{ Shannon, Fano, Huffman }.

Now we know $L \geq H_D(X)$.
but we have to prove

$$\boxed{H_D(X) \leq L \leq H_D(X) + 1}$$

$l_i \geq -\log_D P(x_i)$ } may not be an integer.

$\Rightarrow$ $l_i \geq [-\log_D P(x_i)]$ } for opt: $l_i = [-\log_D P(x_i)]$

$$\sum_i D^{-l_i} = \sum_i D^{-[-\log_D P(x_i)]}$$

$$\leq \sum_i D^{-(-\log_D P(x_i))} = \sum P(x_i) = 1$$

The code with $l_i = [-\log_D(P(x_i))]$ is called shannon code
The above proof shows that such code is possible ($\sum D^{-l_i} \leq 1$)

$$-\log_D P(x_i) \leq l_i \leq -\log_D P(x_i) + 1$$

$$E[-\log_D P(x_i)] \leq E[l_i] \leq E[-\log_D P(x_i) + 1]$$

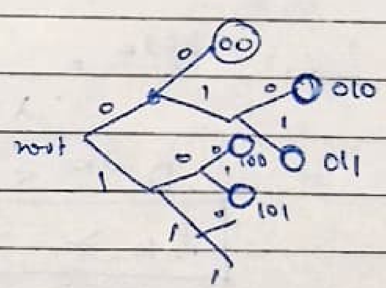
$\boxed{x \leq [x] \leq x+1}$

$$\boxed{H_D(X) \leq L \leq H_D(X) + 1}$$

and

$\boxed{l_i = [-\log_D P(x_i)]}$

Symbol:	A	B	C	D	E
freq:	15	7	6	6	5
Prob	15/39	7/39	6/39	6/39	5/39
$-\log_2 P(x_i)$	1.379	2.480	2.7	2.7	2.963
$[-\log_2 P(x_i)] = l_i$	2	3	3	3	3
codewords	00	010	011	100	101



$H(X) = 0.530 + 0.445 + 0.830 + 0.379 = 2.184$

$L(C) = \dfrac{30 + 21 + 36 + 15}{39} = \dfrac{\cancel{102}^{34}}{\cancel{39}_{13}} = \dfrac{34}{13} = 2.61$

} check

Shannon Code

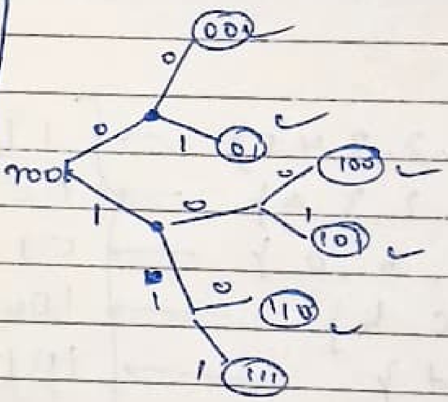
Suppose we have an alphabet $(x_1, x_2, \dots x_n)$. with prob. $(p_1, p_2, \dots p_n)$. The desired codeword lengths are

$$l_i = \lceil -\log_2 p_i \rceil.$$

* · Construct shannon code

x	p(x)
1	0.25
2	0.25
3	0.20
4	0.15
5	0.15

x	1	2	3	4	5
$p(x)$	0.25	0.25	0.20	0.15	0.15
$-\log_2 p(x)$	2	2	2.32	2.736	2.736.
l_i	2	2	3	3	3
Code	00	01	100	101	110



▷ **Huffman Code:**

Suppose there is an alphabet

(Step 1) $A = \{x_1, x_2, \dots x_n\}$. Pick up two x_i & x_j which have smallest prob p_i & p_j. Create a subtree that has two characters x_i and x_j ~~atleast~~ at leaves.

(Step 2). Set up new prob $\longrightarrow$ (ij) x_{ij}

$$p_{ij} = p_i + p_j$$

Remove x_i & x_j and add x_{ij}

$$A' = A \cup \{x_{ij}\} - \{x_i, x_j\}.$$

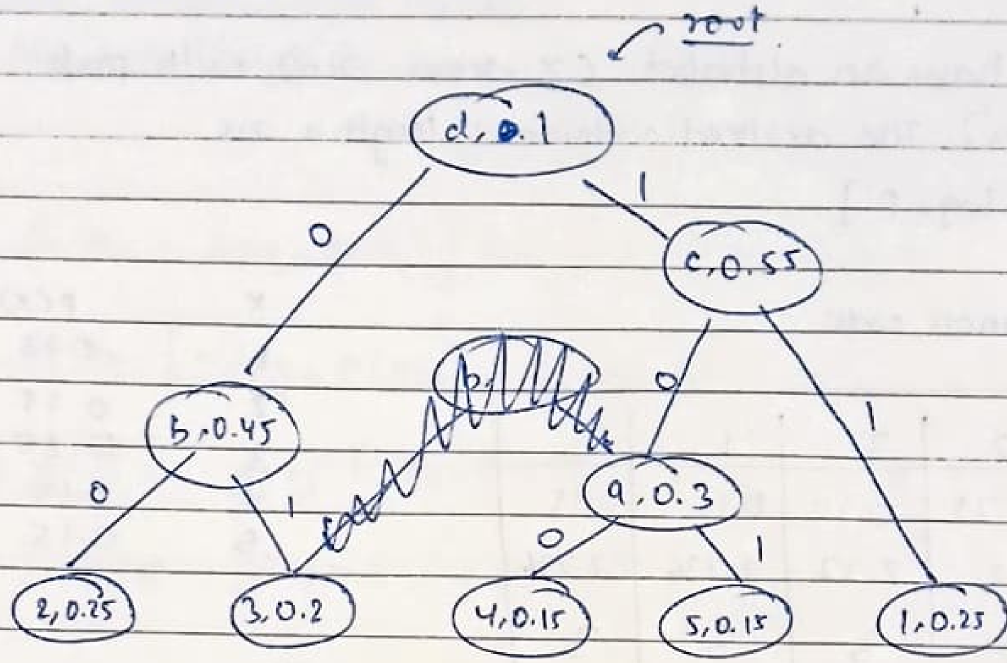
$$|A'| = |A| - 1$$

(Step 3). Repeat this merging until one symbol is left.

$$|A^{(n)}| = 1$$

Q Find Huffman code

x	1	2	3	4	5
p(x)	0.25	0.25	0.2	0.15	0.15



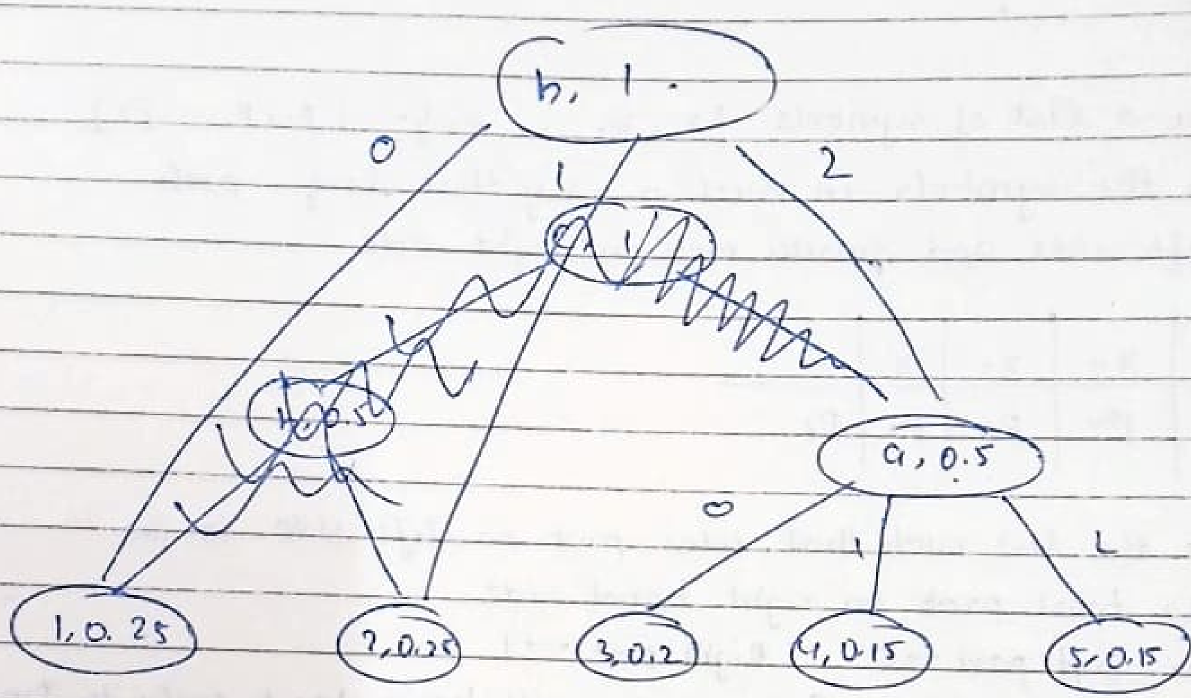
step

step	set		code		x
1	{1, 2, 3, 4, 5}	→	11	→	1
2	{1, 2, 3, a}	→	00	→	2
3	{1, b, a}	→	01	→	3
4	{c, b}	→	100	→	4
5	{d}	→	101	→	5

$$L(c) = \frac{2}{4} + \frac{2}{4} + \cancel{\frac{3}{5}} + \frac{2}{5} + 3 \times 0.15 \times 2$$

$$= 1 + 0.4 + 0.9 = 2.3$$

$$\begin{array}{r} 15 \\ \times\ 6 \\ \hline 0.90 \end{array}$$



Step	A
1	{1, 2, 3, 4, 5}
2	{1, 2, a}
~~3~~	~~{b, a}~~
~~4~~	~~{c}~~
3	{b}

$1 \longrightarrow 0$

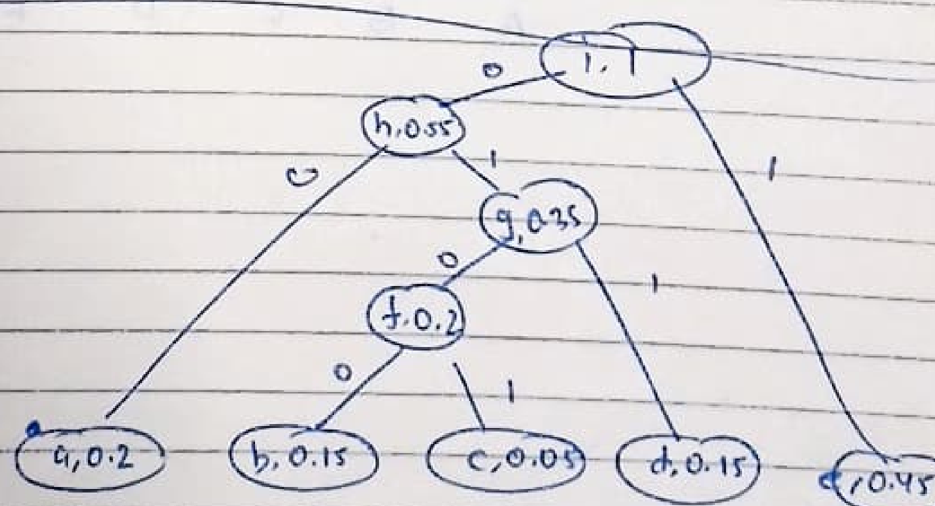
$2 \longrightarrow 1$

$3 \longrightarrow 20$

$4 \longrightarrow 21$

$5 \longrightarrow 22$

A	freq	P
a	20	0.2
b	15	0.15
c	5	0.05
d	15	0.15
e	45	0.45
	100	



00 0100 0101 011 1

▶ Fano code:

(1) → We have a list of symbols $\{x_1, x_2, \dots x_n\} \rightarrow \{p_1, p_2 \dots p_n\}$.

(1) → Divide the symbols in such a way that larger prob on left side and smaller prob on right side

S	x_4	x_3	x_1	x_2
P	P_4	P_3	P_1	P_2

(3) → Divide the list such that total prob on left side is as close as total prob on right hand side

(4) → Assign Left part → 0 Right part → 1

(5) → Apply Recursively → Each symbol will have leaf code in tree

symbol	A	B	C	D	E
prob.	0.385	0.179	0.154	0.154	0.128
	0		1		
	0	1	0	1	
				0	1
	00	01	10	110	111
	A	B	C	D	E