Information Theory. Shannon-Fano and Huffman coding

P .... probability of occurance of a message

I .... Information gained from the message

if $P = 1$ then $I = 0$

and if $P = 0$ then $I = \infty$

$P(x_i)$, $I(x_i)$

$$\boxed{I(x_i) = \log_2 \frac{1}{P(x_i)}} \text{ ... bits}$$

if $P(x_1) < P(x_2) \Rightarrow I(x_1) > I(x_2)$

Entropy: ... average amount of information contained in each message received.

Entropy is defined as a measure of randomness.

... average information per message.

X ... source

$(x_1, x_2, \ldots, x_n)$ ... messages

$$H(X) = E(I(X)) = \sum_{i=1}^{n} P(x_i)\, I(x_i)$$

$$\boxed{\begin{aligned} H(X) &= \sum_{i=1}^{n} P(x_i) \log_2 \frac{1}{P(x_i)} \\ &= -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i) \end{aligned}}$$

... bits/message

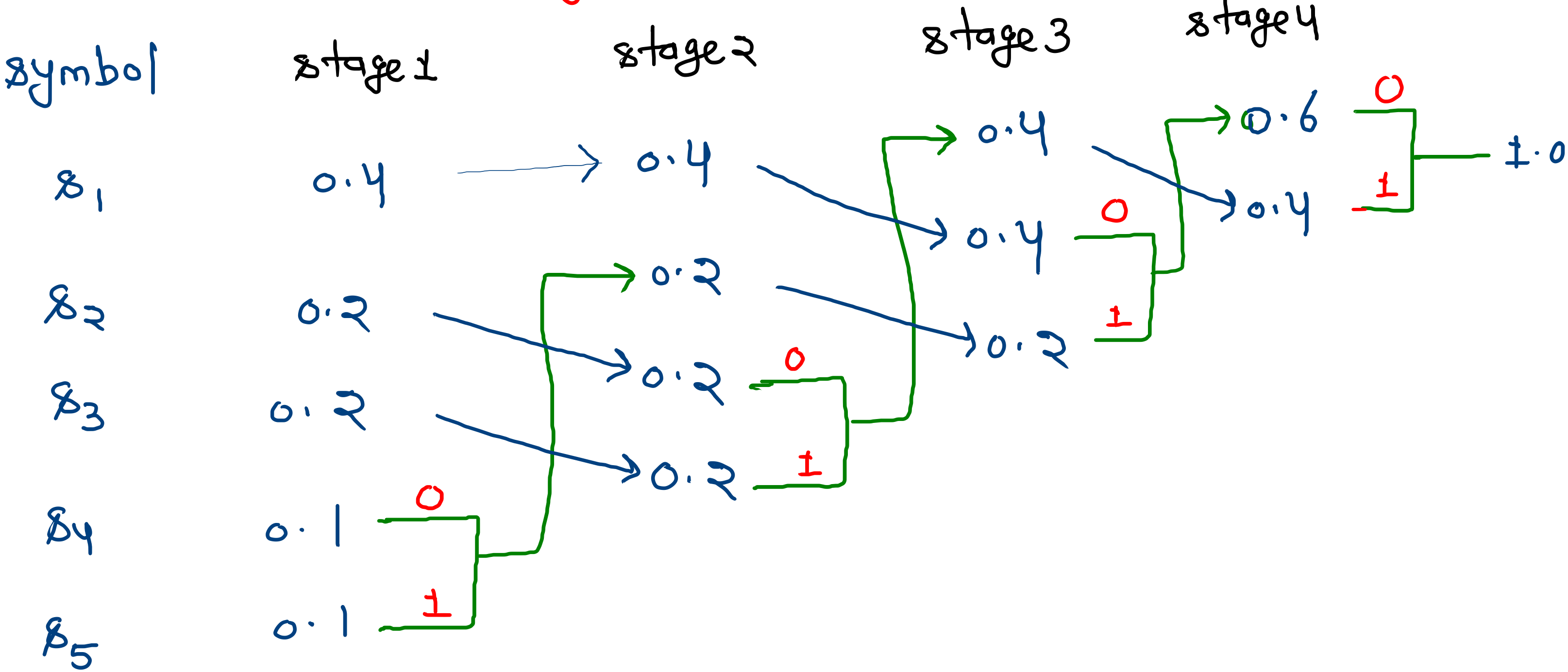
or

bits/symbol.

Entropy is maximum when all messages are equiprobable.

Huffman coding:

Ex A source 'S' emits symbol s_1, s_2, s_3, s_4, s_5 with probability 0.4, 0.2, 0.2, 0.1, 0.1. It is coded with Huffman coding. find its efficiency.



s_i	$p_i = P(s_i)$	code	code length (l_i)
s_1	0.4	00	2
s_2	0.2	10	2
s_3	0.2	11	2
s_4	0.1	010	3
s_5	0.1	011	3

Average length of the code: $\boxed{\bar{L} = \sum_{i=1}^{n} p_i l_i}$

$$= p_1 l_1 + p_2 l_2 + \cdots + p_5 l_5$$

$$= 0.4 \times 2 + \cdots + 0.1 \times 3$$

$$\bar{L} = 2.2$$

min length of code:

$$L_{min} = H(S) = \sum_{i=1}^{5} P(s_i) \log_2 \frac{1}{P(s_i)} = 2.12$$

efficiency

$$\eta = \frac{L_{min}}{\bar{L}} \times 100\,\%$$

$$\eta = \frac{2.12}{2.2} \times 100 = 96.36\%$$

redundancy $= \gamma = 1 - \eta$

Shannon - Fano coding:

Ex

S ... source $\quad s_i \Rightarrow s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8$

$P(s_i) \Rightarrow 0.5, 0.15, 0.15, 0.08, 0.08, 0.02, 0.01, 0.01$

Source	probability						
s_1	0.5	0					
s_2	0.15	1	0	0			
s_3	0.15	1	0	1			
s_4	0.08	1	1	0			
s_5	0.08	1	1	1	0		
s_6	0.02	1	1	1	1	0	
s_7	0.01	1	1	1	1	1	0
s_8	0.01	1	1	1	1	1	1

s_i	$p_i = P(s_i)$	code	length (l_i)
s_1	0.5	0	1
s_2	0.15	100	3
s_3	0.15	101	3
s_4	0.08	110	3
s_5	0.08	1110	4
s_6	0.02	11110	5
s_7	0.01	111110	6
s_8	0.01	111111	6

$$\bar{L} = \sum_{i=1}^{8} p_i l_i$$

$$L_{min} = H(X) = \sum_{i=1}^{8} p_i \log_2 \frac{1}{p_i}$$

$$\eta = \frac{L_{min}}{\bar{L}} \times 100\,\%$$

$$\text{Redundancy} = \gamma = 1 - \eta$$