ITC TUT 4

Q1] Show the source encoding for the message "SAHARA" using the following encoding method. Also calculate the source entropy, Average length per symbol and the symbol coding efficiency in each case:

a) Shannon - Fano coding
b) Huffman Coding
c) Arithmetic coding.

[Assume the symbol probabilities as S = 1/6, A = 3/6, H = 1/6 and R = 1/6]

→ a] Shannon - Fano

Arranging in descending order

$S = \frac{1}{6}$ $A = \frac{3}{6}$ $H = \frac{1}{6}$ $R = \frac{1}{6}$

Sym

A 3/6] 0
S 1/6]
H 1/6] 1
R 1/6]

1/6] 10

1/6]
1/6] 11

1/6] 110

1/6] 111

c_i	n_i
0	1
10	2
110	3
111	3

classmate

Date

Page

$$\text{Entropy} = \frac{3}{6}\log_2\frac{6}{3} + \frac{1}{6}(\log_2)6 \times 3$$

$$= \frac{1}{2}\times 1 + \frac{1}{6}\times 2.5844 \times 3 = 1.792 \text{ bits}$$

$$H = \frac{3}{6}\times 1 + \frac{1}{6}\times 2 + \frac{1}{6}\times 3 + \frac{1}{6}\times 3$$

$$= 1.792481 \text{ bits/symbol}$$

Average coding coord length (N)

$$N = \sum_{k=1}^{a} P_k N_k$$

$$= \left(\frac{3}{6}\right)(1) + \left(\frac{1}{6}\right)(2) + \left(\frac{1}{6}\right)(3) + \left(\frac{1}{6}\right)3$$

$$= \frac{3}{6} + \frac{2}{6} + \frac{3}{6} + \frac{3}{6} = 1.833 \text{ bits/symbol}$$

Coding efficiency (n)

$$n = \frac{H}{N} = \frac{1.792481}{1.8333} = 97.77\%$$

(b) Huffman coding.

S	p_i	I	II	code	n_i
A	0.5	0.5	0.5	1	1
S	0.1667	0.3333	0.5	01	2
H	0.1667	0.1667		000	3
R	0.1667			001	3

$$\text{Entropy } (H) = \sum_{i=1}^{j} P_k \log\left(\frac{1}{P_k}\right)$$

$$= 1.792481 \text{ bits/symbol}$$

$$\text{Efficiency} = e = \frac{H}{\bar{L}\log_2 r}$$

$\bar{L}$ = average code word length =

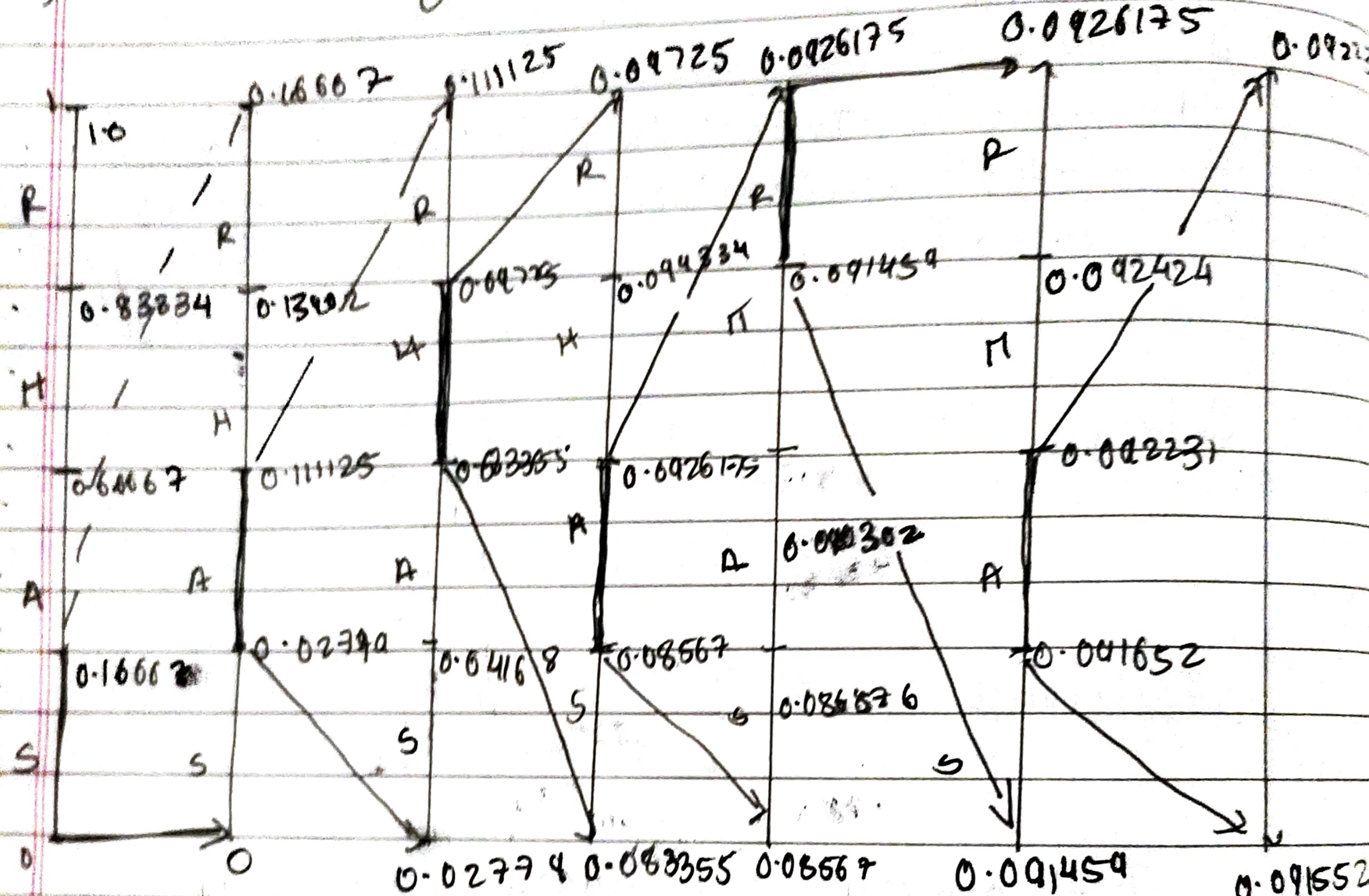
$$\bar{L} = \sum_{i=1}^{N} P_i n_i = \left(\frac{3}{6}\right)(1) + \left(\frac{1}{6}\right)2 + \left(\frac{1}{6}\right)3 + \left(\frac{1}{6}\right)3$$

$$\bar{L} = 1.8333 \text{ bits/symbol}$$

$$e = \frac{H}{\bar{L}\log_2 r} = \frac{1.792481}{1.8333 \times \log_2 2} \qquad e = 97.77\%$$

16010420133

c) Arithmetic coding



$di = UL - LL$

$= 0.16667 - 0$

$= 0.16667$

Range of symbol : LL : LL + di (Prob of symbol)

Range of S : $0 : 0 + 0.16667 (0.16667) = 0.02779$

Range of A : $0.02779 : 0.2779 + 0.16667 (0.5) = 0.11125$

Range of H : $0.11125 : 0.11125 + 0.1666 + (0.16667) = 0.13902$

Range of R : $0.13902 : 0.13902 + 0.16667 (0.16667) = 0.1667$

16010420133.

$d_i = UL - LL = 0.111125 - 0.02779$
$= 0.083335$

Range of S: $0.02779 : 0.08335 + 0.02779 \times 0.1667 = 0.04168$
Range of A: $0.04168 : 0.04168 + 0.08335 \times 0.5 = 0.083355$
Range of H: $0.083355 : 0.083355 + 0.083355 = 0.09725$
Range of R: $0.09725 : 0.09725 + 0.08335 \times 0.16667 = 0.1114$

$d_i = UL - LL$
$= 0.09725 - 0.083355$
$= 0.013895$

Range of S: $0.083355 : 0.083355 + 0.013895 \times 0.16667 = 0.08567$
Range of A: $0.085670 : 0.085670 + 0.013895 \times 0.5 = 0.0926175$
Range of H: $0.0926175 : 0.0926175 + 0.013895 \times 0.16667 = 0.094834$
Range of R: $0.094834 : 0.94834 + 0.013895 \times 0.16667 = 0.097$

$d_i = UL - LL$
$= 0.0926175 - 0.08567$
$= 0.0069475$

Range of S: $0.08567 : 0.08567 + 0.0069475 \times 0.16667 = 0.086828$
Range of A: $0.085828 : 0.086828 + 0.0069475 \times 0.5 = 0.090302$
Range of H: $0.090302 : 0.090302 + 0.0069475 \times 0.96667 = 0.091459$
Range of R: $0.091459 : 0.091459 + 0.0069475 \times 0.16667 =$
0.092617

$d_i = UL - LL = 0.0914590 - 0.0926175 = 0.0011585$

Range of S: 0.91459 : $0.091459 + 0.0011585 \times 0.16667 = 0.0816$

Range of A: $0.091652 : 0.091552 + 0.0011585 \times 0.5 = 0.092281$

Range of H: $0.092281 : 0.092231 + 0.0011585 \times 0.16667 = 0.09242$

Range of R: $0.092424 : 0.092424 + 0.0011585 \times 0.16667 = 0.092617$

$$tag = \frac{UL + LL}{2}$$

$$= \frac{0.091652 + 0.092231}{2} = \frac{0.183883}{2}$$

$$= 0.0919415$$

$\therefore$ The encoded message is 0.0919415

Q2] Explain the advantages & disadvantages for each of the above method.

(1) Shannon-Fano coding

Advantages

(1) we do not need to built the entire codebook instead.
(2) we simply obtain the code for the tag corresponding to a given sequence.
(3) ~~It~~ is entirely feasible to co

Disadvantages

(1) In Shanon-Fano coding, we cannot bo sure about the codes generated. There may be two different codes for the same symbol depending on the way we build our tree.
(2) It does not gurantee optimal code.
(3) There is no unique code, ie code might be a prefix for another code so in case of errors or loss during data transmission we have to start from beginning.

133

Arithmetic coding :- Advantages

1. Arithmetic coding surpasses the Huffman technique in its compression ability.
2. It assigns one log code to the entire input string
3. Arithmetic coding consists of a few arithmetic operations due to its complexity is less.
4. Arithmetic coding is asymptotically better than Huffman coding

Disadvantages

1. whole codeword must be received to start decoding the symbols and if there is a corrupt bit in the codeword the entire message could become corrupt.
2. There is a limit to the precision of the number which can be encoded, thus limiting the number of symbols to encode within a codeword.

Huffman coding

Advantages

1. This encoding scheme results in saving a lot of storage space, since the binary code generated are variable in length.
2. The binary codes generated are prefix-free.

25/3/22

Disadvantages

① Achieve a lower compression ratio compared to lossy technique in its compression ability.
② It assigns one logcode to the entire input string
③ Since length of all binary codes is different, it becomes difficult for the decoding software to detect whether the encoded data is corrupt so it produces wrong output.