Assignment-2

① classes of codes. Consider the code {00, 11, 001}

ⓐ Is it non-singular? why?

ⓑ Is it uniquely decodable? Explain.

ⓒ Is it Instantaneous? why?

Ans ⓐ The codes {00, 11, 001} are non-singular because, every string maps to a different element.

$$\left.\begin{array}{l} a \rightarrow 00 \\ b \rightarrow 11 \\ c \rightarrow 001 \end{array}\right\} \text{non-singular.}$$

ⓑ A code is called uniquely decodable if its extension is non-singular. Here 00, is not the prefix of 11. Hence the dangling suffix is not entered.

ⓒ The instantaneous code is defined when no code is prefix of the other.

{00, 11, 001} is not an instantaneous code because (00) is a prefix of the code (001).

② Huffman coding. Consider the random variable

X.
$$X = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 & x_7 \\ 0.49 & 0.26 & 0.12 & 0.04 & 0.04 & 0.03 & 0.02 \end{bmatrix}$$

ⓐ Find the binary Huffman code for X.

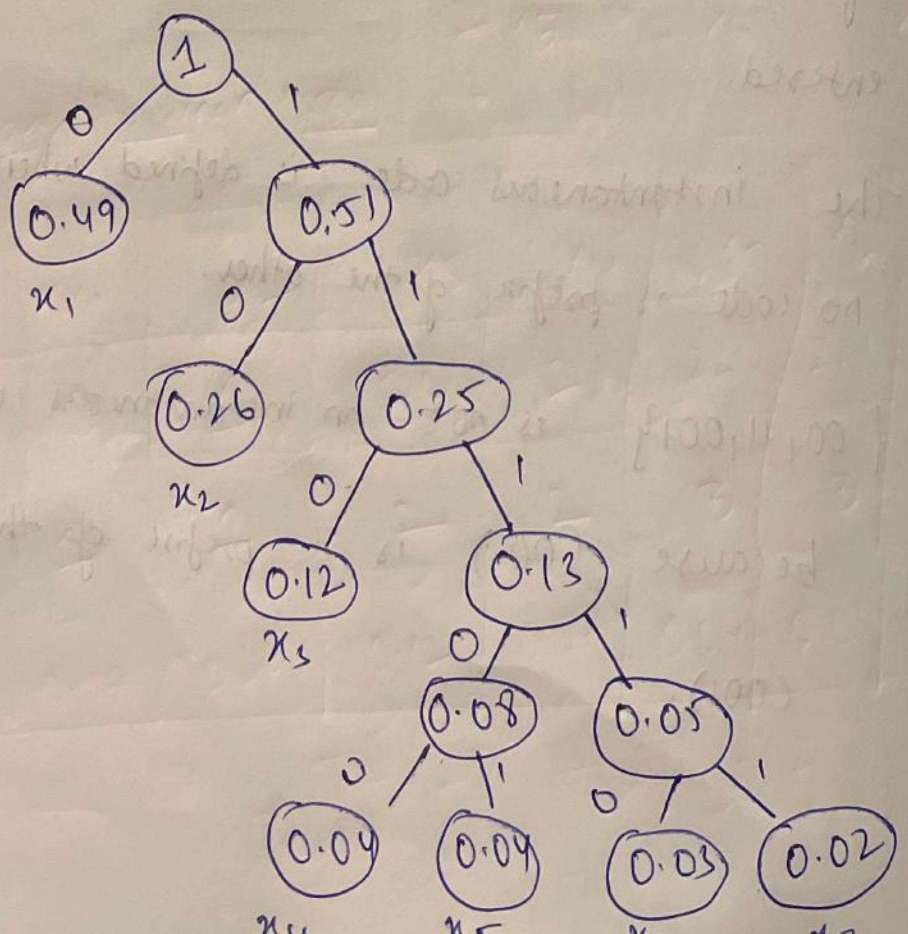
ⓑ Find the expected code length for this encoding

ⓒ Find a ternary Huffman code for X and find the average length.

Sol: Preparing a table,

ⓐ

x	x_1	x_2	x_3	x_4	x_5	x_6	x_7
P	0.49	0.26	0.12	0.04	0.04	0.03	0.02



The Huffman code table is displayed.

X	Codeword	length(l)	Probability $P(x_i)$
x_1	0	1	0.49
x_2	10	2	0.26
x_3	110	3	0.12
x_4	11100	5	0.04
x_5	11101	5	0.04
x_6	11110	5	0.03
x_7	11111	5	0.02

(b) To find the code length of the given Huffman code

We have to find L_{av} (Average Length).

$$L_{av} = \sum_{i=1}^{7} l \times p(x_i)$$

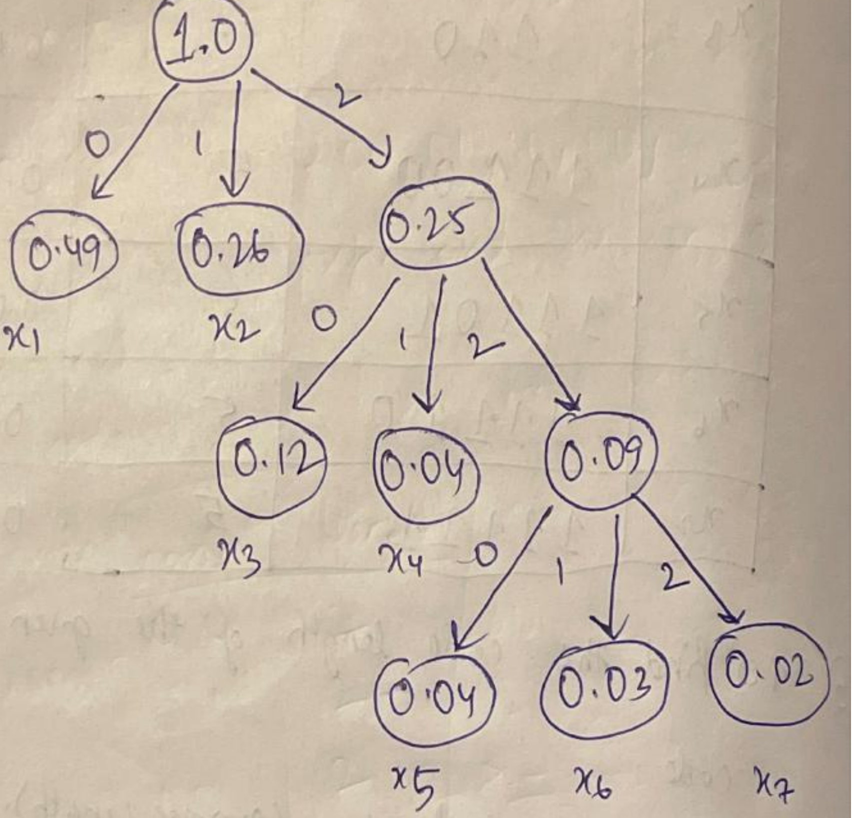
$$\Rightarrow (0.49 \times 1) + (0.26 \times 2) + (0.12 \times 3) + (0.04 \times 5) + (0.04 \times 5) + (0.03 \times 5) + (0.02 \times 5)$$

$$L_{av} = 0.49 + 0.52 + 0.36 + 0.2 + 0.2 + 0.15 + 0.1$$

$$= 2.02$$

Average length (L_{av}) = 2.02 bits

(c) Solving ternary Huffman code for X and finding the average length.



Ternary Huffman code

X	x_1	x_2	x_3	x_4	x_5	x_6	x_7
codeword	0	1	20	21	220	221	222
length	1	1	2	2	3	3	3
Probability	0.49	0.26	0.12	0.04	0.04	0.03	0.02

finding the average length for Ternary Huffman code.

$$L_{av} = \sum_{i=1}^{7} l \times P(x_i)$$

$$= (0.49 \times 1) + (0.26 \times 1) + (0.12 \times 2) + (0.04 \times 2) + (0.04 \times 3) + (0.03 \times 3) + (0.02 \times 3)$$

$$= 0.49 + 0.26 + 0.24 + 0.08 + 0.12 + 0.09 + 0.06$$

$$= 1.34$$

$$\boxed{L_{av} = 1.34 \text{ bits}}$$

---

③ Bad codes. Which of these codes cannot be Huffman codes for any probability assignment?

a) $\{0, 10, 11\}$ b) $\{00, 01, 10, 110\}$ c) $\{01, 10\}$.

Sol:

ⓐ $\{0, 10, 11\}$

The given codes are 0, 10, 11.

let's assume,

$a \rightarrow 0 \quad l_1 = 1$

$b \rightarrow 10 \quad l_2 = 2$

$c \rightarrow 11 \quad l_3 = 2$

$$2^{-l_i} \Rightarrow 2^{-l_1} + 2^{-l_2} + 2^{-l_3}$$

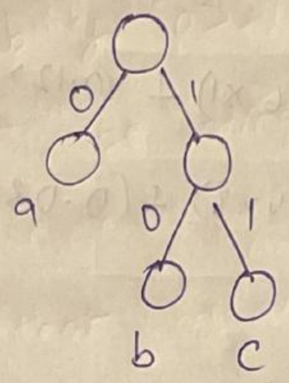
$$\Rightarrow 2^{-1} + 2^{-2} + 2^{-2}$$

$$= \frac{1}{2} + \frac{1}{4} + \frac{1}{4}$$

$$= \frac{1}{2} + \frac{1}{2} \Rightarrow \underline{1}$$

we can construct,

trying to from a Huffman code as $2^{-l_i} = 1$.



codes a, b, c $\{0, 10, 11\}$

can construct Huffman code.

b) given code is $\{00, 01, 10, 110\}$.

$\Rightarrow$ This code is a prefix and the length is minimal.

given code is

code	length	
00	2	l_1
01	2	l_2
10	2	l_3
110	3	l_4

To prove that a code is Huffman which have lengths (l_i), then the sum of 2^{-l_i} should be equal to 1.

adding,

$$\Rightarrow 2^{-l_1} + 2^{-l_2} + 2^{-l_3} + 2^{-l_4}$$

$$\Rightarrow 2^{-2} + 2^{-2} + 2^{-2} + 2^{-3}$$

$$\Rightarrow \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{8} \Rightarrow$$

$$\frac{3}{4} + \frac{1}{8} \Rightarrow \frac{6+1}{8} \Rightarrow \frac{7}{8} < 1.$$

$\therefore 2^{-l_i} < 1$, Huffman code cannot be constructed.

c) $\{01, 10\}$

given,

code	01	10
length	2	2
l_i	l_1	l_2

the sum of 2^{-l_i} should be equal to 1.

$$\Rightarrow 2^{-2} + 2^{-2} \Rightarrow \frac{1}{4} + \frac{1}{4} \Rightarrow \frac{1}{2} < 1.$$

$\therefore 2^{-l_i} < 1$. Huffman code can't be constructed.

---

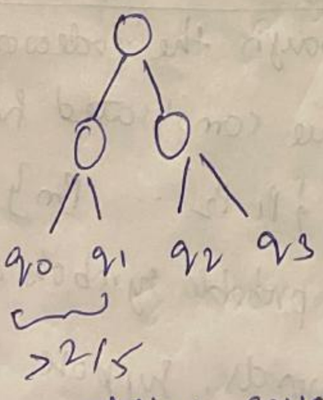
4) Optimal codeword lengths. Although the codeword length. ~~Although the~~ code are complicated function of the message probabilities $\{P_1 P_2 \dots P_m\}$. It can be said that less probable symbols are encoded into longer codewords. Suppose that the message probabilities are given in decreasing order $P_1 \geq P_2 \geq \dots \geq P_m$.

(a) Prove that for any binary Huffman code, if the most probable message symbol has probability $p > 2/5$, then that symbol must be assigned a codeword of length 1.

(b) Prove that for any binary Huffman code, if the most probable message symbol has probability $p < 1/3$, then that symbol must be assigned a codeword of length ≥ 2.

Sol:- Let's assume that the most probable

(a) symbol with $p > 2/5$ has length ≥ 1. length is equal to Lav.



$$q_1 + q_2 > \frac{2}{5} \quad — ①$$

$$\Rightarrow \quad q_2 + q_3 > \frac{2}{5} \quad — ②$$

$$q_1 + q_3 > \frac{2}{5} \quad — ③$$

adding equations ①, ② & ③.

$$2(q_1 + q_2 + q_3) > \frac{2}{5} + \frac{2}{5} + \frac{2}{5}$$

$$q_1 + q_2 + q_3 > \frac{6}{5} \times \frac{1}{2}$$

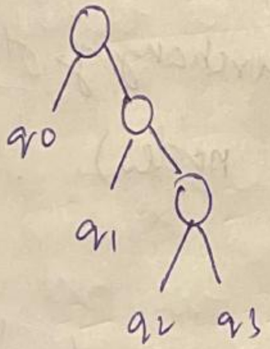
$q_1+q_2+q_3 > \frac{3}{5}$

$q_0+q_1+q_2+q_3 > 1.$

let's assume, $q_2+q_3 \leq 2/5.$

length $= L_{av}'$.

tree $\Rightarrow L_{av}-L_{av}' = q_0-(q_2+q_3) > 0$



$\Rightarrow q_0 > \frac{2}{5}$ & $q_2+q_3 \leq \frac{2}{5}.$

$\Rightarrow$ The newly constructed tree has $P > \frac{2}{5}$

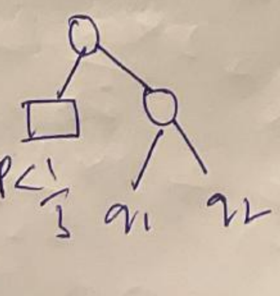
$\Rightarrow$ Thus, It has a better average.

$\Rightarrow$ The symbol with $P > \frac{2}{5}$ should be assigned with a word of length equals to one (1).

(b) Given, most probable symbol of the new probability $< 1/3$.

The length of the codeword should be ≥ 2.

eg:

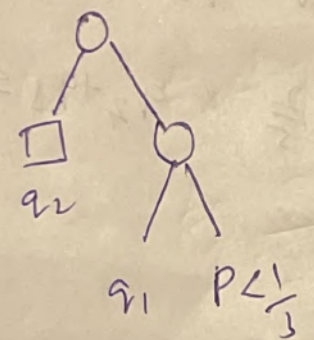


if $P < \frac{1}{3}$ then,

$$q_1 + q_2 > \frac{2}{3}$$

symbol $q_2 > \frac{1}{3}$.

(Average. Length of L_{av}). for $P < \frac{1}{3}$ and $q_1 + q_2 > \frac{2}{3}$



(∴ interchanging the root nodes).

let the ~~les~~ average length be L_{av}'.

$$L_{av} - L_{av}' = q_2 - P > 0$$

this is because $q_2 > \frac{1}{3}$, $P < \frac{1}{3}$.

$$q_2 - P > 0$$

∴ symbol given should be assigned a word of length ≥ 2.

a) Given,
6 bottles of wine.

Good taste wines = 5

bad taste wine = 1.

the probability P_i for

P_1	P_2	P_3	P_4	P_5	P_6
$\frac{8}{23}$	$\frac{6}{23}$	$\frac{4}{23}$	$\frac{2}{23}$	$\frac{2}{23}$	$\frac{1}{23}$
1	2	3	4	5	5

expected number of testing is →

finding average length, $L_{av} = \sum_{i=1}^{6} P_i l_i$

$$\Rightarrow \left(1 \times \frac{8}{23}\right) + \left(\frac{6}{23} \times 2\right) + \left(\frac{4}{23} \times 3\right) + \left(\frac{2}{23} \times 4\right)$$

$$+ \left(\frac{2}{23} \times 5\right) + \left(\quad\right)$$

$$\Rightarrow \left(\frac{8 + 12 + 12 + 8 + 10 + 8}{23}\right)$$

$$\Rightarrow \frac{50}{23} \Rightarrow 2.19$$

$\boxed{L_{av} = 2.17 \text{ bits}}$ 2.17.

The expected number of tastings required are 2.17 bits.

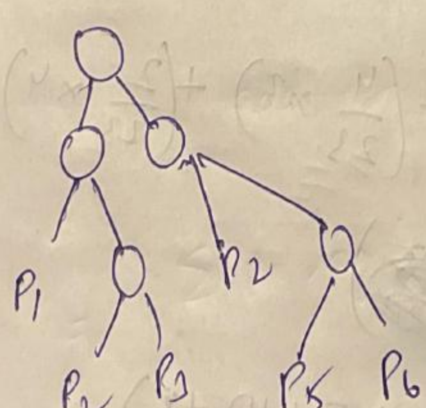
b) Taking the Huffman model as an example, The tasting of the wine should be started in a process.

The first probability is P_1.

where $P_1 = \frac{8}{23}$.

The first probability of tasting should be $\frac{8}{23}$.

c) Using Huffman code, the tree is constructing



average length,

$L_{av} = \sum l_i P(x_i)$.

if the glasses are mixed we get length as 2, 2, 2, 3, 4, 4.

$L_{av} = \sum_{i=1}^{6} P_i l_i$

$\Rightarrow 2 \times \frac{8}{23} + 2 \times \frac{6}{23} + 2 \times \frac{4}{23} + 3 \times \frac{2}{23}$

$+ 4 \times \frac{2}{23} + 4 \times \frac{1}{23} \Rightarrow$

$$\Rightarrow \frac{16+12+8+6+8+4}{23}$$

$$\Rightarrow \frac{54}{23} = 2.35 \text{ bits.}$$

$\boxed{\text{Average number of tastings required are } 2.35 \text{ bits}}$

(6) a) The probability distribution table is

symbol	probability	length	codeword	F_i decimal	F_i binary
1	0.5	1	0	0	0.0
2	0.25	2	10	0.5	0.10
3	0.125	3	110	0.75	0.110
4	0.125	3	111	0.875	0.111

b) since $l_i = \lceil \log \frac{1}{p_i} \rceil$,

we have $\log \frac{1}{p_i} - 1 \leq l_i < \log \frac{1}{p_i} + 1$

which implies that

$$l_i - 1 < \log\frac{1}{p_i} \le l_i$$

$$2^{l_i - 1} < \frac{1}{p_i} \le 2^{l_i}$$

$$\therefore \; 2^{-l_i} \le p_i \le 2^{-l_i + 1}$$

Hence $l_i = \log\frac{1}{p_i}$ //.

$$\Rightarrow \log\frac{1}{p_i} \le l_i < \log\frac{1}{p_i} + 1$$

$$\Rightarrow p_i \log\frac{1}{p_i} \le p_i l_i < p_i \log\frac{1}{p_i} + p_i$$

add $\cancel{\&}$ Σ.

$$= \sum_i p_i \log\frac{1}{p_i} \le \sum_i p_i l_i < \sum_i p_i \log\frac{1}{p_i} + \sum_i p_i$$

$$\Rightarrow H(x) = \sum_i p_i \log\frac{1}{p_i} \qquad L(x) = \sum_i p_i l_i$$

$$\Rightarrow \quad H(x) \le L(x) < H(x) + 1.$$

$$\boxed{\therefore \; H(x) \le L(x) < H(x) + 1}$$ //.

let x_k ~~be~~ be the code for symbol k.

x_k cannot be a prefix for x_i, $i<k$

because $l_i \le l_k$, there is a possibility

that x_i and x_k could be identical, but

this is ~~ca~~ covered by the following case by

swapping the roles of i and k.

assuming that x_k can't be prefix for

x_{k+j}.

Assume x_k is a prefix of x_{k+j}.

Then x_k and x_{k+j} must agree l_k first bits.

Therefore,

$$F_{k+j} - F_k < 2^{-l_k}$$

$$F_{k+j} - F_k < 2^{-l_k}$$

$$\Rightarrow \sum_{i=1}^{k+j-1} p_i - \sum_{i=1}^{k-1} p_i < 2^{-l_k}$$

$$\Rightarrow \sum_{i=k}^{k+j-1} p_i < 2^{-l_k}$$

$$\Rightarrow p_k < 2^{-l_k}.$$

* $l_k = \log \frac{1}{p_k}$

$$l_k \geq \log \frac{1}{p_k}$$

$$\Rightarrow 2^{l_k} \geq \frac{1}{p_k}$$

$$\boxed{\therefore 2^{-l_k} \leq p_k}$$

Therefore x_k cannot be a prefix for x_{k+1}

the Shannon code is prefix.