G. Anitha (ECE).

Five symbols of the alphabet of DMS and their probabilities are given below $S = \{S_0, S_1, S_2, S_3, S_4\}$ $P(S) = \{0.4, 0.2, 0.2, 0.1, 0.1\}$. Code the symbol using Huffman Coding. Find the efficiency of the code.

Huffman Coding

Symbol	Codeword	P(S)				
S_0	1	0.4	0.4	0.4	0.6 (0)	1
S_1	01	0.2	0.2	0.4 (0)	0.4 (1)	
S_2	000	0.2	0.2 (0)	0.2 (1)		
S_3	0010	0.1	0.2 (1)			
S_4	0011	0.1				

To find efficiency of the code.

$$\text{Efficiency } (\eta \%) = \frac{\text{Entropy}}{\text{Average length}} \times 100$$

$$\text{Entropy } H(S) = -\sum_{i=0}^{n} P_i \log_2 P_i$$

$$= -\left[0.4 \times \frac{\log_{10}(0.4)}{\log_{10} 2} + 0.2 \times \frac{\log_{10}(0.2)}{\log_{10} 2} + 0.2 \times \frac{\log_{10}(0.2)}{\log_{10} 2} + 0.1 \times \frac{\log_{10}(0.1)}{\log_{10} 2} + 0.1 \times \frac{\log_{10}(0.1)}{\log_{10} 2}\right]$$

$$= -\left[-0.528 - 0.928 - 0.664\right] = 2.12$$

$\boxed{\text{Entropy} = 2.12 \text{ bits/symbol}}$ $H(S) = 2.12$

Average length (L) $= 1\times0.4 + 2\times0.2 + 3\times0.2 + 4\times0.1 + 4\times0.1$

$= 0.4 + 0.4 + 0.6 + 0.4 + 0.4$

$$\boxed{L = 2.2 \text{ bits/symbol}}$$

Efficiency ($\eta\%$) $= \dfrac{2.12}{2.2} \times 100$

$$\boxed{\eta(\%) = 96.36\%}$$

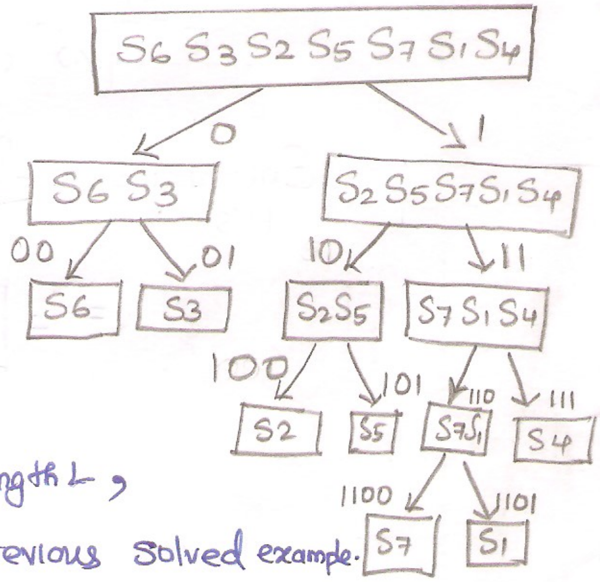
2. Find the Shannon fanno code for the following seven messages with probabilities indicated. $S = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$.

$P(S) = \{0.05, 0.15, 0.2, 0.05, 0.15, 0.3, 0.1\}$.

Step 1:- Writting according to the descending order of probability value.

Symbol	P(S)	Codeword
S_6	0.3	00
S_3	0.2	01
S_2	0.15	100
S_5	0.15	101
S_7	0.1	1100
S_1	0.05	1101
S_4	0.05	111

Step 2:- Finding Code word.



Step 3:

Calculate Entropy $H(S)$, Average length L, Efficiency (η) as explained for previous solved example.

Redundancy = Avg. length − Entropy.

$$\boxed{R = L - H(S)}$$

3. Explain the concept of block codes and coding efficiency. Find the Huffman Code for the following seven message with probabilities indicated. $S = \{S_1, S_2, S_3, S_4, S_5, S_6, S_7\}$.

$P(S) = \{0.05, 0.15, 0.2, 0.05, 0.15, 0.3, 0.1\}$

BLOCK CODES:

The channel encoder accepts information from the Source encoder in successive k bits blocks. For each block it adds (n-k) redundant bits that are algebraically related to the k-message bits thereby producing an overall encoded block of 'n' where $n > k$. The n-bit code is called a Codeword and n is called the block length of the code.

For a block code to be useful, the 2^k codewords must be distinct. Accordingly there should be a one-to-one correspondence b/w the message bits.

The channel encoder produces bits at the rate

$$R_0 = \left(\frac{n}{k}\right) R_s.$$

Where, R_s = Bit rate of the source.

R_0 = Bit rate of the channel.

$r = \frac{k}{n}$ = Code rate where $0 < r < 1$.

So, $R_0 = \frac{R_s}{r}$

Block Codes Types:

1) Linear Block Codes

2) Cyclic Codes.

Coding Efficiency :

Coding efficiency of the source encoder is defined as the ratio of the L_{min} to the average codeword length L.

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

Where, L_{min} = minimum possible value of L.

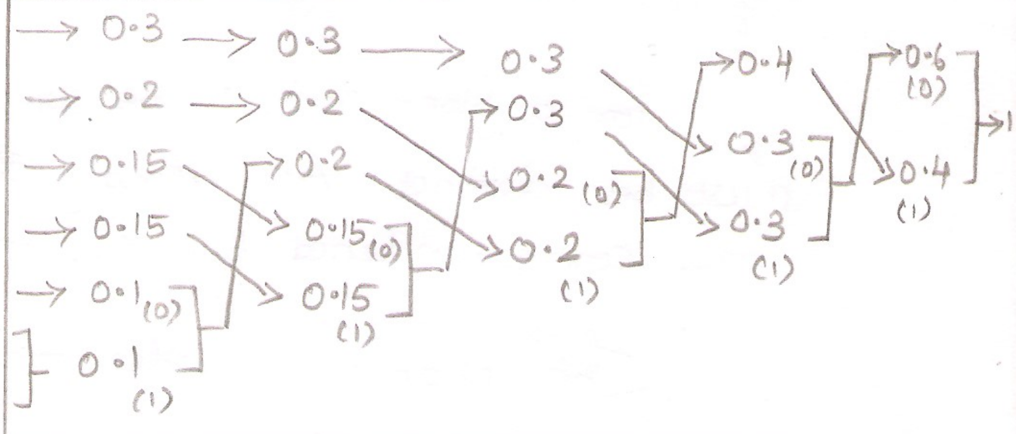
As $L \geq L_{min}$, $\eta \leq 1$.

The source encoder is said to be efficient when $\eta = 1$.

The minimum of L is determined by source coding theorem.

Symbol (S)	Codeword	Prob. P(S)
S_6	00	0.3
S_3	10	0.2
S_2	010	0.15
S_5	011	0.15
S_7	110	0.1
S_1	1110	0.05
S_4	1111	0.05

Huffman Coding.



Similarly calculate entropy, Avg. length, Efficiency, Redundancy as previous example.

4. (i) Write in detail about the procedure of Shannon-Fano coding scheme (8m).

(ii) Define entropy. Explain the properties of entropy (8m).

Coding algorithm:

Step 1: List the symbols in the descending order in accordance with their probabilities.

Step 2: Partition the symbols set into two most equiprobable subsets $\{x_1\}$ and $\{x_2\}$.

Step 3: Assign '0' to each symbol contained in one subset and '1' to each of the symbols in the other ~~set~~ subset.

Step 4: Repeat the same procedure for subsets of $\{x_1\}$ and $\{x_2\}$, until each subset contains single symbol, that is $\{x_1\}$ will be partitioned into the subset $\{x_{11}\}$ and $\{x_{12}\}$. Now the codeword corresponding to message in x_{11} will start with 00 and that corresponding to a message in x_{12} will begin with 01.

This encoding procedure is said to be an "Optimum" procedure to minimize the average length of messages/symbols.

Entropy: It is a measure of the average information content per source symbol. The mean of $I(S_K)$ over the source alphabet X is given by $H(X) = E[I(S_K)]$

H(X) is entropy and its unit is bits/symbol.

$$H(X) = \sum_{K=0}^{K-1} P_K \cdot I(S_K) = \sum_{K=0}^{K-1} P_K \cdot \log_2\left(\frac{1}{P_K}\right)$$

(or) P.T.O

$$H(x) = -\sum_{K=0}^{K-1} P_K \cdot \log_2 P_K \text{ (or) } -\sum_{K=1}^{K} P_K \cdot \log_2 P_K \text{ bits/symbol.}$$

Properties of entropy :-

1) **Symmetry :**

$$H(P_K, P_{K-1}) = H(P_{K-1}, P_K).$$

2) **Additivity :-**

If alphabet x has symbols $X = \{S_0, S_1 \ldots S_n\}$ then partioning of entropy into different subsets does not affect value of entropy $H(x)$.

$$H(x) = H(S_0, S_1, \ldots S_m) + H(S_{m+1}, \ldots S_n).$$

3) **External property :-**

The entropy $H(x)$ of a source is bounded as followes.

$$0 \leq H(x) \leq \log_2 M.$$

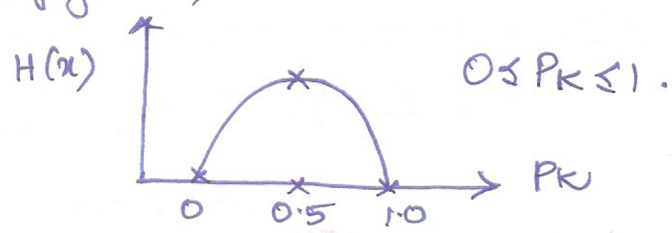
Where m = no. of symbols of alphabet X of source.

(a) $H(x) = 0$, if and only if probability $P_K = 1$ for some 'k' and remaining probabilities in the set are all zero.

(b) $H(x) = \log_2 m$, if and only if $P_K = 1/m$ for all K (ie.) all the symbols are equiprobable.

4) **Continuous Property :**

The entropy $H(x)$ is continuous in the interval $0 \leq P_K \leq 1$



(5) Extension Property:

$$H(X^n) = n \cdot H(X).$$

The no. of distinct symbols emitted by the orginal source is and it lies in the $X = \{S_0, S_1, S_2 \ldots S_k\}$. The extended source may generate a source alphabet X^n which has k^n distinct blocks. Each block consists of 'n' successive source symbols. The probability of a source symbol in X^n is equal to product of probabilities of 'n' source symbols in X, then the entropy of the extended source is equal to 'n' ~~tes~~ times the entropy of orginal source.

5. (i) Compare the coding schemes HDBP and MBnB codes in terms of bandwidth, SNR and transmission efficiency (10m)

(ii) Describe the bandwidth - SNR trade of (6m).

5. (i) Compare the coding schemes HDBP and MBnB codes in terms of bandwidth, SNR and transmission efficiency (10m).

(ii) Describe bandwidth – SNR trade off (6m).

HDBn (High Density Bipolar Substitution)

1) Used to counteract the effects of a long string of binary 0's in the AMI line code.

2) When the number of continuous binary 0's exceeds the n they are replaced by a special code sequence.

3) In HDB3, the fourth zero in a string of zeros is marked ie., forcibly set to 1, but in a way that violates the alternating mark rule.

4) HDB3 is the line code recommended by the ITU-T for PCM systems operating at multiplexed rates of 2, 8 & 34 Mbits/sec. HDB3 is widely used in Europe.

5) Well suited to high data rate transmission

MBnB Codes:-

1) Commonly used as line codes. $n > m$.

2) Known as block codes as they encode the block of data of length m into another block of length n.

3) Redundancy is equal to $2^n - 2^m$. Redundant bit patterns can be used for transmission of control information.

4) Defined in terms of maximum length of one type bit pattern "run length" & ratio of total no of 1's to 0's disparity.

5) Used to adjust the

① DC Component level.

② Shape the Spectrum.

③ Provide timing information.

6) Allows error monitoring by simple rule violation detection logic CKt.

7) Can be constructed as Self framing.

8) Always result in bandwidth increase by (n/m).

6. (i) Discuss the BSC and BEC with their channel diagram and transition matrix (12m).

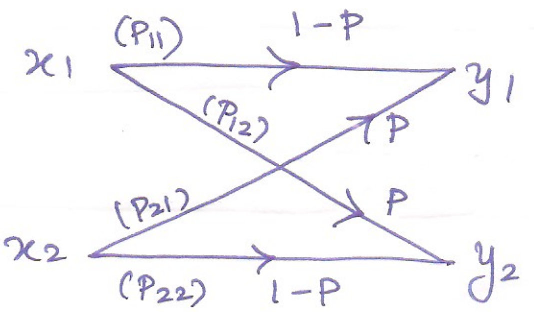
(ii) Draw the polar, unipolar, bipolar and Manchester NRZ line code format for an information {101100} (4m)

(i) BSC (Binary Symmetric Channel)

It is a discrete memoryless channel which has 2 i/p symbols $[x_1 = 0; x_2 = 1]$ and 2 o/p symbols $[y_1 = 0, y_2 = 1]$.

The channel is symmetric b/c, the probability of receiving a '1' if a '0' is sent is same as probability of receiving a '0' if a '1' is sent. This conditional probability of error is denoted by P.

The transition probability diagram is:



The transition matrix of the channel $P(Y/X)$

$$P(Y/X) = \begin{bmatrix} P(y_1/x_1) & P(y_2/x_1) \\ P(y_1/x_2) & P(y_2/x_2) \end{bmatrix}$$

Binary Erase channel (BEC):

In data communication, it is common practice to use ARQ (Automatic Request for Retransmission) techniques for 100% data. In the receiver, an error detecting ckt is used along with some parity checks transmitted with the data and erroneous data is rejected using for retransmiss-ion.