Huffman Coding.

* Huffman suggest a source coding method, in 1952 based on probabilities of the source symbols.

* This method is optimal in the sense that the average number of bits it requires to represent the source symbols is ~~maximum~~ minimum and also meets the prefix condition.

* one important condition for Huffman code is that there must be 'r' maximum length codeword.

Eg:- for radix-2 (r=2) :

* Huffman coding is an example of lossless coding.

Huffman Coding Steps (Radix-2).

(1) Arrange the source symbols with their probabilities in decreasing order of their probabilities.

(2) Take the bottom two symbol probabilities and always assign '1' to the less probable symbol and '0' to more probable symbol between the two.

(3) After assigning '1' and '0' to the symbols add them up.

(4) After this you are having two methods.

Method 1 (OR) Method 2.

Method 1 :-

Put the merged one (addition) as low as possible

(OR)

Method 2 :-

Put the merged one (addition) as high as possible

....

* Now again pick the two smallest probabilities, tie (add) them together. Again assign '1' to less probable symbol and '0' to more probable symbol.

* In case both the symbols have same probability assign '1' to lower symbol and '0' to higher symbol.

* Continue until only one probability is left.

Problem :-

Determine Huffman Code for $P = \{0.4, 0.2, 0.2, 0.1, 0.1\}$. by using Method 1 & Method 2 and also find

(a) Entropy

(b) Average Code length

(c) Efficiency.

(d) Coding tree.

SRM

Solution :-

Formula Required

1. For Calculating Entropy

$$H(X) = -\sum_{i=1}^{n} P_i \log_2 P_i$$

where P_i - Probability of each event

2. For Calculating Average code length

$$\bar{N} = \sum_{i=1}^{n} P_i n_i$$

where n_i - number of code bits assigned to each probability.

3. For Calculating Efficiency

$$\eta = \frac{H(X)}{\bar{N}} \times 100\ \%$$

Method 1 :-

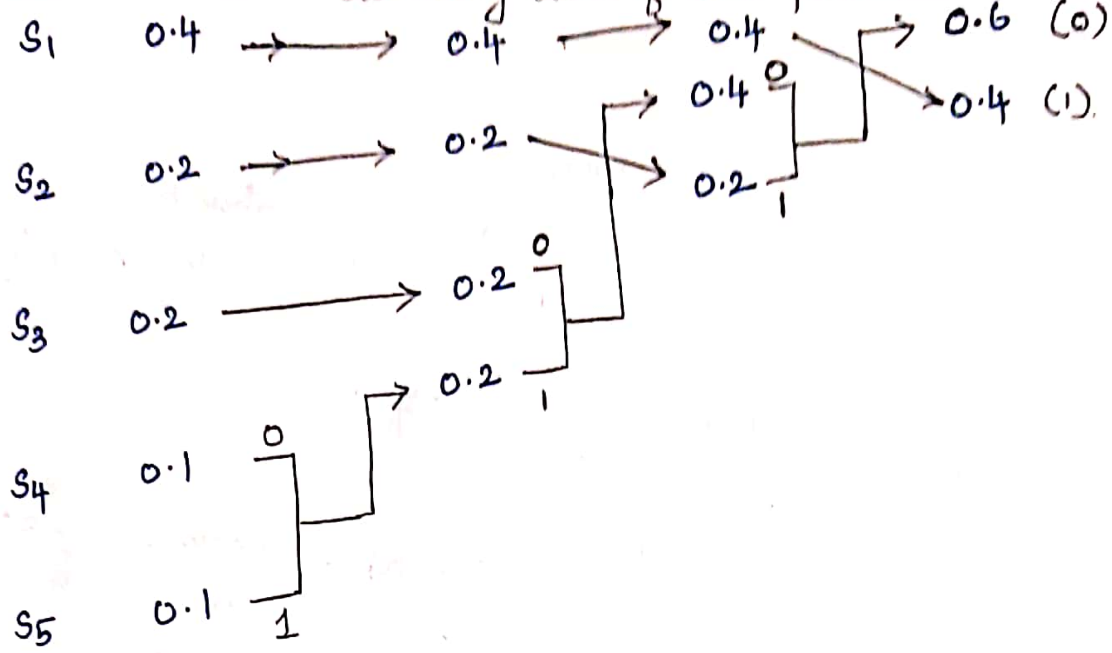
* Given, there are 5 events having probabilities as {0.4, 0.2, 0.2, 0.1, 0.1}.

* Check the summation of given probabilities is '1'

Yes, its '1' (0.4+0.2+0.2+0.1+0.1 = 1).

Assign $S_1 = 0.4$
$S_2 = 0.2$
$S_3 = 0.2$
$S_4 = 0.1$
$S_5 = 0.1$

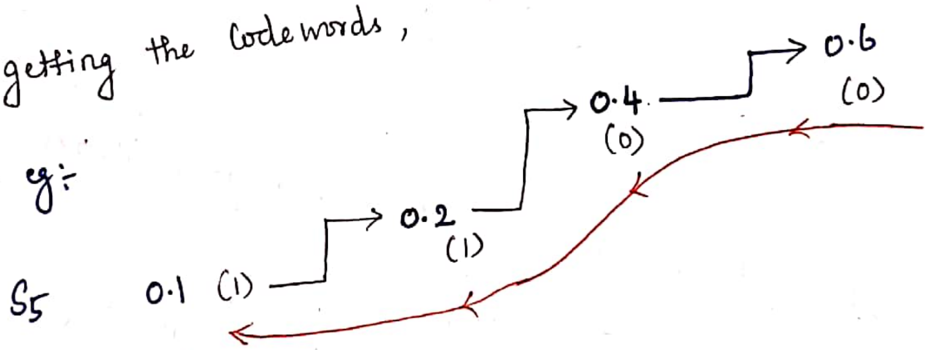
Note: Arrange the source symbols with their probabilities in decreasing order of their probabilities.



S_1	S_2	S_3	S_4	S_5
1	01	000	0010	0011

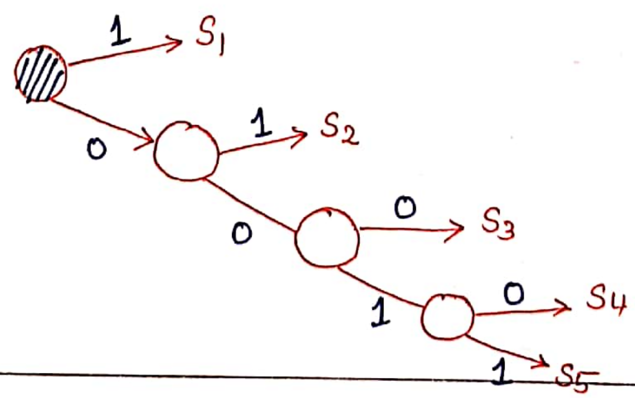
For getting the codewords,

For eg:

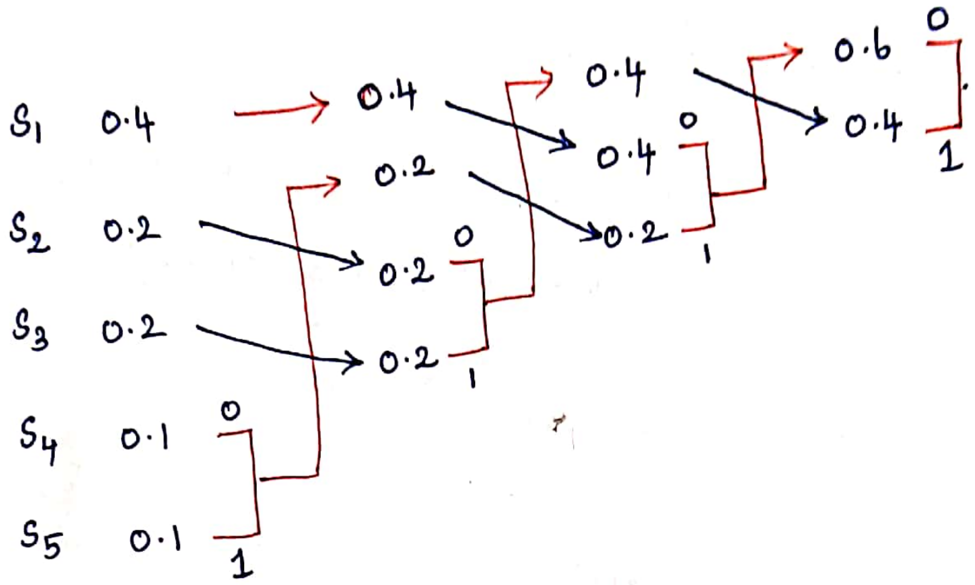


0011 – Codeword for S_5.

Code tree:



Method 2 :



S_1	S_2	S_3	S_4	S_5
00	10	11	010	011.

(a) Entropy :

Same as method 1 $H(x) = 2.12$ bits/symbol.

(b) Avg. code length :

$$\bar{N} = \sum_{i=1}^{n} P_i n_i$$

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

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

$$= 2.2$$

(c) Efficiency :

$$\eta = \frac{H(x)}{\bar{N}} = \frac{2.12}{2.2} = 96.36\%.$$

(a) Entropy :-

$$H(x) = -\sum_{i=1}^{n} P_i \log_2 P_i$$

$$= -\left(0.4\log_2 0.4 + 0.2\log_2(0.2) + 0.2\log_2 0.2 + 0.1\log_2 0.1 + 0.1\log_2 0.1\right)$$

$$= \text{Answer}/\log 2$$

$$= -(-2.1219)$$

$$= 2.12$$

(b) Avg. Code length :-

$$\bar{N} = \sum_{i=1}^{n} P_i n_i$$

$$= 0.4(1) + 0.2(2) + 0.2(3) + 0.1(4) + 0.1(4)$$

$$\bar{N} = 2.2$$

(c) Efficiency :-

$$\eta = \frac{H(x)}{\bar{N}} \times 100\%$$

$$= \frac{2.12}{2.2} \times 100\%$$

$$\eta = 96.36\%$$

SRM

An important term Variance is used to determine which code (Method 1 or Method 2) is preferable.

Variance :-

$$V = \sum_{i=1}^{n} P_i (n_i - \bar{N})^2$$

For method 1 :-

$$V_1 = 0.4(1-2.2)^2 + 0.2(2-2.2)^2 + 0.2(3-2.2)^2 + 0.1(4-2.2)^2 + 0.1(4-2.2)^2$$

$$= 1.36.$$

For method 2 :-

$$V_2 = 0.4(2-2.2)^2 + 2 \times 0.2(2-2.2)^2 + 2 \times 0.1(3-2.2)^2$$

$$= 0.16.$$

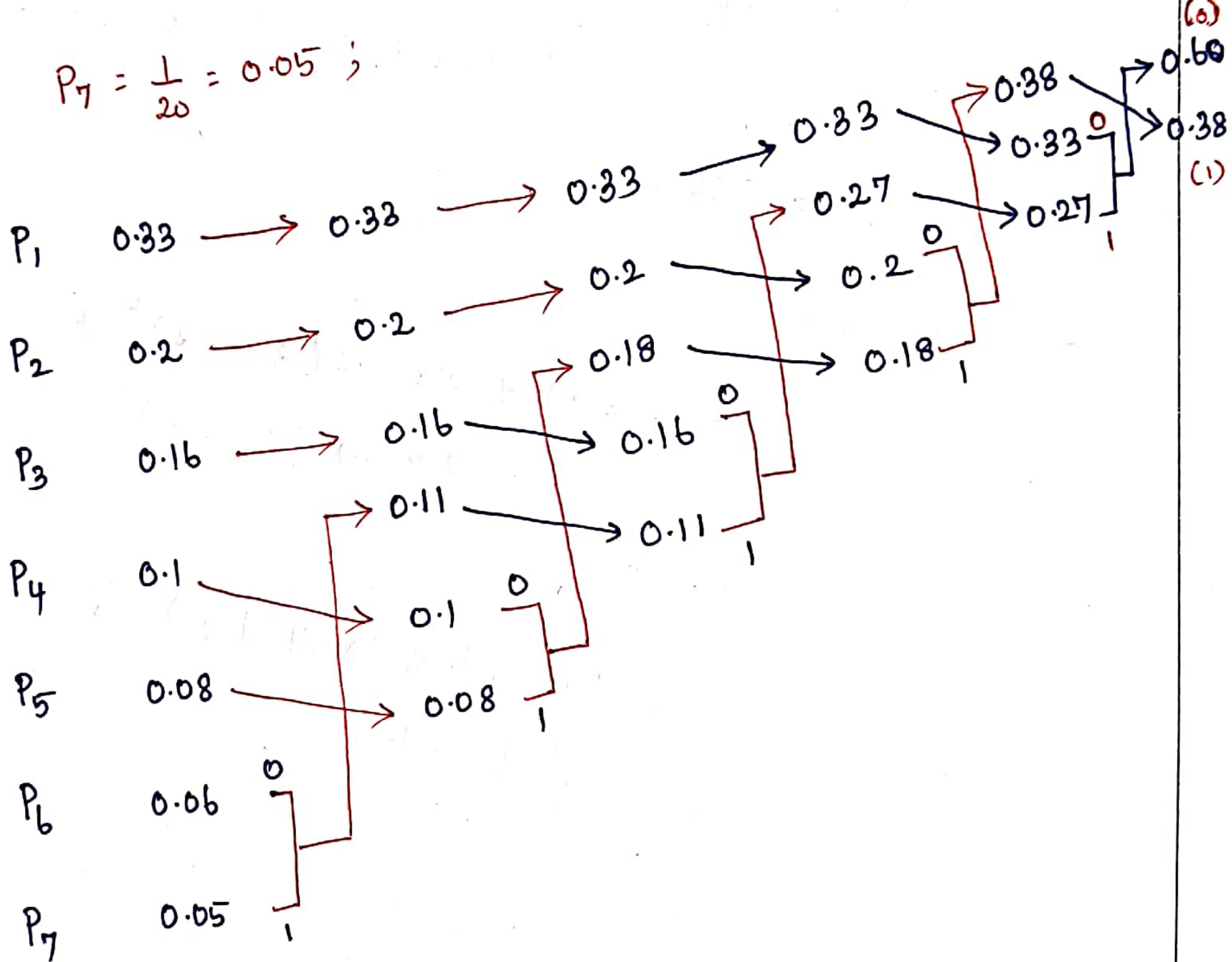
$\therefore V_2 < V_1$, the Code 2 (Method 2) is Preferable.

2) Given $P_1 = 1/3$, $P_2 = 1/5$, $P_3 = 1/6$, $P_4 = 1/10$, $P_5 = 1/12$, $P_6 = 1/15$ and $P_7 = 1/20$. Encode Huffman Code (Method 1) for $r=2$ (Radix-2). Find its Entropy, Average code length and Efficiency.

$P_1 = \frac{1}{3} = 0.33$; $P_2 = \frac{1}{5} = 0.2$; $P_3 = \frac{1}{6} = 0.16$

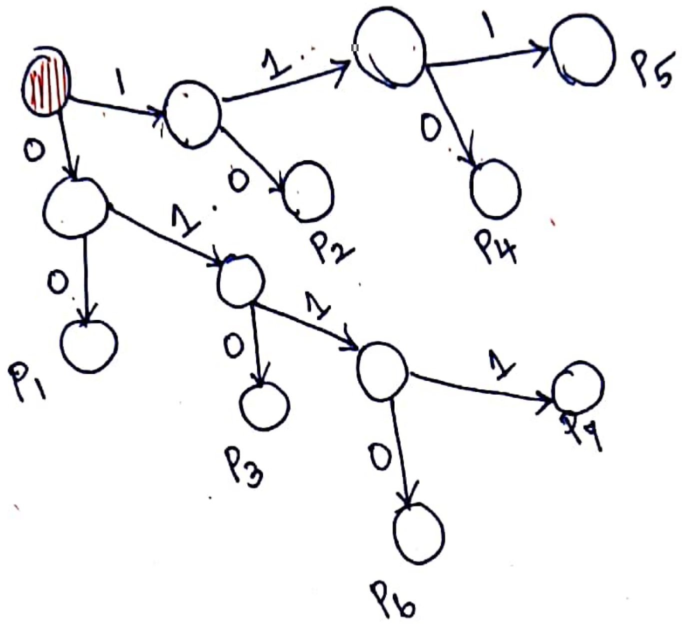
$P_4 = \frac{1}{10} = 0.1$; $P_5 = \frac{1}{12} = 0.08$; $P_6 = \frac{1}{15} = 0.06$

$P_7 = \frac{1}{20} = 0.05$;



P_1	P_2	P_3	P_4	P_5	P_6	P_7
00	10	010	110	111	0110	0111

Coding tree :



As similar to problem 1, find Entropy, Avg. code length and Efficiency.

$H(X) =$