

P1) Let $A = \{a, e, i, o, u, !\}$ be our alphabet, with probability distribution:
 $P_A = \{0.2, 0.3, 0.1, 0.2, 0.1, 0.1\}$. Encode "eaii!" using arithmetic coding.

Sol:

$$1 \cdot (0.2) + 0.3 \cdot 0 + 0.3 \cdot 0.2 \cdot (0.2 + 0.3) + 0.3 \cdot 0.2 \cdot 0.1 \cdot (0.2 + 0.3) + 0.3 \cdot 0.2 \cdot 0.1 \cdot 0.1 \cdot (0.2 + 0.3 + 0.1 + 0.2 + 0.1)$$

So, the lower end of the interval corresponding to "eaii!" is $l = 0.23354$.

$$L = 1 + 0.3 \cdot 0.2 \cdot 0.1 \cdot 0.1 \cdot 0.1$$

The upper end of our interval is $L = 0.2336$.

So the interval corresponding to our sequence is $[0.23354, 0.2336)$.

$$1/2 = 0.5 > \text{add } 0$$

$$1/4 = 0.25 > \text{add } 0$$

$$1/8 - 0.125 < \text{add } 1$$

$$1/8 + 1/2^4 = 0.1875 < \text{add } 1$$

$$0.1875 + 1/2^5 = 0.21875 < \text{add } 1$$

$$0.21875 + 1/2^6 = 0.234375 > \text{add } 0$$

$$0.21875 + 1/2^7 = 0.2265625 < \text{add } 1$$

$$0.2265625 + 1/2^8 = 0.23046875 < \text{add } 1$$

$$0.23046875 + 1/2^9 = 0.232421875 < \text{add } 1$$

$$0.232421875 + 1/2^{10} = 0.2333984375 < \text{add } 1$$

$$0.2333984375 + 1/2^{11} = 0.23388671875 > \text{add } 0$$

$$0.2333984375 + 1/2^{12} = 0.233642578125 > \text{add } 0$$

$$0.2333984375 + 1/2^{13} = 0.2335205078125 < \text{add } 1$$

$$0.2335205078125 + 1/2^{14} = 0.23358154296875 \text{ in intr. add } 1$$

So the encoding of our sequence is 00111011110011.

Let $A = \{a, e, i, o, u, !\}$ be our alphabet, with probability distribution:

$P_A = \{0.2, 0.3, 0.1, 0.2, 0.1, 0.1\}$. Encode "eaii!" using arithmetic coding.

Decode 00111011110011 knowing that 5 letters were sent.

$$0 \cdot 1/2 + 0 \cdot 1/2^2 + 1 \cdot 1/2^3 + 1 \cdot 1/2^4 + 1 \cdot 1/2^5 + 0 \cdot 1/2^6 + 1/2^7 + 1/2^8 + 1/2^9 + 1/2^{10} + 1/2^{13} + 1/2^{14} = 0.23358154296875$$

0.2

e

$$0.2 + 0.3 \cdot (0)$$

a

$$0.2 + 0.3 \cdot (0) + 0.3 \cdot 0.2 \cdot (0.2 + 0.3)$$

i

$$0.2 + 0.3 \cdot (0) + 0.3 \cdot 0.2 \cdot (0.2 + 0.3) + 0.3 \cdot 0.2 \cdot 0.1 \cdot (0.2 + 0.3)$$

i

$$0.2+0.3*(0)+ 0.3*0.2*(0.2+0.3)+$$

$$0.3*0.2*0.1*(0.2+0.3)+0.3*0.2*0.1*0.1*(0.2+0.3+0.1+0.2+0.1)$$

!

So the string was "eaii!".

P2) The source of information A generates the symbols $A= \{a, b, c \}$ with the probabilities $P_A=\{0.5, 0.3, 0.2\}$. Encode the source symbols arithmetic encoder if the message is "cacb".

Sol:

$$(0.5+0.3) + 0.2*0 + 0.2*0.5*(0.5+0.3)+0.2*0.5*0.2*(0.5) = 0.89$$

So, the lower end I of the interval corresponding to the string cacb is 0.89.

$$0.2*0.5*0.2*0.3$$

The upper end L of the interval corresponding to the string cacb is 0.896, and so the interval corresponding to the string is $[0.89, 0.896)$.

- $1/2 = 0.5 <$ add 1
- $0.5+1/2^2 = 0.75 <$ add 1
- $0.75 + 1/2^3 = 0.875 <$ add 1
- $0.875+1/2^4 = 0.9375 >$ add 0
- $0.875 + 1/2^5 = 0.90625 >$ add 0
- $0.875 + 1/2^6 = 0.890625$ in the interval - add 1

The encoding of the string cacb is 111001.

Decode the string 111001 which is an encoding of a string with 4 letters.

$$1/2+1/2^2+1/2^3+0+0+1/2^6 = 0.890625.$$

$A= \{a, b, c \}$ with the probabilities $P_A=\{0.5, 0.3, 0.2\}$

$$0.5+0.3 < 0.890625 \text{ and } 0.5+0.3+0.2 = 1 > 0.890625$$

So, the first letter is c.

$$0.8 < 0.890625 \text{ and } 0.8+0.2*(0.5) > 0.890625$$

So, the next letter is a.

$$0.8 + 0.2*0.5*(0.5+0.3) < 0.890625 \text{ and } 0.9 > 0.890625$$

c

$0.88 + 0.2 \cdot 0.5 \cdot 0.2 \cdot (0.5) < 0.890625$ and $0.88 + 0.2 \cdot 0.5 \cdot 0.2 \cdot (0.5 + 0.3) > 0.890625$

b

And so the decoded string is cacb.

P3) Encode the string AABABBBABAA using Lempel-Ziv algorithm.

Sol:

Let our starting dictionary be A-0 and B-1

Input - Output - Dictionary.

A	-	0	-	AA	-	(10)
A	-	00	-	AB	-	(11)
B	-	01	-	BA	-	(100)
AB	-	011	-	ABB	-	(101)
B	-	001	-	BB	-	(110)
BA	-	100	-	BAB	-	(111)
BA	-	100	-	BAA	-	(1000)
A	-	0000	-		-	

The encoding of AABABBBABAA using Lempel-Ziv algorithm is 000010110011001000000.

Decode 00001011001100100 0000 using Lempel-Ziv algorithm with the starting dictionary. A-0,B-1

Input - Output - Dictionary

0	-	A	-	AA	(10)
00	-	A	-	AB	(11)
01	-	B	-	BA	(100)
011	-	AB	-	ABA	(101)
001	-	B	-	BB	(110)
100	-	BA	-	BAB	(111)
100	-	BA	-	BAA	(1000)
0000	-	A	-		

So the decoded string is AABABBBABAA.

P4) Design a 3-bit Tunstall code for a memoryless source with the following alphabet: A = {A, B, C}, with probabilities p(A) = 0.6, p(B) = 0.3, p(C) = 0.1. Using the designed code, encode AAAAABACB.

Sol:

- 1)
- A - 0.6
- B - 0.3
- C - 0.1

2)

B-0.3

C - 0.1

AA - $0.6 \cdot 0.6 = 0.36$

AB - $0.6 \cdot 0.3 = 0.18$

AC - $0.6 \cdot 0.1 = 0.06$

3)

B-0.3 - 000

C- 0.1 - 001

AB-0.18 - 010

AC - 0.06 - 011

AAA - $0.36 \cdot 0.6 = 0.216$ - 100

AAB - $0.36 \cdot 0.3 = 0.108$ - 101

AAC - $0.36 \cdot 0.1 = 0.036$ - 110

The encoding of the string AAAAABACB is 100101011000.

$$n+k(n-1) \leq 2^N$$

$$3+k \cdot 2 \leq 8 \rightarrow k \leq 5/2 = 2.5$$

1. Binary discrete memoryless source

$$\begin{bmatrix} X \\ p(a) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 \\ p(a_1) & p(a_2) \end{bmatrix}$$

where, $a_1=0, a_2=1$

$$a_{i_k} \in A = (a_1, a_2) = (0, 1), k = 1, 2$$

3.1 The AEP (Asymptotic Equipartition Property)

2. 2nd order extension of binary discrete memoryless source

$$\begin{bmatrix} \mathbf{X} \\ p(\mathbf{\alpha}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 \\ p(\alpha_1) & p(\alpha_2) & p(\alpha_3) & p(\alpha_4) \end{bmatrix}$$

where, $\alpha_1=00, \alpha_2=01, \alpha_3=10, \alpha_4=11$

$$p(\alpha_l) = P(\mathbf{X} = \alpha_l) = \prod_{k=1}^{N=2} P(X_k = a_{i_k}) = \prod_{k=1}^{N=2} p_{i_k}$$

$$a_{i_k} \in A = (a_1, a_2) = (0, 1), k = 1, 2$$

3.1 The AEP (Asymptotic Equipartition Property)

3. 3rd order extension of binary discrete memoryless source

$$\begin{bmatrix} \mathbf{X} \\ p(\mathbf{a}) \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 & \alpha_4 & \alpha_5 & \alpha_6 & \alpha_7 & \alpha_8 \\ p(\alpha_1) & p(\alpha_2) & p(\alpha_3) & p(\alpha_4) & p(\alpha_5) & p(\alpha_6) & p(\alpha_7) & p(\alpha_8) \end{bmatrix}$$

where, $\alpha_1=000, \alpha_2=001, \alpha_3=010, \alpha_4=011$
 $\alpha_5=100, \alpha_6=101, \alpha_7=110, \alpha_8=111$

$$p(\alpha_l) = P(\mathbf{X} = \alpha_l) = \prod_{k=1}^{N=3} P(X_k = a_{i_k}) = \prod_{k=1}^{N=3} p_{i_k}$$

$$a_{i_k} \in A = (a_1, a_2) = (0, 1), k = 1, 2, 3$$

4. Extension of arbitrary discrete memoryless source

If X is a discrete memoryless source with distribution as

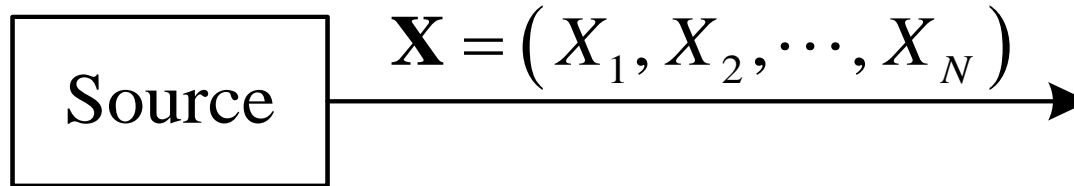
$$\begin{bmatrix} X \\ p(x) \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & \cdots & a_q \\ p_1 & p_2 & \cdots & p_q \end{bmatrix}$$

where, q is number of outcomes from the source, i.e.,
 $p_i = P(X = a_i), i = 1, 2, \dots, q$

X^N denotes the extension of the discrete memoryless source X , which has q^N outcomes

$$\begin{bmatrix} X^N \\ p \end{bmatrix} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_{q^N} \\ p(\alpha_1) & p(\alpha_2) & \cdots & p(\alpha_{q^N}) \end{bmatrix}$$

3.1 The AEP (Asymptotic Equipartition Property)



$$\alpha_l = a_{i_1} a_{i_2} \cdots a_{i_N} \quad a_{i_k} \in A = \{a_1, a_2, \dots, a_q\}$$

$$\left. \begin{aligned} \alpha_1 &= \overbrace{(a_1, a_1, \dots, a_1)}^N \\ \alpha_2 &= (a_2, a_2, \dots, a_2) \\ &\vdots \\ \alpha_q &= (a_q, a_q, \dots, a_q) \end{aligned} \right\} k = 1, 2, \dots, N$$

$$A^N = q^N$$

$$p(\alpha_l) = P(X^N = \alpha_l) = \prod_{k=1}^N p_{i_k}$$

Examples

$X_1X_2X_3$	$P(X_1X_2X_3)$	$H(X)$
000	0.008	0.72
001	0.032	0.72
010	0.032	0.72
011	0.128	0.72
100	0.032	0.72
101	0.128	0.72
110	0.128	0.72
111	0.512	0.72

Examples

ε	$2^{-n(H(X)+\varepsilon)}$	$2^{-n(H(X)-\varepsilon)}$	$A_\varepsilon^{(n)}$
0.7	0.0052	0.9593	011,101 110,111
0.6	0.0643	0.7792	011,101 110,111
0.5	0.0791	0.6329	011,101 110,111
0.4	0.0794	0.5141	011,101 110,111

Examples

ε	$2^{-n(H(X)+\varepsilon)}$	$2^{-n(H(X)-\varepsilon)}$	$A_\varepsilon^{(n)}$
0.3	0.1199	0.4175	011,101 110
0.2	0.1476	0.3392	EMPTY
0.1	0.1817	0.2755	EMPTY

Examples

$X_1X_2X_3X_4$	$P(X_1X_2X_3X_4)$	$H(X)$
0000	0.0016	0.72
0001	0.0064	0.72
0010	0.0064	0.72
0011	0.0256	0.72
0100	0.0064	0.72
0101	0.0256	0.72
0110	0.0256	0.72
0111	0.1024	0.72
1000	0.0064	0.72
1001	0.0256	0.72
1010	0.0256	0.72
1011	0.1024	0.72
1100	0.0256	0.72
1101	0.1024	0.72
1110	0.1024	0.72
1111	0.4096	0.72

Examples

ε	$2^{-n(H(X)+\varepsilon)}$	$2^{-n(H(X)-\varepsilon)}$	$A_{\varepsilon}^{(n)}$
0.7	0.0195	0.9461	0011,0101 0110,0111 1001,1010 1011,1100 1101,1110 1111
0.6	0.0257	0.7170	0011,0101 0110,0111 1001,1010 1011,1100 1101,1110 1111
0.5	0.0340	0.5434	0111,1011 1101,1110 1111

Examples

ε	$2^{-n(H(X)+\varepsilon)}$	$2^{-n(H(X)-\varepsilon)}$	$A_\varepsilon^{(n)}$
0.4	0.0448	0.4118	0111,1011 1101,1110 1111
0.3	0.0591	0.3121	0111,1011 1101,1110
0.2	0.0780	0.2365	0111,1011 1101,1110
0.1	0.1029	0.1792	EMPETY



3.1 The AEP (Asymptotic Equipartition Property)

1. If $(x_1, x_2, \dots, x_n) \in A_\varepsilon(n)$, then:

$$H(X) - \varepsilon \leq -\frac{1}{n} \log p(x_1, x_2, \dots, x_n) \leq H(X) + \varepsilon$$

2. $\Pr\{A_\varepsilon^{(n)}\} > 1 - \varepsilon$ for n sufficiently large.

3. $|A_\varepsilon^{(n)}| \leq 2^{n(H(X) + \varepsilon)}$, where $|A|$ denotes the number of elements in the set A .

4. $|A_\varepsilon^{(n)}| \geq (1 - \varepsilon)2^{n(H(X) - \varepsilon)}$ for n sufficiently large.



Consequences of the AEP: Data compression