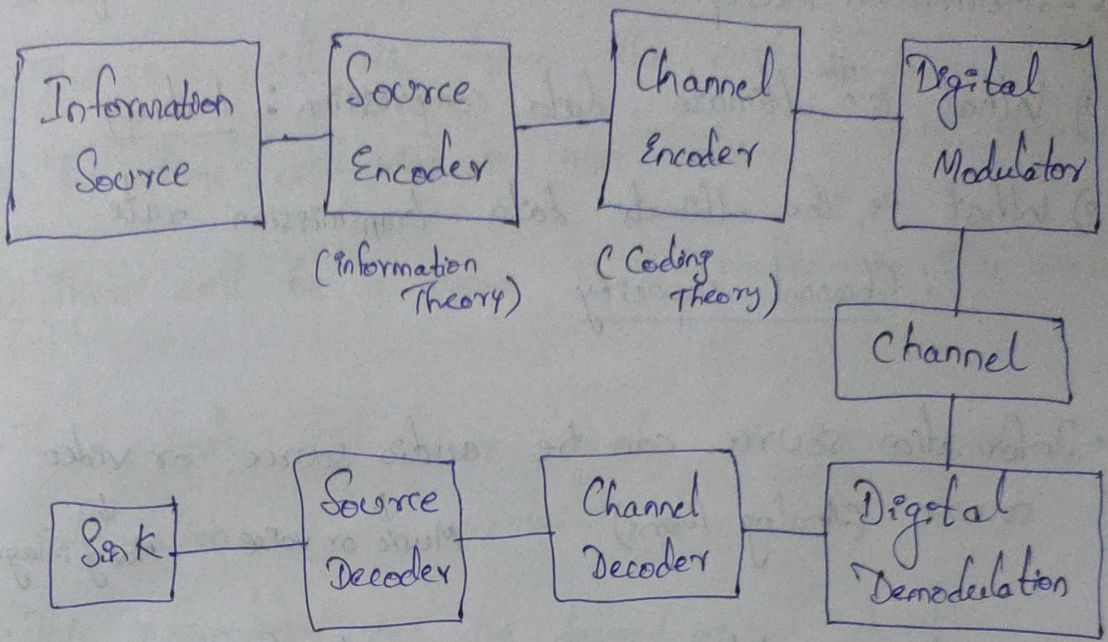
# EC353 - Information Theory & Coding

04/01/202




Block Diagram of communication system

- Information source can be in Analog form or Discrete form.

- Source encoder : will decrease no. of bits → Information Theory
- Channel encoder : add extra bits called redundant bits.
  → inorder to check for error detection at receiver side.
  → Coding Theory.

- For data Compression (or) reducing ~~No.~~ of bits – Information Theory

- For adding redundant bits – Coding Theory

- Do the process in such a way that Probability of Error

~~PE~~ : $P_e \to 0$.

The Information Theory answers two fundamental questions.

1) What is the ultimate data compression : <u>Entropy</u>

2) What is the ultimate data transmission rate.
: <u>Channel capacity</u>

- Information source can be audio source or video source (Analog forms)
  - audio source → Music or voice
  - video → Moving images

- Computer generated sources — Discrete forms

<u>Source - Encoder</u> (purpose)

- To minimize the no. of bits per unit time required to represent the source ~~code~~ output → This process is known as <u>source coding</u> or <u>data compression</u>.

<u>Channel - Encoder</u> (purpose)

- To correct the transmission errors introduced by the channel

<u>Digital Modulator</u> (purpose)

- It will map the code words into waveforms which are then transmitted.

- How to quantify given information?

# Information :

1) The sun will rise from the east. no uncertainity $P(E_1)=1$
↳ no information

2) The phone will ring in next hour.

3) There will be a snow in tadepalligudem this winter.

- The events which occur rarely will have more information. (more uncertainity - more information)

- This information can be defined as :

Self information } $I(x) = \log \frac{1}{P(X=x_i)}$ ; $i=0,1,2,\cdots,n$

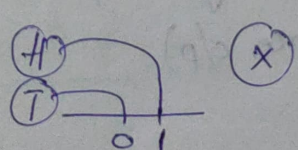
→ Log: Base '2' — units is bits
Base 'e' — unit is nats

'X' maps sample Space to Real-time values
↳ Random variable (continuous)

- 'X' is a random variable. (Discrete one)

Maps Sample Space to Discrete values

Ex: Tossing a Coin (unbiased)




$P(X=0) = \frac{1}{2}$

$P(X=1) = \frac{1}{2}$

- Probability of Head appearing when an unbiased coin tossed

$I(x) = 1$ bit

$\because I(x) = \log_2\left(\frac{1}{1/2}\right)$

$= \log_2 2$

$= 1$

(Mutual information → involvement of 2 or more random variables)

Self-Information - since single random variable involved

# Average Self-Information (Entropy)

$$H(X) = \sum_i P(X=x_i) \log\left(\frac{1}{P(X=x_i)}\right) \text{ bits/symbol}$$

(in discrete)

$$= E\left[\log \frac{1}{P(X=x_i)}\right] \quad \left(\because E(x) = x\,P(x)\right)$$

$$= -\sum_i P(x_i) \log P(x_i)$$

Ex: Tossing an unbiased coin : X

$$H(X) = \frac{1}{2}(1) + \frac{1}{2}(1) = 1 \text{ bit/symbol}$$

05/01/2023

Entropy: $H(X) = -\sum_i P(x_i) \log P(x_i)$

- (Measures) Entropy is a measure of uncertainity of R.V. X
- 'X' is a discrete random variable.
- Discrete Memoryless Source (DMS):

→ Successive symbols are statistically independent of previous symbols. (from the source o/p)

Ex: $\boxed{\text{(DMS) Source}}$

$P(1) = \frac{1}{3}$

$P(0) = \frac{2}{3}$

$$P(E) = \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{1}{3} \times \frac{2}{3} \times \frac{1}{3} \times \frac{2}{3}$$

Emitting 1011010

Ex: Suppose the discrete R.V. 'x' takes

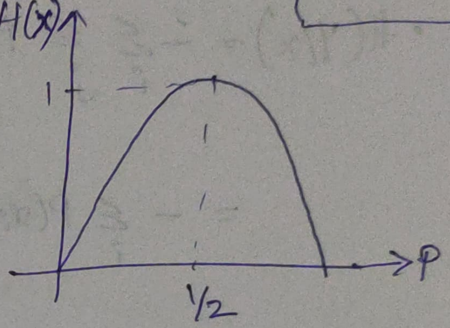
$$X = \begin{cases} 0; & \text{with prob. 'P'} \\ 1; & \text{with prob. '1-P'} \end{cases}$$ (Bernoli R.V)/ (Binary R.V)

$P(x=0) = P$ and $P(x=1) = 1-P$

. Entropy : $H(x) = -P\log P - (1-P)\log(1-P)$ | Binary Entropy function

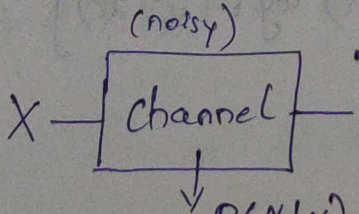
$\triangleq h(p)$

$\triangleq H(P)$




. Max. uncertainity at $P = 1/2$

. At $P=0$, no uncertainity → always $(x=1)$ occurs




Mutual information ✓

Conditional Entropy ✓

$P(Y|X)$ – Channel Matrix (Channel Transition Matrix)

→ Nothing but Conditional Probability.

. Y – noisy observation of X

Conditional Entropy :

. $H(X|Y)$ → It is a measure of uncertainity in 'x' after observing 'Y'.

$i = 1, 2, \ldots, n$ & $j = 1, 2, \ldots, n$

. $H(X|Y) = -\sum_i \sum_j P(x_i, y_j) \log [P(x_i / y_j)]$

. $H(Y|X) = -\sum_i \sum_j P(x_i, y_j) \log [P(y_j / x_i)]$

- Join probability & Conditional probability:

$$P(x/y) = \frac{P(x,y)}{P(y)} \quad \text{and} \quad P(y/x) = \frac{P(x,y)}{P(x)}$$

- $H(Y/X) = -\sum_i \sum_j P(x_i/y_j)\, P(y_j) \log P(y_j/x_i)$

(or)

- $H(Y/X) = -\sum_i \sum_j P(y_j/x_i)\, P(x_i) \log P(y_j/x_i)$

$$= -\sum_i P(x_i) \sum_j P(y_j/x_i) \log P(y_j/x_i)$$

$$= \sum_i P(x_i) \left[-\sum_j P(y_j/x_i) \log P(y_j/x_i)\right]$$

$$= \sum_i P(x_i)\, H(Y/x_i)$$

# $H(Y/X) = \sum_i P(x_i)\, H(Y/x_i)$

# $H(X) = -\sum_i P(x_i) \log P(x_i)$

<u>Joint Entropy</u>: $H(X,Y)$

- It measures how much uncertainity there is in two RVs X & Y taken together.

- $H(X,Y) = -\sum_i \sum_j P(x_i, y_j) \log P(x_i, y_j)$

Joint Entropy:

- $H(X,Y) = -\sum_i \sum_j P(x_i, y_j) \log[P(x_i, y_j)]$

$= -\sum_i \sum_j P(x_i, y_j) \log[P(x_i) P(y_j/x_i)]$

$= -\sum_i \sum_j P(x_i, y_j) \log P(x_i)$

$- \sum_i \sum_j P(x_i, y_j) \log(y_j/x_i)$

- Marginal Probability mass function:

$$\sum_j P(x_i, y_j) = P(x_i)$$

$\Rightarrow H(X,Y) = -\sum_i P(x_i) \log P(x_i) + \sum_i P(\cancel{x_i}) H(\cancel{Y}/\cancel{x_i})$

$= H(X) + H(\cancel{Y}/X)$

$= H(Y) + H(\cancel{X/Y})$

$\therefore H(X,Y) = H(X) + H(Y\cancel{/}X) = H(Y) + H(X/Y)$

Corollary:

- $H(X, Y/Z) = H(X/Z) + H(Y/X, Z)$

[# $\log(ab) = \log a + \log b$
$\log\left(\frac{a}{b}\right) = \log a - \log b$]

## Chain rule for Entropy :

- $H(X_1, X_2, \ldots, X_n) = \sum_{i=1}^{n} H(X_i / X_{i-1}, \ldots, X_1)$

**Proof :**

$$H(X_1, X_2) = H(X_1) + H(X_2 / X_1).$$

$$H(X_1, X_2, X_3) = H(X_1) + H(X_2, X_3 / X_1)$$

$$= H(X_1) + H(X_2 / X_1) + H(X_3 / X_2, X_1)$$

- for 'n' R.V's

$$\therefore H(X_1, X_2, X_3, \ldots, X_n) = H(X_1) + H(X_2/X_1) + H(X_3/X_2, X_1) + \ldots + H(X_n / X_{n-1}, \ldots, X_1)$$

$$= \sum_{i=1}^{n} H(X_i / X_{i-1}, \ldots, X_1)$$

→ Hence, Proved.

## Mutual Information : $I(X;Y)$

- Self-Information : $I(x) = -\log P(x = x_i)$

. Mutual Information: $I(x;y)$

. It measures how much information about channel input 'x' can be obtained after observing the channel output 'y'.

$$I(x;y) = \sum_i \sum_j P(x_i, y_j) \log \left[ \frac{P(x_i, y_j)}{P(x_i)\, P(y_j)} \right]$$

$$= \sum_i \sum_j P(x_i, y_j) \log \left[ \frac{P(x_i / y_j)}{P(x_i)} \right]$$

$$\frac{P(x_i/y_j)}{P(x_i)} = \frac{P(x_i/y_j)\, P(y_j)}{P(x_i)\, P(y_j)}$$

$$\because \frac{P(x,y)}{P(y)} = P(x/y)$$

$$\frac{P(x,y)}{P(x)} = P(y/x)$$

$$= \frac{P(x_i, y_j)}{P(x_i)\, P(y_j)}$$

$$= \frac{P(y_j/x_i)}{P(y_j)}$$

$$\Rightarrow I(x;y) = \sum_i \sum_j P(x_i, y_j) \log \left[ \frac{P(y_j/x_i)}{P(y_j)} \right]$$

$$\boxed{\therefore I(x;y) = I(y;x)}$$

→ The mutual information of a channel is symmetric.

$$I(x;y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i/y_j)}{P(x_i)}$$

$$\Rightarrow I(X;Y) = -\sum_i \underbrace{\left[\sum_j P(x_i, y_j)\right]}_{P(x_i)} \log P(x_i)$$

$$+ \sum_i \sum_j P(x_i, y_j) \log P(x_i / y_j)$$

$$= H(X) - H(X/Y)$$

$$= \underbrace{H(Y)}_{\text{more uncertainity}} - \underbrace{H(Y/X)}_{\text{Low uncertainity}}$$

- Reduction in the uncertainity of one R.V. due to ~~other~~ knowledge of the other R.V.

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

06/01/2023

- $I(X;Y) = H(Y) - H(Y/X) \rightarrow (1)$

- $H(X,Y) = H(X) + H(Y/X) \rightarrow (2)$

From (1) & (2):

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Special Cases:

1) Suppose if X & Y are statistically independent

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Case 1:

- If $x_i, y_j$ are statistically independent then

$$P(x_i, y_j) = P(x_i) \cdot P(y_j)$$

$$P(y_j/x_i) = \frac{P(x_i, y_j)}{P(x_i)} = P(y_j)$$

- $H(Y/X) = -\sum_i \sum_j P(y_j/x_i)\, P(x_i) \log P(y_j/x_i)$

$= -\sum_i \sum_j P(x_i, y_j) \log P(y_j/x_i)$

$= -\sum_i \sum_j P(x_i, y_j) \log P(y_j)$

$= -\sum_j P(y_j) \log P(y_j)$

$= H(Y)$

$\because x_i, y_j$ are statistically independent $P(y_j/x_i) = P(y_j)$

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

$= 0$

→ Occurance of event $Y = y_j$ provides no information about $X = x_i$.

Special case-2:

2) When the occurance of event $Y = y_j$ uniquely determines the event $X = x_i$

$$P(X = x_i / Y = y_j) = 1$$

Case 2:

$$I(X;Y) = \sum_i \sum_j P(x_i, y_j) \log \frac{P(x_i/y_j)}{P(x_i)}$$

$$= \sum_i \sum_j P(x_i, y_j) \log \frac{1}{P(x_i)}$$

$$= -\sum_i P(x_i) \log P(x_i)$$

$$= H(X)$$

$$\rightarrow I(X;Y) = H(X)$$

# $I(X;X) = H(X) - H(X/X)$

$$= H(X)$$

$$\boxed{\cdot\, H(X/X) = 0}$$

By definition:

# $H(X/X) = -\sum_i \sum_j P(x_i, x_j) \log P(x_j/x_i)$

$$= 0$$

$$\boxed{P(x_j/x_i) = \frac{P(x_i, x_j)}{P(x_i)} = \frac{P(x_i)}{P(x_i)} = 1}$$

Chain rule for Information:

- $I(X_1, X_2, \dots, X_n ; Y) = \sum_{i=1}^{n} I(X; Y / X_{i-1}, X_{i-2}, \dots, X_1)$

Proof:

$$I(X_1, X_2, \ldots, X_n ; Y) = H(X_1, X_2, \ldots, X_n) - H(X_1, X_2, \ldots X_n / Y)$$

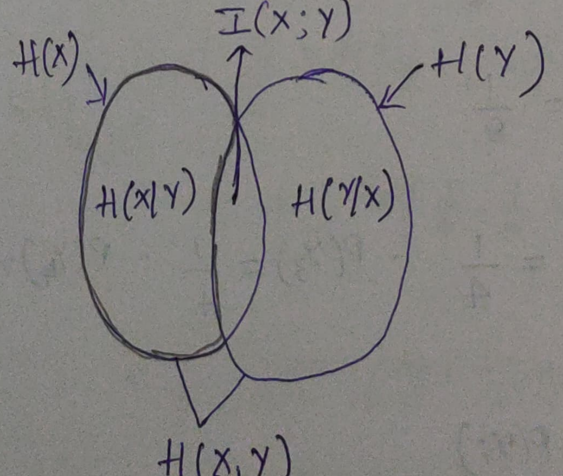
$$= \sum_{i=1}^{n} H(X_i / X_{i-1}, \ldots X_1) - \sum_{i=1}^{n} H(X_i / X_{i-1}, \ldots, X_1, Y)$$

$$= \sum_{i=1}^{n} I(X ; Y / X_{i-1}, \ldots, X_1)$$

$$\boxed{\because I(X;Y) = H(X) - H(X/Y)}$$

→ Hence, Proved.

The interrelationship




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

$$H(X,Y) = H(X) + H(Y/X)$$
$$= H(Y) + H(X/Y)$$

$$I(X;Y) = H(X) + H(Y) - H(X,Y)$$

Q. Let the x, y have the following joint distribution.

| Y \ X | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1/8 | 1/16 | 1/32 | 1/32 |
| 2 | 1/16 | 1/8 | 1/32 | 1/32 |
| 3 | 1/16 | 1/16 | 1/16 | 1/16 |
| 4 | 1/4 | 0 | 0 | 0 |

Joint Distribution Property

$\sum_i \sum_j P(x_i, y_j) = 1$

- $P(x_i) = \sum_j P(x_i, y_j)$

$\Rightarrow P(x_1) = \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \frac{1}{16} = \frac{1}{8} + \frac{1}{8} + \frac{1}{4} = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$

$\Rightarrow P(x_2) = \frac{1}{16} + \frac{1}{8} + \frac{1}{16} = \frac{1}{4}$

$\Rightarrow P(x_3) = \frac{1}{32} + \frac{1}{32} + \frac{1}{16} = \frac{1}{8}$

$\Rightarrow P(x_4) = \frac{1}{32} + \frac{1}{32} + \frac{1}{16} = \frac{1}{8}$

- $P(y_1) = \frac{1}{4}$ • $P(y_2) = \frac{1}{4}$ • $P(y_3) = \frac{1}{4}$ • $P(y_4) = \frac{1}{4}$

- $H(X) = -\sum_i P(x_i) \log P(x_i)$

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

$= \frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8} = 1 + \frac{3}{4} = \frac{7}{4}$ bits/symbol

$\cdot\ H(X) = \frac{7}{4}$ bits/symbol

$\cdot\ H(Y) = 4 \times \frac{1}{4} \log_2 4 = 2$ bits/symbol

$\cdot\ P(x/y) = \dfrac{P(x,y)}{P(y)}$

$$\cdot\ P(X/Y) = \begin{array}{c|c} & X \\ \hline Y & \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 1 & 0 & 0 & 0 \end{bmatrix} \end{array} \quad \text{(Transition matrix)}$$

$P(x,y)$ $\div P(y_1)$ $\div P(y_2)$ $\div P(y_3)$ $\div P(y_4)$

$$\cdot\ H(X/Y) = \sum_j P(y_j)\, H(X/y_j)$$

$\frac{11}{8}$ bits/symbol

$$= \frac{1}{4} \times H\left(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\right) + \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{8}, \frac{1}{8}\right)$$
$$+ \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4}\right) + \frac{1}{4} H(1, 0, 0, 0)$$

$$= \frac{1}{4}\left[\frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{8}\log 8 + \frac{1}{8}\log 8\right]$$
$$+ \frac{1}{4}\left[\frac{1}{4}\log 4 + \frac{1}{2}\log 2 + \frac{1}{8}\log 8 + \frac{1}{8}\log 8\right]$$
$$+ \frac{1}{4}\left[\frac{1}{4}\log 4 \times 4\right] + \frac{1}{4}(0)$$

$$= \frac{1}{4}\left[\frac{1}{2} + \frac{2}{4} + \frac{3}{8} + \frac{3}{8}\right] + \frac{1}{4}\left[\frac{2}{4} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8}\right] + \frac{2}{4} = \frac{1}{4} \times \frac{7}{4} \times 2 + \frac{1}{2} = \frac{7}{8} + \frac{1}{2}$$
$$= 11/8$$
(bits/symbol)

$\cdot\ H(Y/X) = \sum_i P(x_i)\, H(Y/x_i) \rightarrow \frac{13}{8}$ bits/symbol

$\cdot\ H(X,Y) = H(X) + H(Y/X)$ $\Big\}$ $\frac{27}{8}$ bits/symbol

$= H(Y) + H(X/Y)$

$= \frac{7}{4} + \frac{13}{8} = 2 + \frac{11}{8} = \frac{27}{8}$ bits/symbol

$\cdot\ I(X;Y) = H(X) - H(X/Y)$ $\Big\}$ $0.375$ bits/symbol

$= H(Y) - H(Y/X)$

$= \frac{7}{4} - \frac{11}{8} = 2 - \frac{13}{8} = \frac{3}{8} = 0.375$ bits/symbol

$$P(Y/X) = \begin{array}{c} {\scriptstyle Y \diagdown X} \end{array} \begin{bmatrix} 1/4 & 1/4 & 1/4 & 1/4 \\ 1/8 & 1/2 & 1/4 & 1/4 \\ 1/8 & 1/4 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 0 \end{bmatrix}$$

$H(Y/X) = \sum_i P(x_i) H(Y/x_i)$

$\cdot\ H(Y/X) = \frac{1}{2} H\left(\frac{1}{4}, \frac{1}{8}, \frac{1}{8}, \frac{1}{2}\right) + \frac{1}{4} H\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}, 0\right)$

$+ \frac{1}{8} H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0\right) + \frac{1}{8} H\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}, 0\right)$

$= \frac{1}{2}\left[\frac{1}{4}\log 4 + \frac{1}{8}\log 8 + \frac{1}{8}\log 8 + \frac{1}{2}\log 2\right] + \frac{1}{4}\left[\frac{1}{4}\log 4 + \frac{1}{2}\log 2 + \frac{1}{4}\log 4\right]$

$+ \frac{1}{8}\left[\frac{1}{4}\log 4 + \frac{1}{4}\log 4 + \frac{1}{2}\log 2\right] + \frac{1}{8}\left[\frac{1}{4}\log 4 + \frac{1}{4}\log 4 + \frac{1}{2}\log 2\right]$

$= \frac{1}{2}\left[\frac{2}{4} + \frac{3}{8} + \frac{3}{8} + \frac{1}{2}\right] + \frac{1}{4}\left[\frac{2}{4} + \frac{1}{2} + \frac{2}{4}\right] + \frac{2}{8}\left[\frac{2}{4} + \frac{2}{4} + \frac{1}{2}\right]$

$= \frac{1}{2}\left[\frac{7}{4}\right] + \frac{1}{4} \times \frac{3}{2} + \frac{1}{4} \times \frac{3}{2} = \frac{7+3+3}{8} = \frac{13}{8}$ bits/symbol

## Properties:

1) $H(X) \le \log_2 |X|$

how many variables r.v can take?

- $|X|$ is the size of the alphabet
  - for binary - 2

- Equality is achieved when X has uniform distribution.

→ Equality constraint holds when X follows uniform distribution.

2) $I(X;Y) \ge 0$ (non-negative) (positive)

$H(X) \ge H(X/Y)$

$H(Y) \ge H(Y/X)$

11/01/2023

## Channel Capacity

→ The maximum amount of information that can be transmitted via channel.

* → It is the highest rate at which information can be sent arbirarily low probability of error.

→ Measured in bits/channel use

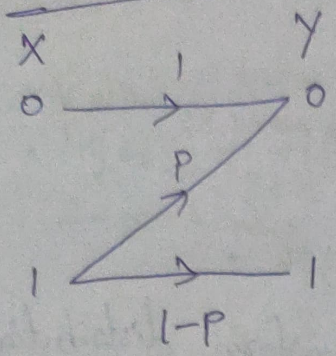
$$C = \max_{P(x_i)} I(X;Y)$$

maximize $I(X;Y)$ (Mutual information) over input distribution $P(x_i)$

## Z-channel:

# Z-channel

11/01/2022




Consider:

$$P(X=0) = 1-\alpha$$

$$P(X=1) = \alpha$$

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

Distribution of Y: $P(Y=0) = P(X=0)\,P(Y=0/X=0) + P(X=1)\,P(Y=0/X=1)$

$$= (1-\alpha)\cdot 1 + (\alpha)(p) = 1-\alpha+\alpha p$$

$\rightarrow P(Y=0) = 1-\alpha+\alpha p$

$P(Y=1) = P(X=1)\,P(Y=1/X=1) = \alpha(1-p)$

$\rightarrow P(Y=1) = \alpha(1-p)$

- $H(Y) = -\sum_j P(y_j) \log P(y_j)$

$$= -(1-\alpha+\alpha p)\log(1-\alpha+\alpha p) - \alpha(1-p)\log \alpha(1-p)$$

$$= H(\alpha(1-p))$$

[$\because$ Binary entropy function: $H(x) = -x\log x - (1-x)\log x$]

Channel Matrix (or) Transition Matrix

$$P(Y|X) = \begin{array}{c|cc} X \backslash Y & 0 & 1 \\ \hline 0 & 1 & 0 \\ 1 & p & 1-p \end{array} = \begin{bmatrix} P[Y=0/X=0] & P[Y=1/X=0] \\ P[Y=0/X=1] & P[Y=1/X=1] \end{bmatrix}$$

$$H(Y/X) = \sum_i P(x_i)\, H(Y/x_i)$$

$$= P(x=0)\, H(1,0) + P(x=1)\, H(p, 1-p)$$

$$= (1-\alpha)(0) + \alpha\left[-p\log p - (1-p)\log(1-p)\right]$$

$$= \alpha H(p)$$

$$\therefore I(x;y) = H(Y) - H(Y/X)$$

$$= H(\alpha(1-p)) - \alpha H(p)$$

Channel Capacity: $C = \max_{\alpha}\left[H(\alpha(1-p)) - \alpha H(p)\right] \rightarrow (1)$

$\frac{\partial}{\partial\alpha}$ $-\left((1-\alpha)+\alpha p\right)\log(1-\alpha+\alpha p) - (\alpha-\alpha p)\log(\alpha-\alpha p) - \alpha H(p)$

On differentiating

$$\Rightarrow -(-1+p)\log(1-\alpha+\alpha p) - (-1+p) - (1-p)\log(\alpha-\alpha p)$$
$$- (1-p) - H(p)$$

$$\Rightarrow (1-p)\log(1-\alpha+\alpha p) - (1-p)\log(1-p)\alpha - H(p)$$

$$\Rightarrow (1-p)\log(1-\alpha+\alpha p) - (1-p)\log(1-p)\alpha + p\log p + (1-p)\log(1-p)$$

$$\Rightarrow (1-p)\log\frac{(1-\alpha+\alpha p)(1-p)}{(1-p)\alpha} + p\log p = 0$$

$$\Rightarrow \log_2\left(\frac{1-\alpha+\alpha p}{\alpha}\right)^{(1-p)} + \log p^p = 0$$

$$\Rightarrow \left(\frac{1-\alpha+\alpha p}{\alpha}\right)^{(1-p)} p^p = 1$$

$$\Rightarrow \left(\frac{1}{\alpha}+(p-1)\right)^{(1-p)} p^p = 1$$

$$\Rightarrow \left(\frac{1}{\alpha}+p-1\right)^{1-p} = \frac{1}{p^p}$$

$$\Rightarrow \frac{1}{\alpha} = \frac{1}{p^{p/1-p}} + 1 - p$$

$$\Rightarrow \alpha = \frac{p^{p/1-p}}{1+(1-p)\,p^{p/1-p}}$$

$$\Rightarrow \alpha = \frac{1}{(1-p)\left[1+2^{H(p)/(1-p)}\right]}$$

or

$$(1-p)\log_2\frac{1-\alpha+\alpha p}{(1-p)\alpha} - H(p) = 0$$

$$\Rightarrow \frac{1-\alpha+\alpha p}{(1-p)\alpha} = 2^{H(p)/1-p}$$

$$\Rightarrow \frac{1}{(1-p)\alpha} - 1 = 2^{H(p)/1-p} \Rightarrow \alpha = \frac{1}{(1-p)\left[1+2^{H(p)/(1-p)}\right]}$$

$$\therefore \alpha = \frac{1}{(1-P)\left[1+2^{H(P)/(1-P)}\right]}$$

Substituting the value of '$\alpha$' in eq①, we

$$\Rightarrow C = H(\alpha(1-P)) - \alpha H(P)$$

$$C = H\left[(1-P)\cdot\frac{1}{(1-P)}\cdot\frac{1}{1+2^{\theta(P)}}\right] - \frac{H(P)}{(1-P)\left[1+2^{\theta(P)}\right]}$$

where $\theta(P) = \frac{H(P)}{1-P}$

$$\Rightarrow C = H\left[\frac{1}{1+2^{\theta(P)}}\right] - \frac{1}{(1-P)\left[1+2^{\theta(P)}\right]} H(P) \quad ; \theta(P) = \frac{H(P)}{1-P}$$

$$= H\left[1/1+2^{\theta(P)}\right] - \theta(P)/\left[1+2^{\theta(P)}\right]$$

- On substituting / simplyfying $H(P) = -P\log P - (1-P)\log(1-P)$

$$\Rightarrow C = \log\left[1+2^{-\theta(P)}\right]$$

$$\Rightarrow C = \log\left[1+2^{\left(\frac{P}{1-P}\log P + \log(1-P)\right)}\right]$$

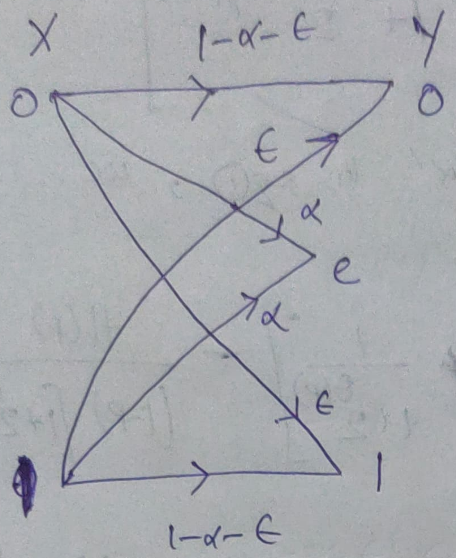
$$\Rightarrow C = \log\left[1+2^{\log_2\left(P^{(P/1-P)}\times(1-P)\right)}\right]$$

$\boxed{\log m + \log n = \log mn}$

$$\Rightarrow C = \log\left[1+(1-P)\,P^{\left(\frac{P}{1-P}\right)}\right]$$

$\boxed{\therefore \text{Channel capacity of Z-channel : } C = \log\left[1+(1-P)\,P^{\left(\frac{P}{1-P}\right)}\right]}$

## Generalized Channel

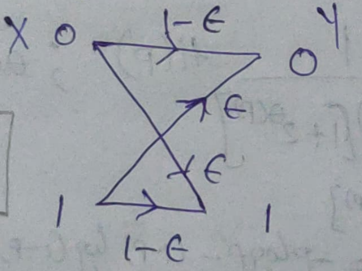



Special cases

1) Binary Symmetric channel
2) Binary Erasure channel
3) Binary Noise less channel

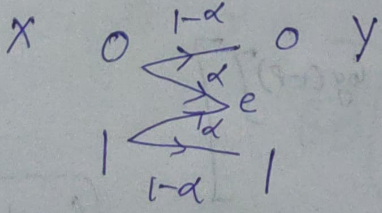
## Binary symmetric channel : (when $\alpha = 0$)

$$\begin{array}{c|cc} X \backslash Y & 0 & 1 \\ \hline 0 & 1-\epsilon & \epsilon \\ 1 & \epsilon & 1-\epsilon \end{array}$$

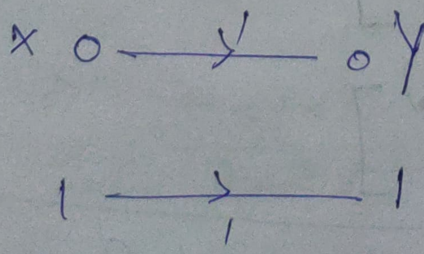



12/2/2023

## Binary Erasure Channel : (when $\epsilon = 0$)




## Binary Noise Less Channel : (when $\alpha = \epsilon = 0$)



# Calculating Channel Capacity

12/01/2023

$P(X=0) = 1-\beta$
$P(X=1) = \beta$

$$P(Y=0) = P(X=0)(1-\alpha-\epsilon) + P(X=1)\,\epsilon$$
$$= (1-\beta)(1-\alpha-\epsilon) + \beta\epsilon$$
$$= -\beta + 2\epsilon\beta + \alpha\beta - \epsilon - \alpha + 1$$

(or)

$P(X=0) = \beta$ & $P(X=1) = 1-\beta$

$$P(Y=0) = P(X=0)(1-\alpha-\epsilon) + P(X=1)\,\epsilon$$
$$= \beta(1-\alpha-\epsilon) + (1-\beta)\,\epsilon$$
$$= \beta - 2\epsilon\beta - \alpha\beta + \epsilon$$

$$P(Y=1) = P(X=0)\,\epsilon + P(X=1)(1-\alpha-\epsilon)$$
$$= \beta\epsilon + (1-\beta)(1-\alpha-\epsilon)$$
$$= 1-\alpha-\epsilon - \beta(1-2\epsilon-\alpha)$$

$$P(Y=e) = P(X=0)\,\alpha + P(X=1)\,\alpha$$
$$= \alpha\beta + (1-\beta)\alpha$$
$$= \alpha$$

Channel matrix

$$P(Y/X) = \begin{array}{c|ccc} X \backslash Y & 0 & e & 1 \\ \hline 0 & 1-\alpha-\epsilon & \alpha & \epsilon \\ 1 & \epsilon & \alpha & 1-\alpha-\epsilon \end{array}$$

$$H(Y) = -\sum_j P(y_j) \log P(y_j)$$

$$= H(\beta - 2\epsilon\beta - \alpha\beta + \epsilon, \alpha, 1-\alpha-\epsilon-\beta(1-2\epsilon-\alpha))$$

$$H(Y/X) = \sum_i P(x_i) H(Y/x_i)$$

$$= \beta H(1-\alpha-\epsilon, \alpha, \epsilon) + (1-\beta) H(\epsilon, \alpha, 1-\alpha-\epsilon)$$

$$= H(1-\alpha-\epsilon, \epsilon, \alpha)$$

$$\rightarrow C = \max_\beta I(X;Y)$$

$$= \max_\beta H(Y) - H(Y/X)$$

$$= \max_\beta H(\beta - 2\epsilon\beta - \alpha\beta + \epsilon, \alpha, 1-\alpha-\epsilon-\beta(1-2\epsilon-\alpha))$$

$$- \underbrace{H(1-\alpha-\epsilon, \epsilon, \alpha)}$$

$\hookrightarrow$ $\beta$ not involved so differentiation $\Rightarrow 0$

$$\Rightarrow \frac{d}{d\beta}\Big[-(\beta - 2\epsilon\beta - \alpha\beta + \epsilon)\log(\beta - 2\epsilon\beta - \alpha\beta + \epsilon) - [(1-\alpha-\epsilon-\beta(1-2\epsilon-\alpha))]$$

$$-\alpha\log\alpha \qquad \log[1-\alpha-\epsilon-\beta(1-2\epsilon-\alpha)]\Big] = 0$$

$$\Rightarrow (1-2\epsilon-\alpha)\log(\beta - 2\epsilon\beta - \alpha\beta + \epsilon) + (1-2\epsilon-\alpha)$$

$$+ (-)(1-2\epsilon-\alpha) + (-)(1-2\epsilon-\alpha) \log(1-\alpha-\epsilon-\beta+2\epsilon\beta+\alpha\beta) = 0$$

$$\Rightarrow (1-2\epsilon-\alpha)\log(\beta-2\epsilon\beta-\alpha\beta+\epsilon) + (1-2\epsilon-\alpha)$$

$$-(1-2\epsilon-\alpha)\log(1-\alpha-\epsilon-\beta+2\epsilon\beta+\alpha\beta)$$

$$-(1-2\epsilon-\alpha) = 0$$

$$\Rightarrow (1-2\epsilon-\alpha)\log\frac{\beta-2\epsilon\beta-\alpha\beta+\epsilon}{1-\alpha-\epsilon-\beta(1-2\epsilon-\alpha)} = 0$$

$$\Rightarrow \beta-2\epsilon\beta-\alpha\beta+\epsilon = 1-\alpha-\epsilon-\beta+2\epsilon\beta+\alpha\beta$$

$$\Rightarrow 2\beta+2\epsilon = 1-\alpha+4\epsilon\beta+2\alpha\beta$$

$$\Rightarrow \beta(2-4\epsilon-2\alpha) = 1-\alpha-2\epsilon$$

$$\Rightarrow \beta = \frac{(1-\alpha-2\epsilon)}{2(1-\alpha-2\epsilon)}$$

$$\boxed{\beta=\frac{1}{2}}$$

Substituting $\beta=\frac{1}{2}$

for a Generalized channel

$$\boxed{\therefore C = -(1-\alpha)\log\left(\frac{1-\alpha}{2}\right) + (1-\alpha-\epsilon)\log(1-\alpha-\epsilon) + \epsilon\log\epsilon}$$

$$\Big[\because C = H\left(\frac{1}{2}-\frac{\epsilon}{\cancel{2}}-\frac{\alpha}{2}+\epsilon, \alpha, 1-\alpha-\epsilon-\frac{1}{2}+\epsilon+\frac{\alpha}{2}\right) - H(1-\alpha-\epsilon, \epsilon, \alpha)$$

$$= H\left(\frac{1-\alpha}{2}, \alpha, \frac{1-\alpha}{2}\right) - H(1-\alpha-\epsilon, \epsilon, \alpha)$$

$$= \cancel{2} H\left(\frac{1-\alpha}{2}\right) - H(1-\alpha-\epsilon, \epsilon)\Big]$$

$$\therefore\ C = -(1-\alpha)\log\left(\frac{1-\alpha}{2}\right) + (1-\alpha-\epsilon)\log(1-\alpha-\epsilon) + \epsilon\log\epsilon$$

① BSC: $\alpha = 0$

$$C = -\log\frac{1}{2} + (1-\epsilon)\log(1-\epsilon) + \epsilon\log\epsilon = \log_2 2 - H(\epsilon)$$

$$= 1 - H(\epsilon)$$

② BEC: $\epsilon = 0$

$$C = -(1-\alpha)\log\left(\frac{1-\alpha}{2}\right) + (1-\alpha)\log(1-\alpha)$$

$$= -(1-\alpha)\log(1-\alpha) + (1-\alpha)\log 2 + (1-\alpha)\log(1-\alpha)$$

$$= (1-\alpha)$$

③ Noise less Binary channel: $\alpha = \epsilon = 0$

$$C = -(1)\log\left(\frac{1}{2}\right) + 1\log 1$$

$$C = \log 2$$

$$C = 1$$

$\therefore$ BSC : $C = 1 - H(\epsilon)$

BEC : $C = 1 - \alpha$

Noise less Binary channel : 1

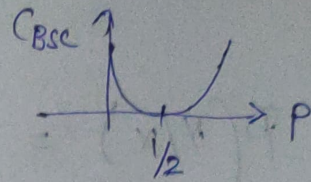
# Channel Capacity for Special cases

① **BSC : Binary Symmetric channel**

↳ when $\alpha = 0$

Binary entropy function

$$C_{BSC} = 1 - H(\epsilon)$$




- $\epsilon \rightarrow$ Cross over probability

② **BEC : Binary eraser channel**

↳ when $\epsilon = 0$

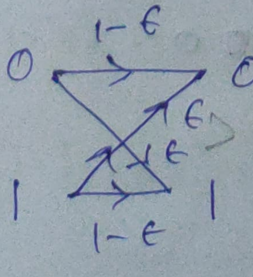
$$C_{BEC} = 1 - \alpha$$

③ **Noise less Binary channel** : when $\alpha = \epsilon = 0$

$$C = 1 \text{ bit / channel use}$$

## Symmetric Channel :

- If the rows of $P(Y/X)$ are permutations to each other and the columns are permutations to each other, then such channel is called as Symmetric Channel.




can have different values $\epsilon_1, \epsilon_2$

$$P(Y/X) = \begin{array}{c|cc} X \backslash Y & 0 & 1 \\ \hline 0 & 1-\epsilon & \epsilon \\ 1 & \epsilon & 1-\epsilon \end{array}$$

$$\Rightarrow C = \log 2 - H(\epsilon, 1-\epsilon) = 1 - H(\epsilon)$$

# Weakly Symmetric Channel

- Rows are permutations to each other and coloumns sums are equal.

Ex:

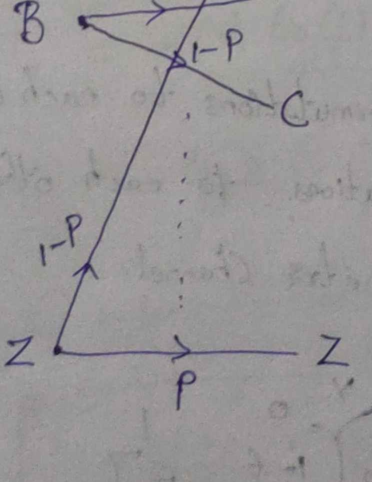
$$P(Y/X) = \begin{array}{c|ccc} X \backslash Y & & & \\ \hline & \left[\begin{matrix} 0.3 & 0.2 & 0.5 \\ 0.5 & 0.3 & 0.2 \\ 0.2 & 0.5 & 0.3 \end{matrix}\right] \end{array}$$

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

Ex: $C = \log 3 - H(0.3, 0.2, 0.5)$    $\cdot |Y|$ – size of the alphabet

18/01/2023

## Noisy type writer Channel




$$P(Y/X) = \begin{array}{c|ccccc} X \backslash Y & A & B & C & \cdots & Z \\ \hline A & p & 1-p & 0 & \cdots & 0 \\ B & 0 & p & 1-p & \cdots & 0 \\ C & 0 & 0 & p & \cdots & 0 \\ \vdots & & & & & \vdots \\ Z & 1-p & 0 & 0 & \cdots & p \end{array}$$

- Rows are permutations to eachother and Coloumns are permutations to eachother.

• It is both symmetric and weakly symmetric channel

$\therefore C = \log_2 2b - H(P, 1-P, \dots\dots)$

$= \log_2 2b - H(P)$

$$\boxed{\therefore C = \log_2(2b) - H(P)}$$

## Differential Entropy:

• For continuous R.V. (till now its for Discrete R.V's)

$$h(X) = -\int_{-\infty}^{\infty} f_X(x) \log_e f_X(x)\, dx \; ; \quad f_X(x) - \text{pdf of R.V "X"}$$

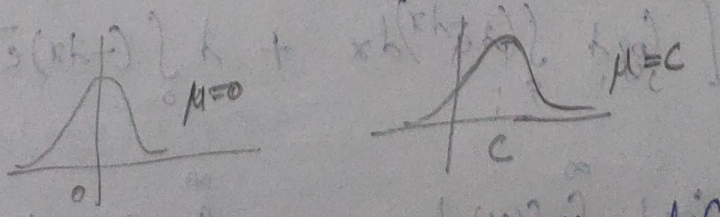
### Properties:

① $h(X_1, X_2, \dots, X_n) = \sum_{i=1}^{n} h\left(X_i / X_1, X_2, \dots, X_{i-1}\right)$

② $h(X+C) = h(X)$ ; C - constant

X

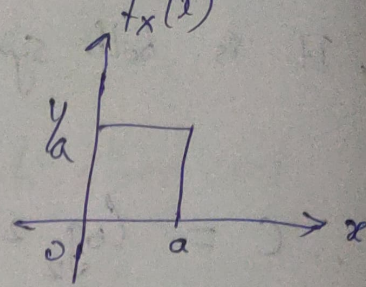
X+C → Translation




• The translation does not alter the differential entropy.

③ $h(aX) = h(X) + \log |a|$

Ex: $f(x) = \frac{1}{a} \; ; \; 0 \le x \le a$




Sol:

$$h(x) = -\int_0^a \frac{1}{a} \log\left(\frac{1}{a}\right) dx$$

Uniform Distribution

$$= \frac{1}{a} \log a \times (a)$$

$$= \log_e a \quad \text{natts}$$

$$\boxed{\therefore h(x) = \log_e a \text{ natts}}$$

Ex: $f_X(x) = \lambda e^{-\lambda x} \; ; \; x > 0$

Exponential distribution

Sol

$$h(x) = -\int_0^\infty f(x) \log_e\left(\lambda e^{-\lambda x}\right) dx$$

$$= -\int_0^\infty f(x) \left[\log_e \lambda - \lambda x\right] dx$$

$$= -\left[\log \lambda \int_0^\infty f(x)\, dx - \lambda \int_0^\infty x f(x)\, dx\right]$$

$$= -\left[\log \lambda \int_0^\infty (\lambda e^{-\lambda x}) dx + \lambda \int_0^\infty (-\lambda x) e^{-\lambda x} dx\right]$$

$$= -\left[\log \lambda \underbrace{\int_0^\infty f(x)\, dx}_{\text{Area under curve} = 1} - \lambda \underbrace{\int_0^\infty x f(x)\, dx}_{E(x) = \frac{1}{\lambda}}\right]$$

(mean/expectation)

$\Rightarrow h(X) = -\left[ \log \lambda - \lambda \times \frac{1}{\lambda} \right]$

$= 1 - \log \lambda$

$$\boxed{\therefore h(X) = (1 - \log_e \lambda) \text{ nats}}$$

Exponential Distribution

Ex: Gaussian Distribution: $X \sim N(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Sol: $$h(X) = -\int_{-\infty}^{\infty} f(x) \log\left( \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right) dx$$

$$= -\log \frac{1}{\sqrt{2\pi}\,\sigma} \underbrace{\int_{-\infty}^{\infty} f(x)\, dx}_{\text{area under curve} = 1} + \frac{1}{2\sigma^2} \underbrace{\int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx}_{\text{Variance}}$$

$$\boxed{\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx}$$

$= \log \sqrt{2\pi\sigma^2} + \frac{1}{2}(1)$

$= \frac{1}{2} \log 2\pi\sigma^2 + \frac{1}{2} \log_e e$

$= \frac{1}{2} \log 2\pi e \sigma^2$

Independent of $\mu$

$$\boxed{\therefore h(X) = \frac{1}{2} \log 2\pi e \sigma^2}$$

Gaussian Distribution

# Kullback-Leibler Divergence : (KL Divergence)

→ It is also called as Relative Entropy.

- Used to measure how similar or dissimilar 2 distributions are.
- Used as a measure of distance between two distributions.

→ The KL distance b/w two distributions $p(x)$ and $q(x)$

$$D(P\|Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)}$$ (for Discrete case) (PMF)

- If $P(x)$ & $Q(x)$ are similar then $D(P\|Q) = D(Q\|P) = 0$

$$D(P\|Q) = \int_{-\infty}^{\infty} P(x) \log\left(\frac{P(x)}{Q(x)}\right) dx$$ (for Continuous time R.V) (pdf)

---

Ex: Gaussian Distributions :

$$P(x) \Rightarrow N(\mu_1, \sigma_1^2) \quad \& \quad Q(x) - N(\mu_2, \sigma_2^2)$$

- $$D(P\|Q) = \int P(x) \log P(x) dx - \int P(x) \log Q(x)\, dx \quad \rightarrow (1)$$

W.K.T

$$\int P(x) \log P(x) = -\frac{1}{2} \log 2\pi\sigma_1^2 - \frac{1}{2} = -\frac{1}{2} \log 2\pi e\sigma_1^2$$

$$\Rightarrow \int p(x) \log p(x) = -\frac{1}{2}\left(\log(2\pi\sigma_1^2) + 1\right)$$

W.K.T

$$\int p(x) \log q(x) = \int_{-\infty}^{\infty} p(x) \log\left(\frac{1}{\sqrt{2\pi}\,\sigma_2} e^{-\frac{(x-\mu_2)^2}{2\sigma_2^2}}\right) dx$$

$$= \log \frac{1}{\sqrt{2\pi}\,\sigma_2} \int_{-\infty}^{\infty} p(x)\,dx - \frac{1}{2\sigma_2^2} \int_{-\infty}^{\infty} (x-\mu_2)^2 p(x)\,dx$$

$$= -\frac{1}{2} \log 2\pi\sigma_2^2 - \frac{1}{2\sigma_2^2}\left(\sigma_1^2 + (\mu_1-\mu_2)^2\right)$$

$$\Big[\because (x-\mu_2)^2 = (x-\mu_1+\mu_1-\mu_2)^2$$

$$= (x-\mu_1)^2 + (\mu_1-\mu_2)^2 + 2(\mu_1-\mu_2)(x-\mu_1)$$

$$\int_{-\infty}^{\infty} (x-\mu_2)^2 p(x)\,dx = \int_{-\infty}^{\infty} (x-\mu_1)^2 p(x)\,dx + (\mu_1-\mu_2)^2 \int_{-\infty}^{\infty} p(x)\,dx + 2(\mu_1-\mu_2) \int_{-\infty}^{\infty} (x-\mu_1)\,p(x)\,dx$$

$$= \sigma_1^2 + (\mu_1-\mu_2)^2 + 2(\mu_1-\mu_2)\left[\int_{-\infty}^{\infty} x\,p(x)\,dx - \mu_1 \int_{-\infty}^{\infty} p(x)\,dx\right]$$

$$= \sigma_1^2 + (\mu_1-\mu_2)^2 + 2(\mu_1-\mu_2)\left[\mu_1-\mu_1\right]$$

$$= \sigma_1^2 + (\mu_1-\mu_2)^2 \Big]$$

$$\therefore D(P\|Q) = -\frac{1}{2}\log(2\pi\sigma_1^2) - \frac{1}{2} - \left[-\frac{1}{2}\log 2\pi\sigma_2^2 - \frac{1}{2\sigma_2^2}\left(\sigma_1^2 + (\mu_1-\mu_2)^2\right)\right]$$

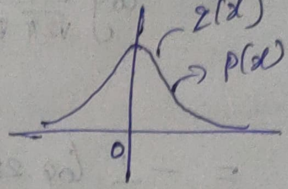
$$= \log\frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1-\mu_2)^2}{2\sigma_2^2} - \frac{1}{2}$$

- For '2' Guassian distributions $p(x)$ & $q(x)$

$$\boxed{D(p||q) = \log \frac{\sigma_2}{\sigma_1} + \frac{\sigma_1^2 + (\mu_1 - \mu_2)^2}{2\sigma_2^2} - \frac{1}{2}}$$

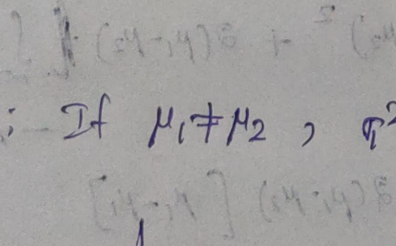
Case 1: If $\mu_1 = \mu_2$ & $\sigma_1 = \sigma_2$

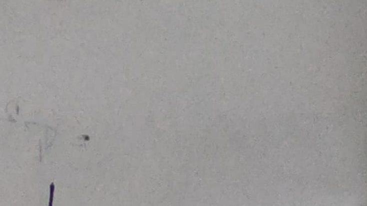
then, $D(p||q) = 0$



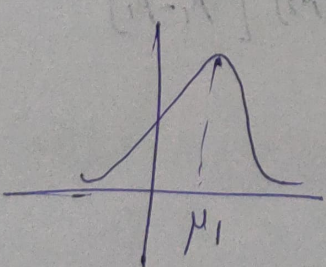

Case 2: If $\mu_1 \neq \mu_2$ & $\sigma_1^2 = \sigma_2^2$ (means distribution is same ⇒ width same)

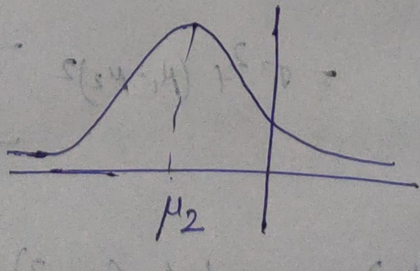
then, $D(p||q) = \dfrac{(\mu_1 - \mu_2)^2}{2\sigma_2^2}$ → minimum (compared to all other cases)




Case 3: If $\mu_1 \neq \mu_2$, $\sigma_1^2 \neq \sigma_2^2$



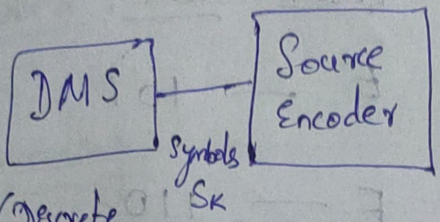




✻ In case 2, the distributions are much similar to each other than the distributions in case 3

19/01/2023

# Efficient representation of symbols : (for Data compression)

Alphabet "26" → 5 bits
characters required




(Discrete Memoryless Source)
(source produces symbols)

code:
A – 00000
B – 00001
C = 00010
⋮
Z = 11001 ← codewords

- The idea to go with this efficient representation of symbols is for data compression.
- Set of codewords is called "code"
- → A code is a set of codewords or vectors.
- Fixed Length code : codewords in code are having same length.
- Variable Length code (VLC) : codewords in code are having variable length.

Example:

Fixed Length code : (8 letters – 3 bits required for representation)

| Letter | Code | Letter | Code |
|---|---|---|---|
| A | 000 | F | 101 |
| B | 001 | G | 110 |
| C | 010 | H | 111 |
| D | 011 | | |
| E | 100 | | |

VLC

(VLC1)

## Variable length code (8 letters)

Code Book

| Letter | Code | | Letter | Code |
|---|---|---|---|---|
| A | 00 (probability of its occurence is high) | | E | 101 |
| B | 010 | | F | 110 |
| C | 011 | | G | 1110 |
| D | 100 | | H | 1111 (Probability of its occurence is low) |

Ex. Suppose, we want to code the series of letters:

A BAD CAB

→ Using Fixed Length Code:

000 001000011 010000001 – Total no. of bits = 21

Variable Length Code: (VLC1)

00 01000100 01100010 – Total no. of bits = 18

## VLC2

| | |
|---|---|
| A – 0 | E – 10 |
| B – 1 | F – ~~000~~ 11 |
| C – 00 | G – 000 |
| D – 01 | H – 111 |

Code Book

A BAD CAB

0 1001 0001 — 9 bits

VLC2 :

Decoding : (At receiver end)

VLC2 : [0][1][0][0][1][0][0][0][1]

A B A A B A A A B

• VLC2 not uniquely decodable.

VLC1 : [0 0][0 1 0][0 0][1 0 0][0 1 1][0 0][0 1 0]

A B A D C A B

• $VLC_1$ is uniquely decodable

• $VLC_1$ — called as a prefix code.

Prefix code :

→ A prefix code is one in which no codeword forms the prefix of any other code word.

→ Also called as Instantaneous Code

Kraft inequality :

→ A necessary and sufficient condition for ~~existing~~ existence of a binary code with codewords having lengths $n_1 \le n_2 \le \ldots \le n_L$ that satisfies the prefix condition is

$\boxed{\text{Prefix Condition}}$ $$\sum_{k=1}^{L} 2^{-n_k} \le 1$$ ; $n_k$ – length of $k^{th}$ codeword

Also; $n_1 \le n_2 \le n_3 \le \cdots \le n_L$

· for VCL$_3$ : A – 00 → $n_1 = 2$

B = 010 → $n_2 = 3$

C – 011 → $n_3 = 3$

D – 100 → $n_4 = 3$

E – 101 → $n_5 = 3$

F – 110 → $n_6 = 3$

G – 1110 → $n_7 = 4$

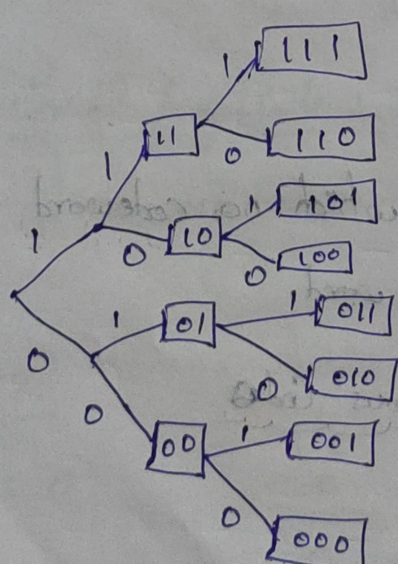
H – 1111 → $n_8 = 4$

## Prefix code for non-binary case :

$\boxed{\text{Prefix Condition}}$ $$\sum_{k=1}^{L} M^{-n_k} \le 1$$ ; M – Alphabet size

→ for binary case ; M=2

## Construction of Binary Prefix code :




VLC for 4 Letters : (Prefix Code) ✓

| | | Length of codeword |
|---|---|---|
| A | 0 | 1 |
| B | 10 | 2 |
| C | 110 | 3 |
| D | 111 | 3 |

## Kraft inequality

$$\sum_{k=1}^{4} 2^{-n_k} = 2^{-1} + 2^{-2} + 2^{-3} + 2^{-3}$$

$$= 0.5 + 0.25 + 0.125 + 0.125$$

$$= 1$$

. Condition satisfied.

## Source Coding Theorem:

→ Let X be the symbol of the letters from a DMS with a finite entropy $H(X)$ and the output symbols are $x_k$; $k = 1, 2, \ldots, L$ occuring with probability $P(x_k)$; $k = 1, 2, \ldots L$.

→ It is possible to construct a code that satisfies the prefix condition and has an average codeword length that satisfies the inequality:

$$\boxed{H(X) \leq \bar{R} < H(X) + 1}$$

(Condition for code to be uniquely decodable)

→ Here, $\bar{R} = \sum_{k=1}^{L} P_k n_k$ — Average codeword length

$= \sum_{k=1}^{L} P(x_k) n_k$ — Average no. of bits required to represent each symbol.

→ This theorem tells us the minimum no. of bits required to represent the source symbols on an average must be atleast equal to the entropy of the source.

20/01/2022

## Source coding Theorem:

**Proof:** (Lower bound)

$$H(X) - \bar{R} = -\sum_k P(x_k) \log P(x_k) - \sum_k P(x_k) n_k$$

$$= \sum_k P(x_k) \log \frac{2^{-n_k}}{P(x_k)} \rightarrow (1)$$

$$= \sum_k P(x_k) \ln \frac{2^{-n_k}}{P(x_k)} \log_2 e$$

$$\boxed{\log_2 x = \ln_e x \log_2 e}$$

$$\le \log_2 e \sum_k P(x_k) \left[ \frac{2^{-n_k}}{P(x_k)} - 1 \right]$$

$$\boxed{\ln_e x \le x - 1}$$

$$\le \log_2 e \sum_k \left[ 2^{-n_k} - \frac{P(x_k)}{\cancel{P(x_k)}} \right]$$

$$\le \log_2 e \left( \sum_k 2^{-n_k} - \sum_k P(x_k) \right)$$

$$\le \log_2 e \left( \sum_k 2^{-n_k} - 1 \right)$$ (By kraft inequality)

$$\le 0 \qquad \left\{ \sum_k 2^{-n_k} \le 1 \right\}$$

$$\therefore H(X) - \bar{R} \le 0$$

$$\boxed{H(X) \le \bar{R}}$$

Hence, proved

(Upper bound)

Proof:

$$2^{-n_k} \leq P(x_k) < 2^{-n_k+1}$$

$$2^{-n_k} \leq P(x_k)$$

$$\sum_k 2^{-n_k} \leq \sum P(x_k)$$

$$\sum_k 2^{-n_k} \leq 1$$

Applying '$\log_2$' on both sides

$$\log P(x_k) < (-n_k+1)$$

$$\Rightarrow n_k < 1 - \log P(x_k)$$

Multiply by $P(x_k)$ on B.S.

$$P(x_k)\, n_k < P(x_k) - P(x_k) \log P(x_k)$$

$$\Rightarrow \sum_k n_k P(x_k) < \sum_k P(x_k) - \sum_k P(x_k) \log P(x_k)$$

$$\Rightarrow \bar{R} < 1 + H(X)$$

$$\boxed{\therefore \bar{R} < H(X) + 1}$$ Hence, proved.

---

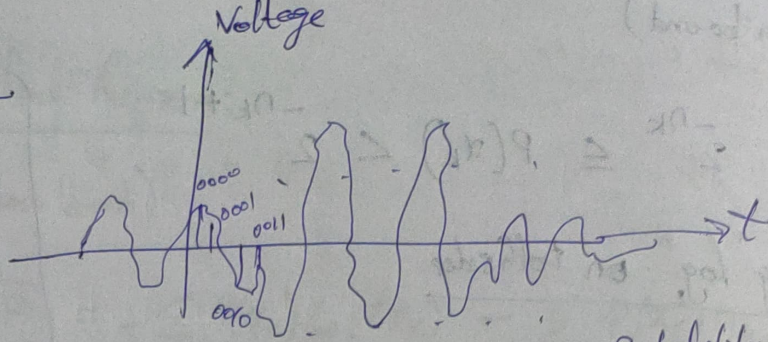
→ The efficiency of a prefix code is given by

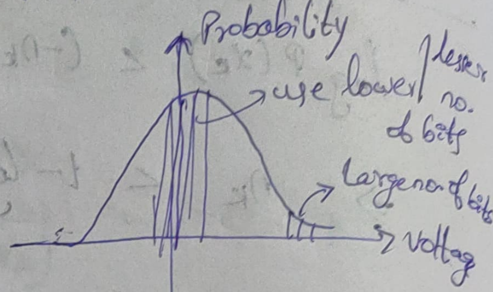
$$\boxed{\eta = \frac{H(X)}{\bar{R}}} \qquad \boxed{\eta \leq 1}$$

- How good the data compression can be quantified by this.

→ Applications: Speech, Text, Image, Video etc.

Speech signal




Example 1:

Consider a source X that generates 4 symbols with probabilities 0.5, 0.3, 0.1, 0.1. Compute the efficiency of the code

Sol: $\eta = \frac{H(X)}{\bar{R}}$

• $H(X) = -\sum_{k=1}^{4} P(x_k) \log P(x_k)$

$= -0.5 \log_2 0.5 - 0.3 \log_2 0.3 - 0.1 \log_2 0.1 - 0.1 \log_2 0.1$

$= 1.685$ bits

• $\bar{R} = \sum_{k=1}^{4} P(x_k)\, n_k$

• Prefix code words from Binary Tree: $\{0, 10, 110, 111\}$

$\downarrow$ 0.5 $\quad\downarrow$ 0.3 $\quad\downarrow$ 0.1 $\quad\downarrow$ 0.1

$\Rightarrow \bar{R} = 0.5 \times 1 + 0.3 \times 2 + 0.1 \times 3 + 0.1 \times 3$

$= 0.5 + 0.6 + 0.3 + 0.3$

$= 1.7$ bits

$\therefore \eta = \frac{H(X)}{\bar{R}} = \frac{1.685}{1.7} = 0.9912$

$\boxed{\therefore \text{Efficiency of the code } (\eta) = 0.9912}$

2Ex: Source X generates 4 symbols with probabilities $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}$. Find out the efficiency of the code.

Sol: $\left(\frac{1}{2}\right) \rightarrow \{0\}$ $\left(\frac{1}{4}\right) \rightarrow \{10\}$ $\left(\frac{1}{8}\right) \rightarrow \{110\}$ $\left(\frac{1}{8}\right) \rightarrow \{111\}$

$H(X) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + \frac{1}{8}\log 8 + \frac{1}{8}\log 8$

$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = 1 + \frac{3}{4} = \frac{7}{4} = 1.75$ bits

$\bar{R} = \frac{1}{2} \times 1 + \frac{1}{4} \times 2 + \frac{1}{8} \times 3 + \frac{1}{8} \times 3$

$= \frac{1}{2} + \frac{1}{2} + \frac{3}{8} + \frac{3}{8} = 1 + \frac{3}{4} = \frac{7}{4} = 1.75$ bits

$\boxed{\therefore \eta = 1}$

Due to
→ Inequalities of assignment of no of bits in Example 1
$\eta = 1$ is not achieved.

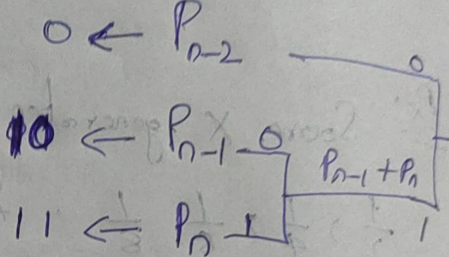
# Huffman Coding

- Variable Length coding
- Author : Huffman, 1952
- follows Prefix code and satisfies Kraft Inequality.

- It is not unique for a given set of probabilities

**Procedure** : ① Write the probabilities in descending order

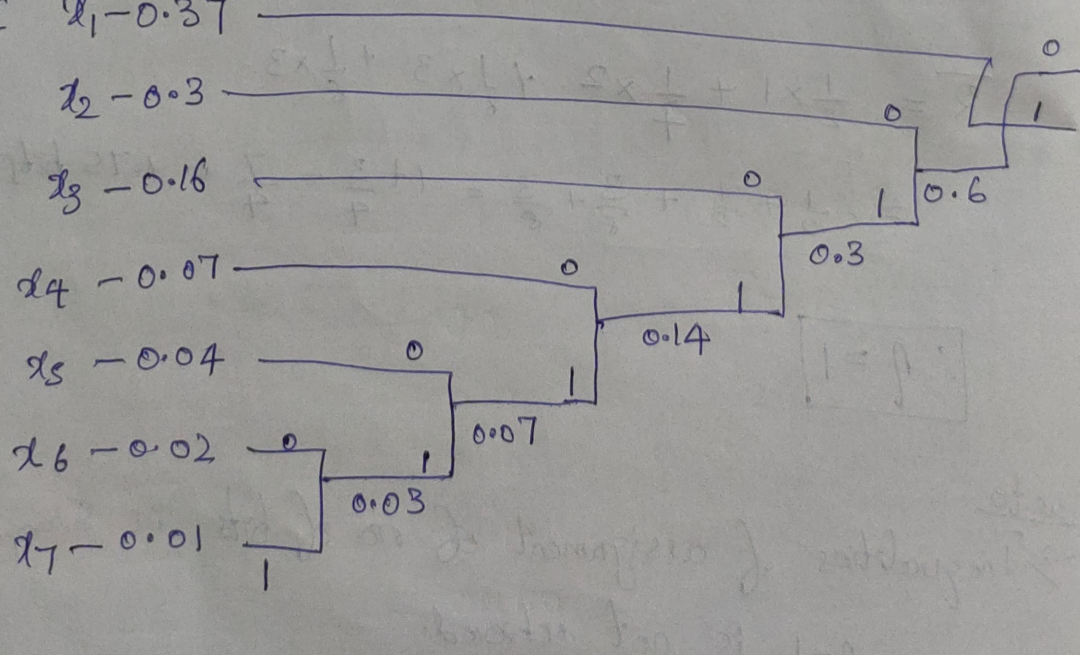
· Write codeword from right to left




---

**Ex:** A source generates 7 symbols with probabilities $x_1 - 0.37$, $x_2 - 0.3$, $x_3 - 0.16$, $x_4 - 0.07$, $x_5 - 0.04$, $x_6 - 0.02$, $x_7 - 0.01$. Compute efficiency.

**Sol:**



Ans

$x_1 - 1$ · Follows prefix code.

$x_2 - 00$

$x_3 - 010$ → $\eta = 1$ not achieved

$x_4 - 0110$ → As $\left.\begin{matrix} x_6 - 0.02 \\ x_7 - 0.01 \end{matrix}\right\}$ but same no. of bits assigned.

$x_5 - 01110$

$x_6 -$ ~~[illegible]~~ $011110$

$x_7 - 011111$ → 2 symbols with unequal probabilities assigned same no. of bits.

· $H(x) = 2.1152$ bits

· $\overline{R} = 1 \times 0.37 + 2 \times 0.3 + 3 \times 0.16 + 4 \times 0.07 + 5 \times 0.04 + 6 \times 0.02 + 6 \times 0.01$

$= 2.17$ bits

· $\eta = 0.9747$

$\boxed{\therefore \text{Efficiency} = 0.9747}$

---

Ex:

$1 \leftarrow x_1 - 0.46$

$00 \leftarrow x_2 - 0.3$

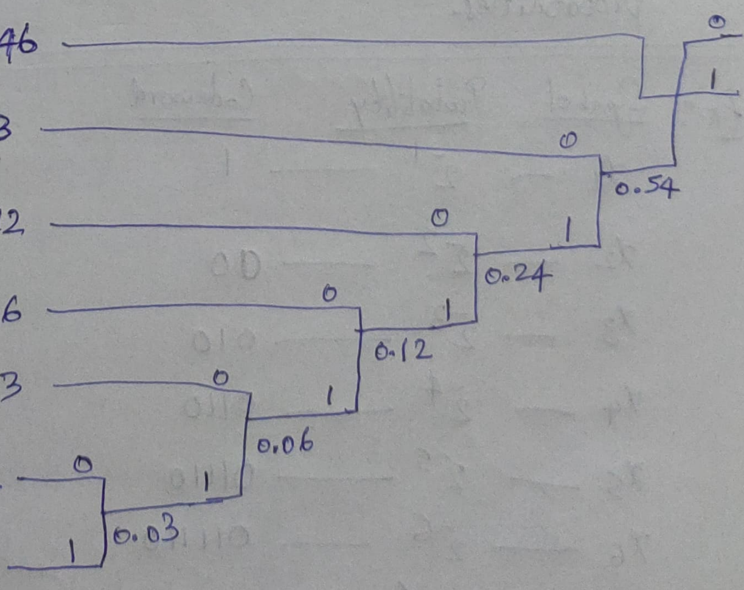
$010 \leftarrow x_3 - 0.12$

$0110 \leftarrow x_4 - 0.06$

$01110 \leftarrow x_5 - 0.03$

$011110 \leftarrow x_6 - 0.02$

$011111 \leftarrow x_7 - 0.01$



| Symbol | Probability | Codeword 1 | Codeword 2 |
| --- | --- | --- | --- |
| $x_1$ | 0.46 | 1 | 1 |
| $x_2$ | 0.3 | 00 | 00 |
| $x_3$ | 0.12 | 010 | 011 |
| $x_4$ | 0.06 | 0110 | 0101 |
| $x_5$ | 0.03 | 01110 | 01001 |
| $x_6$ | 0.02 | 011110 | 010000 |
| $x_7$ | 0.01 | 011111 | 010001 |

- In both cases, $H(X) = 1.9781$ bits

$\overline{R} = 1.99$ bits

- Efficiency $(\eta) = 0.9940$

---

# Huffman coding is not unique for a given set of probabilities.

---

Ex: 

| Symbol | Probability | Codeword |
| --- | --- | --- |
| $x_1$ | $2^{-1}$ | 1 |
| $x_2$ | $2^{-2}$ | 00 |
| $x_3$ | $2^{-3}$ | 010 |
| $x_4$ | $2^{-4}$ | 0110 |
| $x_5$ | $2^{-5}$ | 01110 |
| $x_6$ | $2^{-6}$ | 011110 |
| $x_7$ | $2^{-6}$ | 011111 |

$H(X) = 1.9688$ bits

$\overline{R} = 1.9688$ bits

$\eta = 1$

$x_6$, $x_7$: Same probability $\rightarrow$ same no. of bits

→ This kind of distribution is called D-adic distribution.

→ Probabilities are $D^{-n}$

- Here, $D = 2$

---

## Working Blockwise :

$$B\,H(x) \le \overline{R_B} < B\,H(x) + 1$$ ; B – Length of the Block.

$$H(x) \le \frac{\overline{R_B}}{B} < H(x) + \frac{1}{B}$$

$$H(x) \le \overline{R} < H(x) + \frac{1}{B} \quad ; \quad \overline{R} = \frac{\overline{R_B}}{B}$$

- $B \uparrow \Rightarrow \overline{R} \to H(x) \Rightarrow \eta \to 1$

---

Ex:

| Symbol | Probability | Codeword |
|---|---|---|
| $x_1$ | 0.4 | 1 |
| $x_2$ | 0.35 | 00 |
| $x_3$ | 0.25 | 01 |

$H(x) = 1.5589$ bits

$\overline{R} = 1.6$ bits

$\eta = 0.9743$

Working blockwise:

| Symbol Pairs | Probability | Codeword |
|---|---|---|
| $x_1 x_1$ | 0.16 | 10 |
| $x_1 x_2$ | 0.14 | 001 |
| $x_2 x_1$ | 0.14 | 010 |
| $x_2 x_2$ | 0.1225 | 011 |
| $x_1 x_3$ | 0.1 | 111 |
| $x_3 x_1$ | 0.1 | 0000 |

| Symbol | Probability | Codeword |
|---|---|---|
| $x_2 x_3$ | 0.0875 | 0001 |
| $x_3 x_2$ | 0.0875 | 1100 |
| $x_3 x_3$ | 0.0625 | 1101 |

- Entropy: $2H(x) = 3.1177$ bits

  $H(x) = 1.5589$ bits

- $2\bar{R} = 3.1775$ bits

  $\bar{R} = 1.5888$ bits

  $\eta = 0.9812$

- The new '$\eta$' with block ($\eta = 0.9812$) is greater than the previous $\eta$ ($\eta = 0.9743$)

# Complexity of encoding increased, with $\eta \uparrow$

---

25/01/2023

## Noisy-channel coding Theorem:

Source Encoder — Channel Encoder — Modulator — Channel — Decoder

(compress the no. of bits required to represent each symbol)

(Adds the bits-called as redundancy. required for error detection and correction)

1    101

$\boxed{1011\cdots}$ $\leftarrow k \rightarrow$

$\boxed{1|1|0\cdots}$ $\leftarrow n \rightarrow$

$k < n$

- Coderate : $R = \frac{k}{n}$ ; $k < n$

As long as, $R \leq C$ ; C - channel capacity, we can get/achieve reliable communication with low probability of error.

R - Rate of Transmission

- Shanan theorem :

→ This theorem states that given a Noisy channel with channel capacity C and information transmitted at a rate R, then if $R \leq C$ there exists codes that allow the probability of error at the receiver to be made arbitrarily small.

→ If $R > C$, an arbitrarily small probability of error is not achievable.

Example : Repetation code

1 → [Channel encoder] → $\underbrace{11 \ldots 1}_{n \text{ times}}$ (channel)

At Decoder

0 ← 000...0

0 ← 000...01

0 ← 000...10

⋮

1 ← 111...10

1 ← 111...11

($2^n$ codewords)

- Coderate : $R = \frac{1}{n}$

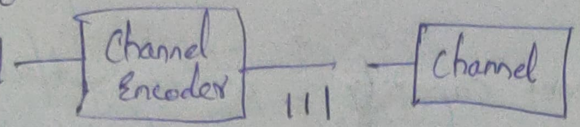
(Majority logic)

- if more no. of zeroes → '0' is considered
- if more no. of ones → '1' is considered

# 3-bit Repetation Code

$X=0$

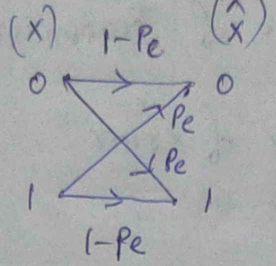
$X=1$




(Appreal probabilities)

$P(x=0) = P(x=1) = \frac{1}{2}$

$(P_e = 10^{-2})$




$C = 1 - H(P_e)$

Decoding $(\hat{x})$

| | |
|---|---|
| 000 | 0 |
| 001 | 0 |
| 010 | 0 |
| 011 | 1 |
| 100 | 0 |
| 101 | 1 |
| 110 | 1 |
| 111 | 1 |

(statistically independent)

$\rightarrow P_{e_{avg}} = P(\hat{x}=0/x=1)\, P(x=1) + P(\hat{x}=1/x=0)\, P(x=0)$

Average - Probability of error

- $P(\hat{x}=1/x=0) \longrightarrow$ 011, 101, 110, 111

$0 \xrightarrow{1-P_e} 0$

$0 \xrightarrow[P_e]{} 1$

$= 3(1-P_e)\, P_e^2 + P_e^3$

- $P(\hat{x}=0/x=1) \Longrightarrow$ 000, 001, 010, 100

- Probability of receiving '0' when '1' transmitted $= P_e$

$= P_e^3 + 3(1-P_e)\, P_e^2$

$\therefore P_{e_{avg}} = \frac{1}{2} \times 2 \left[ P_e^3 + 3(1-P_e) P_e^2 \right]$

$= P_e^3 + 3(1-P_e) P_e^2$

$$\boxed{P_{e_{avg}} = P_e^3 + 3(1-P_e)P_e^2}$$ ; for 3-bit repetation code

$$\# \boxed{P_{e_{avg}} = \sum_{i=\lceil \frac{n}{2} \rceil}^{n} \binom{n}{i} P_e^{i} (1-P_e)^{n-i}}$$ ; for n-bit repetation code

- $P_{e_{avg}} = \sum_{i=\lceil \frac{n}{2} \rceil}^{n} {}^{n}C_i \, P_e^{i} (1-P_e)^{n-i}$ ?

| Code rate (R) | Avg. Probability of Error ($P_e$) | |
|---|---|---|
| 1 | 0.01 | $\rightarrow C = 0.919$ ($R > C$) |
| $\frac{1}{3}$ | $3 \times 10^{-4}$ | $\rightarrow R < C$ |
| $\frac{1}{5}$ | $10^{-6}$ | $\rightarrow R < C$ |
| $\frac{1}{7}$ | $4 \times 10^{-7}$ | $\rightarrow R < C$ |
| $\frac{1}{9}$ | $10^{-8}$ | $\rightarrow R < C$ |
| $\frac{1}{11}$ | $5 \times 10^{-10}$ | $\rightarrow R < C$ |

- $C = 1 - H(P_e)$

$= 1 - H(10^{-2})$

$= 0.919$

# Channel Capacity of Gaussian Channel

$$X \longrightarrow \oplus \longrightarrow Y = X + Z$$

$$\uparrow Z$$

Assume, $X$ & $Z$ are independent

Noise $\sim \mathcal{N}(0, \sigma^2)$

$\sigma^2$ – Noise power

- $C = \max\limits_{f_X(x),\, E[X^2] \le P} I(X; Y)$ ; $P$ – Avg. Power of input signal
  – Avg. Transmitted power of $X$

  $P$ – Transmitted Power

- $I(X;Y) = h(Y) - h(Y/X)$

$$= h(Y) - h(X+Z/X) \quad \text{(As uncertainity lies only in } Z\text{)}$$

$$= h(Y) - h(Z) \quad ; \; h(Z) = \frac{1}{2}\ln(2\pi e \sigma^2)$$

(Gaussian Distribution)

; $Y = X + Z$

$$E[Y] = E[X] + E[Z] = 0 + 0 = 0$$

$$E[Y^2] = E[X^2] + E[Z^2] + 2E[XZ]$$

$$= P + \sigma^2 + 0 \quad ; \text{ as } X \,\&\, Z \text{ are independent and } E[X^2] \le P$$

$$= P + \sigma^2$$

- Maximum capacity achieved, when $X$ is considered to have a Gaussian distribution.

- $I(X;Y)$ can be maximized when $X \sim \mathcal{N}(0, P)$

$$\therefore I(x;y) \le \frac{1}{2}\log_2\left(2\pi e(P+\sigma^2)\right) - \frac{1}{2}\log_2(2\pi e\sigma^2)$$

27/03/2023

$$\Rightarrow I(x;y) \le \frac{1}{2}\log_2\left(1+\frac{P}{\sigma^2}\right)$$

$\rightarrow$ Channel Capacity: $C = \frac{1}{2}\log_2\left(1+\frac{P}{\sigma^2}\right)$ bits/channel wise

- In Continuous time: $y(t) = x(t) + z(t)$

In Discrete time $\Rightarrow y(nT_s) = x(nT_s) + z(nT_s)$

Band Limited Signals: Bandwidth $= W$

$T_s = \frac{1}{2W}$ ; $f_s = 2W$ samples per second

$$\therefore C = \left[\frac{1}{2}\log\left(1+\frac{P}{\sigma^2}\right)\right] 2W \text{ bits/sec}$$

$$= W\log\left(1+\frac{P}{\sigma^2}\right)$$

- $\sigma^2$ = Noise power spectral density $\times$ Bandwidth

$= \frac{N_0}{2} \times 2W$ ($\sigma^2 \Rightarrow$ Noise Power)

$= N_0 W$

$\Rightarrow \sigma^2 = N_0 W$

27/01/2023

$$\Rightarrow C = \omega \log_2\left(1+\frac{P}{\sigma^2}\right)$$

$$C = \omega \log\left(1+\frac{P}{N_o \omega}\right)$$ → Information Capacity Theorem

- P - Avg. transmitted Power : $P = E_b R_b = E_b C$

$E_b$ - Energy per bit

$R_b$ - Bit rate

$$\Rightarrow C = \omega \log\left(1+\frac{E_b C}{N_o \omega}\right) = \omega \log\left(1+\frac{E_b}{N_o}\cdot\frac{C}{\omega}\right)$$

$$\Rightarrow C = \omega \log\left(1+\frac{E_b}{N_o}\cdot\frac{C}{\omega}\right)$$

$$\Rightarrow \frac{E_b}{N_o} = \frac{2^{C/\omega}-1}{C/\omega}$$

$$\left.\frac{E_b}{N_o}\right|_{\omega\to\infty} = \ln 2 = 0.693 = -1.6\,dB$$

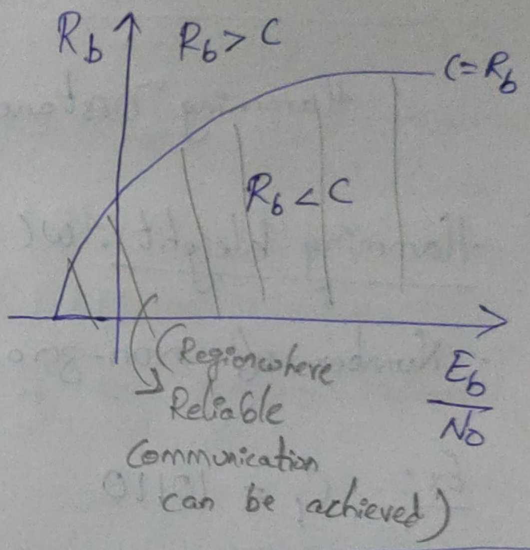
- $\omega \to \infty$ means for a very larger Bandwidth

→ For a very larger B.W reliable communication is possible if signal power < Noise power.

$$C = w\log\left(1+\frac{P}{N_0 w}\right)$$

$$\Rightarrow C\Big|_{w\to\infty} = \frac{P}{N_0}\log_2 e = 1.44\,\frac{P}{N_0}$$

- $C \rightarrow$ goes linearly with transmitted power ($P$) for a larger bandwidth.




## Error Correcting Codes (ECC)

- ECC are used for correcting errors when, messages are transmitted over a noisy channel.
  when, Stored data is retrieved.

- Control Coding or channel coding.

**Codeword**: Sequence of Symbols

**Code**: Set of codewords

$$C = \underbrace{\{0000, \underbrace{0101}_{\text{codeword}}, 1010, 1111\}}_{\text{code}}$$

## Hamming distance : d( )

- The number of places codewords differ.

Ex: $C_1 = 10110$

$C_2 = 11011$

$C_1 - C_2 = 01101$ ; ($C_1 - C_2$ → means places where they differ)

∴ Hamming Distance = 3 ⇒ $d(C_1, C_2) = 3$

## Hamming Weight : w( )

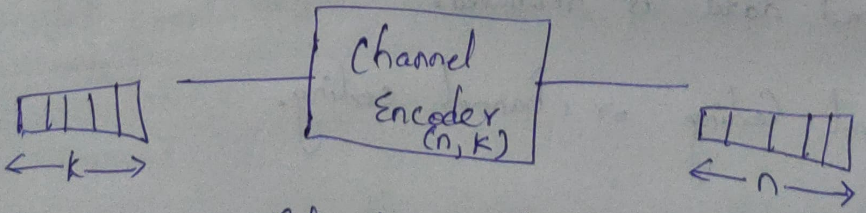
- Number of non-zero elements in the codeword.

Ex: $C_1 = 10110$

$w(C_1) = 3$

# $\boxed{d(C_1, C_2) = w(C_1 - C_2)}$

## Block Code:




(k < n)

Code rate: $r = \frac{k}{n}$

Ex: $S \in \{0,1\}$

$k = 2$

| | | | |
|---|---|---|---|
| 00 | | | 00000 |
| 01 | mapped to | | 10100 |
| 10 | 4 Codewords | | 11110 |
| 11 | | | 11001 |

## Block code :

- It consists of set of fixed length codewords.

## Minimum Hamming Distance : $d^*(\ )$

- The minimum distance between any two codewords.

$$d^* = d(C_i, C_j) \quad ; \text{where } i \neq j$$

Ex: $C = \{C_1, C_2, C_3, C_4\}$

$\min[d(C_1, C_2)\ d(C_1, C_3)\ d(C_1, C_4)\ d(C_2, C_4)\ d(C_3, C_4)]$

## Minimum Hamming weight: $\omega^*(\ )$

- The smallest/lowest weight of any non-zero codeword in a code.

## Linear Block code :

- A Linear Block code has following properties:

(1) Sum of two codewords belonging to the code is also a codeword belonging to the code.

(2) The all-zero code word is always a codeword.

3) The minimum Hamming distance b/w 2 codewords of a linear code is equal to the minimum weight of any non-zero codeword i.e. $d^* = \omega^*$

(4) The presence of all-zero codeword is a necessary not a sufficient condition for Linearity.

# Linear Block code :

Ex : $C = \{0000, 1010, 0101, 1111\}$ ← Code consisting of 4 codewords

- Summing means Ex-OR (exclusively OR)

00 → 0
01 → 1
10 → 1
11 → 0

0000 + 1010 → 1010 ∈ C

1010 + 0101 → 1111 ∈ C

0101 + 1111 → 1010 ∈ C

1010 + 1111 → 0101 ∈ C

1111 + 0000 → 1111 ∈ C

- Presence of all-zero codeword

- $w^* = 2$ ↳ (minimum weight of any non-zero codeword) . $d^* = 2$ i.e. $w^* = d^*$

- As it satisfied all properities of Linear code, given code 'c' is a $\underbrace{\text{Linear Block code}}$

(Field ? Field Theory Sigma Field)

## Galois Field (GF) :

- Field consists of set of elements

→ A field 'F' is a set of elements with 2 operations addition & multiplication satisfies following properties :
(+) (·)

## Galois Field :

### Properties :

① $a, b \in F \Rightarrow a+b, a \cdot b \in F$

② **Commutative Law :** $a+b = b+a$ where $a, b \in F$

$a \cdot b = b \cdot a$

③ **Associative Law :** $(a+b)+c = a+(b+c)$ where $a, b, c \in F$

$(a \cdot b)c = a \cdot (b \cdot c)$

④ **Distributive Law :** $a \cdot (b+c) = a \cdot b + b \cdot c$

⑤ Further, identity elements : 0 and 1 must exist in F satisfying :

(a) $a + 0 = a$ (b) $a \cdot 1 = a$

(c) If $a \in F$, there exists an additive inverse $(-a)$ such that

$a + (-a) = 0$

(d) If $a \in F$, there exists a multiplicative inverse $\left(\frac{1}{a}\right) / (a^{-1})$ such that $a \cdot \left(\frac{1}{a}\right) = 1$ (or) $a \cdot a^{-1} = 1$