

6) Explain the concept of arithmetic coding and illustrate its process with an example of compressing the string

ABABABBB.

Arithmetic coding is a lossless data compression technique used in data compression systems. Unlike fixed length or variable-length codes, arithmetic coding represents an entire message as a single fractional no. b/w 0 and 1.

Instead of assigning separate bit codes to symbols, arithmetic coding progressively narrows a range [low, high] based on symbol probability until the final interval uniquely represents the message.

The cumulative probabilities of symbols are represented on a line from 0.0 to 1.0

Each symbol occupies a sub-range proportional to its probability.

$$\text{New Range} = S + P(c) \times R$$

$S \Rightarrow$  cumulative probability

$P(c) \Rightarrow$  probability of current symbol

$R \Rightarrow$  current range

input string ABABABBB

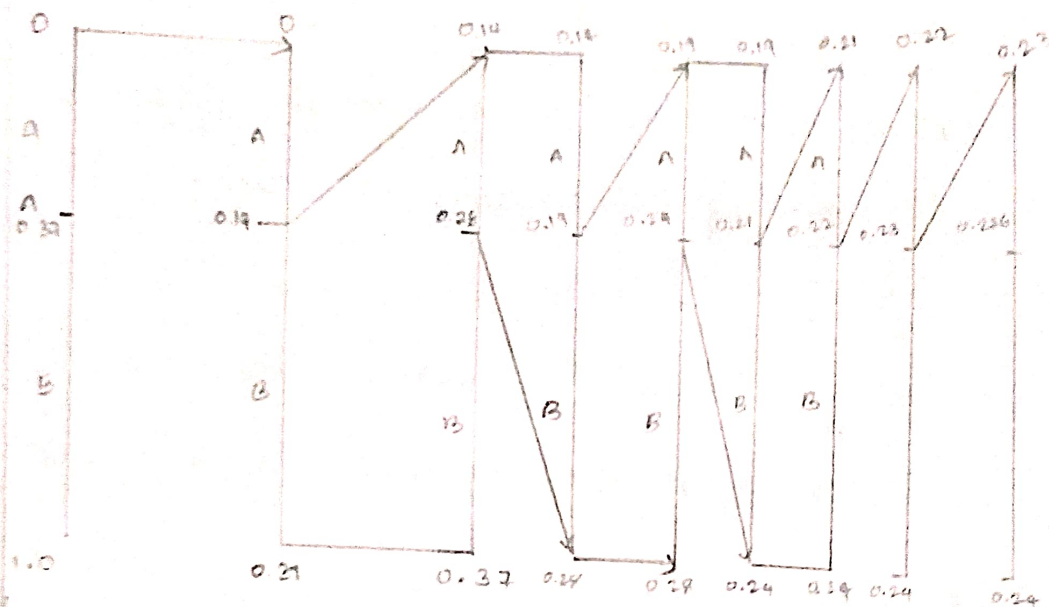
$$\text{No. of A} = 3$$

$$\text{No. of B} = 5$$

$$P(A) = 3/8 = 0.37$$

$$P(B) = 5/8 = 0.62$$

$$F_1(A) = 0.37 \quad F_1(B) = 1.0$$



$$\begin{aligned} \text{i) } S + P(A) * R \\ = 0 + 0.37 * 0.37 \\ = 0.14 \end{aligned}$$

$$\begin{aligned} \text{v) } S + P(A) * R \\ = 0.19 + 0.37 * 0.05 \\ = 0.21 \end{aligned}$$

$$\begin{aligned} \text{ii) } S + P(B) * R \\ = 0.14 + 0.62 * 0.23 \\ = 0.28 \end{aligned}$$

$$\begin{aligned} \text{vi) } S + P(B) * R \\ = 0.21 + 0.62 * 0.03 \\ = 0.22 \end{aligned}$$

$$\begin{aligned} \text{iii) } S + P(A) * R \\ = 0.19 + 0.37 * 0.14 \\ = 0.19 \end{aligned}$$

$$\begin{aligned} \text{vii) } S + P(B) * R \\ = 0.22 + 0.62 * 0.02 \\ = 0.23 \end{aligned}$$

$$\begin{aligned} \text{iv) } S + P(B) * R \\ = 0.19 + 0.62 * 0.09 \\ = 0.24 \end{aligned}$$

$$\begin{aligned} \text{viii) } S + P(B) * R \\ = 0.23 + 0.62 * 0.01 \\ = 0.2362 \end{aligned}$$

$$\Rightarrow 0.23 + 0.24 / 2 = \underline{\underline{0.235}}$$

arithmetic coding generates a unique tag for a sequence without building codes for all sequence of length  $n$  unlike Huffman coding.

The interval is repeatedly subdivided into same proportions as the original range.

The advantages are high compression efficiency than Huffman coding and close to entropy limit.