UNIT-1 INFORMATION THEORY.

Introduction.

→ Purpose of communication system is to carry the information from one place to another over a communication channel.

→ Branch of probability that can be applied to the study of comm systems called Information theory.

→ Generally comm information is statistical in nature & mainly to study the simple statistical communication models.

→ Deals with mathematical modelling & analysis of a communication system rather than with physical sources & channels.

Info source → encoder → tx → channel → decoder → Rx.

↑

Noise

Block diagram of Info System.

Discrete Memoryless Source.

DMS → Present output does not depend on the past Output values at any time.

$$\boxed{\text{DMS}} \xrightarrow{S=(S_1 \ldots S_N)}$$

Mathematical model.

→ O/P emitted by a source during every unit of time, i.e. at unit time interval is called <u>discrete message</u>.

Scanned with CamScanner

Information can be represented using a discrete random variables that takes fixed finite alphabet S.

$$S = \{S_0, S_1, S_2 \dots S_{N-1}\} \qquad i = 1, 2, \dots N-1$$

Consider $S = S_k$ & $i = k$ then

$$P(S = S_k) = P_k$$

↓ symbols

→ probability of source.

Set of probabilities must satisfy the condition

$$\sum_{i=0}^{N-1} p_i = 1.$$

→ The useful information from source is called "Information Content" → sum of entropy & the redundant info.

Information

→ The amt of info / msg transmitted over a channel is defined in statistical terms called probability of occurance

→ If the probability of occurance is more, then there will be less amt of information / if the probability of occurance is less, then there will be more amt of information.

Eg: If a dog bites a man → Prob is more & very less info

If a man bites a dog → Prob is less & very high info

Scanned with CamScanner

The amt of information $I(s_i)$ is related to the logarithm on the inverse of the probability of occurance of an event.

$$I(s_i) = \log_2(1/p_i) \quad i = 0, 1 \ldots N-1.$$

Properties of Info.

1) If more uncertainity, more information carried.

2) If rx knows the msg being transmitted, amt of info carried is zero.

$$I(s_i) = 0 \quad \text{for } p_i = 1$$

Proof

$$I(s_i) = \log_2(1/p_i)$$

$$= \log_2(1/1) = \log_2(1) = \boxed{0 = I(s_i)}$$

3) I_1 = info carried by msg m_1

I_2 = info carried by msg m_2

Then the amt of info carried by msg m_1 & m_2 is $I_1 + I_2$

$$I_k = \log_2(1/p_i)$$

for msg m_1 $\quad I_1 = \log_2(1/p_1)$

for msg m_2 $\quad I_2 = \log_2(1/p_2)$

$$I_{1,2} = \log_2\left(\frac{1}{p_1 p_2}\right) = \log_2\left(\frac{1}{p_1}\right) + \log_2\left(\frac{1}{p_2}\right)$$

$$\boxed{I_{1,2} = I_1 + I_2}$$

Scanned with CamScanner

4) If $m = 2^N$ are equally likely messages, then the amt of info carried by each msg will N bit.

Proof. All m msg are equally likely & independent prob of occurance of each msg be $1/m$.

$$I(m_k) = \log_2(1/P_k)$$

$$= \log_2(1/1/m) = \log_2(m)$$

$$= \log_2(2^N)$$

$$\boxed{I(m_k) = N \text{ bits}}.$$

Entropy. AU: Dec - 06, 07, 09, 15, May - 04, 08, 09, 17 (16)

What is entropy, discuss its properties with suitable example?

→ Entropy of a source is defined as the source which produces avg info per individual msg or symbol in a particular interval.

Let $m_1, m_2 \dots m_k$ be the different msg with $P_1 P_2 \dots P_k$ corresponding prob of occurances.

Let L be the total no of msg transmitted, the info due to msg m_1 be

$$I_1 = \log_2(1/P_1)$$

Total info be $I_1(\text{total}) = P_1 L \log\left(\frac{1}{P_1}\right)$

the info due to msg m_2 be $I_2 = \log_2(1/P_2)$

Total info be $I_2(\text{total}) = P_2 L \log\left(\frac{1}{P_2}\right)$

Scanned with CamScanner

Total info carried

$$I_{(total)} = I_1(total) + I_2(total) + \dots + I_k(total)$$

$$= P_1 L \log_2\left(\frac{1}{P_1}\right) + P_2 L \log_2\left(\frac{1}{P_2}\right) + \dots + P_k L \log_2\left(\frac{1}{P_k}\right)$$

avg info per msg will be

$$H = \text{Total info} / L$$

$$\boxed{H = I_{total} / L}$$

$$H = P_1 \log_2\left(\frac{1}{P_1}\right) + P_2 \log_2\left(\frac{1}{P_2}\right) + \dots + P_k \log_2\left(\frac{1}{P_k}\right)$$

$$\boxed{= \sum_{m=1}^{k} P_m \log_2\left(\frac{1}{P_m}\right) = H}$$

Time duration $T_b = n/\tau$

Avg rate at which the info must be transferred is called "info rate (R)"

$$R = \frac{nH}{n/\tau} = \frac{nH\tau}{n}$$

$$\boxed{R = H\tau \text{ bits/sec}}$$

τ = rate at which msg generated from source.

Scanned with CamScanner

Properties of Entropy.

Entropy of DMS is bounded as $0 \leq H \leq \log_2 k$.

1) Entropy is zero, if the event is sure or it is impossible.

$H = 0$ if $P_m = 0 / P_m = 1$

When $P_m = 0$

$$H = \sum_{m=1}^{k} P_m \log_2 (1/P_m)$$

$$= 0 \log_2 (1/0)$$

$$\boxed{H = 0}$$

When $P_m = 1$

$$H = \sum_{m=1}^{k} P_m \log_2 (1/P_m)$$

$$= 1 \log_2 (1/1)$$

$$\boxed{H = 0}.$$

2) When $P_m = 1/k$ for all k symbols, then the symbols are equally likely. For such source, entropy is $H = \log_2 k$.

Proof. $P = 1/k$

$$P_1 = P_2 = P_3 = \dots P_k = 1/k$$

$$H = \sum_{m=1}^{k} P_m \log_2 1/P_m$$

Scanned with CamScanner

$$= P_1 \log_2 \frac{1}{P_1} + P_2 \log_2 \frac{1}{P_2} + \dots P_k \log_2 \frac{1}{P_k}$$

$$= \frac{1}{k} \log_2 k + \frac{1}{k} \log_2 k + \dots \frac{1}{k} \log_2 k$$

$$= k/k \log_2 k$$

$$\boxed{H = \log_2 k}.$$

Problem.

A source 'S' emits a symbols S_1, S_2, S_3 with probabilities $0.25, 0.5, 0.25$, having bandwidth $B = 250\,Hz$. Find a) amt of info b) Entropy c) info rate.

a) amt of info.

$$I_k = \log_2 (1/P_k) \qquad k = 1, 2, 3$$

$$I_1 = \log_2 (1/P_1) = \log_2 (1/0.25)$$

$$= \frac{\log_{10}(1/0.25)}{\log_{10}(2)} = \frac{0.602}{0.301}$$

$$\boxed{I_1 = 2 \text{ bits}}$$

$$I_2 = \log_2 (1/P_2) = \log_2 (1/0.5)$$

$$= \frac{\log_{10}(1/0.5)}{\log_{10}(2)} = \frac{\log_{10} 2}{\log_{10} 2}$$

$$\boxed{I_2 = 1 \text{ bit}}$$

Scanned with CamScanner

$$I_3 = \log_2\left(\frac{1}{P_3}\right) = \log_2 (1/0.25)$$

$$= \frac{\log_{10}(1/0.25)}{\log_{10}(2)} = \frac{0.602}{0.301}$$

$$\boxed{I_3 = 2 \text{ bits}}.$$

$$\text{Total amt of info} = I_1 + I_2 + I_3$$

$$= 2 + 1 + 2$$

$$\boxed{I_{total} = 5 \text{ bits}}.$$

b) Entropy.

$$H(S) = \sum_{k=1}^{N} P_k \log_2 (1/P_k) = \sum_{k=1}^{3} P_k \log_2 (1/P_k)$$

$$= P_1 \log_2\left(\frac{1}{P_1}\right) + P_2 \log_2\left(\frac{1}{P_2}\right) + P_3 \log_2\left(\frac{1}{P_3}\right)$$

$$= 0.25 \log_2\left(\frac{1}{0.25}\right) + 0.5 \log_2\left(\frac{1}{0.5}\right) + 0.25 \log_2\left(\frac{1}{0.25}\right)$$

$$= 0.25 \frac{\log_{10}(1/0.25)}{\log_{10} 2} + 0.5 \frac{\log_{10}(1/0.5)}{\log_{10}(2)} + 0.25 \frac{\log_{10}(1/0.25)}{\log_{10}(2)}$$

$$= 0.25(2) + 0.5(1) + 0.25(2)$$

$$= 0.5 + 0.5 + 0.5$$

$$\boxed{H(S) = 1.5 \text{ bits/symbol}}.$$

Scanned with CamScanner

c) Info Rate (R).

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

$r = 2B = 2 \times 250$

$r = 500$ levels/sec

$R = 500 \times 1.5$

$$\boxed{R = 750 \text{ bits/sec}}$$

Mutual Information. AU: May - 16 (8m)

Define mutual information & list its properties.

$\longrightarrow$ Mutual Information is defined as the difference b/w the two values $H(X) - H(X/Y)$

$$I(X;Y) = \underbrace{H(X)}_{\text{initial uncertainity}} - \underbrace{H(X/Y)}_{\text{final uncertainity / Cond Entropy}}$$

Conditional Entropy $\longleftrightarrow$ amt of uncertainity about the channel i/p X after the channel o/p Y.

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

Joint probability $P(x_j, y_k) = P(x_j/y_k) \cdot P(y_k)$.

Scanned with CamScanner

Discrete Memoryless Channels. AU: May 14, Dec 15

Define binary symmetry channel & write its channel matrix.

→ DMC is a statistical model with an i/p X & o/p Y.

ie. Noisy version of X both X, Y are random var.

→ Every unit of time, the channel accepts an i/p sig X from x & in response it emits an o/p symbol Y from y.

→ channel is said to be discrete when both x, y have finite sizes.

→ also said to be ~~discrete~~ 'memoryless' when the current o/p depends only on the current i/p & not on the past values.

$$X \left\{ \begin{matrix} x_0 \\ x_1 \\ x_2 \\ \vdots \\ x_{J-1} \end{matrix} \right. \xrightarrow{X} P(y_k/x_j) \xrightarrow{Y} \left. \begin{matrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{k-1} \end{matrix} \right\} Y$$

channel is described in terms of $x = \{x_0 \ldots x_{J-1}\}$ i/p

$y = \{y_0 \ldots y_{k-1}\}$ o/p

Set of transition probabilities

$$P(y_k|x_j) = P(Y = y_k / X = x_j) \text{ for all } j \,\&\, k.$$

Naturally $0 \leq P(y_k/x_j) \leq 1$.

If $k = j$ then $P(y_k|x_j)$ → cond prob of correct reception.

If $k \neq j$ then $P(y_k|x_j)$ → cond prob of error.

Scanned with CamScanner

Channel Matrix → way of describing a discrete memoryless channel is to arrange the various transition probabilities of channel in the form of matrix

$$P = \begin{bmatrix} P(y_0|x_0) & P(y_1|x_0) & \cdots & P(y_{k-1}|x_0) \\ P(y_0|x_1) & P(y_1|x_1) & \cdots & P(y_{k-1}|x_1) \\ \vdots & \vdots & & \vdots \\ P(y_0|x_{J-1}) & P(y_1|x_{J-1}) & \cdots & P(y_{k-1}|x_{J-1}) \end{bmatrix}$$

Above matrix called channel matrix / transition matrix.

each Row → fixed channel i/p

each colomn → fixed channel o/p

- Sum of the elements along any row of the matrix is always equals to one.

$$\sum_{k=0}^{k-1} P(y_k|x_j) = 1 \quad \text{for all } j.$$

If the inputs to a discrete memoryless channels are selected according to the probability distribution $\{P(x_j), j = 0, 1 \ldots J-1\}$ the event that the channel i/p $x = x_j$ occurs with prob

$$P(x_j) = P(x = x_j), \quad j = 0, 1 \ldots J-1$$

o/p $P(y_k) = P(Y = y_k), \quad k = 0, 1 \ldots k-1$

Scanned with CamScanner

The joint probability distribution of X, Y

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

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

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

Marginal probability distribution of O/P Y

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

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

$$= \sum_{j=0}^{J-1} P(y_k / x_j) \underbrace{P(x_j)}_{\text{priori prob.}} \quad \text{for } k = 0, 1, \ldots k-$$

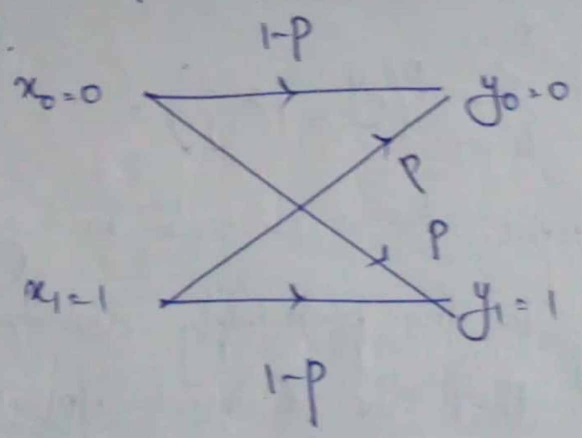
If we give the I/P priori prob $p(x_j)$ & channel matrix then we may calculate the prob of the O/P symbols.

Eg: Binary Sym channel.

$J = k = 2$

I/P $\longrightarrow$ $x_0 = 0$
$x_1 = 1$

O/P $\longrightarrow$ $y_0 = 0$
$y_1 = 1$



Channel is symmetric because the prob of rx 1 if 0 is sent is same as the prob of rx 0 if 1 is sent.

Scanned with CamScanner

Channel capacity

Consider a DMC with i/p x, o/p y & transition prob $p(y_k | x_j)$ where $j = 0, 1 \ldots J-1$ & $k = 0, 1 \ldots k-1$

Mutual info of the channel is defined as

$$I(x:y) = \sum_{k=0}^{k-1} \sum_{j=0}^{J-1} p(x_j, y_k) \log_2 \left[\frac{p(y_k | x_j)}{p(y_k)} \right]$$

$$= I(y:x)$$

here note that

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

also

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

$$p(y_k) = \sum_{j=0}^{J-1} p(Y = y_k | X = x_j)\, p(X = x_j)$$

from these three eqn we see that it is necessary for us to know the i/p prob distribution $\{p(x_j) | j = 0, 1 \ldots J-1\}$ to calcula mutual info $I(x;y)$.

Mutual Info of a channel depends not only on the channel but also on the way in which the channel is used. I/P probability distribution $p(x_j)$ is independent of the channel, we can then maximize the $I(x:y)$ of the channel w.r.to $p(x_j)$.

Channel capacity of a discrete memoryless channel is defined as the max mutual info $I(X:Y)$ in any single use of the channel where the maximization is over all possible i/p prob distributions $P(x_i)$ on X.

$$C = \max I(X:Y) = \max \left[H(X) - H(X/Y) \right].$$

The channel capacity C is measured in bits per channel / bits per txn.

C is the fn only of the transition probabilities $P(y_k/x_j)$ that defines the channel.

2 cond / $P(x_j \geq 0)$ for all j

\ $\sum_{j=1}^{J} P(x_j) = 1$

Channel Efficiency $\eta = \dfrac{\text{Mutual Info}}{\text{Max mutual info}} = \dfrac{I(X:Y)}{\max I(X:Y)}$

$$\boxed{\eta = I(X:Y) / C}$$

Redundancy $R = 1 - \eta$

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

Noise free channel capacity $C = \log_2 N$ bits/symbol.

Scanned with CamScanner

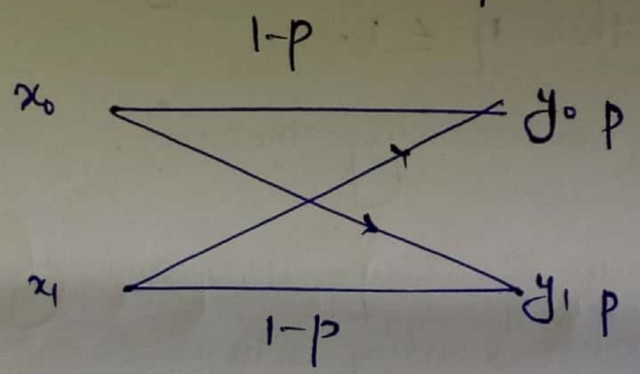
Binary Symmetric Channel (BSC)

→ 2 i/p ($x_1 = 0$, $x_2 = 1$) & 2 o/p ($y_1 = 0$, $y_2 = 1$). Channel is symmetric because the prob of rx 1 if 0 is sent is same as the prob of rx 0 if 1 is sent.

→ Binary Comm Channel is symmetric if

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

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



Binary Sym Channel.

Source Coding Theorem. AU: May 07, 08, 10 Dec 16, 11

Discuss source coding theorem.

Source Coding
- → improve the η of the comm system.
- → process by which the efficient representation of data generated by a discrete source. device → source encoder

2 requirements of the source encoder
- → code words generated by the encoder should be binary.
- → source code should be unique in nature.

$$\text{DMS} \xrightarrow{(s_i)} \text{source encoder} \longrightarrow \text{binary seq } (b_i)$$

→ DMS produces an o/p s_i which is converted by the source encoder into a block of 0's & 1's denoted by b_i.

Scanned with CamScanner

→ Avg codeword length $\bar{L}$ of the source encoder

$$\bar{L} = \sum_{i=0}^{N-1} p_i L_i$$

↓

avg no of bits per source symbol.

→ Coding efficiency $\eta = \frac{L_{min}}{\bar{L}}$ → Shanon's first theorem.

→ Source encoder is efficient when coding efficiency approaches unity. $\bar{L} \geq L_{min}$ then $\eta \leq 1$.

Shanon's first theorem (Source coding theorem).

↓

discrete memoryless source of entropy $H(S)$ and avg codelength $\bar{L}$ for any distortionless source encoding scheme

$$\bar{L} \geq H(S).$$

If $L_{min} = H(S)$ then η is $\eta = H(S) / \bar{L}$

Code Redundancy $\gamma = 1 -$ code efficiency $= 1 - \eta$.

Shanon's second theorem (Channel coding theorem). AU: May 05, 06, Dec 06, 10, 11, 15, 16

Discuss channel coding theorem.

Need → Presence of noise in channel causes discrepancies between the i/p & o/p seq of the digital communication systems.

→ Error probability $= 10^{-6}$, to achieve this we have to use a channel coding.

Channel coding → mapping the incoming data sequence into a channel i/p sequence using encoder at the Tx & decoder at the Rx inverse mapping the o/p sequence into an o/p data sequence in such a way that the overall effect of channel noise is minimized.

Scanned with CamScanner

→ Code rate is the ratio of msg bits in a block to the transmitted bits by the encoder

$$r = \frac{k}{n}$$

k → no of msg bits in a block

n → no of bits in codeword.

→ Channel coding is efficient when the code rate is high.

DMS → avg info rate = $\frac{H(S)}{T_S}$ bits/sec

DMC → channel capacity/unit = $\frac{C}{T_C}$ bits/sec.

II theorem → Let a discrete memoryless source with an alphabet S having entropy H(S) & produce symbols once every T_S seconds. Let a discrete memoryless channel have capacity C & used once every T_S seconds.

$$\frac{H(S)}{T_C} \le \frac{C}{T_C}$$

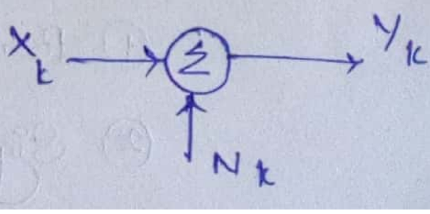
→ $\frac{H(S)}{T_S} > \frac{C}{T_C}$ is not possible to transmit info over the channel & reconstruct it with small prob of error.

Shannon's third theorem (Info Capacity / Shannon's Hartley).

Explain Shannon's Hartley theorem. AU: May 05, 06, 07, 09, 11, 12, 13, 14 Dec 06, 10, 11, 13, 16

Consider the band limited, gaussian channel with zero-mean stationary process x(t). Let X_k, k denote the continuous random variables obtained by uniformly sampling of the process x(t) at the nyquist rate of 2B samples/sec.

No of samples, $k = 2BT$.



$x_k \rightarrow$ samples of the txn signals.

$Y_k \rightarrow$ output by AWGN of zero mean & power spectral density

$$Y_k = x_k + N_k \qquad k = 1, 2 .. k$$

$\hookrightarrow$ noise with 0 mean & variance (gaussian) $\sigma^2 = N_0 B$.

$\rightarrow$ Information capacity of a channel is defined as the maximum of the mutual information between the channel input x_k & channel output Y_k over all distributions on the i/p x_k that satisfy the power constraint of eqn $E[x_k^2] = P$

Let $I(x_k : Y_k)$ be a mutual info then info capacity

$$C = I(x_k : Y_k) : x_k \text{ gaussian } E(x_k^2) = P$$

$$= \frac{1}{2} \log_2 \left(1 + \frac{P}{\sigma^2}\right) \text{ bits/transmission}$$

$\rightarrow$ Info capacity of a continuous channel of BW B hertz with AWGN of power spectral density $N_0/2$ and limited in B is given by

$$C = B \log_2 \left(1 + \frac{P}{N_0 B}\right) \text{ bit/sec}$$

$\rightarrow$ In terms of SNR

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

From the above the channel capacity depends on 2 factors

1. Bandwidth (B) of the channel
2. Signal to Noise Ratio $\left(\frac{S}{N}\right)$ is $N_0 B$.

Scanned with CamScanner

Shannon Fano Coding. AU: Dec -05, 07, 08, 11, 10, 12, 13, 14, 15, 16
May- 09, 11, 12, 13, 14, 15, 16, 17

Explain Entropy coding.

If the prob of occurances of all the messages are not equally likely then the avg information / entropy is reduced which in turn reduces Info rate (R).

→ Shannon Fano is one where less no of bits is used for highly probable msg and more no of bits for rare occurances.

Procedure.

1) Write the message probability in decreasing order
2) ~~Partion~~ Partition the msg set with equi probable subjects
3) Once partitioning is done 0 to 1st set & 1 to 2nd set.
4) Same procedure is followed entirely until the subset contains only one msg.

Problem.

A discrete memoryless source has 6 symbols $s_1, s_2 \ldots s_6$ with probabilities of 0.4, 0.1, 0.2, 0.1, 0.1, 0.1 respectively & construct a Shannon Fano Coding and calculate ?.

Sym	Prob	I	II	III	CV	Length (l)
s_1	0.4	0	0.4] 0	0.4		00	2
s_3	0.2	0	0.2] 1	0.2		01	2
s_2	0.1	1	0.1] 0	0.1] 0	100	3
s_4	0.1	1	0.1] 0	0.1] 1	101	3
s_5	0.1	1	0.1] 1	0.1] 0	110	3
s_6	0.1	1	0.1] 1	0.1] 1	111	3

Scanned with CamScanner

$$\bar{L} = \sum_{i=1}^{N} P_i L_i = \sum_{i=1}^{6} P_i L_i$$

$$= (0.4 \times 2) + (0.2 \times 2) + (0.1 \times 3) + (0.1 \times 3) + (0.1 \times 3) + (0.1 \times 3)$$

$$= 0.8 + 0.4 + 0.3 + 0.3 + 0.3 + 0.3$$

$$\boxed{\bar{L} = 2.4 \text{ bits/symbol}}.$$

Entropy $H(S) = \sum_{i=1}^{N} P_i \log_2\left(\frac{1}{P_i}\right) = \sum_{i=1}^{6} P_i \log_2 (1/P_i)$

$$= 0.4 \log_2\left(\frac{1}{0.4}\right) + 0.2 \log_2\left(\frac{1}{0.2}\right) + 0.1 \log_2\left(\frac{1}{0.1}\right) +$$

$$= 0.4 \times 1.3219 + 0.2 \times 3.3219 + 0.1 \times 3.3219 \times$$

$$= 0.52876 + 0.33219 + [0.33219 \times 4]$$

$$\boxed{H(S) = 2.322 \text{ bits/symbol}}.$$

Code Efficiency $\eta = \frac{H}{\bar{L}} = \frac{2.322}{2.4}$

$$\boxed{\eta = 96.7\%}.$$

Huffmann Coding

→ Uses the same principal as Shano-Fano coding ie. assigning different no of binary digits to the messages according to their probability of occurances.

→ Coding method leads to lowest possible values of $\bar{L}$ for a given msg sequence N resulting is a max η. Also called minimum redundancy code / Optimum code.

Procedure.

1) The source symbols are arranged in decreasing order of probability and lowest prob messages are assigned as 0 & 1 called splitting stage.

2) 2 lowest symbol prob in stage 1 are combined into new source symbol with the probability equal to the sum of 2 original probabilities called a reduction.

3) Steps are repeated until the final list of source symbols are only 2 for which 0, 1 are assigned.

4) Codeword calculation begins at the last reduction to first reduction for each symbols.

Problem.

A DMS has 6 symbols $S_1, S_2 \ldots S_6$ with probabilities 0.4, 0.1, 0.2, 0.1, 0.1 & 0.1 respectively. Construct a Huffmann Code & calculate η.

Symb	I	II	III	IV	V	CW	l
S_1	0.4^{00}	0.4^{00}	0.4^{00}	0.4^{1}	$0.6\ 0$	00	2
S_3	0.2^{11}	0.2^{10}	0.2^{01}	0.4^{00}	$0.4\ 1$	11	2
S_2	0.1^{010}	0.2^{11}	0.2^{10}	0.2^{01}		010	3
S_4	0.1^{011}	0.1^{010}	0.2^{11}			011	3
S_5	0.1^{100}	0.1^{01}				100	3
S_6	0.1^{101}					101	3

$$\bar{L} = \sum_{i=1}^{6} p_i l_i = p_1 l_1 + p_2 l_2 + p_3 l_3 + p_4 l_4 + p_5 l_5 + p_6 l_6$$

$$= 0.4\times 2 + 0.1\times 3 + 0.2\times 2 + 0.1\times 3 + 0.1\times 3 + 0.1\times 3$$

$$= 0.8 + 0.3 + 0.4 + 0.3 + 0.3 + 0.3$$

$$\boxed{\bar{L} = 2.4 \text{ bits/symbol}}$$

$$H(S) = \sum_{i=1}^{6} p_i \log_2\left(\frac{1}{p_i}\right)$$

$$= 0.4\log_2(1/0.4) + 0.1\log_2(1/0.1) + 0.2\log_2(1/0.2) + 0.1\log_2(1/0.1) + 0.1\log_2(1/0.1) + 0.1\log_2(1/0.1)$$

$$= 0.52876 + 0.33219 + 0.46439 + 0.33219 + 0.33219 + 0.33219$$

$$\boxed{H(S) = 2.322 \text{ bits/symbol}}$$

$$\eta = H(S)/\bar{L} = 2.322/2.4$$

$$= 0.967 = \boxed{96.7\% = \eta}$$