ISTA-311

Name - Siddhant Bhardwaj
Collab - Vibhor Mehta

1. $H(X) = \sum p_i \cdot \log_2(1/p_i)$

$= 0.3 \times \log_2\left(\frac{1}{0.3}\right) + 0.05 \times \log_2\left(\frac{1}{0.05}\right) + 0.4 \times \log_2\left(\frac{1}{0.4}\right)$

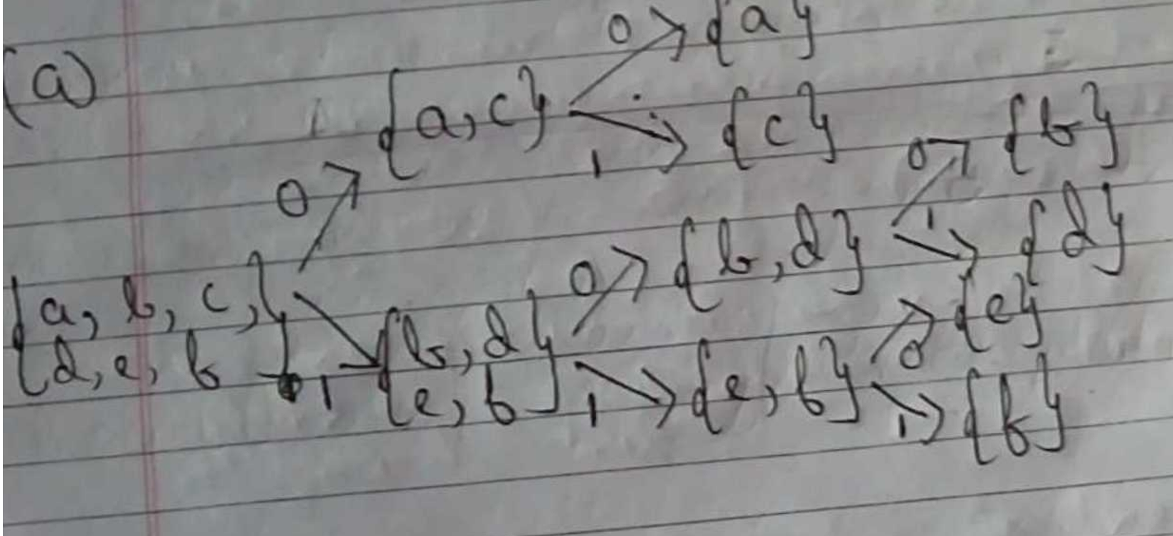
$+ (0.1) \times \log_2(0.1) + (0.15) \times \log_2(0.15)$

$= 2.01$ bits (approx.) (Ans).

2.

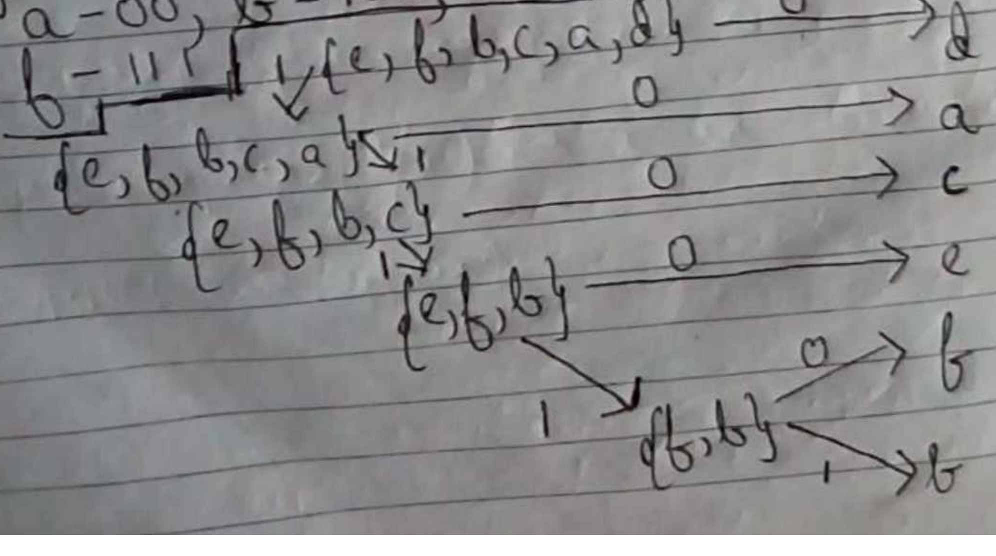
Symbol	a	b	c	d	e	f
Freq.	0.35	0.02	0.15	0.40	0.05	0.03

(a)



Symbol Codes ÷ a - 00, b - 100, c - 01, d - 101, e - 110, f - 111

(b)



Symbol codes :-

a - 10, b - 11111, c - 110, d - 0,
e - 1110, f - 11110

3. Avg. Length by Greedy Algorithm =

$0.35 \times 2 + 0.02 \times 3 + 0.15 \times 2 + 0.4 \times 3$
$+ 0.05 \times 3 + 0.03 \times 3$
$= 0.7 + 0.06 + 0.3 + 1.2 + 0.15 + 0.09$
$= 2.42$ bits.

Avg. Length by Huffman Algorithms =

$0.4 \times 1 + 0.35 \times 2 + 0.15 \times 3 + 0.05 \times 4$
$+ 0.03 \times 5 + 0.02 \times 5$

$= 2.0$ bits.

Now, entropy $= \sum p_i \cdot \log_2(1/p_i)$

$= 0.35 \times \log_2\left(\frac{1}{0.35}\right) + 0.02 \times \log_2\left(\frac{1}{0.02}\right)$

$+ 0.15 \times \log_2\left(\frac{1}{0.15}\right) + 0.4 \times \log_2\left(\frac{1}{0.4}\right)$

$+ 0.05 \times \log_2\left(\frac{1}{0.05}\right) + 0.03 \times \log_2\left(\frac{1}{0.03}\right)$

$= 1.95$ bits (approx).

So, Huffman Algorithm performs closest & best.

4\.

(a) An example of a yes-no question that could be asked is to divide it into 2 626 card' divisions of red & black ones. Thus, the 1st question would be - Is the card red or black?

(b) By Information Theory, shannon information a information content is inversely related to uncertainity. Now, let x_i be the discovery of the particular card. Now, $p(x_i) = 1/52$.

By Information theory,

$\therefore$ information content $= \log_2\left(\frac{1}{\frac{1}{52}}\right)$

$= \log_2(52) = 5.\bar{7} \sim 6$ (approx.).

So, the person would have to ask approx. 6 questions to get knowledge of correct card.

5\. (a) Entropy per character for the given distribution was calculated by the following python code:

```
sum([i*np.log2(1/i) for i in letterdicts.keys()])
```

$= 4.109$ bits (approx.)

(b) Since ASCII uses 8 bits / symbol and assuming that the given distribution is correct, then the size of the text-file can be reduced by a factor of 2 approximately.

6. (a) Information content is related to entropy. If $p(x_i)$ is the probability of an event x_i, then the information content is $\log_2\left(\frac{1}{p(x_i)}\right) = I(x_i)$

$\therefore$ $I(x_i)$ decreases with $p(x_i)$.

$\therefore$ The no. of times Jeffrey will have to check for the poisoned barrel will be

(c) Now, size of hpl cleaned.txt is 404 KB is compressed to a size of 144 KB, which is much less than what a reduction by a factor of 2 would imply. This is not possible, hence the assumption of the random source generating characters being selected independently is wrong since letters / symbols in the English alphabets are not selected independently.

6. (a) Information content is related to entropy. If $p(x_i)$ is the probability of an event x_i, then the information content is $\log_2\left(\frac{1}{p[x_i]}\right) = I(x_i)$.

Hence, $I(x_i)$ decreases with $p(x_i)$.

LOKENATH
Page No.
Date

$\therefore$ The no. of times Joffrey will have to check for the poisoned barrel will be dependent on the number of times he can poison his prisoners. The information content is given by $\log_2\left(\frac{1}{p(x_i)}\right)$, where x_i is discovery of poisoned bottle. $\therefore p(x_i) = 1/1000$

So, $I(x_i) = \log_2\left(\frac{1}{\frac{1}{1000}}\right) = \log_2(1000)$

$= 10$ (approx).

Hence, Joffrey will be able to solve the ~~[illegible]~~ problem since he has the 10 required ~~[illegible]~~ prisoners.

(b)(ii)