EC8395 Communication Engineering

Unit-1-Infromation Theory and Coding

Staff: **Dr. R.Kalidoss**

Session Objectives

•Understand about Huffman coding techniques.

•Solving some problems in Huffman coding.

Session Outcomes

- At the end of the session, students will be able to
  - Understand Huffman coding.
  - To understand the difference between Shannon Fano and Huffman coding.

Huffman coding

- It is a type of source-coding technique in which the average codeword length $\left(\bar{L}\right)$ approaches the fundamental limit set by the entropy of a discrete memoryless source. Hence, we can obtain maximum efficiency from Huffman code. Therefore, it is called as "*Optimum code*".

- ***Steps involved in Huffman coding:***
- Arrange the source symbols or messages in the decreasing probability order.
- The two messages of the lowest probability are assigned 0 and 1. This process is referred to as “*Splitting stage*”
- These two messages are combined into one new message with probability equal to the sum of two original probabilities. The probability of the new message is placed in the list in accordance with its value. This process is referred to as “*Reduction stage*”, since the number of messages or symbols in stage 1 will be reduced by one and this reduction process continues in each stage.
- The procedure is repeated until we are left with a final list of source symbols, for which a 0 and 1 are assigned.
- The codeword calculation or encoding process starts from the last reduction stage (rightmost part) to the first reduction stage (leftmost part).

vi) The coding efficiency which for Huffman coding is given by the expression

$$\eta = \frac{H(S)}{\overline{L}}$$

where $H(S)$ is entropy of the source S. $H(S) = \sum_{k=0}^{K-1} p_k \log_2\left(\frac{1}{p_k}\right)$

and $\overline{L}$ is average code word length.

$$\overline{L} = \sum_{k=0}^{K-1} p_k l_k$$

;where l_k is length of binary code word assigned to symbol.

vi) Redundancy can be calculated by following formula.

Redundancy= $1-\eta$.

Solved Problems

- 1. Find out the Huffman code for a discrete memory less source with probability statistics

$$\{0.1, 0.1, 0.2, 0.2, 0.4\}.$$

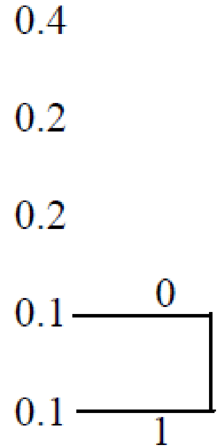
Answer

- Step 1:

Arrange the probabilities in the descending order

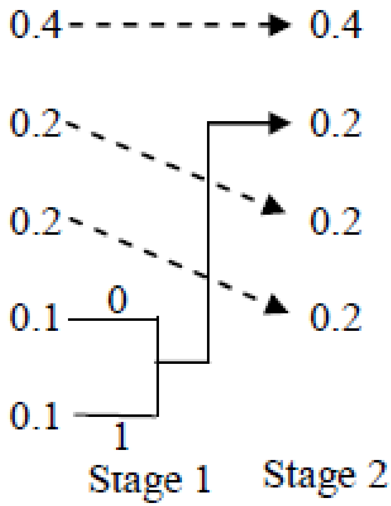
0.4, 0.2, 0.2, 0.1, 0.1

- Step 2:
- At stage 1, assign 0 and 1 to two messages of the lowest probability. Here, the lowest probability of message points are 0.1 and 0.1

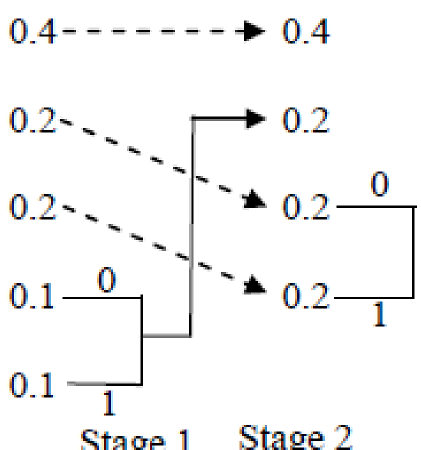


Stage 1

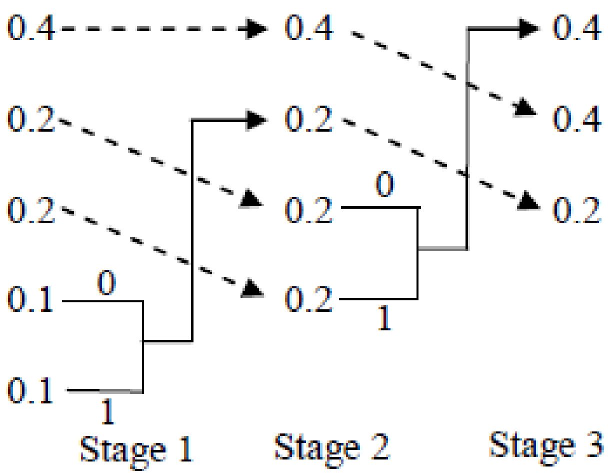
- (ii) Combine these lowest probability messages and rearrange the order in stage 2.



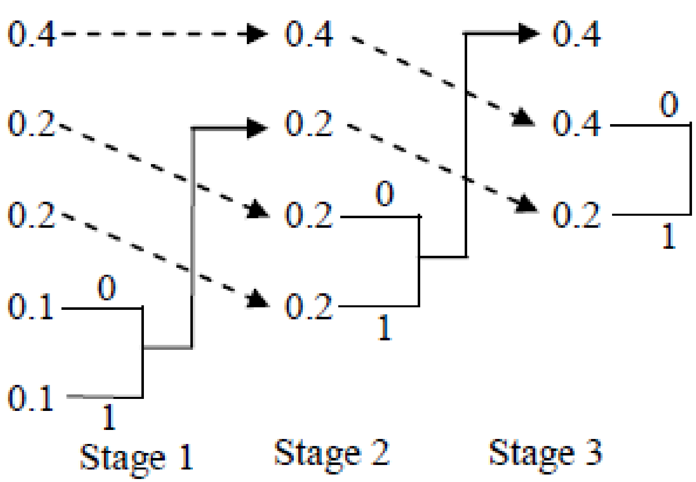
- (iii) At stage 2, assign 0 and 1 to two messages of the lowest probability. Here the lowest probabilities of message points are 0.2 and 0.2.



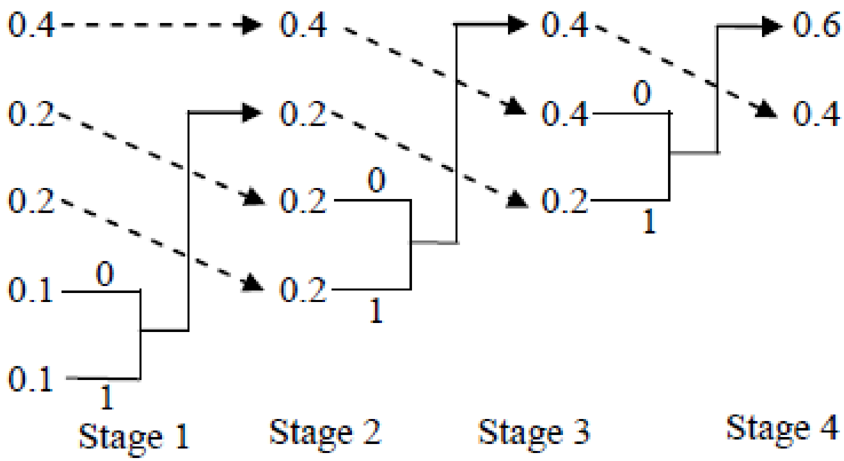
- (iv) Combine these lowest probability messages and rearrange the order in stage 3.



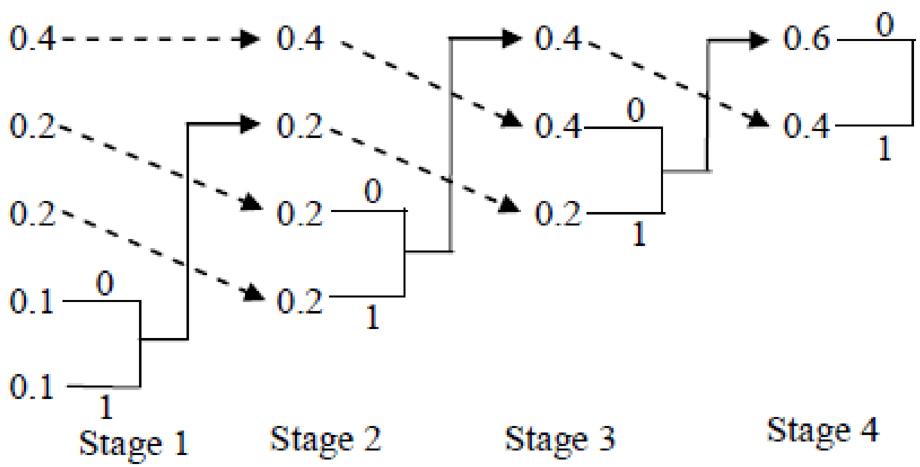
- (v) At stage 3, assign 0 and 1 to two messages of the lowest probability. Here, the lowest probability of message points are 0.4 and 0.2.



- (vi) Combine these lowest probability messages and rearrange the order in stage 4.



- (vii) At stage 4, assign 0 and 1 to two messages of lowest probability. Here, the lowest probability of message points are 0.6 and 0.4.



- Further splitting of stages is not possible. Hence, stop the iteration.

- Step 3: Codeword calculation:
- Symbol 1; whose probability is 0.4:

  0.4 in stage 1 is moved to 0.4 in stage 2 and not assigned any codeword. At stage 3, 0.4 is assigned to codeword 0. At stage 4, 0.4 is modified into 0.6 and codeword for 0.6 is 0 (i.e. combination of 0.4 and 0.2 at stage 3 produces 0.6 at stage 4). Final codewords are obtained from the right direction to the left-most part. Therefore, codeword for symbol 1, whose probability is 0.4 is "0 0".

- Codeword for symbol 2, whose probability is 0.2:

  0.2 in stage 1 is moved to 0.2 in stage 2 with codeword 0. At stage 3, 0.2 is moved to 0.2 with codeword 1. At stage 4, 0.2 is modified into 0.6 and codeword for 0.6 is 0. (i.e. combination of 0.4 and 0.2 at stage 3 produces 0.6 at stage 4). The final codeword for symbol 2 is 1 0. (Written from the right-most direction to the left-most direction.)

- Codeword for symbol 3, whose probability is 0.2:

  0.2 in stage 1 is moved to 0.2 in stage 2 and assigned by a codeword of 1. 0.2 in stage 2 is modified into 0.4 in stage 3 and it is not assigned any codeword (i.e. combination of probability 0.2 and 0.2 in stage 2 produces 0.4 in stage 3). 0.4 at stage 3 is moved to 0.4 in stage 4, with the codeword of 1.

  Therefore, the final codeword for symbol 3 is 1 1.

- Codeword for symbol 4, whose probability is 0.1:

  0.1 in stage 1 is modified into 0.2 in stage 2 (i.e. combination of probabilities 0.1 and 0.1 in stage 1 produces 0.2 in stage 2, and it is not assigned any codeword). 0.2 in stage 2 is moved to 0.2 in stage 3 and assigned by a codeword of 1. 0.2 in stage 3 is modified into 0.6 and assigned by a codeword of 0 (i.e. combination of probabilities 0.4 and 0.2 in stage 3 produces 0.6 in stage 4).

  Final codeword for symbol 4 is 0 1 0.

- Codeword for symbol 5, whose probability is 0.1:

  0.1 in stage 1 is modified into 0.2 in stage 2 and it is not assigned any codeword (i.e. combination of 0.1 and 0.1 in stage 1 produces 0.2 at stage 2). 0.2 in stage 2 is moved to 0.2 in stage 3, with the codeword of 1. 0.2 at stage 3, modified into 0.6 at stage 4, with the codeword of 0 (i.e. combination of probabilities 0.4 and 0.2 at stage 3 produces 0.6 at stage 4).

  Final codeword for symbol 5 is 0 1 1.

- Step 3:

Symbol	Probability	Codeword	Length (l)
S_0	0.4	0 0	2
S_1	0.2	10	2
S_2	0.2	1 1	2
S_3	0.1	0 1 0	3
S_4	0.1	0 1 1	3

- Step 4:

To find out efficiency (η), we must calculate the average codeword length $(\bar{L})$ and entropy $H(S)$.

$$\eta = \frac{H(S)}{\overline{L}}$$

- Where $\overline{L} = \sum_{k=0}^{4} p_k l_k$

$$= p_0 l_0 + p_1 l_1 + p_2 l_2 + p_3 l_3 + p_4 l_4$$

= (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

$$\overline{L} = 2.2\, bits / symbol$$

$$H(S) = \sum_{k=0}^{4} p_k \log_2\left({}^{1}\!/_{p_k} \right)$$

$$= p_0 \log_2\left({}^{1}\!/_{p_0} \right) + p_1 \log_2\left({}^{1}\!/_{p_1} \right) + p_2 \log_2\left({}^{1}\!/_{p_2} \right) + p_3 \log_2\left({}^{1}\!/_{p_3} \right) + p_4 \log_2\left({}^{1}\!/_{p_4} \right)$$

$$= 0.4 \log_2\left({}^{1}\!/_{0.4} \right) + 0.2 \log_2\left({}^{1}\!/_{0.2} \right) + 0.2 \log_2\left({}^{1}\!/_{0.2} \right) + 0.1 \log_2\left({}^{1}\!/_{0.1} \right) + 0.1 \log_2\left({}^{1}\!/_{0.1} \right)$$

$$= 0.4\frac{\log_{10}\left(1/0.4\right)}{\log_{10} 2} + 0.2\frac{\log_{10}\left(1/0.2\right)}{\log_{10} 2} + 0.2\frac{\log_{10}\left(1/0.2\right)}{\log_{10} 2} + 0.1\frac{\log_{10}\left(1/0.1\right)}{\log_{10} 2} + 0.1\frac{\log_{10}\left(1/0.1\right)}{\log_{10} 2}$$

$$= (0.4 \times 1.3219) + (0.2 \times 2.3219) + (0.2 \times 2.3219) + (0.1 \times 3.3219) + (0.1 \times 3.3219)$$

$$= 0.5287 + 0.4643 + 0.4643 + 0.33219 + 0.33219$$

$$= 2.12\, bits / symbol$$

$$H(S) = 2.12\, bits / symbol$$

$$\eta = \frac{H(S)}{\overline{L}}$$

$$= \frac{2.12}{2.2} = 0.96$$

$$\eta = 96\%$$

$$\eta = \frac{H(S)}{\overline{L}}$$

- Where $\overline{L} = \sum_{k=0}^{4} p_k l_k$

$$= p_0 l_0 + p_1 l_1 + p_2 l_2 + p_3 l_3 + p_4 l_4$$

= (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

$$\overline{L} = 2.2\, bits / symbol$$

$$H(S) = \sum_{k=0}^{4} p_k \log_2\left(1/p_k \right)$$

$$= p_0 \log_2\left(1/p_0 \right) + p_1 \log_2\left(1/p_1 \right) + p_2 \log_2\left(1/p_2 \right) + p_3 \log_2\left(1/p_3 \right) + p_4 \log_2\left(1/p_4 \right)$$

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

$$= 0.4\frac{\log_{10}\left(1/0.4\right)}{\log_{10} 2} + 0.2\frac{\log_{10}\left(1/0.2\right)}{\log_{10} 2} + 0.2\frac{\log_{10}\left(1/0.2\right)}{\log_{10} 2} + 0.1\frac{\log_{10}\left(1/0.1\right)}{\log_{10} 2} + 0.1\frac{\log_{10}\left(1/0.1\right)}{\log_{10} 2}$$

$$= \left(0.4\times 1.3219\right) + \left(0.2\times 2.3219\right) + \left(0.2\times 2.3219\right) + \left(0.1\times 3.3219\right) + \left(0.1\times 3.3219\right)$$

$$= 0.5287 + 0.4643 + 0.4643 + 0.33219 + 0.33219$$

$$= 2.12\, bits/symbol$$

$$H\left(S\right) = 2.12\, bits/symbol$$

$$\eta = \frac{H\left(S\right)}{\overline{L}}$$

$$= \frac{2.12}{2.2} = 0.96$$

$$\eta = 96\%$$