Shanon fano Encoding Algorithm

Steps for finding shanon fano coding.

⇒ The given message ~~first~~ written in the decreasing or non increasing probablity.

⇒ Divide the particular message ~~set~~ into two most equiprobable subset $\{x_1\}$ and $\{x_2\}$ so that the probablity ~~set~~ is equal.

⇒ The msg in the 1^{st} set is given bit '0' and the msg in the 2^{nd} set is given bit '1'

⇒ The same procedure is applied for each and every bit seperately and continued until no further seperation is possible.

⇒ At last the codeword for the symbol can easily be find

$$\text{Efficiency } (\eta) = \frac{H}{\hat{H}} \quad \begin{aligned} H &\longrightarrow \text{Entropy} = \sum_{i=1}^{r} p_i \log \frac{1}{p_i} \\ \hat{H} &\longrightarrow \sum_{i=1}^{r} P_i n_i \text{ — code length} \end{aligned}$$

Example find the code word occuring in the proba... $\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$ for symbols S_1, S_2, S_3, and S_4.
find efficiency also.

Answer

$S_1 = \frac{1}{2}$

$S_2 = \frac{1}{4}$

$S_3 = \frac{1}{8}$

$S_4 = \frac{1}{8}$

{ the addition of S_2, S_3 and S_4 is equal to S_1.

→ So the no is changed in the two group.

→ same procedure is followed at per the ~~last~~ digit.

	p	x_1	x_2			Code(w)	length
S_1	1/2	0				0	1
S_2	$\frac{1}{4}$		0			1 0	2
S_3	$\frac{1}{8}$				0	1 1 0	3
S_4	$\frac{1}{8}$		1	1	1	1 1 0 1	3

So the $H = P_i \log \frac{1}{P_i}$

$$H = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{8}\log_2 8 + \frac{1}{8}\log_2 8$$

$$H = \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8}$$

$$\boxed{H = 1.75 \text{ b/s}}$$

$\rightarrow \hat{H} = p_i h_i$ $\qquad h_i =$ Code length

$$\hat{H} = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3$$

$$\hat{H} = 1.75$$

Efficiency $(\eta) = 100\%$ $\qquad \frac{H}{\hat{H}} = \frac{1.75}{1.75}$

$$\boxed{\eta = 100\%}$$

Example

find the code word for the probablity $\frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}, \frac{1}{16}$ for the symbols $S_1 \dots S_8$

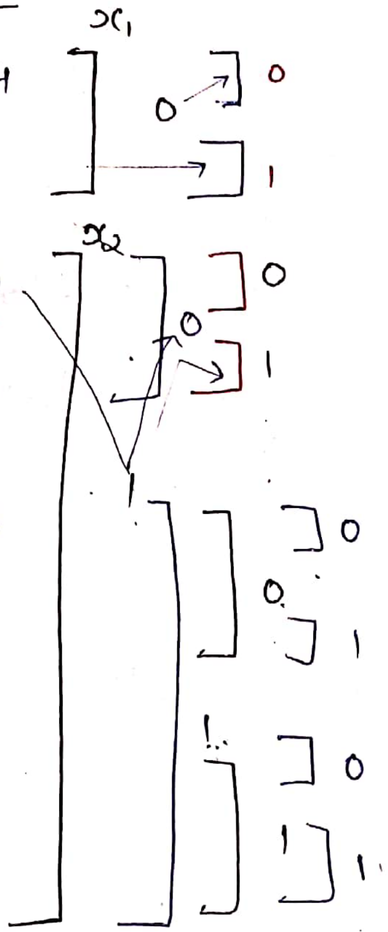
find the code efficiency and Redundancy also?

Solution

Redundancy $R_e = (1 - e)$

$\hookrightarrow$ efficiency

Symbols	P_i
S_1	$1/4$
S_2	$\frac{1}{4}$
S_3	$1/8$
S_4	$1/8$
S_5	$1/16$
S_6	$1/16$
S_7	$1/16$
S_8	$1/16$



Sym.	Codeword	Length
S_1	00	2
S_2	01	2
S_3	100	3
S_4	101	3
S_5	1100	4
S_6	1101	4
S_7	1110	4
S_8	1111	4

$$\eta = \frac{H}{\hat{H}}$$

$$H = 2\times\left(\frac{1}{4}\log_2 4\right) + 2\times\left(\frac{1}{8}\log_2 8\right) + 4\left(\frac{1}{16}\log_2 16\right)$$

$$\boxed{H = 2.75 \text{ b/s}}$$

$$\hat{H} = \frac{1}{4}\times 2 + \frac{1}{4}\times 2 + \frac{1}{8}\times 3 \;\; \frac{1}{8}\times 3 + \frac{1}{16}\times 4 + \frac{1}{16}\times 4 \times \frac{1}{16}\times 4$$

$$\boxed{\hat{H} = 2.75 \text{ b/s}}$$

$$\boxed{\hat{H} = \sum_{i=1}^{n} P_i n_i}$$

$$\text{Efficiency} = \frac{H}{\hat{H}} = \frac{2.75}{2.75} = 100\%$$

$$\text{Redundancy} = 1 - \eta = 1 - 100\% = 0\%$$

Example 2 Problem on ambiguity

Apply the shanon fano encoding

$X = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7\}$

$P = \{0.4, 0.2, 0.12, 0.08, 0.08, 0.08, 0.04\}$

Solution

Symbols	P	Codeword	Length
x_1	0.4	0	1
x_2	0.2	100	3
x_3	0.12	101	3
x_4	0.08	1100	4
x_5	0.08	1101	4
x_6	0.08	1110	4
x_7	0.04	1111	4

in the above example, the solution can be found by using two method.

→ Combine $x_2 \ldots x_7$ with x_1 $(0.4, 0.6)$

→ Combine $x_3 \ldots x_7$ with (x_1, x_2) $(0.6, 0.4)$

$\bar{H} = \Sigma P_i L_i$

$= 0.4 + 0.6 + 0.36 + 3 \times .08 \times 4 + 0.16$

$= \underline{\underline{2.48 \text{ b/s}}}$

Same questions solved by second method

		Codeword	Length
x_1	0.4	0 0	2
x_2	0.2	0 1	2
x_3	0.12	1 0 0	3
x_4	0.08	1 0 1	3
x_5	0.08	1 1 0	3
x_6	0.08	1 1 1 0	4
x_7	0.04	1 1 1 1	4

$\bar{H} = 0.4 \times 2 + 0.2 \times 2 + 0.12 \times 3 + 2 \times 0.08 \times 3$

$+ 4 \times 0.08 + 4 \times .04$

$= \underline{\underline{2.52}} \text{ b/s}$

So the first approach is better than the IInd.

$\eta = \frac{H}{\bar{H}} =$

$H = .4 \log_2 \frac{1}{0.4} + 0.2 \log_2 \frac{1}{0.2} + 0.12 \log_2 \frac{1}{0.12}$

$+ 3 \times .08 \times \log_2 \frac{1}{.08} + 1 \times \log_2 \frac{1}{.04}$

$H = 2.42$

$$\eta = \frac{H}{\bar{H}} = \frac{2.42}{2.48}$$

$$\boxed{\eta = 97.6\%}$$

Lempel Ziv Coding

In Real Time application the probablities was ~~not~~ providing in advance. So ~~it~~ is ~~not~~ recognizable in ~~the~~ real time application. The new scheme was proposed by Lempel and Ziv in 1977. It does ~~not~~ required ~~the~~ source code ~~statistics~~. It is variable to fixed length source coding algorithm and belongs ~~to~~ ~~the~~ ~~the~~ class ~~of~~ universal source coding algorithm.

The working ~~of~~ ~~the~~ coding is as follows:

The compression ~~of~~ any number is decoded by series ~~of~~ 0's and 1's and plus one new bit. These variable length blocks called as phrases. These are exist ~~the~~ pharase and ~~their~~ location.

Question:

~~The string of~~ 10101101110101010111

~~the~~ execution will begin by parsing it into comma seperated each phrases ~~that~~ can be represented by a previous ~~string~~ as the prefix, plus a bit.

The first bit 1 has no predessor so it has null prefix ~~string~~. {: ~~it~~ is assumed ~~that~~ 0 and 1 already ~~stored~~ in 1 and 0 position}

1, 0101101110101010111

same goes for 0

1, 0, 101101110101010111

→ our dictionary contain '1' and '0' so we more further for new word

1, 0, 10, 1101110101010111

Continuing in ~~the~~ way we more forword for completing ~~the~~ bit ~~stream~~.

1, 0, 10, 11, 01, 101, 010, 1011

for visualizing ~~the~~ process the ~~term~~ can be described as

(codeword ~~at~~ location 0, 1), (codeword ~~at~~ location 0, 0),

(codeword ~~at~~ location 1, 0), ...

Scanned by CamScanner

POORNIMA
COLLEGE OF ENGINEERING

DETAILED LECTURE NOTES

Dictionary for the lempel Zive Algorithm

Dictionary Location	Dictionary Content	fixed length Codeword
001	1	0001
010	0	0000
011	10	0010
100	11	0011
101	01	0101
110	101	0111
111	010	1010
	1011	1101

Zive Algorothim proved that for long document the compression of the file approaches the optimum obtainable condition.

Example 2

0 0 0 1 0 1 1 1 0 0 1 0 1 0 0 1 0 1

1	2	3	4	5	6	7	8	9
0	1	00	01	011	10	010	100	101

↳ generate the subsequenle.

null position denote by 000, 001

Number Representation

	1	2	3	4	5	6	7	8	9
	0	1	00	01	011	10	010	100	101
Number Representati			11	12	42	21	41	61	62

Binary value ↓ — location

Binary Encoded block	0010	0011	1001	0101	1000	1100	1101

Question 2

A A B A B B B A B A A B A B B B A B B A B BA

Solution $A \rightarrow 0$ $B \rightarrow 1$

1	2	3	4	5	6	7	8	9
A	AB	ABB	B	ABA	ABAB	BB	ABBA	BBA
∅A	1B	2B	∅B	2A	5B	4B	3A	7A

Binary Code

0 0011 0111 1 0100 1011 1001 0110 1110

Information Capacity Theoram (channel capacity theorem)

Information Capacity theoram is used for band limited and power limited Gaussian Channel.

Let X_k, $k = 1, 2 \ldots k$ denotes the continuos random variable obtained by uniform sampling of the process $X(t)$ at the nyquist rate of 2W samples per second.

The channel output is cumulated by AWGN of zero mean and power spectral density (PSD) $N_0/2$. of the received signal. denote the sample Y_k, $k = 1, 2 \ldots k$

$$\boxed{Y_k = X_k + N_k} \qquad k = 1, 2 \ldots k$$

when N_k is the noise sample with zero mean and variance $\sigma^2 = N_0 W$. It is statistically independent.

$$E(X_k^2) = P, \quad k = 1, 2 \ldots k$$

Power Capacity of the channel

$$C = \max_{f_{X_k}(x)} \{ I(X;Y) \mid E[X_k^2] = P \}$$

$f_{X_k}(x)$ is the probablity density function of X_k.

$$I(X_k; Y_k) = h(Y_k) - h(Y_k \mid X_k)$$

Example 6

$h(Y_k|X_k) = h(N_k)$

$\Rightarrow I(X_k ; Y_k) = h(Y_k) - h(N_k)$

$h(Y_k) = \frac{1}{2}\log_2[2\pi e(P+N_0 W)]$

$h(N_k) = \frac{1}{2}\log_2[2\pi e(N_0 W)]$

subsituting the value in channel capacity

$$C = \frac{1}{2}\log_2\left(1+\frac{P}{N_0 W}\right) \text{ b/s}$$

If 2W sample are transmitted per second.

$$\boxed{C = W\log_2\left(1+\frac{P}{N_0 W}\right) \text{ b/s}}$$

Information Capacity Theoram:

The inf capacity of a continuos channel of bandwidth W herts, peturbed by additive white Gausian noise of power spectral density $N_0/2$ and limited in bandwidth to W, is given by

$$\boxed{C = W\log_2\left(1+\frac{P}{N_0 W}\right)}$$

It is also called channel capacity theoram.

By the use of this theoram it provide the trade off between the channel bandwidth the average transmitted power and noise power spectral density. It provide the channel bandwidth in the term of bits per second.

Channel Capacity (Shannon limit)

B.W → maximum range of frequency occupied by the channel

SNR → signal to noise Ratio

it fix the channel capacity with B.W. and SNR

⇒ maximum information Rate transmitting from sender to receiver in the channel.

Denoted by C

Information Rate can be the value of transmission from sender to Receiver ~~from~~ in the channel

Information Rate → R

(i) R < C → Transmission without error in presence of noise.

(ii) R > C → Transmission without error is ~~not~~ possible in presence of noise

To get noise free transmission error detection and correction code added to the original signal.

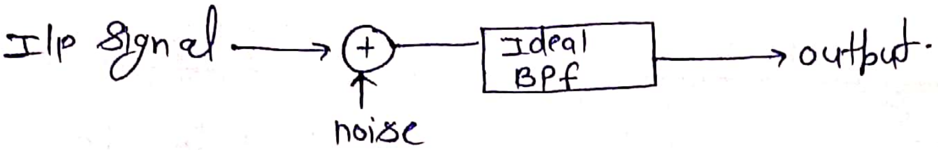
∴ signal length ↑ = circuit Complexity ↑

begining of any other codeword.

Channel Capacity

$$\boxed{C = B \log_2 (1 + S/N)}$$

B.W. — signal to Noise Ratio



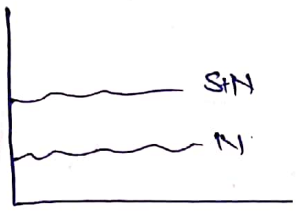
V_s = rms value ~~of~~ signal voltage
V_N = rms value ~~of~~ Noise voltage

3) Signal power $= V_s^2 / R = V_s^2$
) Noise Power $= V_N^2 / R = V_N^2$

If $R = 1$

$S = V_s^2 \Rightarrow V_s = \sqrt{S}$

$N = V_n^2 \Rightarrow V_n = \sqrt{N}$



maximum no ~~of~~ level $= \dfrac{\sqrt{S+N}}{\sqrt{N}} = \sqrt{1 + S/N}$

Information $= \log_2 (m)$

$= \log_2 \sqrt{1 + \frac{S}{N}}$

$$\boxed{B = \frac{1}{2} \log_2 (1 + S/N)}$$

if $S < N$ the condition is ~~not~~ permissible

for k bits /sec

$$\boxed{C = \frac{k}{2} \log_2 \left(1 + \frac{S}{N}\right)}$$

If channel Bandwidth = B
max. info Rate = 2B

$$\boxed{C = B \log_2 \left(1 + \frac{S}{N}\right)}$$

$\Rightarrow \frac{S}{N} \uparrow = C \uparrow \Rightarrow S/N = \infty \Rightarrow C = \infty$

$\Rightarrow$ B.W $\uparrow$ = C $\uparrow$

noise power spectral density = $\eta/2$

noise power (N) = ηB

$$C = B \log_2 \left(1 + \frac{S}{\eta B}\right)$$

$$C = \frac{\eta}{S}\left(\frac{BS}{\eta}\right) \log_2 \left(1 + \frac{S}{\eta B}\right)$$

$$C = \frac{S}{\eta} \log_2 \left(1 + \frac{S}{\eta B}\right)^{\eta B/S}$$

$$\lim_{x \to 0} (1+x)^{1/x} = 0$$

$$C = \frac{S}{\eta} \log_2 e$$

$$\boxed{C = 1.44 \frac{S}{\eta}}$$

As per ...
begining of any other codeword.

Example 1 For a typical telephone line with a signal to noise Ratio of 30dB and an audio Bandwidth 3kHz max data Rate of channl

Ans

$SNR = 30\,dB = 10^3$

$B = 3\,kHz$

$$C = B\log_2\left(1+\frac{S}{N}\right)$$

$$= 3\times10^3 \log_2(1+10^3)$$

$$= 3\times10^3 \frac{\log 1001}{\log 2}$$

$$= 3\times10^6\ bps$$

$$\boxed{C = 3\,mbps}$$

Example A Gaussian Channel has 1MHz bandwidth. Calculate the channel capacity. If the signal power to noise spectral density Ratio S/N is 10^5 Hz. Also find the maximum information Rate.

Ans $B = 1\,MHz = 10^6\,Hz$

$SNR = 10^5$

$C = ?$

$$C = B\log_2(1+S/N)$$

$$= 10^6 \frac{\log(100001)}{\log 2}$$

$$= 16.6\times10^6\ bps$$

$$\boxed{C = 16.6\,mbps}$$