Principles of Data Communications

Reference Book: Data Communications and Networking by Behrouz A. Forouzan

Source Coding

- Aim is to compress data
- Reduces size of data- Compression
- eg)ababbbcca- a occurs 3 times; b occurs 4 times; c occurs 2 times. The number of bits for encoding b should be the least.

Shannon Fano Coding

Procedure

- Messages are written in decreasing order of their probabilities.
- The message set is divided into 2 subsets of equal or nearly equal probabilities.
- 0 is assigned to each message in the subset above the partition line.
- 1 is assigned to each message in the subset below the partition line.
- The procedure is continued until each subset contains only one message.

Example 1

Apply Shannon Fano coding procedure for the following message ensemble.

$[x] = [x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8]$
$[P] = [\frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{4}, \frac{1}{16}, \frac{1}{8}]$

Step 1

Messages are written in **decreasing order of their probabilities**.

Message	Probability	Encoded Message	length(n_k)
x_1	0.25		
x_6	0.25		
x_2	0.125		
x_8	0.125		
x_3	0.0625		
x_4	0.0625		
x_5	0.0625		
x_7	0.0625		

Step 2

The message set is divided into 2 subsets of equal or nearly equal probabilities.

Message	Probability	Encoded Message	length(n_k)
x_1	0.25		
x_6	0.25		
x_2	0.125		
x_8	0.125		
x_3	0.0625		
x_4	0.0625		
x_5	0.0625		
x_7	0.0625		

$0.25 + 0.25 = 0.5$

$0.125 + 0.125 + 0.0625 + 0.0625 + 0.0625 + 0.0625 = 0.5$

Step 3

0 is assigned to each message in the subset above the partition line.

Message	Probability	Encoded Message	length(n_k)
x_1	0.25	0	
x_6	0.25	0	
x_2	0.125		
x_8	0.125		
x_3	0.0625		
x_4	0.0625		
x_5	0.0625		
x_7	0.0625		

Step 4

1 is assigned to each message in the subset below the partition line.

Message	Probability	Encoded Message	length(n_k)
x_1	0.25	0	
x_6	0.25	0	
x_2	0.125	1	
x_8	0.125	1	
x_3	0.0625	1	
x_4	0.0625	1	
x_5	0.0625	1	
x_7	0.0625	1	

Step 4

The procedure is continued until each subset contains only one message.

Message	**Probability**	**Encoded Message**	**length(n_k)**
x_1	0.25	0 0	
x_6	0.25	0 1	
x_2	0.125	1	
x_8	0.125	1	
x_3	0.0625	1	
x_4	0.0625	1	
x_5	0.0625	1	
x_7	0.0625	1	

First Group is split into two (0.25 probability each; 0 above the line and 1 below the line)

Message	**Probability**	**Encoded Message**	**length(n_k)**
x_1	0.25	0 0	
x_6	0.25	0 1	
x_2	0.125	1 0	
x_8	0.125	1 0	
x_3	0.0625	1 1	
x_4	0.0625	1 1	
x_5	0.0625	1 1	
x_7	0.0625	1 1	

Second Group is split into two- x2, x8 in one partition and x3, x4, x5, x7 on the other (0.25 probability each; 0 above the line and 1 below the line)

Message	Probability	Encoded Message	length(n_k)
x_1	0.25	0 0	
x_6	0.25	0 1	
x_2	0.125	1 0 0	
x_8	0.125	1 0 1	
x_3	0.0625	1 1	
x_4	0.0625	1 1	
x_5	0.0625	1 1	
x_7	0.0625	1 1	

x_2, x_8 split into two (0.125 probability each; 0 above the line and 1 below the line)

Message	Probability	Encoded Message	length(n_k)
x_1	0.25	0 0	
x_6	0.25	0 1	
x_2	0.125	1 0 0	
x_8	0.125	1 0 1	
x_3	0.0625	1 1 0	
x_4	0.0625	1 1 0	
x_5	0.0625	1 1 1	
x_7	0.0625	1 1 1	

x_3, x_4, x_5, x_7 split into two; x_3 and x_4 on one partition and x_5, x_7 on the other (0.125 probability each; 0 above the line and 1 below the line)

Message	Probability	Encoded Message	length(n_k)
x_1	0.25	0 0	
x_6	0.25	0 1	
x_2	0.125	1 0 0	
x_8	0.125	1 0 1	
x_3	0.0625	1 1 0 0	
x_4	0.0625	1 1 0 1	
x_5	0.0625	1 1 1	
x_7	0.0625	1 1 1	

x_3, x_4 split into two (0.0625 probability each; 0 above the line and 1 below the line)

Message	Probability	Encoded Message	length(n_k)
x_1	0.25	0 0	
x_6	0.25	0 1	
x_2	0.125	1 0 0	
x_8	0.125	1 0 1	
x_3	0.0625	1 1 0 0	
x_4	0.0625	1 1 0 1	
x_5	0.0625	1 1 1 0	
x_7	0.0625	1 1 1 1	

x_5, x_7 split into two (0.0625 probability each; 0 above the line and 1 below the line)

Message	Probability	Encoded Message	length(n_k)
x_1	0.25	0 0	2
x_6	0.25	0 1	2
x_2	0.125	1 0 0	3
x_8	0.125	1 0 1	3
x_3	0.0625	1 1 0 0	4
x_4	0.0625	1 1 0 1	4
x_5	0.0625	1 1 1 0	4
x_7	0.0625	1 1 1 1	4

x_5, x_7 split into two (0.0625 probability each; 0 above the line and 1 below the line)

Efficiency

$\eta = \frac{H(x)}{L log_2 M}$
where

- $H(x)$ is the entropy
- L is the average length of code
- $M = 2$ (2 symbols- 0,1)
- $log_2 M = log_2 2 = 1$

$\eta = \frac{H(x)}{L}$

- $H(x) = \sum_{k=1}^{8} p_k log_2 \frac{1}{p_k}$
- k number of messages
- $H(x) = \frac{1}{4} log_2 4 + \frac{1}{8} log_2 8 + \frac{1}{16} log_2 16 + \frac{1}{16} log_2 16 + \frac{1}{16} log_1 6 + \frac{1}{4} log_2 4 + \frac{1}{16} log_2 16 + \frac{1}{8} log_2 8$
- $H(x) = 2.75 bits/message$

- $L = \sum_{k=1}^{8} p_k n_k$
- $L = \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{16} \times 4 + \frac{1}{16} \times 4 + \frac{1}{16} \times 4 + \frac{1}{4} \times 2 + \frac{1}{16} \times 4 + \frac{1}{8} \times 3$
- $L = 2.75 letters/message$
- $\eta = \frac{H(x)}{L} = \frac{2.75}{2.75} = 1$
- $Efficiency(\%) = 1 \times 100 = 100\%$

Example 2

Apply Shannon Fano coding procedure for the following message ensemble.

$[x] = [x_1, x_2, x_3, x_4, x_5, x_6, x_7]$

$[P] = [0.4, 0.2, 0.12, 0.08, 0.08, 0.08, 0.04]$

- Messages are already assigned in decreasing order of probabilities
- Two ways of partitioning:
  - $[x_1, x_2]$ & $[x_3, x_4, x_5, x_6, x_7]$ - 0.6,0.4 - Difference 0.2
  - $[x_1]$ & $[x_2, x_3, x_4, x_5, x_6, x_7]$- 0.4, 0.6 - Difference 0.2
- Out of the two possibilities, choose the one with the least L value (as L is in the denominator of eff.) -Aim is to maximize efficiency.

THANK YOU