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 number between 0 and 1

Instead of assigning seperate bit codes to symbols, arithmetic coding progressively narrows a range [low, high) based on symbol probabilities 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 = cumulative probability of previous symbols

P(c) = probability of current symbol

R = current range (high - low)

input string ABABABBB

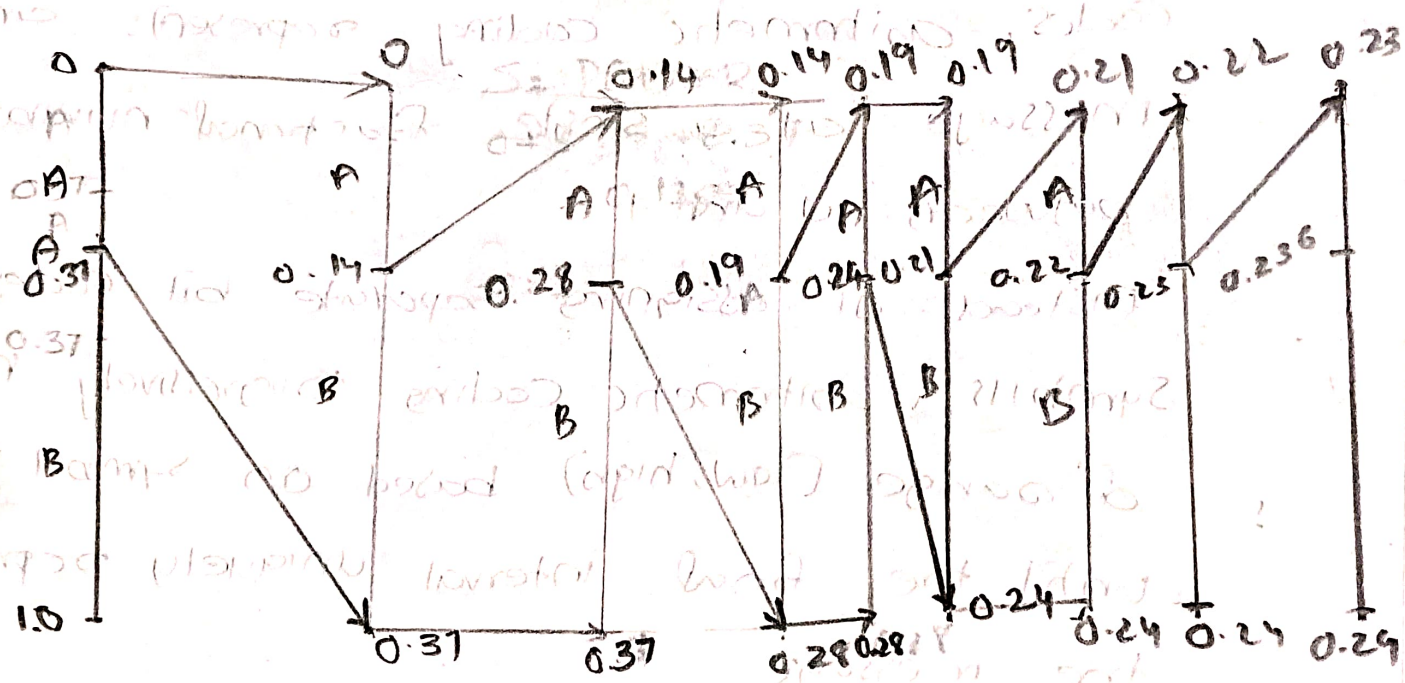
Number of A = 3

Number of B = 5

probability of A = 3/8 = 0.37

probability of B = 5/8 = 0.62

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



i) S+P(A)*R

=0+0.37*0.37

=0.14.

ii) S+P(B)*R

0.14+0.62*0.23

=0.28

iii) S+P(A)*R

0.14+0.37*0.14

=0.19

iv) S+P(B)*R

=0.19+0.62*0.09=0.24.

v) S+P(A)*R

0.19+0.37*0.05

=0.21

vi) S+P(B)*R

0.21+0.62*0.03

=0.22

vii) S+P(B)*R

0.22+0.62*0.02=0.23

viii) S+P(B)*R

0.23+0.62*0.01

=0.2362.

tap = (0.23+0.24)/3 = 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 the same proportions as the original range.

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