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

$$p(x_j, y_k) = p(x_j|y_k)p(y_k)$$

$$= \sum_{k=0}^{K-1}\sum_{j=0}^{J-1} p(x_j, y_k) \log_2\left(\frac{1}{p(x_j|y_k)}\right)$$

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

- The conditional entropy, *H(X|Y),* is the average amount of uncertainty remaining about the channel input after the channel output has been observed
- The conditional entropy *H(X|Y)* relates the channel output Y to the channel input X. The entropy *H(X)* defines the entropy of the channel input X by itself. Given these two entropies, *I(X;Y)* which is called the mutual information of the channel.

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

- To add meaning to this new concept, we recognize that the entropy $H(X)$ accounts for the uncertainty about the channel input before observing the channel output and the conditional entropy $H(X|Y)$ accounts for the uncertainty about the channel input after observing the channel output.
- We may, therefore, go on to make the statement:
- The mutual information $I(X;Y)$ is a measure of the uncertainty about the channel input, which is resolved by observing the channel output.
- The mutual information $I(Y;X)$ is a measure of the uncertainty about the channel output that is resolved by sending the channel input.

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

Properties of Mutual Information

- *1. The mutual information of a channel is symmetric*
- *I(X;Y) = I(Y;X)*
- *The input entropy* $H(X) = \sum_{j=0}^{J-1} p(x_j) \log_2\left(\frac{1}{p(x_j)}\right)$

$$= \sum_{j=0}^{J-1} p(x_j) \log_2\left(\frac{1}{p(x_j)}\right) \sum_{k=0}^{K-1} p(y_k|x_j) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(y_k|x_j) p(x_j) \log_2\left(\frac{1}{p(x_j)}\right)$$

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

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

$$I(X;Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2\left(\frac{p(x_j|y_k)}{p(x_j)}\right)$$

- To further confirm this property, we may use Bayes' rule for conditional probabilities,

$$\frac{p(x_j|y_k)}{p(x_j)} = \frac{p(y_k|x_j)}{p(y_k)}$$

- interchanging the order of summation, we get

$$I(X;Y) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} p(x_j, y_k) \log_2\left(\frac{p(y_k|x_j)}{p(y_k)}\right)$$

$$= I(Y;X)$$

- 2.*The mutual information is always nonnegative; that is*;
- $I(X;Y) \geq 0$

$$I(X;Y) = \sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2\left(\frac{p(x_j|y_k)}{p(x_j)}\right)$$

$$p(x_j|y_k) = \frac{p(x_j, y_k)}{p(y_k)}$$

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

- We cannot lose information, on the average, by observing the output of a channel.

- 3. *The mutual information of a channel is related to the joint entropy of the channel input and channel output by*

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

- where the joint entropy H(X, Y) is defined by

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

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

- Second term That is $\sum_{j=0}^{J-1} \sum_{k=0}^{K-1} p(x_j, y_k) \log_2\left(\frac{1}{p(x_j)p(y_k)}\right)$

$$= \sum_{j=0}^{J-1} \log_2\left(\frac{1}{p(x_j)}\right) \sum_{k=0}^{K-1} p(x_j, y_k) + \sum_{k=0}^{K-1} \log_2\left(\frac{1}{p(y_k)}\right) \sum_{j=0}^{J-1} p(x_j, y_k)$$

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

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

$$\sum_{k=0}^{K-1} p(x_j, y_k) = p(y_k)$$

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

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