Entropy rate

- Average per symbol entropy in an information source.

- Random walk on graph

Coin Tossing versus poker

Toss a fair coin and see the sequene

Head, Tail, Tail, Head . . .

$$(x_1, x_2, \ldots x_n) \approx 2^{-nH(X)}$$

- Play card games with friend and see a sequence

output depend on previous card

not independent.

$$(x_1, x_2, x_3 \ldots x_n) = ?$$

How to model dependence : Markov chain

- A stochastic process $X_1, X_2, \ldots$
  - State $\{x_1, \ldots x_n\}$, each state $X_i \in \mathcal{X}$
  - Next step only depend on the previous state

$$p(x_{n+1} | x_n, \ldots x_1) = p(x_{n+1} | x_n)$$

principle of Markov chain

- Transition Probability
    - Probability of moving from one state to another state

$$P_{i,j} \text{ : the transition probability of } i \to j$$

$$P(x_{n+1}) = \sum_{x_n} p(x_n)\, p(x_{n+1} \mid x_n)$$

$$P(x_1, x_2 \cdots x_n) = P(x_1)\, P(x_2 \mid x_1) \cdots P(x_n \mid x_{n-1}).$$

Hidden Markov Model (HMM)

- used extensively in speech recognition, hand writing recognition, machine learning.

not observable

unobservable

- Markov process $x_1, x_2 \cdots x_n$

- Observe a random process $Y_1, Y_2, Y_3 \cdots Y_n$ such that

emission probability $Y_i \sim P(y_i \mid x_i)$

$\boxed{\begin{cases} Y - \text{observable.} \\ X - \text{hidden state.} \end{cases}}$

- We can build a probability model

$$P(x^n, y^n) = P(x_1) \prod_{i=1}^{n-1} P(x_{i+1} \mid x_i) \prod_{i=1}^{n} P(y_i \mid x_i)$$

Time invariance Markov Chain

- A Markov chain is time invariant if the conditional probability $p(x_n | x_{n-1})$ does not depend on n.

$$p(X_{n+1} = b \mid X_n = a) = p(X_2 = b \mid X_1 = a) \quad \text{for all } a, b \in \mathcal{X}.$$

- For this kind of Markov chain, define transition matrix

$$P = \begin{bmatrix} P_{11} & \cdots & P_{1n} \\ \vdots & & \\ P_{n1} & \cdots & P_{nn} \end{bmatrix}.$$

time independent
↓
Transition take equal probability.

Simple weather Model

Satitionary distribution

μ

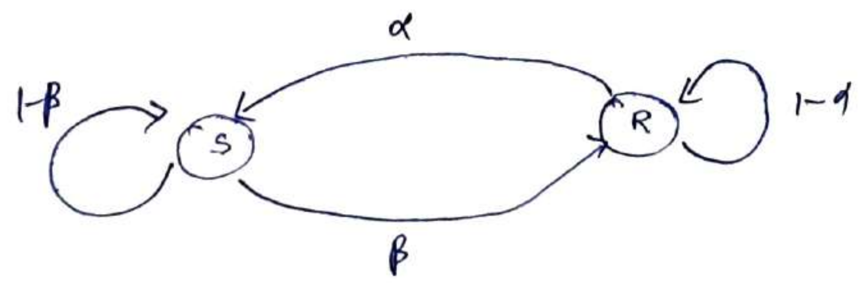
$n+1$ n

same distribution at both time

$\mathcal{X} = \{ \text{Sunny}; S, \text{Rainy } R \}.$

$p(S|S) = 1-\beta$, $p(R|R) = 1-\alpha$

$p(R|S) = \beta$ $p(S|R) = \alpha$



$$P = \begin{bmatrix} 1-\beta & \beta \\ \alpha & 1-\alpha \end{bmatrix}.$$ transition probability matrix.

Stationary distribution

$$\boxed{\mu P = \mu} \quad \text{— (i)}$$

$$\boxed{\mu(S) + \mu(R) = 1.} \quad \text{— (ii)}$$

$$\mu(S) = \frac{\alpha}{\alpha+\beta} \qquad \mu(R) = \frac{\beta}{\alpha+\beta}$$

$$\mu(S)\,\beta = \mu(R)\,\cancel{\alpha}$$

$$\mu(S) = \mu(R)\frac{\alpha}{\beta}$$

from eq (ii)

$$\mu(R)\frac{\alpha}{\beta} + \mu(R) = 1$$

$$\mu(R) = \beta/\alpha+\beta.$$

Probability of seeing a sequence SSRR:

$$p(SSRR) = p(S)\; p(S|S)\; p(R|S)\; p(R|R)$$

$$= p(S)\,(1-\beta)\;\beta\;(1-\alpha)$$

↳ what will this sequence behave after long day?

Stationary distribution

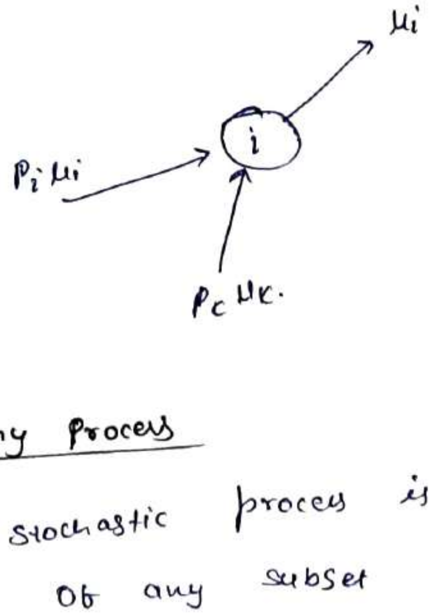
- a distribution μ on the states such that the distribution at time $n+1$ is same as at time n.

How to calculate stationary distribution

- μ_i $i = 1, 2 - |\chi|$ satisfy

$$\mu_i = \sum_j \mu_j p_j, \left(\mu P = \mu P\right) \quad \text{and} \quad \sum_{i=1}^{|\chi|} \mu_i = 1.$$

"Detailed balancing":



Stationary Process

A stochastic process is stationary if the joint distribution of any subset is invariant to time shift

$$p(X_1 = x_1, \ldots X_n = x_n) = p(X_2 = x_1 \ldots, X_{n+1} = x_n).$$

e.g. coin tossing

$$p(x_1 = \text{head}, x_2 = \text{tail}) = p(x_2 = \text{head}, x_3 = \text{tail}) = p(1-p).$$

Entropy rate

- When x_i are iid, entropy

$$H(x^n) = H(x_1, \cdots x_n) = \sum_{i=1}^{n} H(x_i)$$
$$= n H(x).$$

- with dependent sequence x_i, how does $\{H(x^n)\}$ grow with n? $\boxed{\text{Still linear ?}}$

- Entropy rate characterizes the growth rate.

- **Definition 1 :-**
  Average entropy per symbol

$$H(X) = \lim_{n \to \infty} \frac{H(x^n)}{n}$$

- **Definition 2 !**
  rate of information innovation

$$H'(X) = \lim_{n \to \infty} H(x_n | x_{n-1}, \cdots, x_1).$$

$H'(x)$ exists, for x_i sationary

$$H(x_n | x_1 \cdots x_{n-1}) \leq H(x_n | x_2 \cdots x_{n-1}) \quad \text{— (1)}$$
$$\leq H(x_{n-1} | x_1 \cdots x_{n-2}). \quad \text{— (2)}$$

- $H(x_n | x_1 \cdots x_{n-1})$ decreases as n increases.
- $H(x) \geq 0$
- The limit must exist.

$\mathcal{X}$ - set of sequence.

AEP for Stationary process

$$-\frac{1}{n} \log p(x_1 \ldots x_n) \rightarrow H(\mathcal{X})$$

- $p(x_1, \ldots x_n) \approx 2^{-nH(\mathcal{X})}$
- Typical sequence in typical set of size $2^{-nH(\mathcal{X})}$
- We can use $nH(\mathcal{X})$ bits to represent typical sequence.

Entropy rate for Markov chain

- For Markov chain

$$H(\mathcal{X}) = \lim H(X_n | X_{n-1}, \ldots X_1)$$

$$= \lim H(X_n | X_{n-1})$$

$$= H(X_2 | X_1)$$

By ~~Markov chain~~ definition

$$p(X_2 = j | X_1 = j) = P_{ij}$$

① Find sationary distribution μ_i

② Use transition probability P_{ij}

$$H(\mathcal{X}) = -\sum_{ij} \mu_i p_{ij} \log p_{ij}$$

Entropy rate of weather model

$\mu(S) = \frac{\alpha}{\alpha+\beta}$ $\quad \mu(R) = \frac{\beta}{\alpha+\beta}$.

$$H(\alpha) = \frac{\beta}{\alpha+\beta} \left\{ \alpha \log \alpha + (1-\alpha) \log(1-\alpha) \right)$$

$$= \frac{\alpha}{\alpha+\beta} H(\beta) + \frac{\beta}{\alpha+\beta} H(\alpha)$$

05|03|2024

Source Coding

P_i	code1	code-2
	000	0
$1/2$		10
	001	
$1/4$		110
	1	
$1/8$		1110
		111100
$1/16$		111101
$1/64$		
$1/64$		111110
$1/64$		111111
$1/64$	111	
		2
$E l_i$	3	

Min^m code word length
⇓
and transmit max^m information.

Morse's code (1836)

Codes

Block codes

00	1
01	01
10	110
11	111
fixed length	variable length.

Non-singular codes

- If all of them are distincts
  ⇓
  Then the block codes are called ~~distinct~~ non-singular codes

Uniquely decodable codes

If its nth extension is also a non-singular then it is known as uniquely decodable code

~~code~~.

$C \rightarrow \quad 0 \rightarrow 00$

$\quad\quad\quad 1 \rightarrow 11$

2nd extension C^2 0000   0001

0100   0101

Prefix codes

instantaneous

Source	Coherent
A	0
B	01
C	11
D	10

Not this arise scenario

'AB CAD'

0 01 11 0 10 → Rx.

A A C D D → Not instaneous as o/p is not uniquely ~~defined~~ decoded in receiver.

~~word length~~

word length.

$$\sum_{k=1}^{N} r^{-l_k} \leq 1.$$

r - base

Total no. of codes

Kraft - Mcmillan inequality.

code A	code B	code C	code D
0	1	00	10
10	01	110	111
110	111	1110	110
1110	10	001	01
111	00	011	00
X	X	X	✓

Prefix code.

Requirement of code construction

e.g.

code A	code B	code C	word length
1	2	3	1
1	1	1	2
2	2	1	3
2	1	2	4

Use Kraft - McMillan inequality to verify.

$$\sum_{k=1}^{N} 2^{-l_k} \leq 1.$$

code A

X

$= (1 \times 2^{-1}) + (1 \times 2^{-2}) + (2 \times 2^{-3}) + (2 \times 2^{-4})$

$= 9/8 > 1$

So, code A can't be ~~code~~ to construct a prefix code

code B

$$\sum_{k=1}^{N} 2^{-l_k} \leq 1$$

$$= (2 \times 2^{-1}) + (1 \times 2^{-2}) + (2 \times 2^{-3}) + (1 \times 2^{-4})$$

X $= 1 + (\quad) \geq 1$

code-C

$$\sum_{k=1}^{N} 2^{-l_k} \leq 1$$

$$= (1 \times 2^{-1}) + (1 \times 2^{-2}) + (1 \times 2^{-3}) + (2 \times 2^{-4})$$

$$+ \frac{1}{2^2} + \frac{1}{2^2}$$

$$\frac{1}{2} + \frac{1}{2}$$

Prefix code $= 1$

~~0~~	0
01	10
001	110
0000	1110

Therefore, If the give code satisfy Kraft-Mcmillan Theorem them it can be use to construct prefix code.

Q.

Symbol	Prob.	Codeword	Length
x	0.5	0	1
Y	0.3	10	2
z	0.2	110	3

(b) Consider a 2nd order extension of the source Recompute the codewords and the efficiency comment on both code

H.W

N

Symbols	Prob.	
xx	0.25	0.5 × 0.5
xy	0.15	
xz	0.10	
yx	0.15	
yy	0.09	
yz	0.06	
zx	0.10	
zy	0.06	
zz	0.04	

Shannon - Fano Algorithm

① Arrange the probabilities in decreasing order

② Group the prob. in exactly two sets

↳ such that sum of the two sets is approx. equal.

③ Assign bit 0 to all elements of group 1 and 1 to all elements of group 2.

④ Repeat the above steps until no further division is possible.

E.g.

$S = (A, B, C, D, E, F)$

$P = (0.10, 0.15, 0.25, 0.35, 0.08, 0.07)$

Use Shannon - Fano encoding:-

$P = (0.35, 0.25, 0.15, 0.10, 0.08, 0.07)$

(D, C, B, A, E, F)

0.10	0.35
0.15	0.08
0.25	0.07
0.50	0

Set 2

$(0.35, 0.08, 0.07)$ set 1 (D, E, F)

$(0.25, 0.15, 0.10)$ (C, B, A)

0.25 $(0.15, 0.10)$

0.35 0
0.25 0
0.15 1
0.10 1
0.08 1
0.07 1

0.35 0
0.25 1

0.15 - 0
0.10 - 1
0.08 - 1
0.07 - 1

0.10 0
0.08 1
0.07 1

0.08 - 0
0.07 - 1

Symbol	Codeword	Prob.	Length
D	0 0	0.35	2
C	0 1	0.25	2
B	1 0	0.15	2
A	1 1 0	0.10	3
E	1 1 1 0	0.08	4
F	1 1 1 1	0.07	4

Efficicient $= \frac{H(S)}{L} \times 100\%$

$H(S) = -0.35 \log_2 0.35 + \ldots$

$= 2.33.$

$L = \Sigma p_i l_i$

$L = \Sigma p_i l_i$

$= 0.35 \times 2 + 0.25 \times 2 + 0.15 \times 2 + 0.10 \times 3 + 0.08 \times 4 + 0.07 \times 4$

$= 2.4$

Efficiency $= \frac{2.33}{2.4} \times 100$

$= 97.08\%$.

Huffman's Code $\longrightarrow$ consider as optimal code.

(compact code).

(r-ary codeword)

Step 1 : compute the number of stages required for the encoding operation.

$$n = \frac{N-r}{r-1}$$; N : total no. of symbols in the source alphabet.

; n has to be ~~zero~~ integer

⇓

If it is not we have to append dummy symbol.

Step 2 :- If n is not integer then append minimum no. of dummy symbol with probability zero.

For binary, n is always integer.

Step 3 : Arrange the prob. in descending order.

Step 4. Combine the last r. probabilities in the set by summing as a single prob. and place the sum in the appropriate position in the set by recording it.

For r = 2 } combine last two probabilities.

Step 5 :- Continue step 4 till we reach the position where we have only r elements

19/03/2?

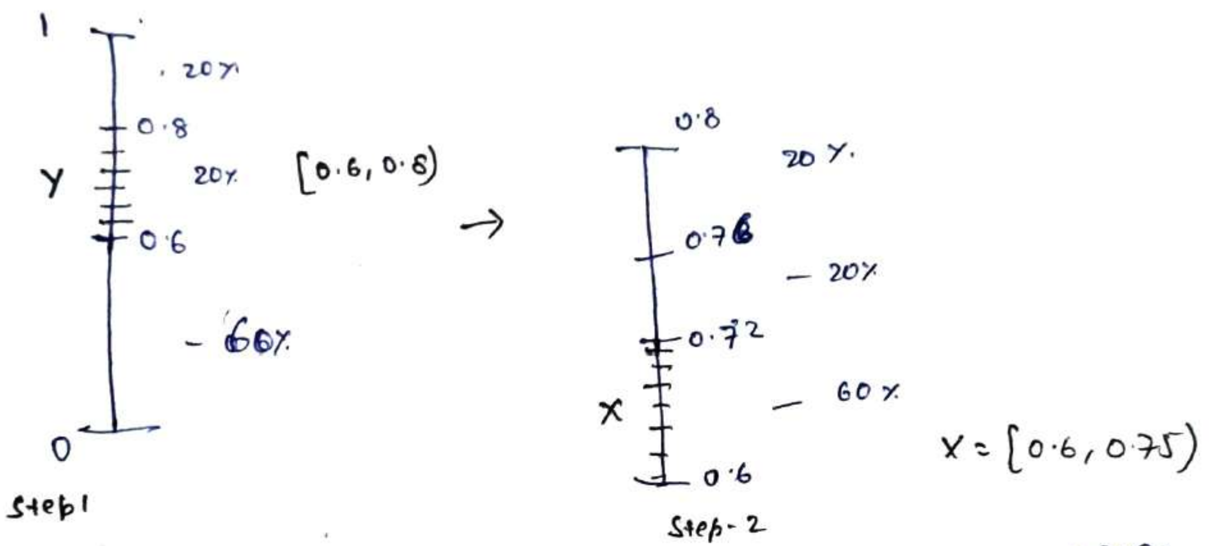
Arithmetic coding

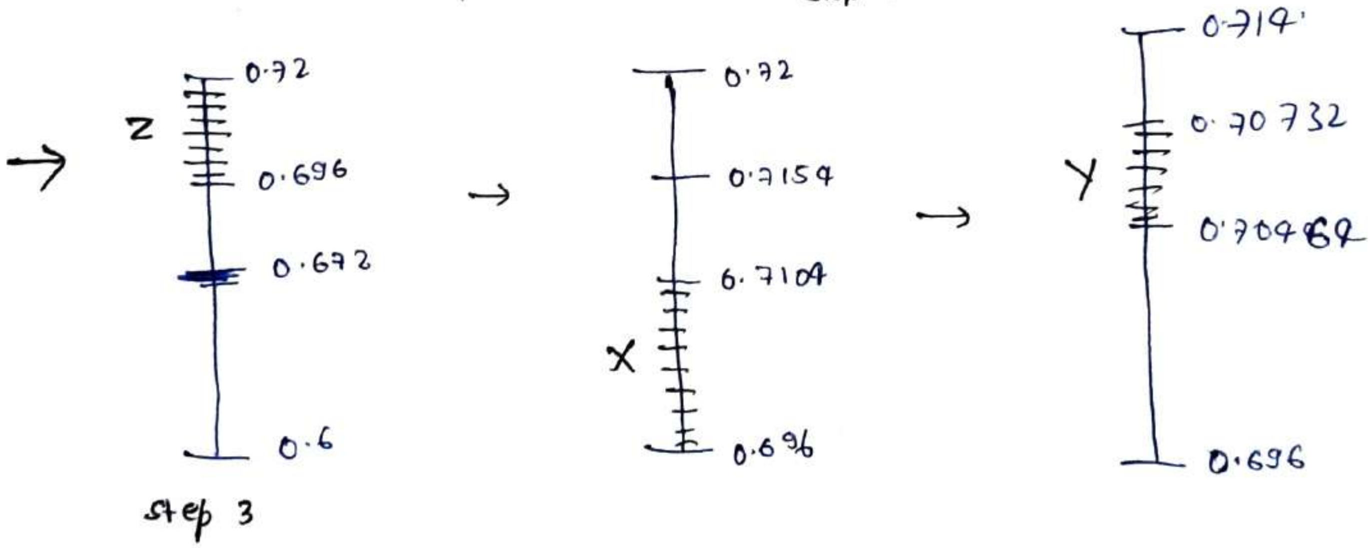
Q. Consider a discrete memoryless source with $S = \{x, y, z\}$ with respective probabilities $P = \{0.6, 0.2, 0.2\}$. Find the codeword for the message 'yxzxy' using arithmetic coding

[0 1)

'yxzxy' $P = \{0.6, 0.2, 0.2\}$.

Step.1 :- Divide the range (





yxzxy → 0.70464 (receive lower bit tag.

Lempol - ziv Algorithm

< index , code >

' THIS _ IS _ HIS _ HIT '

Dictionary

index	Symbol
1	T
2	H
3	I
4	S
5	_
6	IS
7	_H
8	IS_
9	HI

Encoding Scheme

Symbol	Encoding
T	(0, code(T))
H	(0, code(H))
I	(0, code(I))
S	(0, code(S))
_	(0, code(_))
IS	(3, code(S))
_H	(5, code(H))
IS_	(6, code(_))
HI	(2, code(I))
T	(0, code(T))

Run - length Encoding

00000 11111 000000 111 0000 1111 000 111

50 51 60 31 40 41 30 41

JPEG

- Joint photographic Experts Groups.

- Standard specification for lossy compression for digital images.

- lossy compression means that JPEG images lose data when saved. This data is lost forever, and the original can never be re-formed.

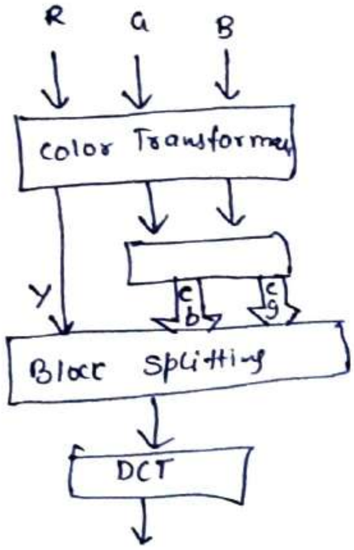
JPEG is not a file format

- JFIF (JPEG file interchange format file) is a wrapper that holds the compressed data created by the JPEG compression.

- metadata

JPEG vs JPG

- no difference b/w JPEG and JPG extensions.
- Before window 95 "only 3. characters" allowed.



PEG encoding : Color Transformation

- Colorspace
-

RGB - format

R: 229
G: 102
B: 102

CMYK format

C: 229
M: 59
Y: 59
K: 12

- RGB (additive) — monitor.
- CMYK (substractive)
  - print industry

YCbCr Color model

- human eye is more sensitive to fine variations in brightness (luminance) than to changes in color (chroma)

- JPEG can take advantage of this by converting to the YCbCr color model which splits the liuminance.
  - Y : Luminance
  - Cb : Chroma Blue (RGB blue - luminance)
  - Cr : Chroma red (RGB red - luminance).

Downsampling

Chroma downsampling is where the color information in an image's Cb and Cr channels is sampled at a lowe resolution than original.

(J : a : b)

J – horizontal sampling reference.

a –

b –

JPEG Encoding

Step1 : Recentre around zero.

↳ subtract (128)

Step2 : Calculate the Discrete Cosine Transform coefficients

-415.38 — DC component.

Quantization

- Quantization process aims to reduce the overall size of the DCT coefficients so that they can be more efficiently compressed in the final Entropy encoding scheme.

↳ Quantization matrix

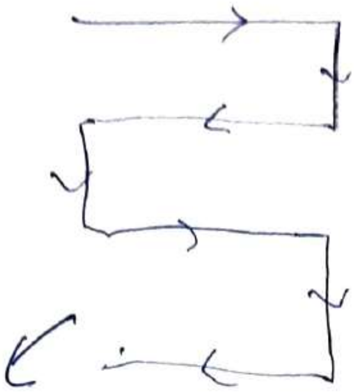
- determines the compression ratio

- To calculate the quantized DCT coefficients we divide the

$$\text{round}\left(\frac{-415.37}{16}\right) = \text{round}(-25.96) = -26.$$

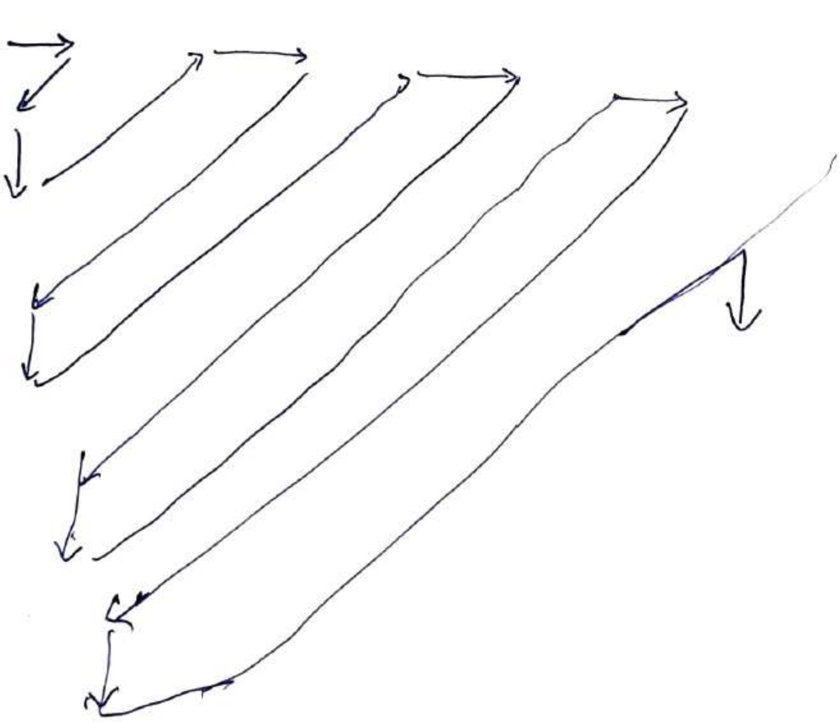
JPEG coding

- Arrange the quantized matrix in zig-zag



Apply run-length encoding.

Or



~~-26 -3 0 -3 -8~~

-26 -3 1 0 -3 -2 -6 2 -4 1 1 -3 5 1 1

2 -1 1 1 -1 2 5 0 -1 -1 38 0

Apply Run-length encoding and Huffman coding on AC coefficients

(runlength, size) (amplitude)

(0, 2) (-3) ; (1, 2) (-3) ; (0, 1) (-2) ; (0, 2) (-6) ;

(0, 1) (2) ; (0, 1) (-4) ; ( ) ( ) ;

If number of zeros exceed 15 then we can denote (15, 0) (0) or (0, 0) (0).

-26 -3 0 -3 -2 -6 2 -4 1 -3 1 1 5 1 2

-1 1 -1 2 0 0 0 0 0 -1 -1 $\underbrace{0\ 0 \ - \ - \ - \ 0\ 0}_{38}$

21/03/24

Maximum Entropy

$$\begin{cases} p(B) + p(C) + p(F) + p(T) = 1 \\ \$1\ p(B) + \$2\ p(C) + \$3\ p(F) + \$8\ p(T) = \$2.5 \end{cases}$$

Cannot be determined the frequency of each item

Maximum entropy principles

- If nothing is known about a distribution except that it belongs to a certain class.

- Distribution with the largest entropy should be chosen as the default.

Formulation

Maximize entropy

$$H(p) = -\sum_{i=1}^{n} p_i \log p_i$$

$$p_i \geq 0 \qquad \text{——} \qquad (1)$$

$$\sum_{i=1}^{n} p_i = 1 \qquad (2)$$

$$\sum_{i=1}^{n} p_i r_{ij} = \alpha_j \quad \text{for} \quad 1 \leq j \leq m \qquad (3).$$

Form Lagrangian

- $$J(p) = -\sum_{i=1}^{n} p_i \log p_i + \lambda_0 \left(\sum_{i=1}^{n} p_i - 1 \right) + \sum_{j=1}^{m} \lambda_j \left(\sum_{i=1}^{n} p_i r_{ij} - \alpha_j \right)$$

- Take derivative w.r.t. p_i :

$$-1 - \log p_i + \lambda_0 + \sum_{j=1}^{m} \lambda_j r_{ij}$$

Set this to 0, this solution is maximum entropy distribution

$$p_i^* = \frac{e^{\sum_{j=1}^{m} \lambda_j r_{ij}}}{e^{1-\lambda_0}}$$

Tare r_{ij} = as price or calories

Dice, no constraint

$$\mathcal{X} = \{1, 2, 3, 4, 5, 6\}.$$

Max^m entropy distribution

$$p_i = 1/6.$$

$$p_i^* = e^{d_i} \Big/ \sum_{i=1}^{6} e^{d_i}$$

Maximum entropy minimizes the amount of prior information.

Channel capacity (wireless)

$$I(x;y) = H(x) - H(x|y)$$

$$H(x|y) = H(x|y = \text{rain})\, p(\text{rain}) + H(x|$$

→ "Information" channel capacity.

$$C = \max_{p(x)} I(x;y)$$

We have proved, for fixed $p(y/x)$, $I(x;y)$ is a concave function in $p(x)$.

Duality

Data compression :- remove redundancy

Data transmission :- add redundancy

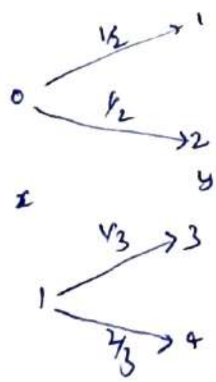
→ Why channel capacity ?

- Shannon propose to focus on information then Computation.

Semantic communication

→ Shannon's Secret of success

→ Binary noiseless channel

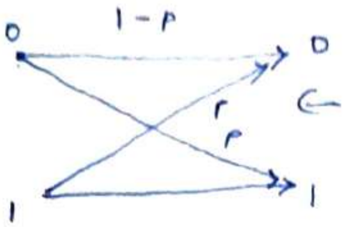


$C = \log 2 = 1 \text{ bit.}$

- Binary symmetric channel. (BSC).

Binary Symmetric Channel

2/04/24



channel transition matrix.

$$C = 1 - H(p) \text{ bits}$$

$$I(X;Y) = H(Y) - H(Y|X)$$

$$= H(Y) - \sum p(x)\, H(Y|X=x)$$

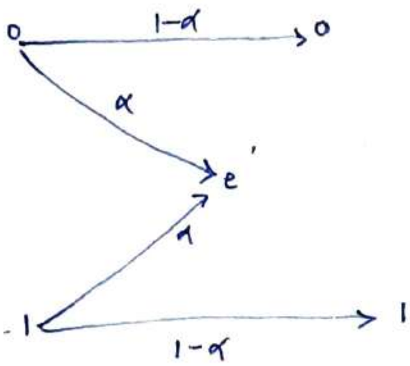
$$= H(Y) - \sum p(x)\, H(p)$$

$$I(X;Y)_{max} \le 1 - \sum p(x)\, H(p)$$

$$\le 1 - H(P)$$ when i/p is uniform.

Therefore $C \le \underbrace{1 - H(P)}$ — $I(X;Y)_{max}$.

Binary erasure Channel



$X = \{0,1\}$

$Y = \{0, e, 1\}$

Some bits are lost, can be use as a model for DNA sequencing

$$C = 1 - \alpha$$

$$C = \max_{p(x)} H(Y) - H(Y|X)$$

$$= \max_{p(x)} H(Y) - H(\alpha)$$

Transition Probability matrix

$$X = \{ x_0, x_1, x_2 \cdots x_{J-1} \}.$$

$$Y = \{ y_0, y_1, y_2, \cdots y_{K-1} \}.$$

$$P(Y|X) = \begin{bmatrix} p(y_0|x_0) & p(y_1|x_0) & \cdots & p(y_{K-1}|x_0) \\ p(y_0|x_1) & p(y_1|x_1) & \cdots & p(y_{K-1}|x_1) \\ \vdots & & & \\ p(y_0|x_{J-1}) & p(y_1|x_{J-1}) & \cdots & p(y_{K-1}|x_{J-1}) \end{bmatrix}$$

/

$p(y_k|x_j)$

$$\sum_{i=0}^{K} p(y_i|x_l) = 1 \quad \forall\, l.$$

Rows are input, columns are output

each rows and columns are permutation and combination of each other.

Symmetric Channel

$$p(y|x) = \begin{bmatrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{bmatrix}$$

Let r be a row of the transition matrix

$$I(X;Y) = H(Y) - H(Y|X)$$
$$= H(Y) - H(r)$$
$$\leq \log|\mathcal{Y}| - H(r)$$

with equality if $p(x) = 1/|\mathcal{X}|$

$$p(y) = \sum_{x \in \mathcal{X}} p(y|x)\, p(x) = \frac{c}{|\mathcal{X}|}$$

Weakly Symmetric Matrix

All the rows are p & c of every other rows and all the columns sums are equal

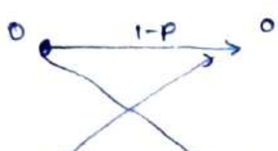
$$p(y|x) = \begin{bmatrix} 1/3 & 1/6 & 1/2 \\ 1/3 & 1/2 & 1/6 \end{bmatrix}.$$

$$C = \log|Y| - H(\text{row of transition matrix})$$

Properties of channel capacity :-

1. $C \geq 0$ since $I(X;Y) \geq 0$
2. $C \leq \log|X|$ since $C = \max I(X;Y) \leq \max H(X) = \log|X|$
3. $C \leq \log|Y|$ — same reason
4. $I(X;Y)$ is a continuous function of $p(x)$
5. $I(X;Y)$ is a concave function of $p(x)$

Transition probability matrix for binary Symmetric Channel



$$\begin{array}{cc} & \begin{array}{cc} 0 & 1 \end{array} \\ \begin{array}{c} 0 \\ 1 \end{array} & \begin{bmatrix} 1-p & p \\ p & 1-p \end{bmatrix} \end{array}$$

Lagrangian and KKT condition.

$\xrightarrow{w}$ [Encoder] $\xrightarrow{x^n}$ [Channel $p(y|x)$] $\xrightarrow{y^n}$ [Decoder] $\xrightarrow{\hat{w}}$

message.

Communication channel

$\{1, 2, \dots M\} \rightarrow x^n(w)$

$$Y^n \sim p(y^n | x^n)$$

Channel output

$\hat{w}$

$\hat{w} \neq w$ — error

Def^n 1. $(x, \; p(y|x), \; y)$ Discrete channel

$$\sum_y p(y|x) = 1$$

Def^n-2

The nth extension of the DMC is

$(x^n, \; p(y^n|x^n), \; y^n)$ where,

$$p(y_k | x^k, y^{k-1}) = p(y_k | x_k) \quad ; k = 1, 2, \dots$$

$$p(y^n | x^n) = \prod_{i=1}^{n} p(y_i | x_i)$$

4/09/24

→ Channel coding Theorem

For a DMC

(i) all rates below capacity $R < C$ are achievable.

(ii) converse :- any sequence

code rate (missing)

K- bit long

↳ 2^K possible information symbols.

- add $n-K$ redundant bits.

⇓

bits length becomes n

⇓

2^n possible information symbol

Then 9b $2^n - 2^K$ error pattern

Types of codes

(i) Error detecting code

(ii) Error correcting code

ARQ

Automatic repeat request.

BER can be large

Bit error rate.

If

→ FEC (Forward Error Correction)

Error correcting.

(i) Block code* (n, k) — k-tuple binary, n-tuple codeword

(ii) Convolution code

e.g. $(7, 4)$ block code

↳ 4-bit information
convert into
7-bit code word

It has a memory element. present input depend on past input.

Linear Block Codes

(n, k)

k-dimensional subspace of a n-dimensional vector space V. over the field F_2 such that a linear combination of any two vectors in the subspace will lead to another vector in subspace.

Mathematically

$\forall \; v_1, v_2 \in C \qquad v_1 \oplus v_2 \in C$

k-tuple message

2^k

C is k-dimensional subspace of a vector space V.

k n-tuple vectors in C which forms a basis set of C

$$G = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \qquad \left.\begin{aligned} g_1 &= g_{10} \quad g_{11} \; \cdots \; g_{1\,n-1} \\ g_2 &= g_{20} \quad g_{21} \; \cdots \; g_{2\,n-1} \\ &\vdots \\ g_k &= g_{k0} \quad g_{k1} \; \cdots \; g_{k\,n-1} \end{aligned}\right\} \text{n-tuple vectors.}$$

linear combination of k n-tuple vectors.

$$v = u_0 g_1 + u_1 g_2 + \cdots + u_{k-1} g_k$$

$$= \begin{bmatrix} u_0 & u_1 & \cdots & u_{k-1} \end{bmatrix} \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_k \end{bmatrix}$$

$$v = u \cdot \underbrace{G}_{\text{Generator matrix}}$$

Systematic form

k-bit information	(n-k) Check bits

LBC

$$G = [I_k : P]$$

or $[\underbrace{P}_{k \times (n-k)} : \underbrace{I_k}_{k \times k}]$

I_k - Identity matrix.

dimension - $k \times k$.

P - $k \times (n-k)$

Q.

Q. Obtain the encoded information for the given message and generator matrix.

$$u = \begin{bmatrix} 1 & 0 & 1 & 1 \end{bmatrix} \quad \text{and} \quad G = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$V = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Q. Obtain all possible code vectors for a (7,4) LBC in its systematic form for the generator matrix

$$G = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$R_1 = R_1 + R_2$

$r_3 = r_3 + r_1$

$R_1 \quad R_1 + R_2$

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$R_3 = r_3 + r_1$

$$G = \left[\underbrace{\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}}_{\text{Identity matrix.}} \quad \begin{matrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{matrix} \right].$$

Parity Check Matrix

$V H^T = 0$ if $\neq 0$ then error.

$u \cdot G H^T = 0$

$G H^T = 0$

Remaining.

$$G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}$$

Q. Determine the parity check matrix.

Step- (n, k) - identify

$(6, 3)$ - block code.

$K = 3$
$n = 6$

$$H = [P^T : I_{n-k}] = [P^T : I_3]$$

$$P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \qquad = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$G H^T = 0$$

$$\begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}^T_{3\times 6}$$

$$\Big[\qquad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}_{6\times 3}$$

Q. Determine the parity check matrix for a (7,4) Systematic LBC whose parity equations are:

$K = 4$
$n = 3$

$$P_1 = u_0 + u_1 + u_3$$

$$P_2 = u_1 + u_2 + u_3$$

$$P_3 = u_0 \; u_2 + u_3.$$

$$P = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$ [u_0, u_1, u_2, u_3]

$$P^T = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}_{3\times 4}$$

$$H = [P^T : I_{n-k}] = [P^T ; I_3]$$

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

u - input.
H - parity check.
V - output.
G - Gener

Q. Find H for given G.

$$G = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

H is in the form

$$G = [P \,;\, I_{n-k}] \longleftrightarrow H = [I_{n-k} \,;\, P^T].$$

$$H = [P^T \,;\, I_{n-k}]$$

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}$$

Encoding circuit for a (n, k) LBC

$$v = uG$$

$$[v_0 \; v_1 \; \cdots \; v_{n-1}] = [u_0 \; u_1 \; \cdots \; u_{k-1}] \begin{bmatrix} 1 & 0 & \cdots & 0 & P_{11} & P_{12} & \cdots & P_{1\,n-k} \\ 0 & 1 & \cdots & 0 & P_{21} & P_{22} & \cdots & P_{2\,n-k} \\ \vdots & & & & & & & \\ 0 & \cdots & \cdots & 1 & P_{k1} & P_{k2} & \cdots & P_{k\,n-k} \end{bmatrix}$$

$v_0 = u_0$

$v_1 = u_1$

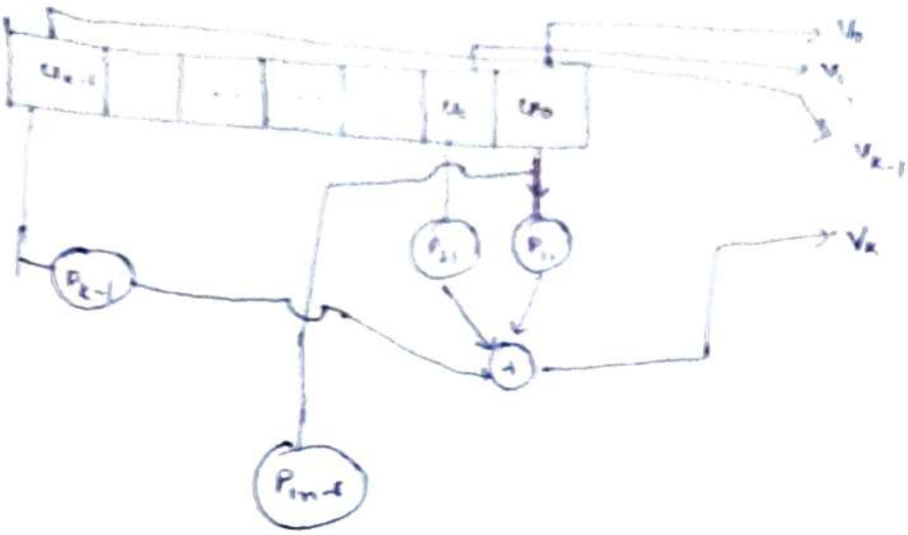
$\vdots$

$v_{k-1} = u_{k-1}$

$v_k = u_0 P_{11} + u_1 P_{21} + \cdots + u_{k-1} P_{k-1}$

$\vdots$

$v_{n-1} = u_0 P_{1\,n-k} + u_1 P$



Q. Consider a Systematic (8,4) LBC whose parity check equations are.

$$V_4 = u_1 + u_2 + u_3$$

$$V_5 = u_0 + u_1 + u_2$$

$$V_6 = u_0 + u_1 + u_3$$

$$V_7 = u_0 + u_2 + u_3.$$

Q (i) write the Generator and parity check matrices.

(ii) Draw the encoder diagram.

$I_{n-k} = I_4.$

$$P = \begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{bmatrix}$$

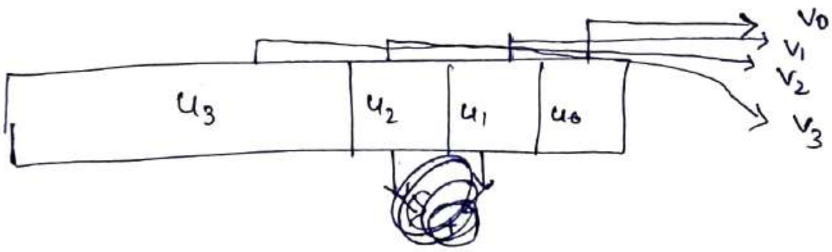
$G = [\, I_{n-k} \; ; P \,]$

$n = 8$
$k = 4.$

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix}$$

$H = [\, P^T \; ; \; I_{n-k} \,]$

$$H = \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$



$$V = [\, u_0 \; u_1 \; u_2 \; u_3 \,] \begin{array}{c} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{bmatrix} \\ \begin{array}{cccccccc} v_0 & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 \end{array} \end{array}$$

$V_0 = u_0$

$V_1 = u_1$

$V_2 = u_2$

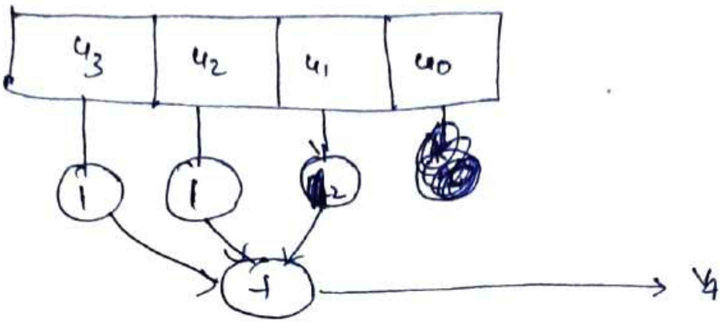
$V_3 = u_3$

$V_4 = u_1 + u_2 + u_3$

$V_5 = u_0 + u_1 + u_3$

$V_6 = u_0 + u_1 + u_3$

$V_7 = u_0 + u_2 + u_3$



09/04/24

Syndrome Calculation and Error detection

$vH^T = 0$

$r = v + e$ (e- error)

$rH^T \neq 0$; $\boxed{S = rH^T}$ (Syndrome)

$(v+e)H^T = vH^T + eH^T$

$= eH^T$

for any non-zero error pattern syndrome will be non-zero.

non-zero syndrome indicate.

• If both errors are same then it will have same ~~error~~ syndrome. (doesn't depend on input).

$$H = [P^T : I_{n-k}]$$

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 & 1 \end{bmatrix}$$

$$H^T = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad 6$$

$$VH^T = [0\ 1\ 1\ 0\ 1\ 0\ 0] \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$0 + 1 \cdot 1 + 0 + 0$$

$$= [0\ 1\ 0] \quad T_4$$

Syndrome calculation circuit

$$S = rH^T = [r_0\ r_1 \cdots r_{n-1}] \begin{bmatrix} P_{11} & P_{12} & \cdots & P_{1\,n-k} \\ P_{21} & P_{21} & \cdots & P_{2\,n-k} \\ \vdots & & & \\ P_{n\,k-1} & P_{n,k-2} & \cdots & P_{n\,n-k} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & \\ 0 & & & \\ 0 & & & 0 \end{bmatrix}.$$

$S_0 = r_0 P_{11} + r_1 P_{21} +$

S_1

S_2

Q. Draw the syndrome circuit for (7, 4) LBC.

Properties of LBC

① Hamming distance:

$v_1 = 1101010$

$v_2 = 1011100$

$d(v_1, v_2) = (v_1 \oplus v_2).$

② Hamming weight

$(110110 1)$

$H_w = 5 \rightarrow 5$

$$\boxed{d(v_1, v_2) = H_w(v_1 \oplus v_2)}$$

③ Minimum Distance

$d_{min} = \min\{d(v_i, v_j)\}.$ ~~and i≠j~~ $\forall\ i, j$ and $i \neq j$

Theorem :

The minimum distance of a LBC is equal to the minimum weight of a non-zero code vector.

Theorem :

For any code vector of weight l in an LBC there exist 'l' columns in H matrix whose sum is zero.

H.W. prove

corollary :

Corollary ① A LBC is said to have a dmin of d if there are no fewer d columns in H matrix whose sum is zero.

corollary ②. If there are d colums in H - matrix whose sum is zero and no fewer d columns whose sum is zero ; then the LBC is said to have a minimum distance of d.

Theorem A (n, K) LBC with a minimum distance of dmin is capable of detecting $(d_{min} - 1)$ bit errors.

Th

Theorem

A (n, k) LBC is capable of correcting upto $\left\lfloor (d_{min}-1)/2 \right\rfloor$ numbers of errors bits for a minimum distance d_{min}.

Q.

$$G = \begin{bmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

(A) Find n, k value

(B) G in its systematic form.

(C) Find all codewords.

(4) Find d_{min}.

(5) Find the error detecting