

Sol: 1. Composite symbol placed "as low as possible".

$$q = r + (r-1)\alpha ; \text{ Given } q=7 \text{ and } r=2,$$

$$7 = 2 + \alpha \rightarrow \underline{\alpha=5} ; \text{ An integer.}$$

Here we have to proceed '5' stages to reduce the given sources to 2 symbols.

S			S _a		S _b		S _c		S _d		S _e	
Source	p	code	p	code	p	code	p	code	p	code	p	code
S ₁	0.4		0.4		0.4		0.4		0.4		0.6	0
S ₂	0.2		0.2		0.2		0.2		0.4		0.4	1
S ₃	0.1		0.1		0.2		0.2		0.2			
S ₄	0.1		0.1		0.1		0.2					
S ₅	0.1		0.1		0.1							
S ₆	0.05		0.1									
S ₇	0.05											

Code Table :

Source symbol	prob	n. code	code	length
S ₁	0.4	1	1	1
S ₂	0.2	10	01	2
S ₃	0.1	0100	0010	4
S ₄	0.1	1100	0011	4
S ₅	0.1	0000	0000	4
S ₆	0.05	01000	00010	5
S ₇	0.05	11000	00011	5

$$\text{Avg lgf; } L = \sum_{i=1}^7 p_i \lg p_i$$

$$= 0.4 \times 1 + 0.2 \times 2 + 0.1 \times 4 \times 3 +$$

$$0.05 \times 5 \times 2$$

$$= \underline{2.5 \text{ bits/msg}}$$

$$H(x) = - \sum_{i=1}^7 p_i \log p_i$$

$$= \underline{2.4219 \text{ bits/msg}}$$

$$\eta_c = \frac{H(x)}{L} = \frac{2.4219}{2.5} \times 100 = \underline{96.88\%}$$

Variance, $\text{Var}(x) = E[(x - \mu)^2]$ where $\mu \rightarrow$ average value.

Then the variance of the word length;

$$\text{Var}(L^0) = \sum_{i=1}^7 p_i^0 (L_i^0 - \mu)^2$$

$$= 0.4(1-2.5)^2 + 0.2(2-2.5)^2 + 0.1(4-2.5)^2 \times 3 + 0.05(5-2.5)^2 \times 2$$

$$= \underline{\underline{2.25}}$$

2. Composite symbol placed "as high as possible".

	S	S_a	S_b	S_c	S_d	S_e
Source	p	code	p	code	p	code
S_1	0.4	0.4	0.4	0.4	0.4	0.6
S_2	0.2	0.2	0.2	0.2	0.4	0.4
S_3	0.1	0.1	0.2	0.2	0.2	0.2
S_4	0.1	0.1	0.1	0.2	0.2	0.2
S_5	0.1	0.1	0.1	0.1	0.2	0.2
S_6	0.05	0.1	0.1	0.1	0.2	0.2
S_7	0.05	0.1	0.1	0.1	0.2	0.2

Code Table :

Source symbol	prob	x. code word	code word	length
S_1	0.4	00	00	2
S_2	0.2	11	11	2
S_3	0.1	110	011	3
S_4	0.1	001	100	3
S_5	0.1	101	101	3
S_6	0.05	0010	0100	4
S_7	0.05	1010	0101	4

$$L = \sum_{i=1}^7 p_i L_i^0$$

$$= 0.4 \times 2 + 0.2 \times 2 + 0.1 \times 3 \times 3 + 0.05 \times 4 \times 2$$

$$= \underline{\underline{2.5 \text{ bits/msg}}}$$

$$\text{Hes} = \underline{\underline{2.4219 \text{ bits/msg}}}$$

$$\eta_c = \frac{H(s)}{L} = \frac{2.4219}{2.5} \times 100 = \underline{\underline{96.88\%}}$$

$$\begin{aligned} \text{Var}(L^o) &= \sum_{p=1}^7 p_i (L^o - L)^2 \\ &= 0.4(2 - 2.5)^2 + 0.2(2 - 2.5)^2 + 0.1(3 - 2.5)^2 \times 3 \\ &\quad + 0.05(4 - 2.5)^2 \times 2 \\ &= \underline{\underline{0.45}} \end{aligned}$$

When the composite symbol is moved as high as possible, the variance of the word length over source symbols would become smaller, which is desirable.

Q9. Consider a source with 8 alphabets A to H with respective probabilities of 0.22, 0.20, 0.18, 0.15, 0.10, 0.08, 0.05, 0.02.

(a) Construct a binary compact code using Huffman coding and determine the code efficiency.

(b) Construct a ternary compact code using Huffman coding and determine the code efficiency.

(c) Construct a quaternary compact code and determine the code efficiency.

Ans. (a) Binary code.

Code table:

Msg	prob	Code word	Length
A	0.22	10	2
B	0.20	11	2
C	0.18	000	3
D	0.15	001	3
E	0.10	011	3
F	0.08	0100	4
G	0.05	01010	5
H	0.02	01011	5

$$L = \underline{\underline{2.8 \text{ bits/msg}}}$$

$$H(s) = \underline{\underline{2.7535 \text{ bits/msg}}}$$

$$\eta_c = \underline{\underline{98.34\%}}$$

⑥ Huffman Ternary Code :

Consider $q = r + (r-1)\alpha$

where $q = 8$; $r = 3$.

Substitute $r \rightarrow q = 3 + 2\alpha$

$$\alpha = \frac{q-3}{2}$$

Then q should have the value 5, 7, 9, 11 ... to ' α ' be an integer.

if $q = 8$; $\alpha = \frac{8-3}{2} = \frac{5}{2} = \underline{2.5}$; not an integer

if $q = 9$; $\alpha = \frac{9-3}{2} = \underline{3}$.

So add "one dummy symbol" 'I' with zero probabilities. So $q = 9$; with code alphabet $x = \{0, 1, 2\}$.

Source "s"			"sa"		"sb"		"sc"	
Source	prob	code	prob	code	prob	code	prob	code
A	0.22		0.22		0.25		0.53	0
B	0.2		0.2		0.22		0.25	1
C	0.18		0.18		0.2	0	0.22	2
D	0.15		0.15		0.18	1		
E	0.1		0.1	0	0.15	2		
F	0.08		0.08	1				
G	0.05	0	0.07	2				
H	0.02	1						
I	0	2						

Discard

Dummy Symbol

code Table :

Msg	prob	n.code	code	length
A	0.22	2	2	1
B	0.2	00	00	2
C	0.18	10	01	2
D	0.15	20	02	2
E	0.10	01	10	2
F	0.08	11	11	2
G	0.05	021	120	3
H	0.02	121	121	3

$$L^{(3)} = 1.85 \text{ bits/msg}$$

$$H(s) = 2.7585 \text{ bits/msg}$$

$$H_3(s) = \frac{H(s)}{\log 3} = \frac{1.7373 \text{ ternary}}{\text{unit/msg}}$$

$$\eta_c^{(3)} = \frac{H_3(s)}{L^{(3)}} = \frac{1.7373}{1.85} = 93.91\%$$

(c) Huffman Quaternary Code

Here $q = 8$ and $r = 4$.

We know that $q = r + (r-1)\alpha$

$$q = 4 + 3\alpha$$

$$\alpha = \frac{q-4}{3};$$

q should be 7, 10, 13, 16

Given that $q = 8$, so take new $q = 10$ by adding two "dummy" symbols 1 and 2 each with probabilities zero.

if $q = 10$; $\alpha = \frac{10-4}{3} = 2$, An integer.

with code alphabet $x = \{0, 1, 2, 3\}$.

Some "s"			"sa"		"sb"	
Source	prob	code	prob	code	prob	code
A	0.22		0.22		0.4	0
B	0.2		0.2		0.22	1
C	0.18		0.18		0.2	2
D	0.15		0.15		0.18	3
E	0.1		0.1	1		
F	0.08		0.08	2		
G	0.05		0.07	3		
H	0.02	1				
I	0	2				
J	0	3				

Dummy symbols

Discard

Code Table

Msg	prob	codeword	codeword	lgf
A	0.22	1	1	1
B	0.2	2	2	2
C	0.18	3	3	1
D	0.15	00	00	2
E	0.10	10	01	2
F	0.08	210	012	2
G	0.05	030	030	3
H	0.02	130	031	3

$$L^{(4)} = 1.47 \text{ quaternary digits/msg}$$

$$H(s) = 2.7535 \text{ bits/msg}$$

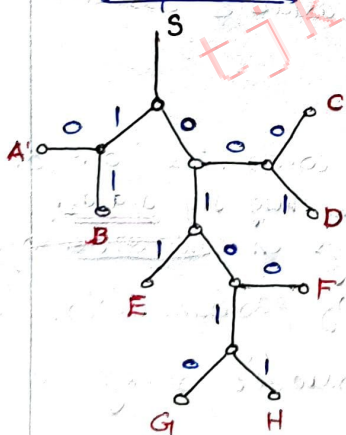
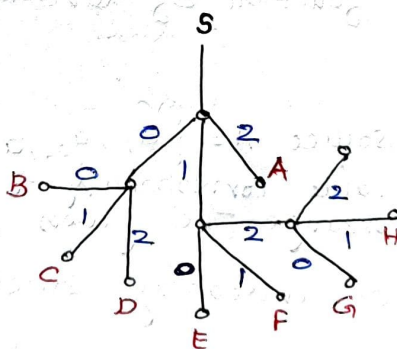
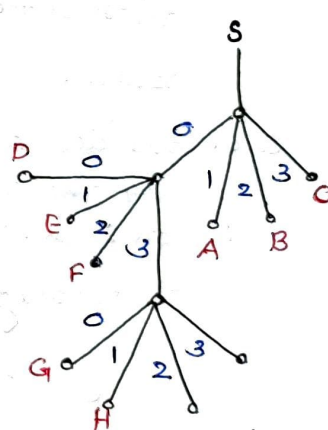
$$H_4(s) = \frac{H(s)}{\log 4} = 1.37675$$

$$\text{quaternary units/msg}$$

$$\eta^{(4)} = \frac{H_4(s)}{L^{(4)}} = 93.66\%$$

Comparison

Type of Coding	Code efficiency
Binary coding	98.34%
Ternary coding	93.91%
Quaternary coding	93.66%

Code tree① Binary code② Ternary code③ Quaternary Code

Q10. Apply Huffman encoding procedure for the following set of msgs and determine the η_c of the binary code formed.

$$\begin{array}{ccc} x_1 & x_2 & x_3 \\ 0.7 & 0.15 & 0.15 \end{array}$$

Also apply Huffman coding to the second order extension for the above msg and calculate how much the efficiency be improved?