→ A channel can atmost make 1 error, else it is a bad channel.

16/09/2025

Capacity Channel

X — Random variable → $\hat{X}$

Source of Messages → Source Encoder → Channel coder → Channel → Channel coder → Source Decoder → Message

A
B
C
D

→ Adding Redundancy

$\underbrace{H(X) \leq L \leq H(X)+1}_{\text{Information Theory}}$ $\quad$ $\underbrace{\qquad\qquad\qquad}_{\text{Coding Theory}}$

→ Generally $X \neq \hat{X}$ (noises are included)

Goal :- $P(X \neq \hat{X}) \to 0$ (reduce prob. of error)

- We need our message to be recieved as it was produced (i.e. $X = \hat{X}$, but this is not practically possible) so we minimize the prob. of error.

- For a channel $\boxed{p \leq 10^{-6} \text{ or } 10^{-5}}$

Bits we need to send → ① Yes = 1
No = 0

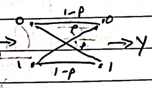
K (no. of bits) = 1 $\xrightarrow{\text{Adding Redundancy}}$ $n=3$ (2 extra bits) → Redundancy (helpful while decoding)
111
000

How much redundancy is tolerable? → r (Data rate)

$\boxed{\text{Rate} = \frac{K}{n}}$ $\Rightarrow$ rate $= \frac{1}{3}$ for ①

Shanon's First Theorem $\Rightarrow$ source coding Theorem

$H(X) \le L < H(X)+1$

$\boxed{\text{Channel Capacity } (C) \triangleq \max_{P_X(x)} I(X;Y)}$ $X \to$  $\to Y$

Channel Encoding Theorem (Shanon's 2nd Theorem)

$\boxed{R \le C} \Rightarrow P_e \to 0$

If the rate is less than channel capacity then it will always be perfectly decodable.

$$C = \max_{P_X(x)} I(X;Y)$$

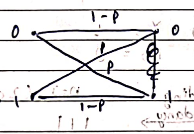
$$I(X;Y) = \sum_x \sum_y p(x,y) \log \frac{P(x,y)}{P(x)P(y)}$$

$$= P(0,0) \log \frac{P(0,0)}{P_x(0)P_y(0)} + P(0,1) \log \frac{P(0,1)}{P_x(0)P_y(1)} + P(1,0) \log \frac{P(1,0)}{P_x(1)P_y(0)}$$

$$+ P(1,1) \log \frac{P(1,1)}{P(1)P_y(1)}$$

$P_x(0)$ $= q$

$P_x(1)$ $= 1-q$



$\longrightarrow$ $\begin{array}{c|cc} & 0 & 1 \\ \hline 0 & 1-p & p \\ 1 & p & 1-p \end{array}$ = Channel Transition Matrix

$$P_Y(0) = P_x(0)P(0|0) + P_x(1)P(0|1) = P_x(0)(1-p) + P_x(1)p$$
$$= q(1-p) + (1-q)p$$
$$= p+q-2pq = r$$

$$P_Y(1) = P_x(0)P(1|0) + P_x(1)P(1|1)$$
$$\cancel{q(1-p)}\ qp + (1-q)(1-p) = 1-r$$

$$P(x,y) = P(x)\,P(y|x)$$

$P(0,0) = P_x(0)P(0|0) = q(1-p)$
$P(1,0) = P_x(1)\,P(0|1) = qp$
$P(0,1) = P_x(0)\,P(1|0) = qp$
$P(1,1) = P_x(1)\,P(1|1) = q(1-p)$

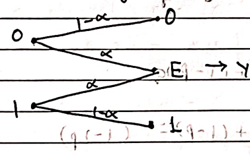
Put all these values in $I(X;Y)$

After simplifying

$\boxed{I(X;Y) = H(Y) - H(p)}$, where $r = p+q-pq$

$H(p) = -[p\log p + (1-p)\log(1-p)]$

Practice :- Binary Eraser Channel (BEC)



$X \to$ … $\to Y$ $C = ?$ (Answer $= \alpha$)

Date - 17/09/2025

$$\Rightarrow \left(q(1-p) \log \frac{q(1-p)}{q[\underbrace{q(1-p)+(1-q)p}_{r}]} + qp \log \frac{qp}{q[\underbrace{qp+(1-p)(1-q)}_{1-r}]} + \right.$$

$$qp \log \frac{qp}{\cancel{[\ldots]}\ [(1-q)[q(1-p)+(1-q)p]]} +$$

$$C = \max_{P_x(x)} I(X,y)$$

$$I(x,y) = H(r) - H(p)$$

$\Rightarrow$ let $f(q) = H(r) - H(p)$ as $(H(r) - H(p))$ is depending only on q not p

$$f'(q) = H'(r) \frac{dr}{dq} = 0$$

But $H(r) = -r\log r - (1-r)\log(1-r)$

$$H'(r) = -\cancel{r}\frac{1}{\cancel{r}} - \log r + \frac{\cancel{(1-r)}\,1}{\cancel{(1-r)}} + \log(1-r)$$

$$H'(r) = \cancel{-1} - \log r \cancel{+1} + \log(1-r)$$

$$H'(r) = \log\left(\frac{1-r}{r}\right)$$

$$r = p(1-q) + (1-p)q$$

$$\frac{dr}{dq} = -p + (1-p) = (1-2p)$$

We have $f'(q) = H'(r)\frac{dr}{dq} = 0$

$$\log\left(\frac{1-r}{r}\right)(1-2p) = 0$$

$\therefore \log\left(\frac{1-r}{r}\right) = 0$

$\because p$ can't be $1/2$ as it will be a bad channel in such case.

$$\left(\frac{1-r}{r}\right) = 1$$

$$\Rightarrow r = \frac{1}{2}$$

$$\Rightarrow q(1-p) + (1-q)p = \frac{1}{2}$$

$$\Rightarrow q + p - 2pq = \frac{1}{2}$$

$$\Rightarrow q(1-2p) = \frac{1}{2} - p$$

$$\Rightarrow q(\cancel{1-2p}) = \frac{\cancel{1-2p}}{2}$$

$$\Rightarrow \boxed{q = \frac{1}{2}}$$

$$\therefore I(X,Y) = H(r) - H(p)$$

$$\Rightarrow \max_{P_x(x)} I(X,Y) = \max_q \{H(r) - H(p)\}$$

$$= H(\tfrac{1}{2}) - H(p) \text{ at } q = \frac{1}{2}$$

$$\therefore C = H\left(\frac{1}{2}\right) - H(p) \text{ at } q = \frac{1}{2}$$

$\boxed{C = 1 - H(p)}$ for binary symmetric channel.

$I(X,Y) \Rightarrow H(Y) - H(Y|X)$

$H(X) - H(X|Y)$

Channel Coding Theorem

If the rate R of a channel is less than the capacity C, then a code such that $P_e \to 0$ converse also true.

$$\boxed{R < C \Rightarrow P_e \to 0}$$

Rate Distortion Theory

source coding

$X \longrightarrow$ [digital] $\longrightarrow$ [ ] $\longrightarrow$ [digital] $\longrightarrow \hat{X}$

goal $\Rightarrow X = \hat{X}$

$X \neq \hat{X}$

$d(X, \hat{X}) \rightarrow$ distortion

Date - 8/10/25

Continuous Random Variable

A r.v. X is said to be continuous if $\exists$ a function $f_X(x) \geq 0$ such that $P(X \in S)$

$= \int_S f_X(x)\,dx$, if integral exist

The $f_X(x)$ is called a pdf

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

if it is defined b/w (a,b)

$$\int_a^b f(x)\,dx = 1$$

→ Uniform continuous r.v. X is said to be uniform on $[a,b]$ if

$$f_X(x) = \frac{1}{b-a} \quad \forall\, x \in [a,b]$$

$$\left\{ \int_a^b f(x) = 1 \right.$$

→ Normal r.v. (or) Gaussian r.v.

A ct. r.v is said to be normal or gaussi
iff

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

Note - Unlike entropy, diff. entropy can be -ve.

Differential Entropy

Let X be a cts r.v. with a pdf $f_X(x)$.
Then the differential entropy

$$\boxed{h_X(x) = -\int_X f_X(x) \log f_X(x)\, dx}$$

① eg:- let X be a uniform r.v. on $[a,b]$

$$h_X(x) = -\int_a^b f_X(x) \log f_X(x)\, dx$$

$$= -\int_a^b \frac{1}{b-a} \log\left(\frac{1}{b-a}\right) dx$$

$$= \log(b-a) \begin{cases} -ve & \text{iff } 0<b-a<1 \end{cases}$$

② eg:- find the entropy of normal r.v.

$$h_X(x) = -\int_X f_X(x) \log f(x)\, dx$$

$$= -\int_X f_X(x) \left[\log\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}\right)\right] dx$$

$$\left(-\int_X f_X(x) \log \frac{1}{\sqrt{2\pi}\sigma}\, dx + \int_X f_X(x) \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 dx\right.$$

$$\sigma^2 = var(X) = E[(x-\mu)^2]$$
$$= \int f(x)(x-\mu)^2\, dx$$

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

$$= \log\sqrt{2\pi\sigma^2} + \frac{1}{2}\log_2 e = \frac{1}{2}(\log 2\pi\sigma^2 + \log e)$$

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

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

Date - 14/10/2025

AWGN Channel

Noisy

Additive White Gaussian

$$C = \frac{1}{2}\log(1 + SNR)$$

$$W \log\left(1 + \frac{P}{N}\right)$$

$$Z \downarrow$$
$$X \rightarrow \boxed{+} \rightarrow Y$$

$Y = X + Z$,

$Z \sim \mathcal{N}(0,N)$

$X \rightarrow$ normal random variable $N(\mu,\sigma^2)$

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

Discrete r.v.

$$0 \le H(X) \le \underbrace{\log|X|}_{-\phi \log \frac{1}{|X|}}$$

$$\sum_{X} \frac{1}{|X|} \log \frac{1}{|X|}$$

Continuous r.v.

$$0 \le D_L(\underset{\text{distribution}}{f} \,||\, \underset{\text{normal}}{\phi}) = \int f \log \frac{f}{\phi}\, dx$$

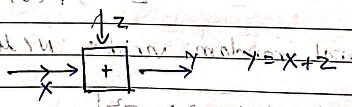
$$= \int f \log f - \int f \log \phi$$

$$\le h(f) - \int \phi \log \phi$$

$$= -h(f) + h(\phi)$$

$$\therefore h(\phi) - h(f) \ge 0$$

$$\Rightarrow \boxed{\underset{\text{arbitrary r.v.}}{h(f)} \le \underset{\text{nr.v.}}{h(\phi)}}$$



$$C = \max_{P_X(x)} I(X;Y)$$

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

$$h(Y) - h(X+Z|X)$$

$$= \boxed{\underset{\text{arbitrary r.v.}}{h(Y) - h(Z)} \le \underset{\text{normal r.v.}}{h(Y) - h(Z)}}$$

$$Y \sim N(0, var)$$

$$var(Y) = var(X+Y)$$

$$= E((X+Z) - E[X+Z])^2$$

$$= E[(X - E[X])^2] + E[(Z - E[Z])^2]$$

$$var(X) + var(Z)$$

$$= P + N$$

$$var(Y) = P+N$$

$$\therefore h(Y) = \frac{1}{2} \log 2\pi e (P+N)$$

$$\therefore I(X;Y) = h(Y) - h(Z)$$

$$= \frac{1}{2} \log 2\pi e (P+N) - \frac{1}{2} \log 2\pi e N$$

$$= \frac{1}{2} \log \left(\frac{P+N}{N}\right)$$

$$\therefore \boxed{I(X;Y) = \frac{1}{2} \log \left(1 + \frac{P}{N}\right)}$$

★ $PSNR = 10\log_{10}\left(\frac{max^2}{MSE}\right)$ dB

$SNR = \frac{P}{N}$

$\boxed{\Rightarrow \frac{1}{2}\log_2(1+SNR) \text{ bits/channel}}$

Nyquist - Shannon Sampling "W - bandwidth"

→ Once you open the channel, you can push 2W samples per channel.

$$\boxed{C = 2W\left(\frac{1}{2}\log(1+SNR)\right) \text{ bit/sec}}$$

where W - bandwidth

Rate distortion theory

goal ⇒ $X = \hat{X}$

$X \neq \hat{X}$

$d(X, \hat{X})$

$d: X \times \hat{X} \rightarrow R$

① $d(x, \hat{x}) = (x - \hat{x})^2$ → square error

$d(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{n}(x_i - \hat{x}_i)^2$ → mean square error

② $d(x, \hat{x}) = |x - \hat{x}|$ → absolute

$d(x, \hat{x}) = \frac{1}{N}\sum_{i=1}^{N}|x_i - \hat{x}_i|$ → absolute mean error

③ $d(x, \hat{x}) = \begin{cases} 0, & \text{if } x = \hat{x} \\ 1, & \text{if } x \neq \hat{x} \end{cases}$ Hamming metric

	x	$\hat{x}$	$d(x, \hat{x})$
A	0	0	0
D	1	0	1
C	10	01	0
D	11	10	1

★ Higher the rate → less is the distortion & vice-versa.

Date - 28/10/25

eg:-

X	Y	Z	$\bar{x}$	$d(\bar{x}, x) = (\bar{x} - x)^2$
0	0000	000	0	0
1	0001	000	0	1
2	0010	001	1	1
3	0011	001	1	4
4	0100	010	2	4
5	0101	010	2	9
6	0110	011	3	9
7	0111	011	3	16
8	1000	100	4	16
9	1001	100	4	25
10	1010	101	5	25
11	1011	101	5	36
12	1100	110	6	36
13	1101	110	6	49
14	1110	111	7	49
15	[illegible]	111	7	64

Rate = 3

$D = \frac{1}{N}\sum d(x, \bar{x}) = \frac{344}{16} = 2$

Z_2 (Compressing by 2 bits):

	$\hat{x}_2$	$d(\hat{x}_2, x)$
00	0	0
00	0	1
00	0	4
00	0	9
01	1	9
01	1	16
01	1	25
01	1	36
10	2	36
10	2	49
10	2	64
10	2	81
11	3	81
11	3	100
11	3	121
11	3	144

$D_2 = \frac{776}{16} = 48.5$, similarly $D_3 = 66.5$ → Z_3 compress by 3 bits

$R_2 = 2$ $R_3 = 1$

Two worst case scenario:
- & $R_4 = 0$, $D_4 = \infty$ ⇒ Rate ↑ → Distortion ↓
- $D = 0$, $R = 4$

Data Processing Inequality

<u>Markov Chain</u>

$X \rightarrow Y \rightarrow Z$ forms a markov chain iff → $\boxed{P(Z/x,y) = P(Z/y)}$

$$P(x, z/y) = P(x/y)\,P(z/y)$$

Let X, Y, Z be three r.v. form a markov chain then $\boxed{I(X;Y) \geq I(X;Z)}$ ↳ Data processing inequality

<u>Proof</u>:

Bayes Rule -
$$P(x,y) = P(x)\,P(y/x)$$
$$P(x,y,z) = P(x)\,P(y/x)\,P(z/x,y)$$

Chain Rule for MI

$$I(X;Y,Z) = I(X;Y) + I(X;Z/y)$$

Also,

$$I(X;Y,Z) = I(X;Z) + I(X;Y/Z)$$

$$\Rightarrow I(X;Y) + I(X;Z/y) = I(X;Z) + I(X;Y/Z)$$

But $I(X;Z/Y) = 0$

$P(x,z/y) = P(x/y)P(z/y)$

$$I(X;Z/y) = \sum P(x,z/y) \log \frac{P(x,z/y)}{P(x/y)P(z/y)} = 0$$

Due to markov chain

$$I(X;Y) + 0 = I(X;Z) + I(X;Y/Z) \quad \text{--- (1)}$$

$$\Rightarrow \boxed{I(X;Y) \geq I(X;Z)}$$ → Data processing inequality

$$X \rightarrow Y \rightarrow Z$$

$$\boxed{I(X;Y) \geq I(X;Y/Z)}$$ → Always true if r.v.s form markov chain but in general r.v's this can change

$I(X;Y) < I(X;Y|Z)$ {general r.v's}

eg $X \rightarrow$ Tossing $= \begin{cases} 0, & 1/2 \\ 1, & 1/2 \end{cases}$

$Y \rightarrow$ Tossing $= \begin{cases} 0, & 1/2 \\ 1, & 1/2 \end{cases}$

$I(X;Y) = 0$

$I(X;Y|Z) \neq 0$

$Z = X + Y = \begin{cases} 0 & 1/4 \\ 1 & 1/2 \\ 2 & 1/4 \end{cases}$

Date = 29/10/2025

$X \rightarrow Y \rightarrow Z$

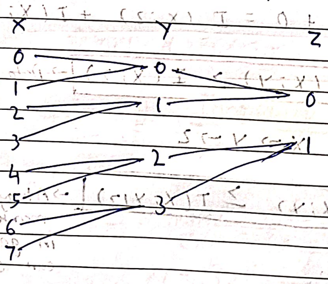
$P(Z|x,y) = P(z|y)$

Z only depends on y but not on X

$P(x, z/y) = P(x|y)\, P(z|y) \Rightarrow X \rightarrow Y \rightarrow Z$

$\rightarrow P(z; x|y) = P(z|y)\, P(x|y) \Rightarrow Z \rightarrow Y \rightarrow X$

Markov chain $\rightarrow X \leftrightarrow Y \leftrightarrow Z$



$P(X, z/y) = P(x|y)\, P(z|y)$
But $P(x,z) \neq p(x)p(z)$

* $X \rightarrow Y \rightarrow Z$ M.C.

$I(X;Y) \geq I(X;Z)$

⇒ example from last class

$I(X;Y) = 0 \rightarrow$ tossing coins $\rightarrow$ independent event $\rightarrow I = 0$

$I(X;Y|Z) > 0$

$\{ I(X;Y) = H(X) - H(X|Y) \}$

$I(X;Y|Z) = H(X|Z) - H(X|Y,Z)$
$= \underbrace{H(Y|Z)}_{\text{1 bit needed to represent}} - \underbrace{H(Y|X,Z)}_{\text{x \& z are given so no bits needed}}$
$= H(Y|Z) - 0$
$= 1 - 0$

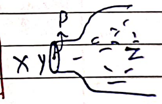
Rate Distortion

$X \rightarrow Y \rightarrow Z$

Min $I(X;Y)$

Max $I(Y;Z)$
$= \min -I(Y;Z)$

$$\boxed{L = \min \left[I(X;Y) - \beta\, I(Y;Z) \right]}$$



Information Bottle Neck.

relating to rate distortion :-

$$R(D) = \min I(\underset{\uparrow \text{original img.}}{X} ; \underset{\text{compressed img.}}{Y})$$

$$\boxed{E[d(\bar{x}, x)] \le D}$$ ↳ how much we want to compress

Min $I(X;Y)$ s.t. $E[d(\bar{x}, x)] \le D$

Max D given R (rate)

$$E[d(\underset{\text{compressed}}{\bar{x}}, \underset{\text{og}}{x})] = \sum P(\bar{x}, x) d(\bar{x}, x)$$

$$= \sum P(x) p(\bar{x} \mid x) \, d(\bar{x}, x)$$

$E[d(\bar{x}, x)] = \sum \frac{1}{N} (\bar{x} - x)^2$

$\rightarrow p(\bar{x}, x) = \frac{1}{N}$

$\rightarrow d(\bar{x}, x) = (\bar{x} - x)^2$

$= \frac{1}{N} \sum (\bar{x} - x)^2$

↳ MSE

↓

Distortion

$$R(D) = \min I(X;Y)$$

$$E[d(\bar{x}, x)] \le D$$

$$= \min_{P(\bar{x} \mid x)} I(X;Y), \text{ where } \sum_{\bar{x}} P(\bar{x} \mid x) = 1$$

Lagrange ⇒ $\boxed{L = I(\bar{X}; X) + \beta [E[d(\bar{x}, x)]] + \lambda(x) [\sum P(\bar{x} \mid x) - 1]}$

$$\frac{\partial L}{\partial P(\bar{x} \mid x)} = \frac{\partial L}{\partial q}$$

$P(\bar{x} \mid x) = q$

$$L = \sum_{x, \bar{x}} P(\bar{x}, x) \log \frac{P(\bar{x}, x)}{P(\bar{x}) P(x)} + \beta \sum_{x, \bar{x}} P(\bar{x}, x) d(\bar{x}, x) + \lambda(x) \left[\sum_{\bar{x}} P(\bar{x} \mid x) - 1 \right]$$

$P(\bar{x} \mid x) = q$

$$\Rightarrow L = \sum_{x, \bar{x}} P(x) P(\bar{x} \mid x) \log \frac{P(x) P(\bar{x} \mid x)}{P(x) P(\bar{x})} +$$

$$\beta \sum_{x, \bar{x}} P(x) P(\bar{x} \mid x) d(\bar{x}, x) + \lambda(x) [P(\bar{x} \mid x) - 1]$$

$$\Rightarrow L = \sum_{x, \bar{x}} P(x) q \log \frac{q}{P(x) P(\bar{x})} + \beta \sum_{x, \bar{x}} p(x) q d(\bar{x}, x) + \lambda(x) [q - 1]$$

$$\Rightarrow \frac{\partial L}{\partial q} = \sum_{x, \bar{x}} p(x) \left(\log \frac{q}{P(\bar{x})} + 1 \right) + \beta \sum_{x, \bar{x}} P(x) d(x, \bar{x}) + \lambda(x) = 0$$

$$\Rightarrow \sum_{x, \bar{x}} P(x) \left(\log \frac{q}{P(x)} + 1 \right) + \beta \sum_{x, \bar{x}} P(x) d(\bar{x}, x) + \lambda(x) = 0$$

$$\log \frac{q}{P(x)} + 1 = -\beta d(\bar{x}, x) \frac{\lambda(x)}{P(x)}$$

$$\therefore \log \frac{q}{P(x)} = -1 - \beta d(\bar{x}, x) - \frac{\lambda(x)}{P(x)}$$

$$= -1 - \beta d(\bar{x}, x) - \frac{\lambda(x)}{P(x)}$$

$$q = P(\bar{x}) e^{-1 - \beta d(\bar{x}, x) - \frac{\lambda(x)}{P(x)}} = \frac{P(\bar{x}) e^{-\beta d(\bar{x}, x)}}{Z(x)}$$

$$Z(x) = e^{1 + \frac{\lambda(x)}{P(x)}}$$

But $\sum_{\bar{x}} q = 1$

$$\Rightarrow \sum_{\bar{x}} \frac{p(\bar{x}) e^{-\beta d(x,\bar{x})}}{z(x)} = 1$$

$$\Rightarrow \sum_{\bar{x}} p(\bar{x}) e^{-\beta d(x,\bar{x})} = z(x)$$

$$\Rightarrow \boxed{q = p(\bar{x} \mid x) = \frac{p(\bar{x}) e^{-\beta d(\bar{x},x)}}{\sum_{\bar{x}'} p(\bar{x}') e^{-\beta d(\bar{x}',x)}}}$$

Date - 3/11/2025

Balahut (1972) - Arimoto (≈1978) Algorithm : (2015 in ML)

↳ Measurement of Rate distortion.

$$R(D) = \min_{E(d(x,\hat{x})) \le D} I(X;\hat{X})$$

$$= \min_{P(\hat{X} \mid X)} I(X;\hat{X})$$

Input = $P_X(x)$

Distortion measurement = MSE ; $d(x,\hat{x}) = (x-\hat{x})^2$

Output = $P_{t+1}(\hat{x} \mid x)$

$$\boxed{P_{t+1}(\hat{x} \mid x) = \frac{P_t(\hat{x}) e^{-\beta d(x,\hat{x})}}{\sum_{\hat{x}'} P_t(\hat{x}) e^{-\beta d(x,\hat{x})}} , \quad t = 0, 1, 2 \cdots}$$