
KTU
NOTES

MODULE- 2

Noiseless Coding Theorem (OR) Shannon's First Theorem

Shannon suggested that the length l_i is

$$l_i = \log_r \frac{1}{P_i}$$

If l_i happens to be a fraction, then it is rounded off to next integer.

ie $$\log_r \frac{1}{P_i} \le l_i \le 1 + \log_r \frac{1}{P_i}$$

$$\frac{\log_2 \frac{1}{P_i}}{\log_2 r} \le l_i \le 1 + \frac{\log_2 \frac{1}{P_i}}{\log_2 r} \quad [\because \log \cdot \text{property}]$$

Multiply throughout by P_i and take summation for all i varying from 1 to q

$$\frac{1}{\log_2 r} \sum_{i=1}^{q} P_i \log_2 \frac{1}{P_i} \leq \sum_{i=1}^{q} P_i l_i \leq \sum_{i=1}^{q} P_i + \frac{1}{\log_2 r} \sum_{i=1}^{q} P_i \log_2 \frac{1}{P_i}$$

We know $H(S) = \sum_{i=1}^{q} P_i \log_2 \frac{1}{P_i}$

$$L = \sum_{i=1}^{q} P_i l_i$$

$$\sum_{i=1}^{q} P_i = 1$$

Substituting these in above equation

$$\frac{H(S)}{\log_2 r} \leq L \leq 1 + \frac{H(S)}{\log_2 r}$$

We know $\frac{H(S)}{\log_2 r} = H_r(S)$

So $H_r(S) \leq L \leq 1 + H_r(S)$

$\therefore \quad L \geq H_r(S)$

For n^{th} extension source, equation becomes

$$\frac{H(S^n)}{\log_2 r} \leq L_n \leq 1 + \frac{H(S^n)}{\log_2 r}$$

L_n = average length of code words of n^{th} extended source symbols

We know $H(S^n) = n\,H(S)$

So $$\frac{n\,H(S)}{\log_2 r} \leq L_n \leq 1 + \frac{n\,H(S)}{\log_2 r}$$

Dividing throughout by n

$$\frac{H(S)}{\log_2 r} \leq \frac{L_n}{n} \leq \frac{1}{n} + \frac{H(S)}{\log_2 r}$$

$$H_r(S) \leq \frac{L_n}{n} \leq \frac{1}{n} + H_r(S)$$; Hence proved.

Taking limits $n \to \infty$

$$\lim_{n \to \infty} \frac{L_n}{n} = H_r(S)$$

When L is the average length of code words for the basic source S, then for all values of extension n,

$$\frac{L_n}{n} \leq L$$

Noiseless coding theorem states that "given a code alphabet with 'r' symbols and source alphabet of 'q' symbols, the average length of code words can be made as close to H(s) as possible by increasing the extension.

OR

"By increasing the value of n, the term $\frac{L_n}{n}$ can be reduced to as small a value as we wish, but we can never reduce it below $H_r(S)$ ie by increasing 'n' the code efficiency η_c can be increased to as large a value as we wish".

$$\eta_c = \frac{H_r(S)}{L}$$

For n^{th} extension $\eta_c = \frac{H_r(S)}{(L_n/n)}$.

Shannon - fano Encoding Algorithm :

It is used to get a compact code with minimum redundancy.

Procedure : -

1. The symbols are arranged according to non increasing probabilities
2. The symbols are divided in to two groups so that the sum of probabilities in each group is approximately equal.
3. All the symbols in the I group are designated by '1' and 2nd group by '0'.
4. The I group is again subdivided in to two sub groups such that each sub group probabilities are approximately same.
5. All the symbols of the I group are designated by a '1' and II group by '0'
6. The II group is subdivided in to two more subgroups and step 5 is repeated.

7. The process is continued till further sub division is impossible.

Q. Given the messages x_1, x_2, x_3, x_4, x_5 and x_6 with respective probabilities 0.4, 0.2, 0.2, 0.1, 0.07 and 0.03, construct a binary code by applying Shannon – Fano encoding procedure. Determine code efficiency and redundancy of the code.

	P_i							
x_1	0.4	1	0.4	1				
x_2	0.2	1	0.2	0				
x_3	0.2	0	0.2	1				
x_4	0.1	0	0.1	0	0.1	1		
x_5	0.07	0	0.07	0	0.07	0	0.07	1
x_6	0.03	0	0.03	0	0.03	0	0.03	0

Code	l_i in binits
11	2
10	2
01	2
001	3
0001	4
0000	4

Average length $L = \sum_{i=1}^{6} p_i l_i$

$$= 0.4\times 2 + 0.2\times 2 + 0.2\times 2 + 0.1\times 3 + 0.07\times 4 + 0.03\times 4$$

$= 2.3$ binits / message symbol

Entropy $H(s) = \sum_{i=1}^{6} p_i \log \frac{1}{p_i}$

$$= 0.4 \log \frac{1}{0.4} + 2 \times 0.2 \log \frac{1}{0.2} + 0.1 \log \frac{1}{0.1} + 0.07 \log \frac{1}{0.07} + 0.03 \log \frac{1}{0.03}$$

$= 2.209$ bits / message symbol

Code efficiency $\eta_c = \frac{H(s)}{L} = \frac{2.209}{2.3}$

$= 96.04\ \%$

Code redundancy $R_{\eta c} = 3.96\ \%$

2^{nd} ~~method~~ way

	p_i							
x_1	0.4	1						
x_2	0.2	0	0.2	1	0.2	1		
x_3	0.2	0	0.2	1	0.2	0		
x_4	0.1	0	0.1	0	0.1	1		
x_5	0.07	0	0.07	0	0.07	0	0.07	1
x_6	0.03	0	0.03	0	0.03	0	0.03	0

Code	l_i
1	1
011	3
010	3
001	3
0001	4
0000	4

Average Length $L = \sum_{i=1}^{6} P_i l_i$

$= 0.4 \times 1 + 0.2 \times 3 + 0.2 \times 3$

$+ 0.1 \times 3 + 0.07 \times 4 +$

0.03×4

$= 2.3$ binits / message symbol

Entropy $H(S) = 2.209$ bits / message symbol

Code efficiency $\eta_c = \frac{H(S)}{L} = 96.04\%$.

Code efficiency remains same for both ways of coding.

Q) You are given 4 messages x_1, x_2, x_3 and x_4 with respective probability 0·2, 0·3, 0·4.
~~with respective prob~~

(1) Device a code with prefix property (Shannon- fano code) for these messages

(2) Calculate the efficiency and redundancy of the code

(3) Calculate the probabilities of 0's and 1's in the code.

	p_i					
x_4	0·4	1				
x_3	0·3	0	0·3	1		
x_2	0·2	0	0·2	0	0·2	1
x_1	0·1	0	0·1	0	0·1	0

Average length, $L = \sum_{i=1}^{4} P_i l_i$

$= 0.4 \times 1 + 0.3 \times 2 + 0.2 \times 3 + 0.1 \times 3$

$= 1.9$ binits / message symbol

Entropy $H(s) = \sum_{i=1}^{4} P_i \log \frac{1}{P_i}$

$= 0.4 \log \frac{1}{0.4} + 0.3 \log \frac{1}{0.3} +$

$0.2 \log \frac{1}{0.2} + 0.1 \log \frac{1}{0.1}$

$= 1.846$ bits / message symbol

Code efficiency $\eta_c = \frac{H(s)}{L} = \frac{1.846}{1.9} = 97.15\%$

Code redundancy $R_{\eta_c} = 2.85\%$

Probability of '0', $P(0) = \frac{1}{L} \sum_{i=1}^{4}$ [Number of '0's in the code for x_i] [P_i]

$= \frac{1}{1.9} [3 \times 0.1 + 2 \times 0.2 + 1 \times 0.3 + 0 \times 0.4]$

$P(0) = 0.5263$

Probability of '1's in the code

$P(1) = \frac{1}{L} \sum_{i=1}^{4}$ [Number of 1's in the code for x_i] [P_i]

$= \frac{1}{1.9} [0 \times 0.1 + 1 \times 0.2 + 1 \times 0.3 + 1 \times 0.4]$

$= 0.4737$

Q. Consider a source $S = [S_1, S_2]$ with probabilities $3/4$ and $1/4$ respectively. Obtain shannon-fano code for source S, its 2^{nd} and 3^{rd} extension. Calculate efficiencies for each case.

		Code	l_i
S_1	$3/4$	1	1
S_2	$1/4$	0	1

Average length $L = \sum_{i=1}^{2} P_i l_i$

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

$= 1$ binits / message symbol

Entropy $H(S) = \sum_{i=1}^{2} P_i \log \frac{1}{P_i}$

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

$= 0.8113$ bits / message symbol.

Code efficiency $\eta_c^{(1)} = \frac{H(S)}{L} = \frac{0.8113}{1} = 81.13\%$

For 2^{nd} extension

2^{nd} extension will have $2^2 = 4$ symbols ~~S_1, S_2~~, S_3 S_1S_1, S_1S_2, S_2S_1 and S_2S_2 with probabilities $9/16$, $3/16$, $3/16$ and $1/16$ respectively.

	p_i					
$S_1\ S_1$	$9/16$	1				
$S_1\ S_2$	$3/16$	0	$3/16$	1		
$S_2\ S_1$	$3/16$	0	$3/16$	0	$3/16$	1
$S_2\ S_2$	$1/16$	0	$1/16$	0	$1/16$	0

Code	l_i
1	1
01	2
001	3
000	3

Average length $L_2 = \sum_{i=1}^{4} P_i l_i$

$= \frac{9}{16} \times 1 + \frac{3}{16} \times 2 +$

$\frac{3}{16} \times 3 + \frac{1}{16} \times 3$

$= 1.6875$ bits/msg symbol

Entropy $H(S^2) = 2H(S)$ $[\because H(S^n) = nH(S)]$

$= 2 \times 0.8113$

$= 1.6226$ bits/msg symbol

$\eta_c^{(2)} = \frac{H(S^2)}{L_2} = \frac{1.6226}{1.6875} = 96.15\%$

3rd extension

3rd extension have $2^3 = 8$ symbols.

$S_1S_1S_1$	27/64	1	27/64	1				
$S_1S_1S_2$	9/64	1	9/64	0				
$S_1S_2S_1$	9/64	0	9/64	1	9/64	1		
$S_2S_1S_1$	9/64	0	9/64	1	9/64	0		
$S_1S_2S_2$	3/64	0	3/64	0	3/64	1	3/64	1
$S_2S_1S_2$	3/64	0	3/64	0	3/64	1	3/64	0
$S_2S_2S_1$	3/64	0	3/64	0	3/64	0	3/64	1
$S_2S_2S_2$	1/64	0	1/64	0	1/64	0	1/64	0

Code	l_i
11	2
10	2
011	3
010	3
0011	4
0010	4
0001	4
0000	4

Average length $L_3 = \sum_{i=1}^{8} P_i l_i$

$$= \frac{27}{64} \times 2 + \frac{9}{64} \times 2 + \frac{9}{64} \times 3 + \frac{9}{64} \times 3 + \frac{3}{64} \times 4 + \frac{3}{64} \times 4 + \frac{3}{64} \times 4 + \frac{1}{64} \times 4$$

$= 2.59375$ binits / message symbol

$H(S^3) = 3H(S)$

$= 3[0.8113]$

$= 2.4339$ bits / message symbols.

$$\eta_c^{(3)} = \frac{H(S^3)}{L_3} = \frac{2.4339}{2.59375}$$

$= 93.84\%$

Q. A discrete memoryless source has an alphabet of seven symbols with probabilities for its output, as described below

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

Compute shannon-fano code for this source. Find coding efficiency.

S_0	0.25	1	0.25	1				
S_1	0.25	1	0.25	0				
S_2	0.125	0	0.125	1	0.125	1		
S_3	0.125	0	0.125	1	0.125	0		
S_4	0.125	0	0.125	0	0.125	1		
S_5	0.0625	0	0.0625	0	0.0625	0	0.0625	1
S_6	0.0625	0	0.0625	0	0.0625	0	0.0625	0

Code	l_i
11	2
10	2
011	3
010	3
001	3
0001	4
0000	4

The average length $L = \sum_{i=0}^{6} P_i l_i$

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

$= 2.625$ binits /msg symbol

Entropy $H(s) = \sum_{i=0}^{6} P_i \log \frac{1}{P_i}$

$= \left[0.25 \log \frac{1}{0.25}\right] * 2 + \left[0.125 \log \frac{1}{0.125}\right] \times 3$

$+ \left[0.0625 \log \frac{1}{0.0625}\right] \times 2$

$= 2.625$ bits /message symbol

Coding efficiency, $\eta_c = \frac{H(S)}{L} = \frac{2.625}{2.625} = 100\%$.

Q. A source produces two symbols S_1 and S_2 with probabilities $7/8$ and $1/8$ respectively. Device a coding scheme using shannon-fano encoding procedure to get a coding efficiency of atleast 75%.

	(P_i)	(Code)	(l_i)
S_1	$7/8$	1	1
S_2	$1/8$	0	1

Average length $L = \sum_{i=1}^{2} P_i l_i$

$= \frac{7}{8} \times 1 + \frac{1}{8} \times 1$

$= 1$ binits / message symbol

Entropy $H(S) = \sum_{i=1}^{2} P_i \log \frac{1}{P_i}$

$= \frac{7}{8} \log \frac{8}{7} + \frac{1}{8} \log 8$

$= 0.54356$

Coding efficiency $\eta_c = \frac{H(S)}{L} = \frac{0.54356}{1}$

			Source S_A		Source S_B		Source S_C		Source S_D		Source S_E	
Source symbol	P_i	Code	P_i	Code	P_i	Code	P_i	Code	P_i	Code	P_i	Code
X_1	0.4	1	0.4	1	0.4	1	0.4	1	0.4	1	0.6	0
X_2	0.2	01	0.2	01	0.2	01	0.2	01	0.4	00	0.4	1
X_3	0.1	0010	0.1	0010	0.2	000	0.2	000	0.2	01		
X_4	0.1	0011	0.1	0011	0.1	0010	0.2	001				
X_5	0.1	0000	0.1	0000	0.1	0011						
X_6	0.05	00010	0.1	0001								
X_7	0.05	00011										

Coding efficiency $\eta_c = \frac{H(S)}{L} = \frac{0.54356}{1}$

$= 54.356\%$

Coding efficiency is less than given efficiency.

Let us consider 2^{nd} extension of the given code.

Symbol	Probability					
$S_1 S_1$	49/64	1				
$S_1 S_2$	7/64	0	7/64	1		
$S_2 S_1$	7/64	0	7/64	0	7/64	1
$S_2 S_2$	1/64	0	1/64	0	1/64	0

Code	length l_i
1	1
01	2
001	3
000	3

$$\text{Average length } L_2 = \sum_{i=1}^{4} P_i l_i$$

$$= \frac{49}{64} \times 1 + \frac{7}{64} \times 2 + \frac{7}{64} \times 3 + \frac{1}{64} \times 3$$

$$= 1.359375 \text{ binits/message symbol}$$

Entropy of second source,

$$H(S^2) = 2H(S)$$

$$= 2 \times 0.54356$$

$$= 1.08712 \text{ bits/message symbol}$$

Code efficiency, $\eta_c^{(2)} = \dfrac{H(S^2)}{L_2}$

$$= \frac{1.08712}{1.359375}$$

$$= 80\%.$$

Shannon-Fano Ternary Code

Procedure :-

1. The symbols are arranged according to non-increasing probabilities.

2. The symbols are divided into three groups so that sum of probabilities in each group is approximately equal

3 All the symbols in the I group are designated by a '2', the 2nd group by a '1' and 3rd group by '0'.

4 The I group is again subdivided into three more sub groups such that each sub group probabilities are approximately same.

5 All the symbols of the I subgroup are designated by a '2' the 2nd subgroup by a '1' and 3rd subgroup by '0'.

6 The 2nd and 3rd groups are subdivided in to three more subgroups each and step 5 is repeated.

7. This process is continued till further subdivision is impossible.

Q. Construct a Shannon-fano ternary code for the following ensemble and find code efficiency and redundancy.

$$S = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7\}$$

$$P = \{0.3, 0.3, 0.12, 0.12, 0.06, 0.06, 0.04\}$$

with $X = \{0, 1, 2\}$

S_1	0.3	2				
S_2	0.3	1				
S_3	0.12	0	0.12	2		
S_4	0.12	0	0.12	1		
S_5	0.06	0	0.06	0	0.06	2
S_6	0.06	0	0.06	0	0.06	1
S_7	0.04	0	0.04	0	0.04	1

Code	l_i
2	1
1	1
02	2
01	2
002	3
001	3
000	3

The average length $L = \sum_{i=1}^{7} P_i l_i$

$= 0.3 \times 1 + 0.3 \times 1 + 0.12 \times 2$

$+ 0.12 \times 2 + 0.06 \times 3$

$+ 0.06 \times 3 + 0.04 \times 3$

$= 1.56$ trinits/message symbol

Entropy $H(S) = \sum_{i=1}^{7} P_i \log \frac{1}{P_i}$

$= 0.3 \log \frac{1}{0.3} + 0.3 \log \frac{1}{0.3} + 0.12 \log \frac{1}{0.12}$

$+ 0.12 \log \frac{1}{0.12} + 0.06 \log \frac{1}{0.06} +$

$0.06 \log \frac{1}{0.06} + 0.04 \log \frac{1}{0.04}$

$= 2.4491$ bits/msg symbol.

The entropy, in r-ary units / message symbol

is $H_r(S) = \dfrac{H(S)}{\log_2 r}$

$= \dfrac{2.4419}{\log_2 3}$

$= 1.5452$ ternary units / msg symbol

Ternary coding efficiency

$\eta_c = \dfrac{H_3(S)}{L}$

$= \dfrac{1.5452}{1.56}$

$= 0.9905$ %

Huffman Coding

This is an encoding procedure for obtaining a compact ~~w~~ code with least redundancy

Procedure.

1. Source symbols are listed in the nonincreasing order of probabilities.
2. If q = no. of source symbols
   r = no. of different symbols used in the code alphabet

   Then $q = r + (r-1)\alpha$

   where α should be an integer. If it is not, add dummy symbols with zero probability, to q symbols, to make the variable 'α' an integer.

Note: for binary code, α will always be an integer. So no need to check this condition.

3. The last 'r' symbols are combined into a single composite symbol by adding their probabilities to get a reduced source S_A. The symbols of this reduced source S_A are now arranged in the order of non-increasing probabilities

4. The last 'r' symbols of source 'S_A' are combined to form another symbol by adding their probabilities to get further reduced source S_B. The symbols of S_B are now arranged in the order of non-increasing probabilities

5. This process of combining last 'r' symbols is continued till we arrive at a last source having r symbols.

6. The last source with 'x' message symbols are now encoded with 'r' different code symbols ie 0, 1, 2 ... (r − 1)

7. In binary coding, the last source with 2 symbols are encoded with two words '0' and '1'. As we coming backwards (to the source) with 3 symbols, either '0' may be recomposed as '00' and '01' or '1' may be recomposed as '10' and '11' depending on which two out of three have been combined to get the last reduced source symbol.

In ternary coding '0' is recomposed as 00, 01 and 02, 1 as 10, 11 and 12 and 2 as 20, 21 and 22 .. etc.

8. As we pass from source to source, the recomposition of one code-word each time is done ~~to~~ to form new code-word

9. This procedure is continued till we assign code-words to all the source symbols of alphabet of source S.

10. If any dummy symbol are used, they are discarded.

Q Given the messages x_1, x_2, x_3 and x_4 with probabilities 0.4, 0.3, 0.2 and 0.1 Construct binary code by applying Huffman encoding procedure. Determine efficiency and redundancy of the code.

Step 1

Symbols are arranged in the nonincreasing order of probabilities.

Source symbol	P_i	Code
x_1	0.4	
x_2	0.3	
x_3	0.2	
x_4	0.1	

Step 2. For a binary code; no need to add dummy variable i.e α is integer always.

Step 3. The last two probabilities are combined to get a composite symbol probability of 0.3. The reduced source X_A has now three probabilities as 0.4, 0.3 and new probability 0.3. These three probabilities of source X_A are arranged in the non increasing order.

Source symbol	P_i	Code	Source X_A P_i	code
x_1	0.4		0.4	
x_2	0.3		0.3	
x_3	0.2		0.3	
x_4	0.1			

Note:

we have 2 options for placing the composite symbol.

method 1: Composite symbol is placed "as low as possible".

method 2: composite symbol is placed "as high as possible".

In the above table, we use method 1.

Step 4:

The last two probabilities of X_A are combined to get a composite symbol probability of 0.6. The reduced source X_B has now two probabilities viz. 0.4 and 0.6.

These two probabilities of source X_B are now arranged in the nonincreasing order.

Source Symbol	P_i	Code	Source X_A P_i	Code	Source X_B P_i	Code
X_1	0.4		0.4		0.6	
X_2	0.3		0.3		0.4	
X_3	0.2		0.3			
X_4	0.1					

Step 5 :

In step No. 4 itself, we arrived at a last source having 'r' symbols (ie here 2 symbols)

Step 5 :

In step No. 4 itself, we arrived at a last source having 'r' symbols (ie here 2 symbols) since it is a binary code

Step No. 6

The last two symbols of source X_B are now encoded with ~~to~~ 0 and 1

Step No. 7

The code corresponding to the pointer which is '0' is now recomposed as 00 and 01. corresponding to the probability of 0.3 and 0.3 of source X_A. The code for 0.4 which is 1 in source X_B is brought backwards and retained as 1 corresponding to probability of 0.4 in X_A

Source symbols	P_i	Code	Source X_A P_i	Source X_A Code	Source X_B P_i	Source X_B Code
x_1	0.4		0.4	1	0.6	0
x_2	0.3		0.3	00	0.4	1
x_3	0.2		0.3	01		
x_4	0.1					

Step No 8 and 9

The arrow mark is now pointing towards the code 01 in source X_A. This code is now recomposed as 010 and 011. These two code words are then assigned to X_3 and X_4. The code 1 corresponds to 0.4 and code 00 corresponds to 0.3 of source X_A are now retained as code words for x_1 and x_2 of the given source. The code written in 3rd column [ie I column for under the label 'code'] is Huffman code.

Source Symbol	P_i	Code	Source X_A P_i	Source X_A Code	Source X_B P_i	Source X_B Code
x_1	0.4	1	0.4	1	0.6	0
x_2	0.3	00	0.3	00	0.4	1
x_3	0.2	010	0.3	01		
x_4	0.1	011				

Average length $L = \sum_{i=1}^{4} p_i l_i$

$= (0.4)(1) + (0.3)(2) + (0.2)(3) + (0.1)(3)$

$= 1.9$ binits / message symbol

Entropy $H(S) = \sum_{i=1}^{4} p_i \log \frac{1}{p_i}$

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

$= 1.846$ bits/message symbol

Code efficiency $\eta_c = \frac{H(S)}{L} = \frac{1.846}{1.9} = 97.15\%$

Code redundancy $R\eta_c = 2.85\%$

Method 2 : As high as possible.

Source symbol	P_i	Code	Source X_A P_i	Source X_A Code	Source X_B P_i	Source X_B Code
x_1	0.4	1	0.4	1	0.6	0
x_2	0.3	01	0.3	00	0.4	1
x_3	0.2	000	0.3	01		
x_4	0.1	001				

(It is left to the students to compute efficiency and redundancy)

Q. Given the message x_1, x_2, x_3, x_4, x_5 and x_6 with respective probabilities of 0.4, 0.2, 0.2, 0.1, 0.07 and 0.03. Construct a binary code by applying Huffman coding procedure. Determine the efficiency and redundancy of the code.

Source Symbol	P_i	Code	Source X_A P_i	Source X_A Code	Source X_B P_i	Source X_B Code	Source X_C P_i	Source X_C Code	Source X_D P_i	Source X_D Code
x_1	0.4	1	0.4	1	0.4	1	0.4	1	0.6	0
x_2	0.3	01	0.2	01	0.2	01	0.4	00	0.4	1
x_3	0.2	000	0.2	000	0.2	000	0.2	01		
x_4	0.1	0010	0.1	0010	0.2	001				
x_5	0.07	00110	0.1	0011						
x_6	0.03	00111								

2) Consider a zero memory source with $S = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$ and $P = \{0.4, 0.2, 0.1, 0.1, 0.1, 0.05, 0.05\}$

(a) Construct binary Huffman code by placing composite symbol as low as possible

(b) Compute the code by moving the symbol as high as possible.

(c) Compute the variance in each case.

The average length $L = \sum_{i=1}^{7} P_i l_i$

$= (0.4)(1) + (0.2)(2) + (0.1)(4) + (0.1)(4) + (0.1)(4) + (0.05)(5) + (0.05)(5)$

$= 2.5$ binits / message symbol

Variance, $Var(x) = E\left[(x-\mu)^2\right]$

μ = average value

Variance of word length, $Var(l_i) = E\left[(l_i - L)^2\right]$

$= \sum_{i=1}^{7} P_i (l_i - L)^2$

$= (0.4)(1-2.5)^2 + (0.2)(2-2.5)^2 + (0.1)(4-2.5)^2 + (0.1)(4-2.5)^2 + (0.1)(4-2.5)^2 + (0.05)(5-2.5)^2 + (0.05)(5-2.5)^2$

$= 2.25$

2. As high as possible.

Average length $L = \sum_{i=1}^{7} P_i l_i$

$= (0.4)(2) + (0.2)(2) + (0.1)(3) + (0.1)(3)$
$+ (0.1)(3) + (0.05)(4) + (0.05)(4)$

$= 2.5$ bints / message symbol

$Var(l_i) = E[(l_i - L)^2]$

$= (0.4)(2-2.5)^2 + (0.2)(2-2.5)^2 +$
$(0.1)(3-2.5)^2 + (0.1)(3-2.5)^2 +$
$(0.1)(3-2.5)^2 + (0.05)(4-2.5)^2$
$+ (0.05)(4-2.5)^2$

$= 0.45$

			Source SA		Source SB		Source SC		Source SD		Source SE	
Source symbol	P_i	Code	P_i	Code	P_i	Code	P_i	Code	P_i	Code	P_i	Code
S_1	0.4	00	0.4	00	0.4	00	0.4	00	0.4	1	0.6	0
S_2	0.2	11	0.2	11	0.2	10	0.2	01	0.4	00	0.4	1
S_3	0.1	011	0.1	010	0.2	11	0.2	10	0.2	01		
S_4	0.1	100	0.1	011	0.1	010	0.2	11				
S_5	0.1	101	0.1	100	0.1	011						
S_6	0.05	0100	0.1	101								
S_7	0.05	0101										

Note:
When the composite symbol is moved as high as possible, the variance of the word lengths over the ensemble of source symbols would become smaller, which is desirable.

3. Construct a source with 8 alphabet. A to H with respective probabilities of 0.22, 0.20, 0.18, 0.15, 0.10, 0.08, 0.05, 0.02.

(a) Construct a binary compact (Huffman code) and determine the code efficiency

(b) Construct a ternary compact code and determine the efficiency

(c) Construct a quartenary compact code and determine efficiency. Compare and comment on the result

Source symbol	P_i	Code	Source S_A P_i	Source S_A Code	Source S_B P_i	Source S_B Code	Source S_C P_i	Source S_C Code	Source S_D P_i	Source S_D Code	Source S_E P_i	Source S_E Code	Source S_F P_i	Source S_F Code
A	0.22	10	0.22	10	0.22	10	0.25	01	0.33	00	0.42	1	0.58	0
B	0.20	11	0.20	11	0.20	11	0.22	10	0.25	01	0.33	00	0.42	1
C	0.18	000	0.18	000	0.18	000	0.20	11	0.22	10	0.25	01		
D	0.15	001	0.15	001	0.15	001	0.18	000	0.20	11				
E	0.10	011	0.10	011	0.15	010	0.15	001						
F	0.08	0100	0.08	0100	0.10	011								
G	0.05	01010	0.07	0101										
H	0.02	01011												

Average length $L^{(2)} = \sum_{i=1}^{8} P_i l_i$

$= (0.22)(2) + (0.20)(2) + (0.18)(3) + (0.15)(3) + (0.10)(3) + (0.08)(4) + (0.05)(5) + (0.02)(5)$

$= 2.8$ binits / message symbol

$H(S) = \sum_{i=1}^{8} P_i \log \frac{1}{P_i}$

$= 0.22 \log \frac{1}{0.22} + 0.20 \log \frac{1}{0.20} + 0.18 \log \frac{1}{0.18} + 0.15 \log \frac{1}{0.15} + 0.10 \log \frac{1}{0.10} + 0.08 \log \frac{1}{0.08} + 0.05 \log \frac{1}{0.05} + 0.02 \log \frac{1}{0.02}$

$= 2.7535$ bits / message symbol

Code efficiency $\eta_c = H(S)/L = \frac{2.7535}{2.8} = 98.34\%$

(b) Huffman Ternary code.

We know $q = r + (r-1)\alpha$ [$\because$ For binary, this eqn automatically satisfies, so no need to check]

For ternary code $r = 3$

$$q = 3 + 2\alpha$$

$$\alpha = \frac{q-3}{2}$$

α will be an integer when $q = 5, 7, 9, 11, 13, \ldots$

In the problem we have $q = 8$ then the select new q > old q

So $q = 9$ (select)

So we let us add " one dummy variable I ' with zero probability to the message. With this dummy symbol I, q becomes 9.

Source symbol	P_i	Code	Source S_1 P_i	Code	Source S_2 P_i	Code	Source S_3 P_i	Code
A	0.22	2	0.22	2	0.25	1	0.53	0
B	0.20	00	0.20	00	0.22	2	0.25	1
C	0.18	01	0.18	01	0.20	00	0.22	2
D	0.15	02	0.15	02	0.18	01		
E	0.10	10	0.10	10	0.15	02		
F	0.08	11	0.08	11				
G	0.05	120	0.07	12				
H	0.02	121						
I	0	122						

Average length $L^{(3)} = \sum_{i=1}^{8} P_i l_i$

$= (0.22)(1) + (0.20)(2) + (0.18)(2) + (0.15)(2) + (0.10)(2) + (0.08)(2) + (0.05)(3) + (0.02)(3)$

$= 1.85$ trinits / message symbol.

Entropy when $r = 3$, $H_3(S) = \dfrac{H(S)}{\log_2 3}$ $\left[\because H_r(S) = \dfrac{H(S)}{\log r}\right]$

$= \dfrac{2.7535}{\log_2 3} = 1.7373$ ternary units / message symbol

Code efficiency $\eta_{c(3)} = \dfrac{H_3(S)}{L^{(3)}} = \dfrac{1.7373}{1.85} = 93.91\%$

(c) Huffman quarternary code. $(r = 4)$

$$q = r + (r-1)\alpha$$

$$q = 4 + 3\alpha$$

$\alpha = \dfrac{q-4}{3}$; α is an integer when $q = 7, 10, 13, 16 \ldots$

Take $q = 10$

We have $q = 8$ which makes $\alpha = 1\frac{1}{3}$. So add two dummy variables I and J, each with zero probability to the message.

with $q = 10$, $\alpha = \dfrac{10-4}{3} = 2$ = an integer.

			Source Sa		Source Sb	
Source Symbol	P_i	Code	P_i	Code	P_i	Code
A	0.22	1	0.22	1	0.40	0
B	0.20	2	0.20	2	0.22	1
C	0.18	3	0.18	3	0.20	2
D	0.15	00	0.15	00	0.18	3
E	0.10	01	0.10	01		
F	0.08	02	0.08	02		
G	0.05	030	0.07	03		
H	0.02	031				
I	0	032				
J	0	033				

Average length $L^{(4)} = \sum_{i=1}^{8} P_i l_i$

$= (0.22)(1) + (0.20)(1) + (0.18)(1) + (0.15)(2)$
$+ (0.10)(2) + (0.08)(2) + (0.05)(3)$
$+ (0.02)(3)$

$L^{(4)} = 1.47$ quaternary digits/message symbol

Entropy is quaternary units/message symbol

$$H_4(S) = \frac{H(S)}{\log_2 4} = \frac{2.7535}{2}$$

$= 1.37675$ quaternary units/message symbol.

Code efficiency $\eta_{c(4)} = \frac{H_4(S)}{L^{(4)}} = \frac{1.37675}{1.47}$

$= 93.66\ \%$

Comparison

Design a quaternary and binary source code for the source $S = (s_1, s_2, s_3, s_4, s_5, s_6, s_7)$ with $P = \left\{ \frac{9}{32}, \frac{3}{32}, \frac{2}{32}, \frac{9}{32}, \frac{3}{32}, \frac{3}{32} \right\}$. & the code alphabet $X = \{0, 1, 2, 3\}$ and $X = \{0, 1\}$

Find coding efficiency.

<u>Quaternary code</u>

$$q = r + (r-1)\alpha$$

Given $q = 7$ and $r = 4$

$$\alpha = \frac{q-r}{r-1} = \frac{7-4}{4-1} = 1 = \text{an integer}$$

~~So~~ So dummy symbols are not necessary.

Source symbol	p_i	Quaternary Code	Source p_i	SA Code
S_1	9/32	1	11/32	0
S_5	9/32	2	9/32	1
S_2	3/32	3	9/32	2
S_3	3/32	0 0	3/32	3
S_6	3/32	0 1		
S_7	3/32	0 2		
S_4	2/32	0 3		

Average length $L^{(4)} = \sum_{i=1}^{7} P_i l_i$

$$= \left(\frac{9}{32}\right)(1) + \left(\frac{9}{32}\right)(1) + \left(\frac{3}{32}\right)(1) + \left(\frac{3}{32}\right)(2) + \left(\frac{3}{32}\right)(2) + \left(\frac{3}{32}\right)(2) + \left(\frac{2}{32}\right)(2)$$

$$= \frac{43}{32} = 1.34375$$ quaternary digits / message symbol

Entropy in quaternary units / message symbol with $r = 4$

$$H_4(S) = \frac{H(S)}{\log_2 4} = \frac{2.56}{\log_2 4} = 1.28$$ quaternary unit / message symbol

(∵ binary code entropy)

Code efficiency $\eta_c^{(4)} = \frac{H_4(S)}{L^{(4)}} = \frac{1.28}{1.34375} \times 100$

$= 95.26\ \%$

Source symbol	P_i	Binary code	Source S_A P_i	Code	Source S_B P_i	Code	Source S_C P_i	Code	Source S_D P_i	Code	Source S_E P_i	Code
S_1	9/32	00	9/32	00	9/32	00	9/32	00	14/32	1	18/32	0
S_5	9/32	01	9/32	01	9/32	01	9/32	01	9/32	00	14/32	1
S_2	3/32	101	5/32	100	6/32	11	8/32	10	9/32	01		
S_3	3/32	110	3/32	101	5/32	100	6/32	11				
S_6	3/32	111	3/32	110	3/32	101						
S_7	3/32	1000	3/32	111								
S_4	2/32	1001										

$$\text{Average length} = \sum_{i=1}^{7} P_i l_i = \left(\frac{9}{32}\right) \times 2 + \frac{9}{32} \times 2 + \frac{3}{32} \times 3 + \frac{3}{32} \times 3 + \frac{3}{32} \times 3 + \frac{3}{32} \times 4 + \frac{2}{32} \times 4$$

$= 2.59375$ binits/message symbol.

Entropy $H(s) = \sum_{i=1}^{7} P_i \log \frac{1}{P_i}$

$$= \left(\frac{9}{32} \log \frac{32}{9}\right) \times 2 + \left(\frac{3}{32} \log \frac{32}{3}\right) \times 4$$

$$+ \frac{2}{32} \log \frac{32}{2}$$

$H(s) = 2.56$ bits / message symbol

Coding efficiency $\eta_c = \frac{H(s)}{L} = \frac{2.56}{2.59375}$

$\eta_c = 98.7\%$

Q The five symbols S_1, S_2, S_3, S_4 and S_5 of a source have probabilities 0.4, 0.2, 0.2, 0.1 and 0.1 respectively. Find code words using Huffman algorithm.

Source Symbol	P_i	Binary Code	Source SA P_i	Source SA Code	Source SB P_i	Source SB Code	Source SC P_i	Source SC Code
S_1	0.4	1	0.4	1	0.4	1	0.6	0
S_2	0.2	01	0.2	01	0.4	00	0.4	1
S_3	0.2	000	0.2	000	0.2	01		
S_4	0.1	0010	0.2	001				
S_5	0.1	0011						

Design a ternary code for the source $S = \{s_1, s_2, s_3, s_4, s_5, s_6\}$ with $P = \left\{\frac{1}{3}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{12}, \frac{1}{12}\right\}$ Using Huffman Coding. Code alphabet $X = \{0, 1, 2\}$

We know $q = r + (r-1)\alpha$

For ternary code $r = 3$

$q = 6$ (given)

$$\alpha = \frac{q - r}{r-1} = \frac{6-3}{3-1} = 3/2$$

α becomes an integer for $q = 5, 7, 9, 11 \ldots$

So next higher integer to $q = 6$ in the sequence is 7. So add one dummy symbol.

			Source S_A		Source S_B	
Source Symbol	P_i	Ternary code	P_i	Code	P_i	Code
s_1	1/3	1	1/3	1	5/12	0
s_2	1/4	2	1/4	2	1/3	1
s_3	1/8	0 1	1/6	0 0	1/4	2
s_4	1/8	0 2	1/8	0 1		
s_5	1/12	0 0 0	1/8	0 2		
s_6	1/12	0 0 1				
s_7	0	0 0 2				

Q. Apply Huffman encoding procedure for the following set of messages and determine the efficiency of the binary code.

x_1	x_2	x_3
0.7	0.15	0.15

If the same technique is applied to the 2^{nd} extension for the above messages, how much will the efficiency be improved?

Source Symbol	P_i	Code	Source P_i	SA Code
x_1	0.7	0	0.7	0
x_2	0.15	10	0.3	1
x_3	0.15	11		

Average length $L = \sum_{i=1}^{3} P_i l_i$

$= (0.7)(1) + (0.15)(2) + (0.15)(2)$

$= 1.3$ bits / message symbol.

Entropy $H(s) = \sum_{i=1}^{3} P_i \log \frac{1}{P_i}$

$= 0.7 \log \frac{1}{0.7} + 2 \times 0.15 \log \frac{1}{0.15}$

$= 1.18 13$ bits / message symbol

Code efficiency $\eta_c = \frac{H(s)}{L} = \frac{1.1813}{1.3}$

$= 90.87\ \%$

2^{nd} extension

The 2^{nd} extension will have $3^2 = 9$ symbols. ie x_1x_1, x_1x_2, x_1x_3, x_2x_1, x_2x_2, x_2x_3, x_3x_1, x_3x_2 and x_3x_2 which ~~are~~ should be arranged in nonincreasing order of probabilities.

Source symbol	P_i	Code	Source S_A P_i	Code	Source S_B P_i	Code	Source S_C P_i	Code	Source S_D P_i	Code	Source S_E P_i	Code	Source S_F P_i	Code	Source S_G P_i	Code
x_1x_1	0.49	1	0.49	1	0.49	1	0.49	1	0.49	1	0.49	1	0.49	1	0.51	0
x_1x_2	0.105	001	0.105	001	0.105	001	0.105	001	0.195	000	0.21	01	0.30	00	0.49	1
x_1x_3	0.105	010	0.105	010	0.105	010	0.105	010	0.105	001	0.195	000	0.21	01		
x_2x_1	0.105	011	0.105	011	0.105	011	0.105	011	0.105	010	0.105	001				
x_3x_1	0.105	0000	0.105	0000	0.105	0000	0.105	0000	0.105	011						
x_2x_2	0.0225	000110	0.045	00010	0.045	00010	0.090	0001								
x_2x_3	0.0225	000111	0.0225	000110	0.045	00011										
x_3x_2	0.0225	000100	0.0225	000111												
x_3x_3	0.0225	000101														

Average length $L_2 = \sum_{i=1}^{9} P_i l_i$

$= (0.49)(1) + (0.105)(3) + (0.105)(3) + (0.105)3 + (0.105)4 + (0.025)(6 + (0.0225)(6)) + (0.0225)(6) + (0.0225)(6)$

$= 2.395$ binits/message symbol

Entropy $H(S^2) = 2H(S)$

$= 2(1.1813)$

$= 2.3626$ bits/message symbol

Code efficiency $= \eta_c^{(2)} = \frac{2.3626}{2.395} = 98.65\%$

The efficiency improves by 98.65% ⟵

90.87%

= $\underline{\underline{7.78\%}}$

Channel Capacity

The capacity of a discrete memoryless noisy channel is defined as the max. possible rate of information transmission over the channel

$$C = \text{Max}\{R_t\}$$

$$C = \text{Max}\left[H(A) - H(A/B)\right] r_s$$

Proof

We have the entropy of the input symbol is

$$H(A) = \sum_{i=1}^{r} p(a_i) \log \frac{1}{p(a_i)} \text{ bits/msg symbol}$$

If the discrete memoryless channel accepting symbol at the rate of r_s message-symbol/sec. Then the average rate at which information is going into the channel is given by

$$R_{in} = H(A)\, r_s \text{ bits/sec}$$

Some amount of information is lost in the channel due to noise. This lost information is $H(A/B)$.

The net amount of information is called mutual information

We have $I(A,B) = H(A) - H(A/B)$

bits/message symbol

The average rate of information transmission

$$R_t = I(A,B) \cdot r_s \text{ bits/sec}$$

$$R_t = \left[H(A) - H(A/B)\right] r_s \text{ bits/sec}$$

We know $I(A,B) = I(B,A)$

$$R_t = \left[(H(B) - H(B/A)\right] r_s \text{ bits/sec.}$$

Shannon's Theorem on channel capacity (Shannon's II Theorem)

"When the rate of information transmission $R_t \leq C$, then there exist a coding technique which enables transmission over a channel with a small probability of error as possible, even in the presence of noise in the channel".

ie For $R_t \leq C$; transmission without error even in the presence of noise.

If $R_t > C$ then reliable transmission of information is not possible without errors. Thus when $R_t > C$, then the errors can't be controlled by any coding technique and probability of error of receiving the correct message becomes close to unity.

Redundancy and efficiency of a channel

Channel efficiency $\eta_{ch} = \frac{R_t}{C} \times 100\%$

We have $R_t = [H(A) - H(A/B)]\, r_s$

and $C = \text{Max}\, [(H(A) - H(A/B)]\, r_s$

$$\eta_{ch} = \frac{[H(A) - H(A/B)]\, r_s}{\text{Max}\,([H(A) - H(A/B)]\, r_s)} \times 100\%$$

$$\eta_{ch} = \frac{I(A,B)}{\text{Max}\,(I(A,B))} \times 100\%$$

Channel redundancy $R\eta_{ch} = 1 - \eta_{ch}$

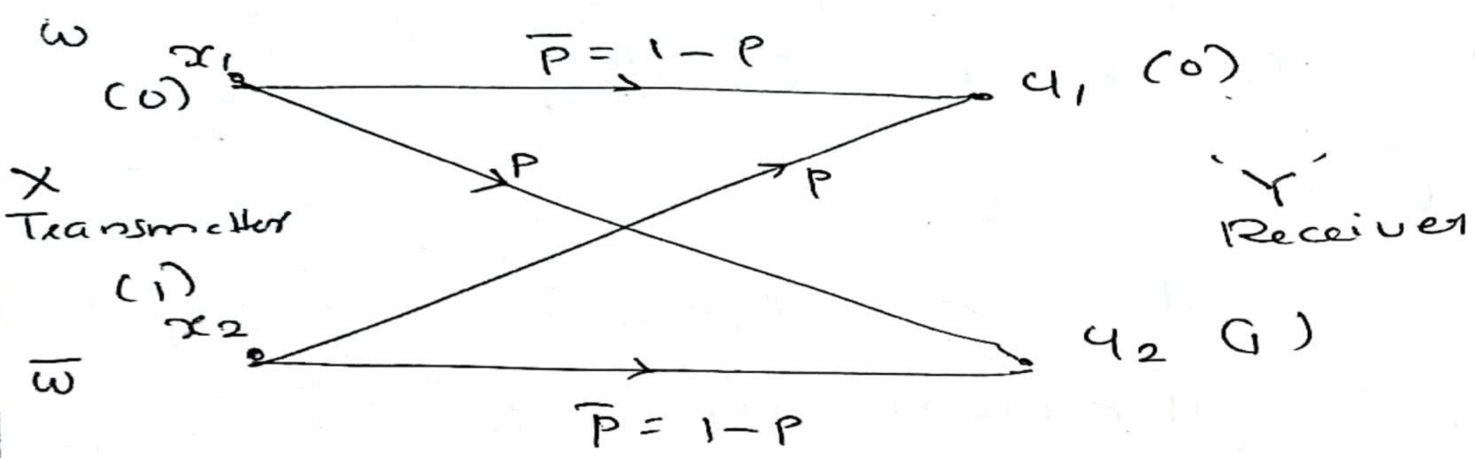
Special Channels

1. Symmetric / Uniform channels
2. Binary symmetric channel (BSC)
3. Binary Erasure channel (BEC)
4. Noiseless channel
5. Deterministic channel
6. Cascade channel.

Binary Symmetric Channel (BSC)

A channel is said to be symmetric or uniform channel, if the second and subsequent rows of the channel matrix contains the same elements as that of first row, but in a different order

Channel diagram



Let p = probability of error

= probability of reception of '1' when '0' is transmitted

or

probability of reception of '0' when '1' is transmitted

Let $P(x_1) = \omega$ & $P(x_2) = 1 - \omega$
$= \bar{\omega}$

The symbol x_1 is encoded as '0' and x_2 as '1'

Channel matrix from channel diagram

$$P(y/x) = \begin{array}{c|cc} & y_1 & y_2 \\ \hline x_1 & P(y_1/x_1) & P(y_2/x_1) \\ x_2 & P(y_1/x_2) & P(y_2/x_2) \end{array}$$

$$P(y/x) = \begin{bmatrix} \bar{p} & p \\ p & \bar{p} \end{bmatrix}$$

$\Rightarrow$ symmetric matrix ($\because$ both row contains same elements in a different order)

For a symmetric channel

$$H(Y/X) = h = \sum_{j=1}^{s} P_j \log \frac{1}{P_j} =$$

a constant which remains same for all rows since all the rows consist of same elements as that of the I row.

For a binary symmetric channel, $S=2$

$$H(Y/X) = h_2 \sum_{j=1}^{2} P_i \log \frac{1}{P_j}$$

$$= P_1 \log \frac{1}{P_1} + P_2 \log \frac{1}{P_2}$$

$$H(Y/X) = \bar{p} \log \frac{1}{\bar{p}} + p \log \frac{1}{p}$$

Entropy of output symbol,

$$H(Y) = \sum_{j=1}^{2} P(y_j) \log \frac{1}{P(y_j)}$$

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

We know, output probability can be found using total probability theorem.

~~$H(X) =$~~

$$P(y_1) = p(y_1/x_1)\,p(x_1) + p(y_1/x_2)\,p(x_2)$$

$$= \bar{p}\omega + p\bar{\omega}$$

$$P(y_2) = p(y_2/x_1)\,p(x_1) + p(y_2/x_2)\,p(x_2)$$

$$= p\omega + \bar{p}\bar{\omega}$$

Substitute $p(y_1)$ and $p(y_2)$ in $H(Y)$.

$$H(Y) = (\bar{p}\omega + p\bar{\omega})\log\frac{1}{(\bar{p}\omega + p\bar{\omega})} + (p\omega + \bar{p}\bar{\omega})\log\frac{1}{(p\omega + \bar{p}\bar{\omega})}$$

Mutual information of a BSC is,

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

$$= (\bar{p}\omega + p\bar{\omega}) \log \frac{1}{(\bar{p}\omega + p\bar{\omega})} + (p\omega + \bar{p}\bar{\omega})$$

$$\log \frac{1}{(p\omega + \bar{p}\bar{\omega})} - \left[\bar{p} \log \frac{1}{\bar{p}} + p \log \frac{1}{p}\right]$$

We know $C = \log s - h$

Since BSC is a symmetric channel, channel capacity with $r_s = 1$ message-symbol per sec is found by,

$$C = \log s - h \text{ bits/sec}$$

$$= \log 2 - h$$

$$= 1 - h$$

$$C = 1 - \left[\bar{p} \log \frac{1}{\bar{p}} + p \log \frac{1}{p}\right] \text{ bits/sec}$$

when $\omega = \bar{\omega} = \frac{1}{2}$

$$H(Y) = \frac{1}{2}(p+\bar{p}) \log \frac{1}{\frac{1}{2}(p+\bar{p})} + \frac{1}{2}(p+\bar{p}) \log \frac{1}{\frac{1}{2}(p+\bar{p})}$$

$$= \frac{1}{2} \log 2 + \frac{1}{2} \log 2 \qquad [\because p+\bar{p} = 1$$

$$H(Y) = 1$$

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

$$= 1 - h$$

$$I(X,Y) = 1 - \left[\bar{p} \log \frac{1}{\bar{p}} + p \log \frac{1}{p}\right]$$

ie when input symbols become equiprobable the mutual information maximizes and becomes equal to channel capacity C

Q) A binary symmetric channel has the following noise matrix with source probabilities $p(x_1) = 2/3$ and $p(x_2) = 1/3$

$$p(y/x) = \begin{array}{c c} & \begin{array}{cc} y_1 & y_2 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \end{array} & \begin{bmatrix} 3/4 & 1/4 \\ 1/4 & 3/4 \end{bmatrix} \end{array}$$

Q) 1. Determine $H(x)$, $H(y)$, $H(x,y)$, $H(y/x)$, $H(x/y)$ and $I(x,y)$

2. Determine channel capacity

3. Find channel efficiency and redundancy.

From channel ~~di~~ diagram of BSC, we know

$$p = \frac{1}{4} \quad \bar{p} = 3/4 \quad \omega = 2/3 \text{ and } \bar{\omega} = 1/3$$

$$H(x) = \sum_{i=1}^{2} p(x_i) \log \frac{1}{p(x_i)}$$

$$= \frac{2}{3} \log 3/2 + 1/3 \log 3$$

$$= 0.9183 \text{ bits/message symbol}$$

$$H(y) = (\bar{p}\omega + p\bar{\omega}) \log \frac{1}{(\bar{p}\omega + p\bar{\omega})} + (p\omega + \bar{p}\bar{\omega})$$

$$\log \frac{1}{(p\omega + \bar{p}\bar{\omega})}$$

$$\bar{p}\omega + p\bar{\omega} = \left(\frac{3}{4}\right) \cdot \left(\frac{2}{3}\right) + \left(\frac{1}{4}\right) \cdot \left(\frac{1}{3}\right)$$

$$= 7/12$$

$$p\omega + \bar{p}\bar{\omega} = \left(\frac{1}{4}\right)\left(\frac{2}{3}\right) + \left(\frac{3}{4}\right)\left(1/3\right)$$

$$= 5/12$$

$$H(Y) = \frac{7}{12} \log \frac{12}{7} + \frac{5}{12} \log \frac{12}{5}$$

$$= 0.9799 \text{ bits / message symbol}$$

$$H(y/x),\ h = \bar{p}\log\frac{1}{\bar{p}} + p\log\frac{1}{p}$$

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

$$= 0.8113 \text{ bits /message symbol}$$

$$H(x,y) = H(x) + H(y/x)$$

$$= 0.9183 + 0.8113$$

$$= 1.7296 \text{ bits /message symbol}$$

$$H(y/x) = H(x/y) = H(y) + H(x/y)$$

$$H(x/y) = H(x,y) - H(y)$$

$$= 1.7296 - 0.9799$$

$$= 0.7497 \text{ bits /message symbol}$$

$$I(x,y) = H(x) - H(x/y) \text{ or } H(y) - H(y/x)$$

$$= 0.9183 - 0.7497$$

$$= 0.1686 \text{ bits/message symbol}$$

Channel capacity, $C = 1 - h$

$$= 1 - H(y/x)$$

$$= 1 - 0.8113$$

$$= 0.1887 \text{ bits/message symbol}$$

Channel efficiency $= \dfrac{I(x,y)}{C}$

$$= \frac{0.1686}{0.1887}$$

$$\eta_{ch} = 89.35\%$$

Channel redundancy $R_{ch} = 10.65\%$

Q. A message source produces two independent symbols A and B with probabilities $P(A) = 0.4$ and $P(B) = 0.6$. Calculate the efficiency of the source and hence its redundancy. If the symbols are received in average with 4 in every 100 symbols in error, calculate transmission rate of the system.

Given, transmitter X produces two symbols A and B with probabilities 0.4 and 0.6 respectively.

Entropy of source X is,

$$H(X) = P(A)\log\frac{1}{P(A)} + P(B)\log\frac{1}{P(B)}$$

$$= 0.4\log\frac{1}{0.4} + 0.6\log\frac{1}{0.6}$$

$$= 0.97095 \text{ bits/message symbol.}$$

$$H(X)_{max} = \log q$$

$$= \log 2$$

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

Source efficiency = $\dfrac{H(x)}{H(x)_{max}}$

$= \dfrac{0.97095}{1}$

$= 97.095\,\%$

Source redundancy $R_{\eta c} = 2.905\,\%$

$P =$ probability of error

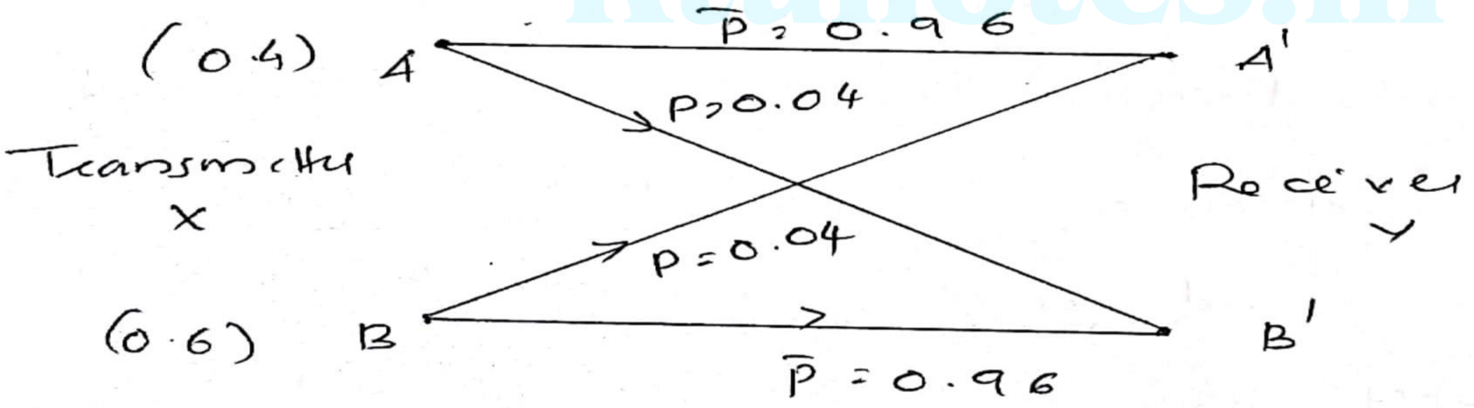
$$= \frac{4}{100}$$

$$P = 0.04$$

$$\bar{P} = 1 - 0.04$$

$$= 0.96$$

Channel diagram



$\omega = P(A) = 0.4$ and $\bar{\omega} = P(B) = 0.6$

$$I(x,y) = (\bar{p}\,\omega + p\,\bar{\omega}) \log \frac{1}{(\bar{p}\omega + p\,\bar{\omega})} + (p\,\omega + \bar{p}\,\bar{\omega}) \log \frac{1}{(p\omega + \bar{p}\,\bar{\omega})} - \left[\bar{p} \log \frac{1}{\bar{p}} + p \log \frac{1}{p}\right]$$

$$\bar{p}\,\omega + p\,\bar{\omega} = (0.96 \times 0.4) + (0.04 \times 0.6)$$
$$= 0.408$$

$$p\,\omega + \bar{p}\,\bar{\omega} = (0.04 \times 0.4) + (0.96 \times 0.6)$$
$$= 0.592$$

$$I(x,y) = 0.408 \log \frac{1}{0.408} + 0.592 \log \frac{1}{0.592} - \left[0.04 \log \frac{1}{0.04} + 0.96 \log \frac{1}{0.96}\right]$$

$= 0.7331$ bits / message symbol.

Transmission rate $R_t = I(x,y) \cdot r_s$

$= 0.7331 \times 100$

$R_t = 73.31$ bit/sec.

Q. A binary channel has the following characteristics

$$P(y/x) = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$$

If input symbols are transmitted with probabilities 3/4 and 1/4 respectively. find entropies $H(x)$, $H(x,y)$ and $H(y/x)$.

From general Channel diagram, we have

$$p = 1/3 \qquad \bar{p} = 2/3$$

$$\omega = 3/4 \qquad \bar{\omega} = 1/4$$

$$H(x) = \sum_{i=1}^{2} p(x_i) \log \frac{1}{p(x_i)}$$

$$= \omega \log \frac{1}{\omega} + \bar{\omega} \log \frac{1}{\bar{\omega}}$$

$$= 3/4 \log 4/3 + \frac{1}{3} \log 4$$

$$= 0.8113 \text{ bits/message symbol}$$

$$H(y/x) = h = \bar{p} \log \frac{1}{\bar{p}} + p \log 1/p$$

$$= 2/3 \log 3/2 + 1/3 \log 3$$

$$= 0.9183 \text{ bits/message symbol}$$

$$H(x, y) = H(x) + H(y/x)$$

$$= 0.8113 + 0.9183$$

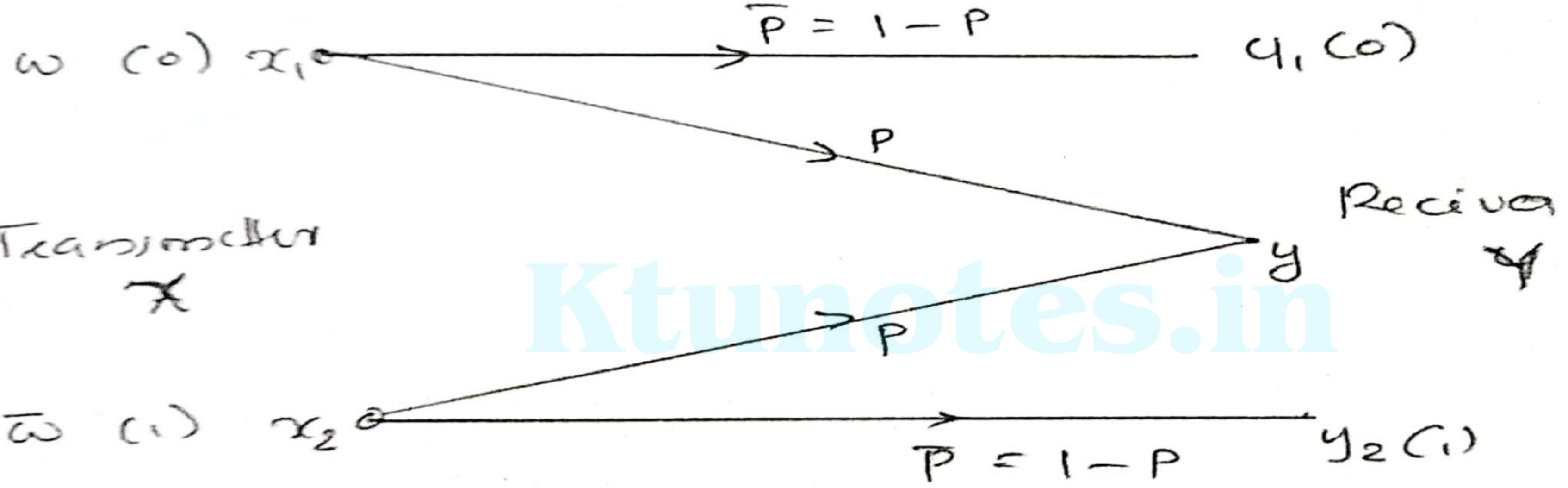
$$= 1.7296 \text{ bits/message symbol}$$

Binary Erasure Channel (BEC)

This is the most important channel used in digital communication. Whenever an error occurs, the symbol will be received as 'y' and no decision will be made about the information, but an immediate request will be made through a receiver channel for retransmission of the transmitted signal till a correct symbol is received at the output. This ensures 100% correct data recovery. Since the error is totally erased in this type of channel, it is called binary erasure channel.

Disadv: A reverse channel is required.

Channel diagram



Channel matrix

From Channel diagram,

$$p(y/x) = p(y_j/x_i) = \begin{array}{c} \\ x_1 \\ x_2 \end{array} \begin{array}{c} \begin{array}{ccc} y_1 & y & y_2 \end{array} \\ \begin{bmatrix} \bar{p} & p & 0 \\ 0 & p & \bar{p} \end{bmatrix} \end{array}$$

Let $p(x_1) = \omega$ $\quad p(x_2) = \bar{\omega}$

$$\omega + \bar{\omega} = 1$$

$$p + \bar{p} = 1$$

$$H(y/x) = h = \sum_{j=1}^{s} p_j \log \frac{1}{p_j}$$

$$h = \bar{p} \log \frac{1}{\bar{p}} + p \log \frac{1}{p}$$

$$H(x) = \sum_{i=1}^{2} p(x_i) \log \frac{1}{p(x_i)}$$

$$= \omega \log \frac{1}{\omega} + \bar{\omega} \log \frac{1}{\bar{\omega}}$$

We know $p(x_i, y_j) = p(x_i) \cdot p(y_j / x_i)$

ie Multiply first row of matrix $p(y_j/x_i$ with $p(x_1) = \omega$ and 2^{nd} row with $p(x_2), \bar{\omega}$, we get JPM $p(x,y)$

$$p(x,y) = p(x_i, y_j) = \begin{array}{c|ccc} & y_1 & y & y_2 \\ \hline x_1 & \bar{p}\,\omega & p\omega & 0 \\ x_2 & 0 & p\bar{\omega} & \bar{p}\,\bar{\omega} \end{array}$$

$$p(x/y) = p(x_i/y_j) = \begin{bmatrix} \dfrac{\bar{p}\,\bar{\omega}}{\bar{p}\,\bar{\omega}} & \dfrac{p\,\omega}{p} & 0 \\ 0 & \dfrac{p\,\bar{\omega}}{p} & \dfrac{\bar{p}\,\bar{\omega}}{\bar{p}\,\bar{\omega}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \omega & 0 \\ 0 & \bar{\omega} & 1 \end{bmatrix}$$

We know $H(X/Y) = \sum_{i=1}^{2} \sum_{j=1}^{2} p(x_i, y_j) \log \dfrac{1}{p(x_i/y_j)}$

$$= \bar{p}\,\omega \log 1 + p\,\omega \log \frac{1}{\omega} + p\,\bar{\omega} \log \frac{1}{\bar{\omega}} + \bar{p}\,\bar{\omega} \log 1$$

$$H(X/Y) = p\left[\omega \log \frac{1}{\omega} + \bar{\omega} \log \frac{1}{\bar{\omega}}\right]$$

$$p(x/y) = p(x_i/y_j) = \begin{bmatrix} \frac{\bar{p}\,\bar{\omega}}{\bar{p}\,\bar{\omega}} & \frac{p\,\omega}{p} & 0 \\ 0 & \frac{p\,\bar{\omega}}{p} & \frac{\bar{p}\,\bar{\omega}}{\bar{p}\,\bar{\omega}} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \omega & 0 \\ 0 & \bar{\omega} & 1 \end{bmatrix}$$

We know $H(x/y) = \sum_{i=1}^{2} \sum_{j=1}^{2} p(x_i, y_j) \log \frac{1}{p(x_i/y_j)}$

$$= \bar{p}\,\omega \log 1 + p\,\omega \log \frac{1}{\omega} + p\,\bar{\omega} \log \frac{1}{\bar{\omega}} + \bar{p}\,\bar{\omega} \log 1$$

$$H(x/y) = p\left[\omega \log \frac{1}{\omega} + \bar{\omega} \log \frac{1}{\bar{\omega}}\right]$$

Mutual information $I(x,y) = H(x) - H(x/y)$

$$= \omega \log \frac{1}{\omega} + \bar{\omega} \log \frac{1}{\bar{\omega}} - p\left[\omega \log \frac{1}{\omega} + \bar{\omega} \log \frac{1}{\bar{\omega}}\right]$$

$$= [1-p]\left[\omega \log \frac{1}{\omega} + \bar{\omega} \log \frac{1}{\bar{\omega}}\right]$$

$$I(x,y) = \bar{p}\, H(x)$$

Channel capacity $C = \operatorname{Max}\left[I(x,y)\right]$

$$= \max\left[\bar{p}\, H(x)\right]$$

$$= \bar{p}\, H(x)_{max}$$

$$= \bar{p} \log 2 \quad \left[\because H(x)_{max} = \log q = \log 2 = 1\right]$$

$$= \bar{p} \cdot 1$$

$$C = \bar{p}$$

[∵ Here we can't use $I(x,y) = H(y) - H(y/x)$, because $H(y)$ involves 'y' which is rejected @ receiver.