Latitha Miam : Office hours, 5pm-6pm Friday

Grading Scheme :-

Quiz 1	10 %	
Quiz 2	10 %	
Midsem	20 %	
final	30 %	3 H1
Assign	15 %	3 H2
Proj/Pap	15 %	

Books :-

Probability and measure — P. Billingsley

Probability and Random Variables

2/1/25

Information and Communication

→ A Communication System :-

- Basics of Communication will be covered.
- One important type of communication system is a Point to Point system :-

Tx ⟶ Rx (P2P Communication)

Transmitter Reciever

- In this "channel" between the Tx and Rx, Nature will add some noise to the data transmitted.

∴ The channel aids in transmission but also adds noise

- If the problem of noise is not there, then ∞ data rates can be obtained, since the precision of the data is maintained.
- Also the amplitude of the signal being sent, depends on the power of the transmitter.
- SNR : Signal to Noise Ratio
- Due to these 2 points, only a finite amount of data can be transmitted at any time.

- $SNR \propto \frac{1}{Noise}$, Celioreless Communication has low SNR
  ∴ It is only used in last-mile connectivity.

- Wireless Communication happens at around 900 MHz. The bandwidth is much smaller than what is possible in wired medias.

- Fundamental Limit of Communication :- (for P2P)
  Defined by Claude Shannon in 1948 (A Mathematical Theory Of Communication).

  - Claude Shannon came up with a method of quantizing different kinds of information like text, images, etc.

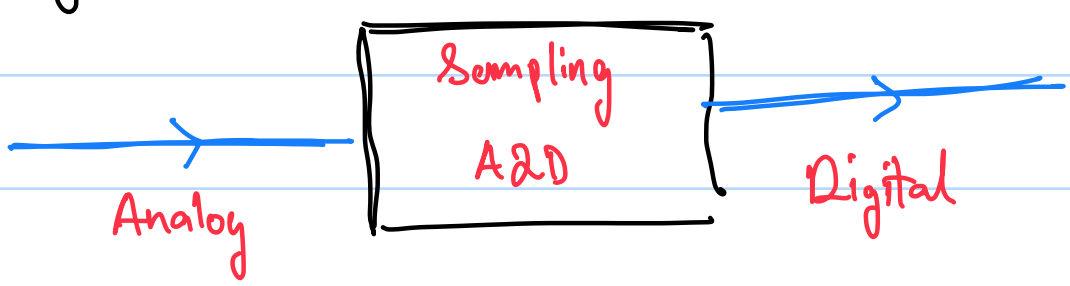
  - Channel Capacity :-
    It is a function of SNR (and power), that decides the limit of how much information can be sent through the channel.

    - 5G is reaching channel capacity, for P2P at the fundamental level.

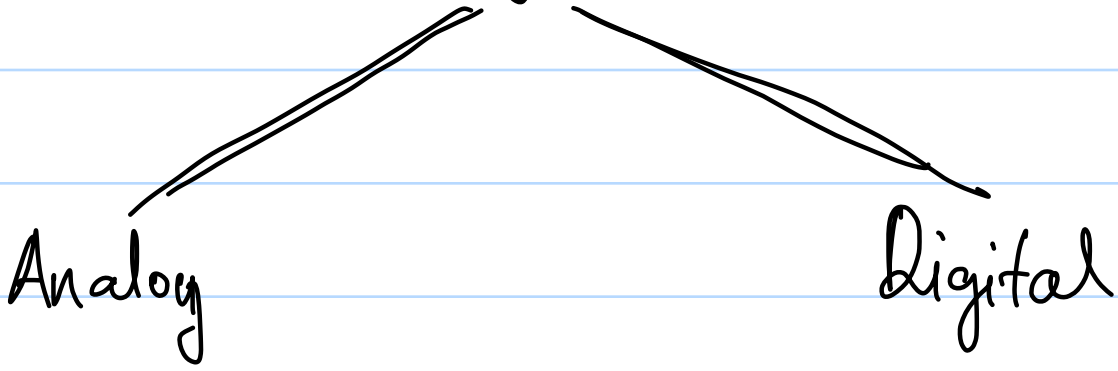
    - Networks also have a channel capacity.

6/1/25

→ Communication System :-



Signals



- Most signals are a function of one independent variable, ie, 1D signals. $f(x)$
- Images are signals in 2 independent variables (x & y coordinates). $f(x,y)$
- Videos are signals in 3 independent variables (x, y and time) $f(x,y,t)$
- Analog Signals: Signals that are continuous in an independent variable and are continuous in amplitude.
  - Highly vulnerable to noise.

- **Digital signals:** Signals that are discrete in the independent variable and also discrete in amplitude.
  - Resistant to noise.

- When an Analog signal is converted to Digital, extra precautions must be taken to prevent loss of information.

- **A2D Conversion :-**

1) **Sampling :-**

Discretizes the signal in the independent variable.

$x(t) \longrightarrow x[nT]$, $n \in \mathbb{N}$, $T \in \mathbb{R}$ is constant

$1/T = f_s$ = rate of sampling

continuous signal → discrete signal

- There will definitely be a loss of information in this step, ie, this step is irreversible and not invertible.

Not invertible since 2 different signals can produced the same sampled output.

- Nyquist's Theorem :-

  - Contrary to the prev. point, using Nyquist's theorem, it is possible to reconstruct the continuous signal from the sampled output.

  - The signal must be band-limited.

Signal Transformation

Laplace Transform	Fourier Transform
Used in Analysis of Systems	Used in Analysis of Signals

- Fourier transform of a signal gives you the component / parameter of a signal in each frequency.

$$X(f) = \int_{-\infty}^{\infty} x(t)\, e^{-j2\pi ft}\, dt$$

- $x(t)$ is said to be band limited to $[-B, B]$, if

$$X(f) = 0 \quad \forall \quad |f| > B$$ where $B \in \mathbb{R}$ constant

B = max. frequency

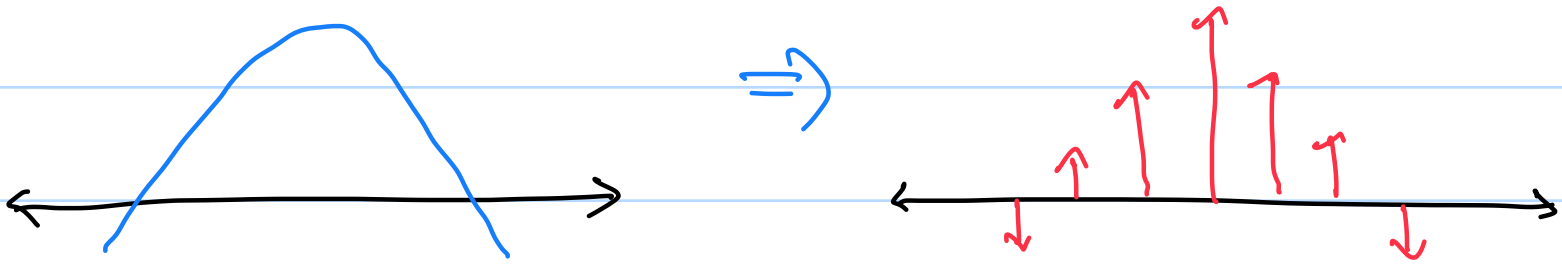
- In the sampled output, if $f_s \geq 2B$, then the original signal can be reconstructed from the sampled output

f_s = rate of sampling

Intuitively, if the samples are much closer together, the loss of information is minimized.

$f_s < 2B \rightarrow$ Aliasing

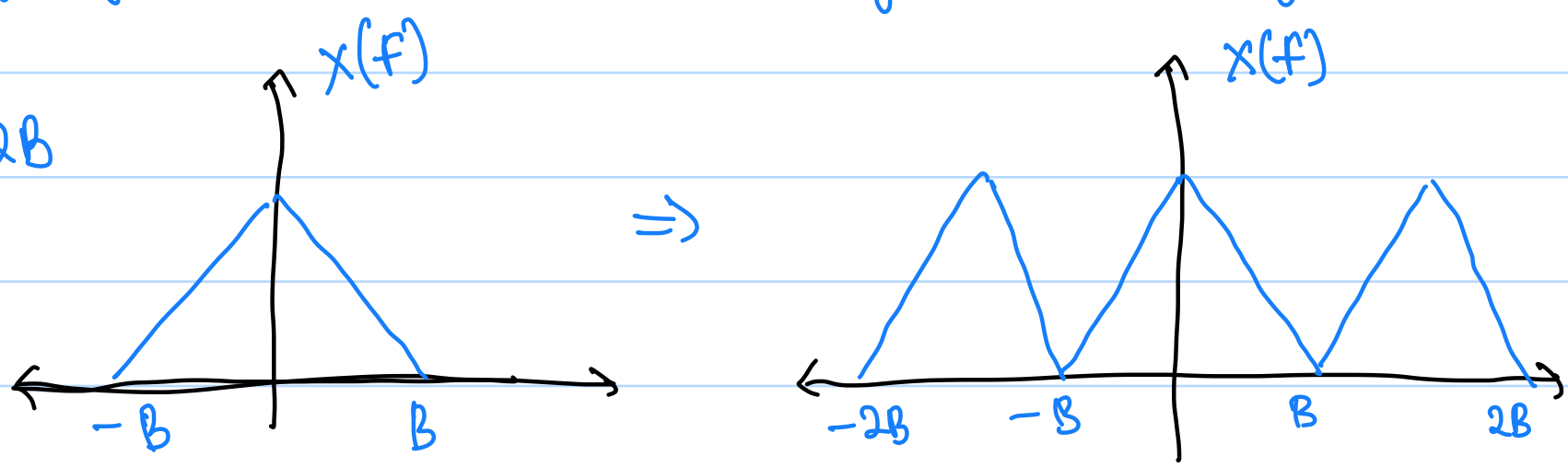
- The sampled signal is an impulse train signal.



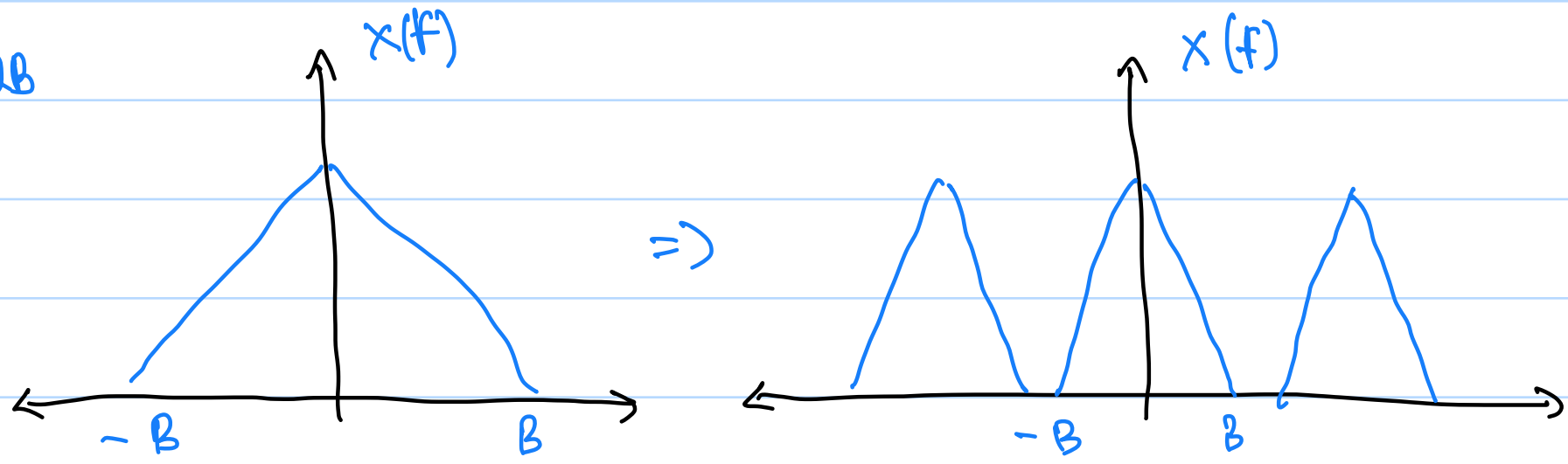
$$x_d(t) = \sum_{n=-\infty}^{\infty} x(nT)\,\delta(t - nT)$$

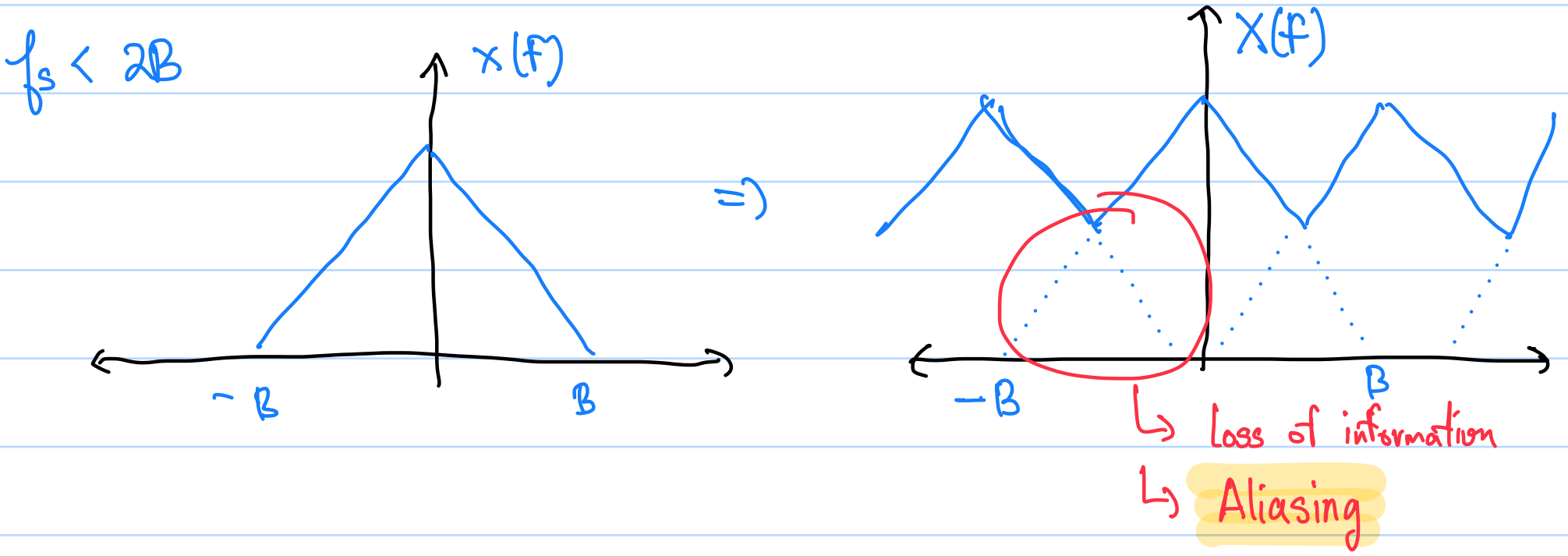
- Discretization in the time domain makes the signal periodic in the frequency domain, since the sampling occurs at regular

$f_s = 2B$



$f_s > 2B$

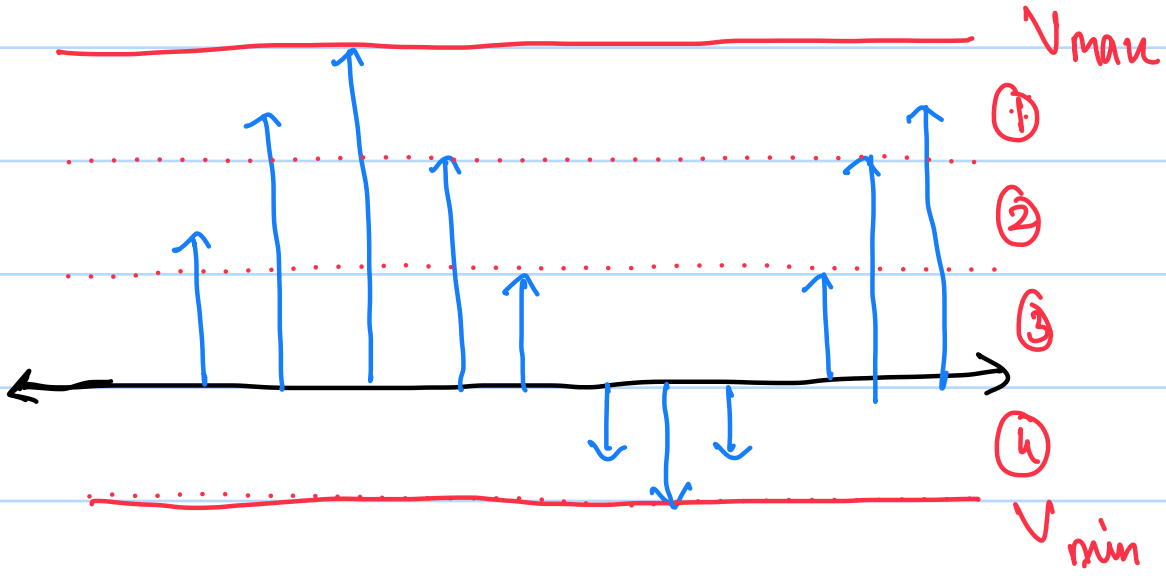




2) Quantization :

Discretizes the amplitude of the signal.

- Let V_{min} and V_{max} be the min. and max value of the signal, after sampling.



- An n-bit A2D module divides the interval (V_{min}, V_{max}) into 2^n intervals, and sets the mid-value of each interval as a reference point of each interval. (say $n = 2$ here)

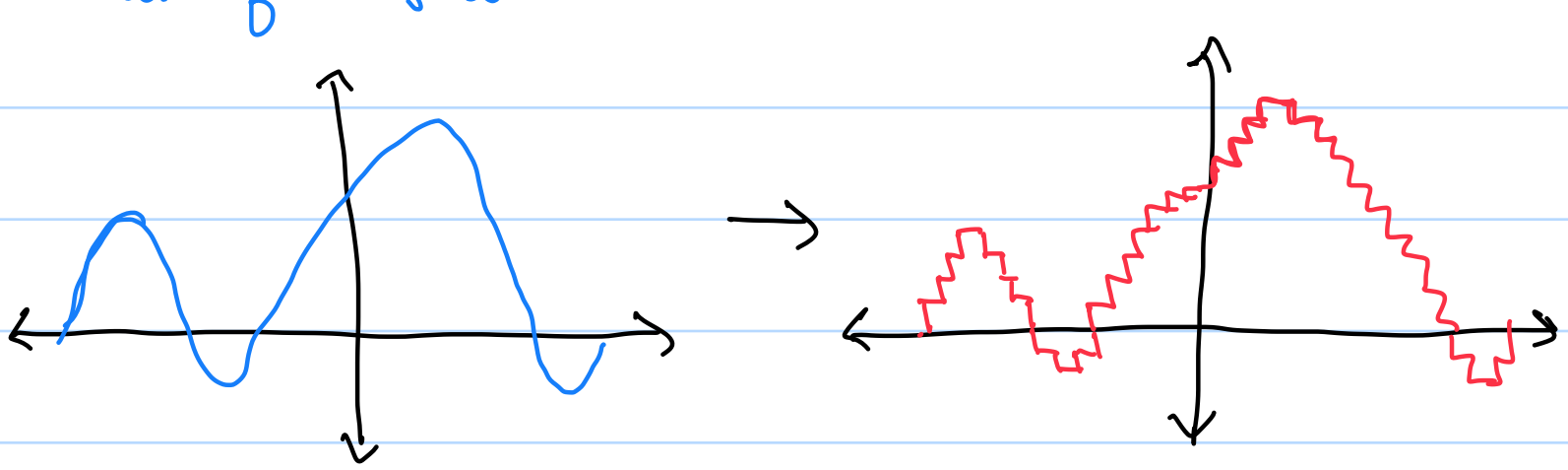
- At each interval, the value of the signal is set to the reference value.

- $$\text{Max error} = \frac{V_{max} - V_{min}}{2^{n+1}}$$

- Each reference value is given a binary code. Depending on the signal, the analog signal is now compressed into a digital binary stream.

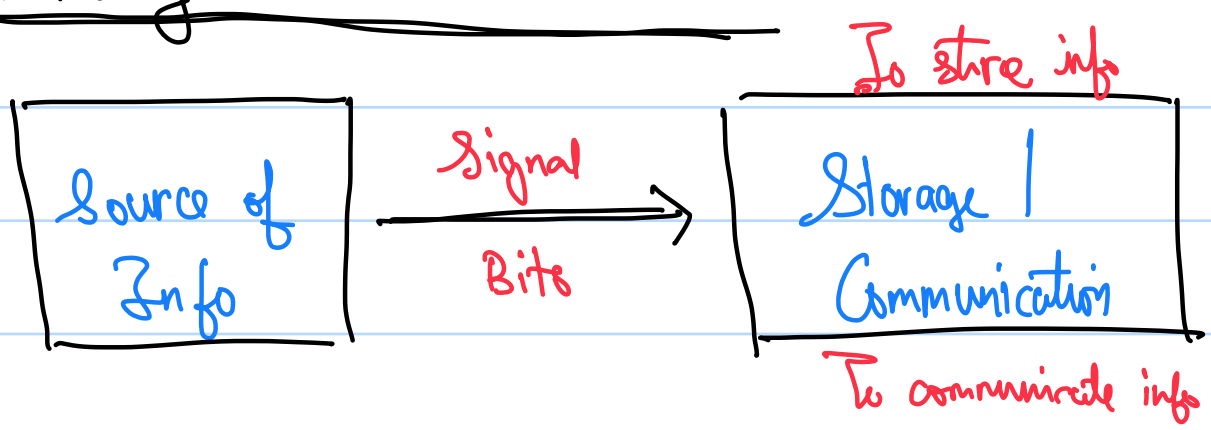
- Sampling Circuit :-

- A sampling circuit reads the value of the signal at a point and holds it for T_s time.



- A voltage divider block will converts this signal into a impulse train.

⇒ Probability in Random Variables :-



- To store / communicate data efficiently, we need some information about the source of information.
- The source of info may have some random variables that we may need to process.

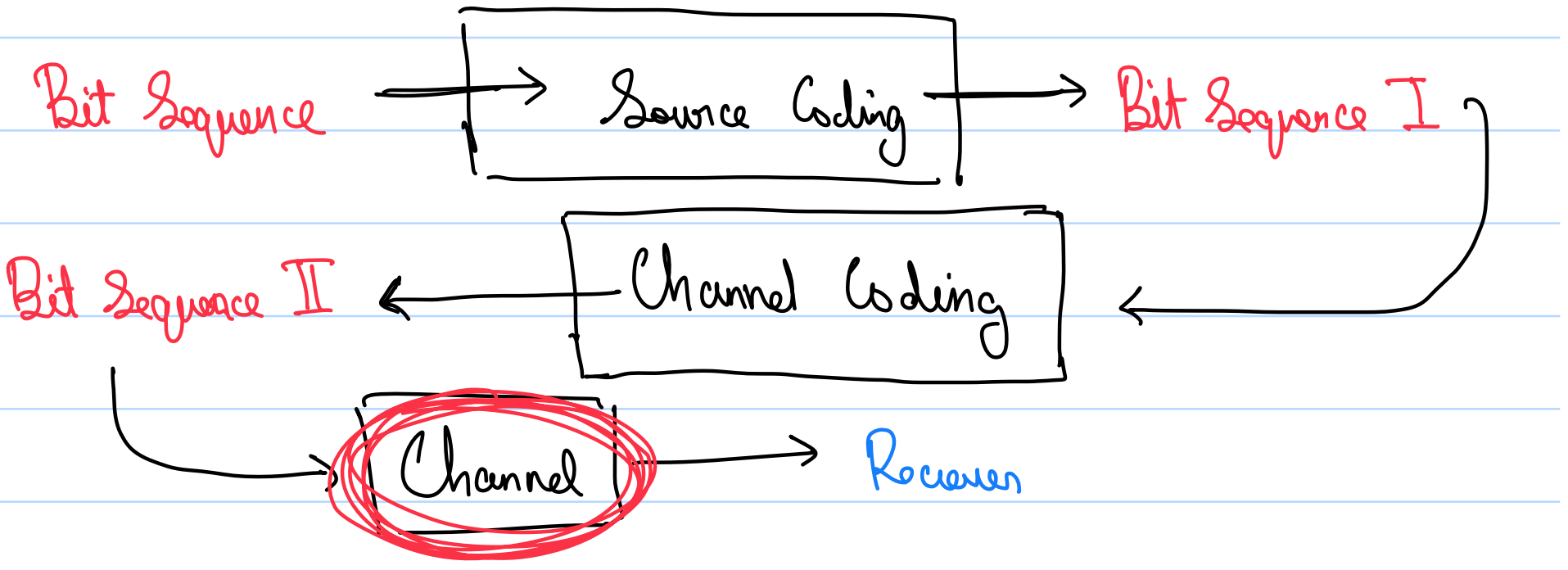
- To apply the concept of probability in this analysis, we use the axiomatic approach. (Ω, F, P)

  Sample space ↑ ↑ Event Space $\subseteq \Omega$

  Tossing a Coin, $\Omega = \{H, T\}$

- The axiomatic approach allows us to forego the empirical part of finding probability of an event.

- Analog → Sampling → Quantization → Bit Sequence.



In the above shown communication pathway, the "Channel" is the main culprit when it comes to distortion.

To combat this distortion we use the principles of probability.

- Need of probability: Ability to say "something" about the outcome of an event in a deterministic space.

◦ Probability Space :-

	Set Theory	Probability Space
–	Universe	Sample space
	Subset	Event
	Element	Outcome
	Simpleton set	Simple event
	Null set	Impossible event
	Disjoint set	Mutually exclusive event

– Probability:

A value assigned to each event. In function notation,

$P(E) : \mathfrak{I} \longmapsto [0,1]$

it's a measure of how likely E is.

– A probability space is a formally defined space that is used for calculating the probability of an event within a sample space.

– In a probability space we have 3 things : (Ω, F, P)

1) Sample Space (Ω) : Set of all possible outcomes

2) Event Space (F) : Collection of Desirable Events from Ω

3) Probability measure (P)

- A probability space gives us a formal model of a random experiment.

- Event space must be closed under complement, ie, $A^c \in F$, and under countable union, ie, if $A_1, A_2, A_3 \in F$. then $A_1 \cup A_2 \cup A_3 \in F$

Same can be said for intersection through De-Morgan's laws.

- Event Space must contain Ω and ϕ.

Example: Experiment of rolling a six sided dice, there are 2 events,

$$A = \{1,2,3\}, \quad B = \{1\}$$

Event Space generated by A is

$$F_A = \{\phi, \Omega, \{1,2,3\}, \{4,5,6\}\}$$

Event Space generated by B is

$$F_B = \{\phi, \Omega, \{1\}, \{2,3,4,5,6\}\}$$

Event Space generated by A and B is,

$$F_{A,B} = \{\phi, \Omega, \{1\}, \{2,3,4,5,6\}, \{1,2,3\}, \{4,5,6\}\}$$

- Any random event can be summarized by a probability space.

20/1/25

• Random Variables :-

- Suppose our experiment is the rolling of 2 dice, in which we are interested in their sums.

$$\Omega = \{ (1,1), (1,2) \ldots (1,6) \\ \vdots \\ (6,1), (6,2) \ldots (6,6) \}$$

- Suppose we have X = sum of the 2 rolls. If we define another probability space for the possibilities of X,

$$\Omega' = \{2, 3, 4, \ldots, 12\}$$

- Let F and F' be the power sets / event space of Ω and Ω'. and P and P' be their probability measure.

$$\therefore (\Omega, F, P) \xrightarrow{X} (\Omega', F', P')$$

The probability space has been transformed because of the variable X.

- Define $X : (\Omega, F, P) \longrightarrow (\Omega', F', P')$. Still our work is the same as we have gone from one probability space to another.

- Now, let $\Omega' = \mathbb{R}$, $F' = B(\mathbb{R})$, $P' = P_X$ — a special probability space.

$B(\mathbb{R})$ = Borel σ-Algebra (Highly advanced topic)

- This concept is best used in very large Ω.

- **Borel σ-Algebra** :-

If $\Omega = \mathbb{R}$, then $B(\mathbb{R})$ is the event space generated by open sets of the form (a, b), $a, b \in \mathbb{R}$ and $a \leq b$.

Defn of event space:

1) $\phi, \Omega \in F$
2) Complement is closed
3) Union is closed.

In Borel σ Algebra, each event is represented by an interval in $\mathbb{R}$.
ie, $(-10, 10)$ is an event in BoA, so is $(-\infty, 0)$.

∴ $B(\mathbb{R})$ is the collection of all possible subsets in $\mathbb{R}$

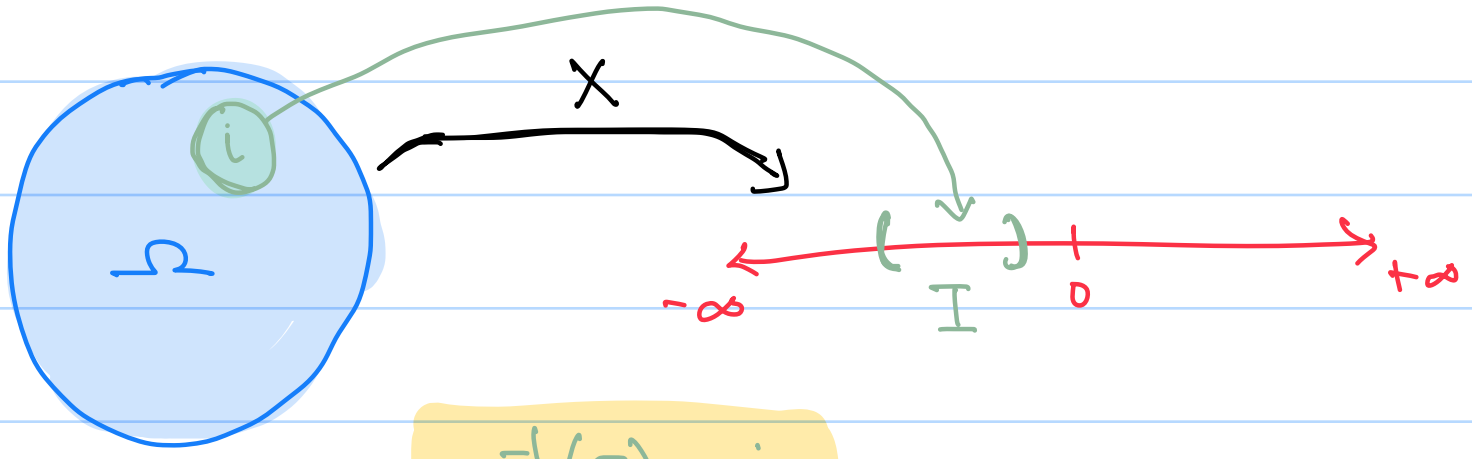
- **Defn of Random Variable** :-

It is a map from a probability space to the real line,

$$X : (\Omega, F, P) \longrightarrow (\mathbb{R}, B(\mathbb{R}), P_X)$$

$$\Omega \xrightarrow{X} \mathbb{R}\ , \quad \mathcal{F} \xrightarrow{X} \mathcal{B}(\mathbb{R})\ , \quad P \xrightarrow{X} P_X$$

- Any valid event in $\mathcal{B}(\mathbb{R})$ are also valid events in $\mathcal{F}$.



$$X^{-1}(I) = i$$

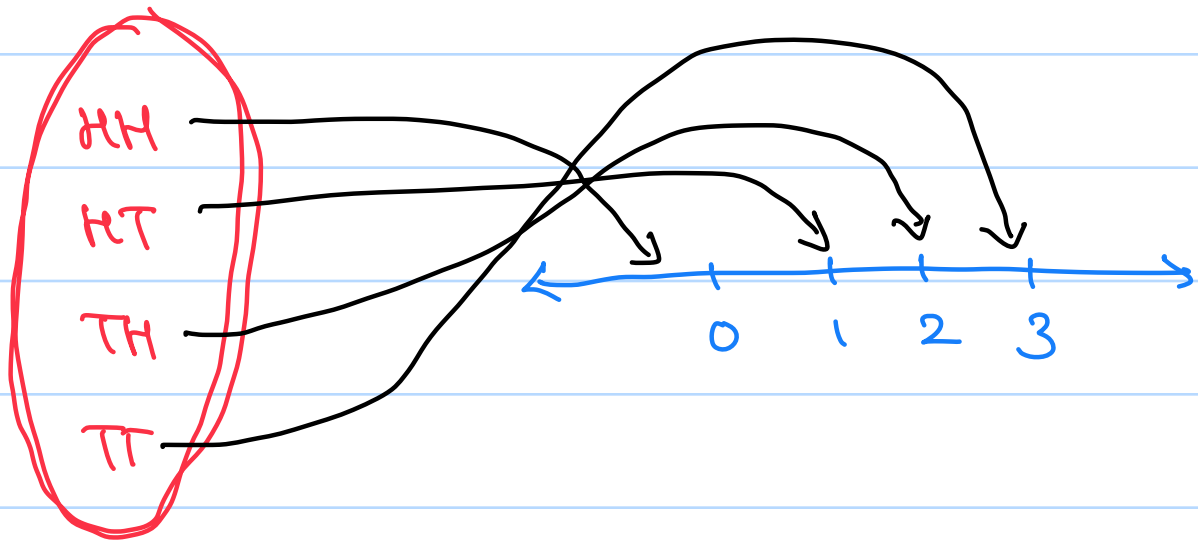
$$\Rightarrow \quad X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

$\therefore$ A random variable X is a map $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$ such that $\forall\ B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ such that

$$X^{-1}(B) \in \mathcal{F} \text{ and}$$

$$P_X(B) = P_r(\omega \in \Omega : X(\omega) \in B)$$

Example: Consider the map, $HH \xrightarrow{X} 0$, $HT \xrightarrow{X} 1$, $TH \xrightarrow{X} 2$, $TT \xrightarrow{X} 3$



Let $\mathcal{F} = \{\phi, \Omega, (HH \cup TT), (HT \cup TH)\}$, $B_1 = [-10, 5]$,

$B_2 = [2, 3] \cup [9, 10]$

$X^{-1}(B_1) = \{HH, HT, TH, TT\} = \Omega$

$X^{-1}(B_2) = \{TH, TT\} \longrightarrow$ This is not a valid event in the given event space.

∴ The given mapping for the given probability space is not valid.

Example: $\Omega = \{1,2,3,4\}$, $\Omega' = \{a,b,c\}$

F, F' are power sets of Ω, Ω'

Let $X(1) = X(4) = a$, $X(2) = b$, $X(3) = c$

Mapping to another probability space.

$X^{-1}(a) = \{1,4\}$, $X^{-1}(b) = \{2\}$, $X^{-1}(c) = \{3\}$

$X^{-1}(a \cup b) = \{1,4,2\}$, $X^{-1}(b \cup c) = \{2,3\}$

Since any union / intersection of events in Ω' will have a pre-image in Ω, the given RV is valid.

Example: Let the experiment be choosing a random number b in $[-1,1]$

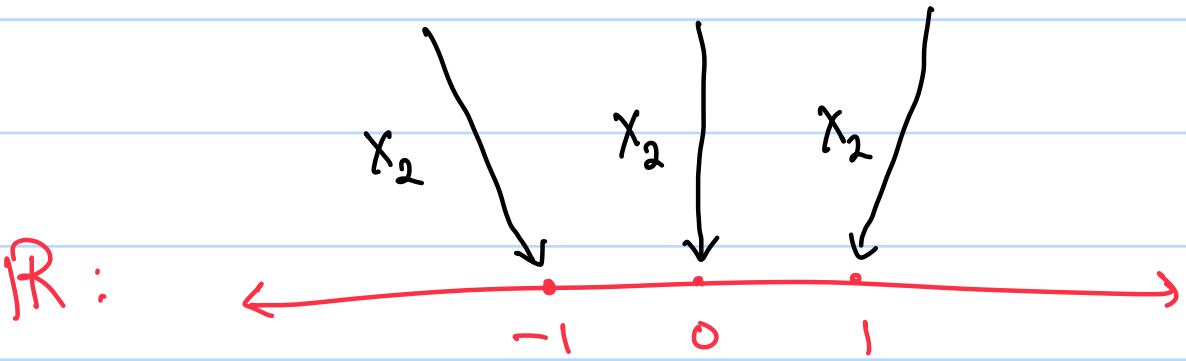
There are uncountably many elements in Ω.

Let $X_1(b) = b^2$ Continuous RV

$$X_2(b) = \begin{cases} -1, & b < 0 \\ 0, & b = 0 \\ 1, & b > 0 \end{cases}$$

Discrete RV.

Ω : $[-1, 0) \cup \{0\} \cup (0, 1]$



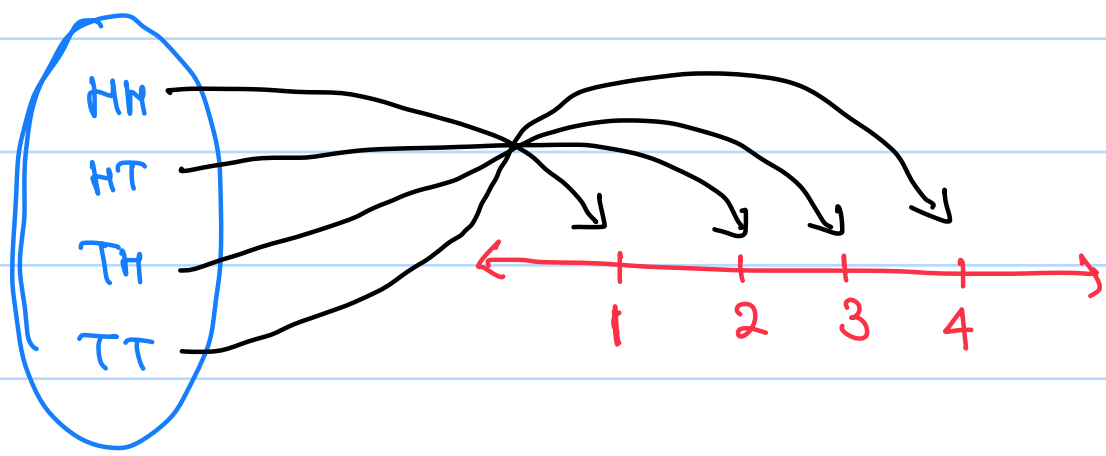
Example:- $\Omega = \{a, b, c\}$, $\Omega' = \{0, 1\}$

$F = \{\phi, \Omega, \{a\}, \{b, c\}\}$, F' = power set of Ω'

$X(a) = 1$, $X(b) = X(c) = 0$

Here X is an **indicator Rv**. Output of any event except a is 0.

- Let X be a Rv with **support set** $\mathcal{X}$. $\mathcal{X}$ is defined as the set of all possible outputs of the Rv.



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

- x is said to be the **realization** of an event E if $X(E) = x$.

$$x \in \mathcal{X}$$

- **Discrete RV :-**

X is a finite / countably infinite set.

ex: $\Omega = \{HH, HT, TH, TT\}$

Suppose $P(HH) = 0.2$, $P(HT) = 0.3$, $P(TH) = 0.35$, $P(TT) = 0.15$

If we define X st $X(HH) = 1$, $X(HT) = 2$, $X(TH) = 3$, $X(TT) = 4$, then

$P_X(X=1) = 0.2$

$P_X(X=2) = 0.3$

$P_X(X=3) = 0.35$

$P_X(X=4) = 0.15$

- $P(X = x)$ is a probability mass function.

- CDF, $F(x) = P_X(X \leq x)$ Cumulative Distribution Function

27/11/25

- X essentially is a map from $\Omega \to \mathbb{R}$, where any closed interval in $\mathbb{R}$, corresponds to a subset / event in Ω.

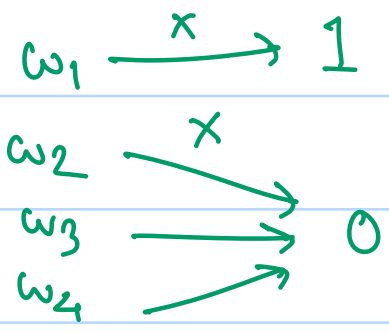
$$X^{-1}([-\infty, x]) \in F \quad \forall \quad x \in \mathbb{R}.$$

- **Example:**

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

$$F = \{\phi, \Omega, \{\omega_1\}, \{\omega_2, \omega_3, \omega_4\}\}$$

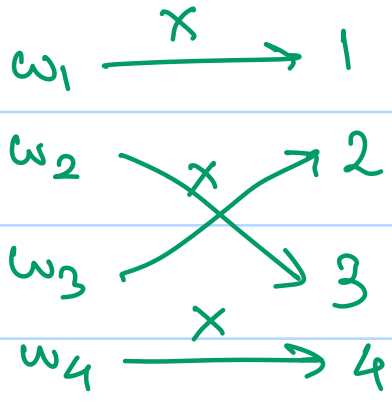
X_1,



is a valid Rv,

$$\chi = \{0, 1\}$$

X_2,



is an invalid Rv since

$$X^{-1}([2, \infty]) = \{\omega_2, \omega_3\} \notin F$$

- $P_X((-\infty, x]) = P(\{\omega \mid X(\omega) \in (-\infty, x]\})$

  $P_X([x, \infty)) = P(\{\omega \mid X(\omega) \in [x, \infty]\})$

- If the original sample set Ω is finite / countably infinite, then the χ for any Rv is going to be finite / countably infinite, ie, a discrete Rv.

- Representation of a closed set using ∞ open sets,

$$[x, \infty) = \bigcup_{n=1}^{\infty} \left(x + \frac{1}{n}, \infty\right) = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, \infty\right)$$

- **Probability Mass Function :-** (PMF)

  Probability mass function of a discrete Rv is given by $P_X(x)$, $x \in \chi$, where χ is the support set of the Rv.

  - Total PMF = 1

- If $n(\chi) = 2$, the PMF is termed as a Bernoulli Mass Function. $P_X(0) = p$, $P_X(1) = 1-p$.

- Example: n coin tosses,

Ω = { $< 2^n$ possibilities $>$ }

F = Power set of Ω

$X: \Omega \longmapsto \mathbb{R}$

Let $X(\omega)$ = No. of heads in ω

ex: $X(\{\underset{1}{H}, \underset{2}{H}, T, T, \underset{3}{H}\}) = 3$

$X(\omega)$ can range from 0 to n.

Each coin toss is independent, and in each toss, define $P(H) = p$.

$$P(X=0) = (1-p)^n$$

$$P(X=1) = (1-p)^{n-1} \cdot p \cdot {}^nC_1$$

$$\vdots$$

$$P(X=m) = {}^nC_m\, p^m (1-p)^{n-m}$$

Since the probability mass function is similar to the n^{th} term expression of a binomial expansion, the random variable is termed as a **binomial random variable**.

$$\sum_{m=0}^{n} P(X=m) = (1-p+p)^n = \underline{1}$$

- **Example**: Geometric Random Variable,

Experiment of flipping a coin until we get heads

$\Omega = \{H, TH, TTH, TTTH, \ldots\}$

$X: \Omega \to \mathbb{R}$

$X(\omega)$ = No. of tails in ω. $P(H) = p$

$P(X=0) = p$

$P(X=1) = (1-p)\cdot p$

$P(X=2) = (1-p)^2 \cdot p$

$\vdots$

$P(X=n) = (1-p)^n \cdot p$

GP

$$\sum_{n=0}^{\infty} P(X=0) = \frac{p}{1-(1-p)} = \underline{1}$$

Since the PMF follows a GP, the random variable is a Geometric Random Variable.

- Poisson RV, $P(X=n) = \frac{e^{-\lambda}\lambda^n}{n!}$ Used for modelling rare events.

$$\sum P(X=n) = \frac{e^{-\lambda}\lambda}{1!} + \frac{e^{-\lambda}\lambda^2}{2!} \cdots = e^{-\lambda}\left(\frac{\lambda}{1!} + \frac{\lambda^2}{2!} \cdots\right)$$

$$= e^{-\lambda} e^{\lambda} = \underline{1}$$

λ is a parameter that depends on the average outcome of the experiment.

3|2|27

- **Mean of A Rv / Expectation :-**

$$E(X) = \sum_i x_i P_X(x_i)$$

For a Bernoulli Rv,

$$E(X) = 0\cdot(1-p) + 1\cdot p$$

$$= \underline{\underline{p}}$$

- **Variance Of A Rv :-** The spread of the Rv about the mean/ expectation.

$$V(X) = E\left((X - E(X))^2\right)$$

↳ $()^2$ is taken instead of $||$ since it is differentiable.

$$\Rightarrow V(X) = \sum (x_i - E(X))^2 P_X(x_i)$$

↳ Since $(X - E(X))^2$ is a Rv in itself so, $P((X-E(X))^2) = P(X)$

$$= V(x) = \sum (x_i^2 - 2E(x)\cdot x_i + E(x)^2) P_X(x_i)$$

$$= V(x) = \sum (x_i^2 P_X(x_i) - 2E(x)\cdot x_i P_X(x_i) + E(x)^2 P_X(x_i))$$

$$= \sum (x_i^2 P_X(x_i)) - 2E(X)\sum (x_i P_X(x_i)) + E(x)^2 \sum (P_X(x_i))$$

$$= V(x) = E(x^2) - 2E(X)^2 + E(x)^2 \sum P_X(x_i)$$

$\sum P_X(x_i) = 1$

$$= V(n) = E(X^2) - E(X)^2$$

- Variance is always greater than / equal to 0.

- For a Geometric Rv,

$$P(x) = (1-p)^{x-1} p$$

$$E(X) = 1p + 2p(1-p) + 3p(1-p)^2 \cdots\cdots$$

$$= \sum_{k=1}^{\infty} kp(1-p)^{k-1}$$

$$= p \sum_{k=1}^{\infty} k(1-p)^{k-1}$$

$$= -p \sum_{k=1}^{\infty} \frac{d}{dp}(1-p)^k$$

$$= -p \frac{d}{dp} \sum_{k=1}^{\infty} (1-p)^k = -p \frac{d}{dp}\left(\frac{1-p}{p}\right)$$

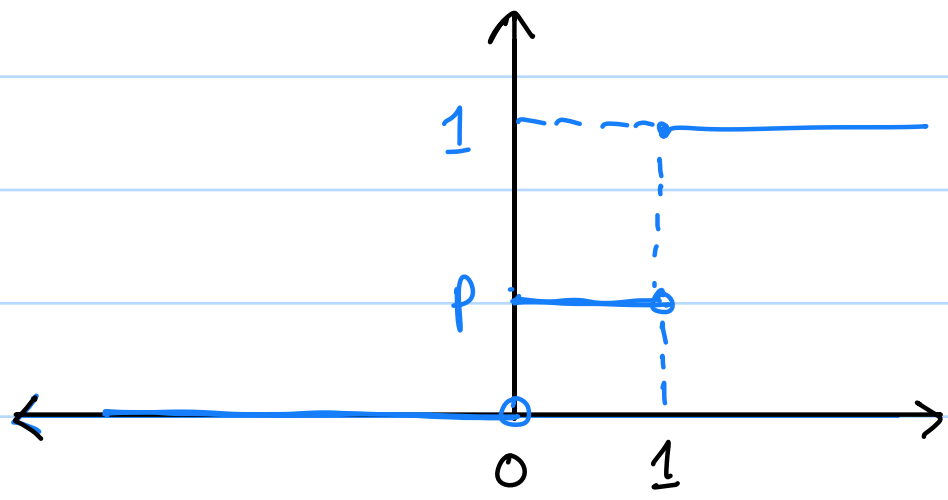
$$= -p \frac{d}{dp}\left(1 - \frac{1}{p}\right) = p\left(\frac{1}{p^2}\right)$$

$$\Rightarrow E(X) = \frac{1}{p}$$

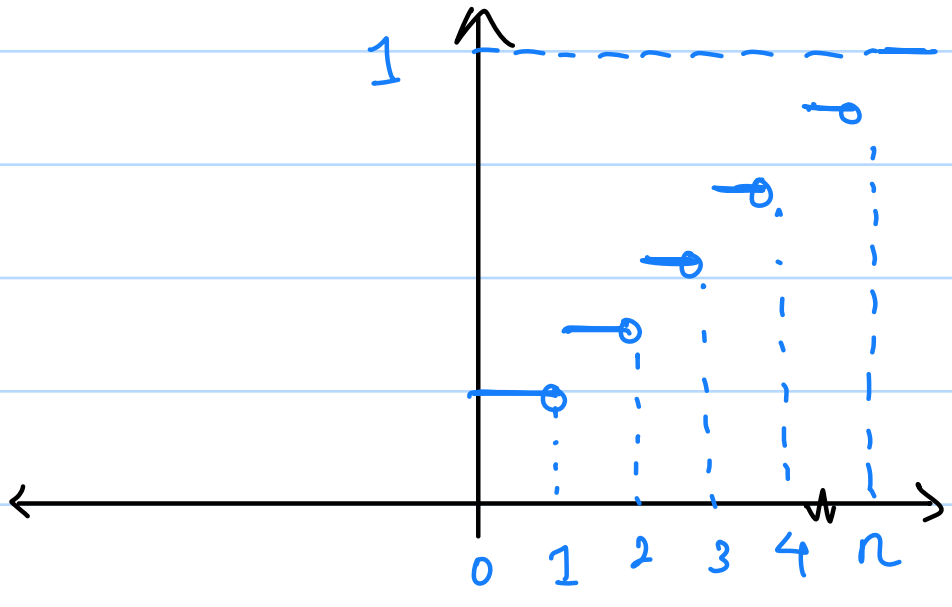
→ Continuous Rv :-

- $F(x) = P([-\infty, x])$ ⟶ Cumulative Distribution Frequency
- CDF of a Discrete Rv is discontinuous, but right continuous.

Bernoulli :-



Binomial :-



- Valid CDF :-
  1) $f_x(-\infty) = 0$
  2) $F_x(\infty) = 1$
  3) f_x is non decreasing
  4) f_x is right continuous

- A Rv is discrete if $n(X)$ is finite or countably infinite.

- A Rv is continuous if $\exists$ a function (density function) such that

$$P(X \in B) = \int_{x} P_x(x)\, dx$$

- Valid density function:

1) $p(x) \geq 0 \quad \forall\ x$

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

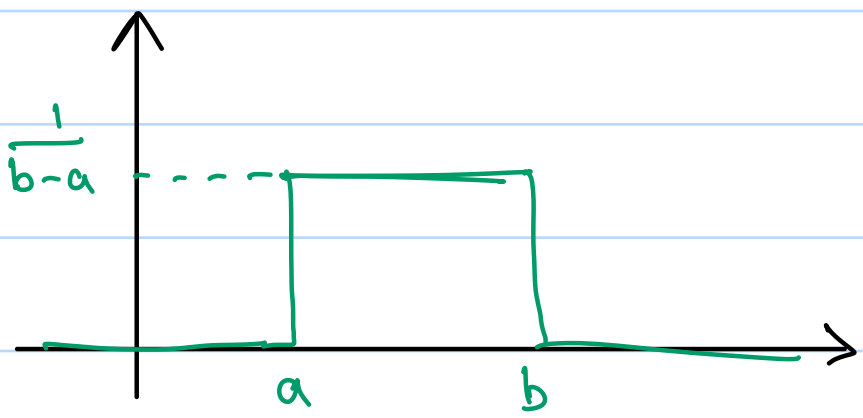
- CDF, $F(x) = \int_{-\infty}^{x} p(x)\, dx$

$$\Rightarrow \quad p(x) = \frac{d}{dx} F(x)$$

Example: Uniform Rv.

$$p(x) = \begin{cases} \frac{1}{b-a}, & [a,b],\ b > a \\ 0, & \text{otherwise} \end{cases}$$

$p(x)$:



10/11/25

→ Joint PMF of 2 Random Variables :-

Capital letters - Denote the whole Rv.
Small letters - Denote the values of the Rv.

- $P_{X_1,X_2}(x_1, x_2)$

- Example: Two Dice,

1 \ 2	1	2	3	4	5	6
1	1,1	1,2	1,3	1,4	1,5	1,6
2	2,1	2,2	2,3	2,4	2,5	2,6
3	3,1	3,2	3,3	3,4	3,5	3,6
4	4,1	4,2	4,3	4,4	4,5	4,6
5	5,1	5,2	5,3	5,4	5,5	5,6
6	6,1	6,2	6,3	6,4	6,5	6,6

Joint Events

$$P_{X_1}(X_1 = 1) = \sum_{i=1}^{6} P_{X_1,X_2}(X_1 = 1, X_2 = i)$$

Marginalized Pmf

Marginalizing

Can be extended to Joint Rvs of higher dimensions.

- In a Joint PMF system, the Pmf associated with the indiv. rv's is termed as marginalized PMF.

- The individual rv's PMF can be calculated by marginalization of the joint PMF.

- ## Conditional Probability as Joint PMFs :-

- $P_{X_1,X_2}(X_1 = x_1 \mid X_2 = x_2)$

Define 2 events A, B do be

$$A = \{ \omega : X_1(\omega) = x_1 \ \& \ X_2(\omega) = x_2 \}$$

$$B = \{ \omega : X_2(\omega) = x_2 \}$$

A is the intersection of the 2 events.

B is the given event.

$$P_{X_1,X_2}(X_1 = x_1 \mid X_2 = x_2) = \frac{P(A)}{P(B)} = \frac{P_{X_1,X_2}(X_1 = x_1, X_2 = x_2)}{P_{X_2}(X_2 = x_2)}$$

- ## Independent Rv's :-

Joint pmf = Product of marginal pmf for independent rv's.

$$\Rightarrow P_{X_1,X_2}(X_1 = x_1 \mid X_2 = x_2) = P_{X_1}(X_1 = x_1)$$

Example: 2 Bernoulli Rv's that are independent

$$\chi(X_1) = \{0,1\}$$
$$\chi(X_2) = \{0,1\}$$

1 \ 2	0	1
0	(0,0)	(0,1)
1	(1,0)	(1,1)

similar to 2 coins experiment

$$P_{X_1,X_2}(X_1 = 1 \mid X_2 = 0) = P_{X_1}(X_1 = 1)$$ since independent rv's.

- **IID Rv's :-**

  Independent and identically distributed Rv's.

  1) The set of Rv's are independent.
  2) The probability distribution is the same.

  (rv's of the same situation)

- **Expectation and Variance in Joint Rvs :-**

  - $$E(X_1 \mid X_2 = x_2) = \sum_{x_1 \in X_1} x_1 P(X_1 = x_1 \mid X_2 = x_2)$$

  - The above expectation is a function in X_2.

  - $$E(X_1, X_2) = \sum_{x_2 \in X_2} E(X_1 \mid X_2 = x_2) P_{X_2}(X_2 = x_2)$$

- **Linearity of Expectations :-**

$$E\left(\Sigma(X)\right) = \Sigma\left(E(X)\right)$$

- **Variance of sums of Expectations :-**

$$Var\left(\Sigma(X)\right) = \Sigma\, Var(X)$$

if $\forall X$ are independent

20 | 2 | 25

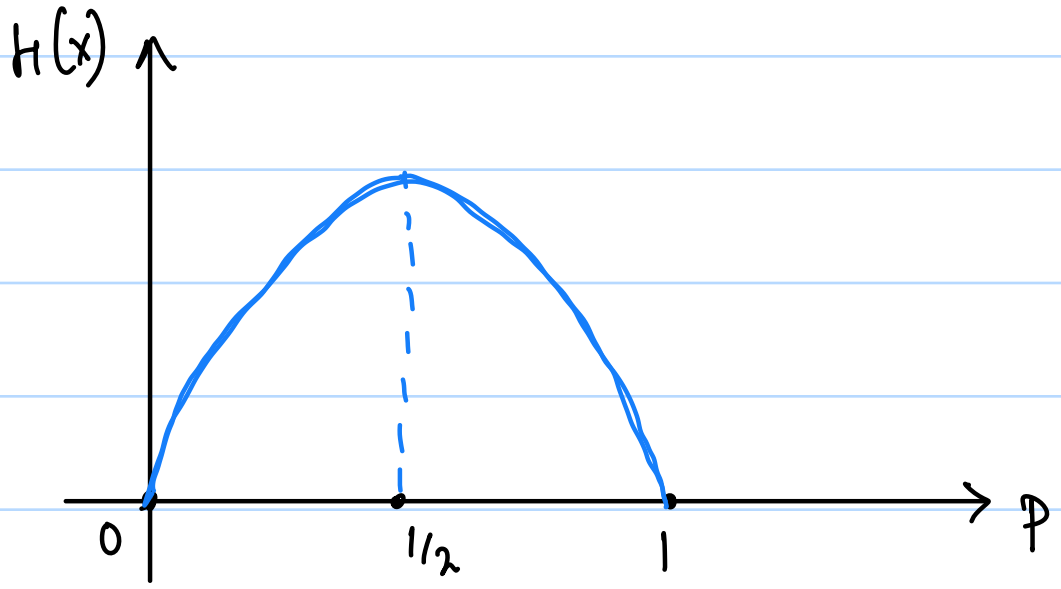
→ Entropy :-

$$H(X) = \sum_{x \in X} P_X(x) \log_2 \frac{1}{P_X(x)}$$

- Entropy of a r.v is the average value of $\log \frac{1}{P_X(x)}$
- It can be thought of as the amount of information stored or the measure of uncertainity.
- Or it can be thought of as the no. of bits / symbols (on average) required to encode an r.v.
- $H(X) \geq 0$ and $H(X) = 0$ iff $\exists\ x \in X$ st $P_X(x) = 1$
- Binary Entropy Function :-

  Entropy of a Bernoulli r.v, ($P_X(0) = 1-p$, $P_X(1) = p$)

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



- **Entropy of a Joint PMF :-**

$$H(X_1, X_2) = \sum_{x_1 \in X_1} \sum_{x_2 \in X_2} P_{X_1 X_2}(x_1, x_2) \log_2 \frac{1}{P_{X_1 X_2}(x_1, x_2)}$$

- **Conditional Entropy :-**

Marginal entropy

$$H(X_2 | X_1 = x_1) = \sum_{x_2 \in X_2} P_{X_2|X_1}(x_2|x_1) \log_2 \frac{1}{P_{X_2|X_1}(x_2|x_1)}$$

$$\Rightarrow \quad H(X_2|X_1) = \sum_{x_1 \in X_1} P_{X_1}(x_1)\, H(X_2 | X_1 = x_1) \rightarrow \text{definition}$$

↳ Average over all consideration of X_1.

and
$$\left.\begin{aligned} H(X_1|X_1) &= 0 \\ H(g(X_1)|X_1) &= 0 \end{aligned}\right\}$$ Once X_1 is given there is no additional information in X_2 (which is $X_1, g(X_1)$ here).

- **Chain Rule of Entropy :-**

- W.k.t $\quad P_{X_1, X_2}(x_1, x_2) = P_{X_2|X_1}(x_2|x_1)\, P_{X_1}(x_1)$

$$\Rightarrow \quad \begin{aligned} H(X_1, X_2) &= H(X_2|X_1) + H(X_1) \\ &= H(X_1|X_2) + H(X_2) \end{aligned}$$

- **Proof:**

$$H(X_1, X_2) = \sum_{x_2} \sum_{x_1} P_{X_2, X_1}(x_2, x_1) \log_2 \frac{1}{P_{X_2 X_1}(x_2, x_1)}$$

$$= \sum_{x_2} \sum_{x_1} P_{X_1|X_2}(x_1|x_2)\, P_{X_2}(x_2) \log \frac{1}{P_{X_1|X_2}(x_1|x_2)\, P_{X_2}(x_2)}$$

$$= \sum_{x_2} P_{X_2}(x_2) \sum_{x_1} P_{X_1|X_2}(x_1|x_2) \left(\log \frac{1}{P_{X_1|X_2}(x_1|x_2)} + \log \frac{1}{P_{X_2}(x_2)} \right)$$

$$= \sum_{x_2} P_{X_2}(x_2) \sum \left(P_{X_1|X_2}(x_1|x_2) \log \frac{1}{P_{X_1|X_2}(x_1|x_2)} + P_{X_1|X_2}(x_1|x_2) \log \frac{1}{P_{X_2}(x_2)} \right)$$

$$= \sum_{x_2} P_{X_2}(x_2) \left(H(X_1|X_2 = x_2) + \sum P_{X_1|X_2}(x_1|x_2) \log \frac{1}{P_{X_2}(x_2)} \right)$$

$$= \sum_{x_2} P_{X_2}(x_2) H(X_1|X_2 = x_2) + \sum_{x_2} \log \frac{1}{P_{X_2}(x_2)} \sum_{x_1} P_{X_1|X_2}(x_1|x_2)$$

$$= H(X_1|X_2) + \sum \log \frac{1}{P_{X_2}(x_2)} P_{X_2}(x_2)$$

$$P_{X_2}(x_2) = \sum P_{X_1|X_2}(x_1|x_2)$$

$$= \underline{H(X_1|X_2) + H(X_2)}$$ Hence Proved

- $H(X_1, X_2, X_3) = H(X_1) + H(X_2|X_1) + H(X_3|X_2, X_1)$

- If 2 r.v's X_1, X_2 are independent,

$$H(X_1|X_2) = H(X_1)$$
$$H(X_2|X_1) = H(X_2)$$

] Information about one does not affect the information about another.

$$\Rightarrow \quad H(X_1, X_2) = H(X_1) + H(X_2)$$

Example: Given below is the table of Joint pmf of X and Y.

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

Find $H(X)$, $H(Y)$, $H(X|Y)$, $H(Y|X)$

W.k.t $P_X(X=x) = \sum_{y \in Y} P_{X,Y}(X=x, Y=y)$

$P_X(X=1) = 1/8 + 2/16 + 1/4 = 1/2$

$P_X(X=2) = 1/8 + 2/16 = 1/4$

$P_X(X=3) = 1/16 + 2/32 = 1/8$

$P_X(X=4) = 1/16 + 2/32 = 1/8$

} Sum of these is 1 so valid.

$H(X) = \sum_{x \in X} P(x) \log_2 \frac{1}{P_X(x)} = 1/2 \log_2 2 + 1/4 \log 4 + 1/8 \log 8 + 1/8 \log 8$

$= 1/2 + 1/4 \cdot 2 + \frac{1}{8} \cdot 3 + 1/8 \cdot 3$

$= 7/4$

$P_Y(Y=1) = 1/8 + 1/16 + 2/32 = 1/4$

$P_Y(Y=2) = 1/8 + 1/16 + 2/32 = 1/4$

$P_Y(Y=3) = 4/16 = 1/4$

$P_Y(Y=4) = 1/4$

} Sum is 1 is valid.

$$H(Y) = \sum_{y \in Y} P_Y(y) \log_2 \frac{1}{P_Y(y)}$$

$$= 4 \times \frac{1}{4} \log_2 4 = 1 \times 2$$

$$= \underline{\underline{2}}$$

$$H(X,Y) = \sum_x \sum_y P_{X,Y}(x,y) \log_2 \frac{1}{P_{X,Y}(x,y)}$$

$$= \frac{1}{8} \log 8 \times 2 + \frac{1}{16} \log 16 \times 6 + \frac{1}{32} \log 32 \times 4 + \frac{1}{4} \log 4$$

$$= 3/4 + 6/4 + 5/8 + 1/2$$

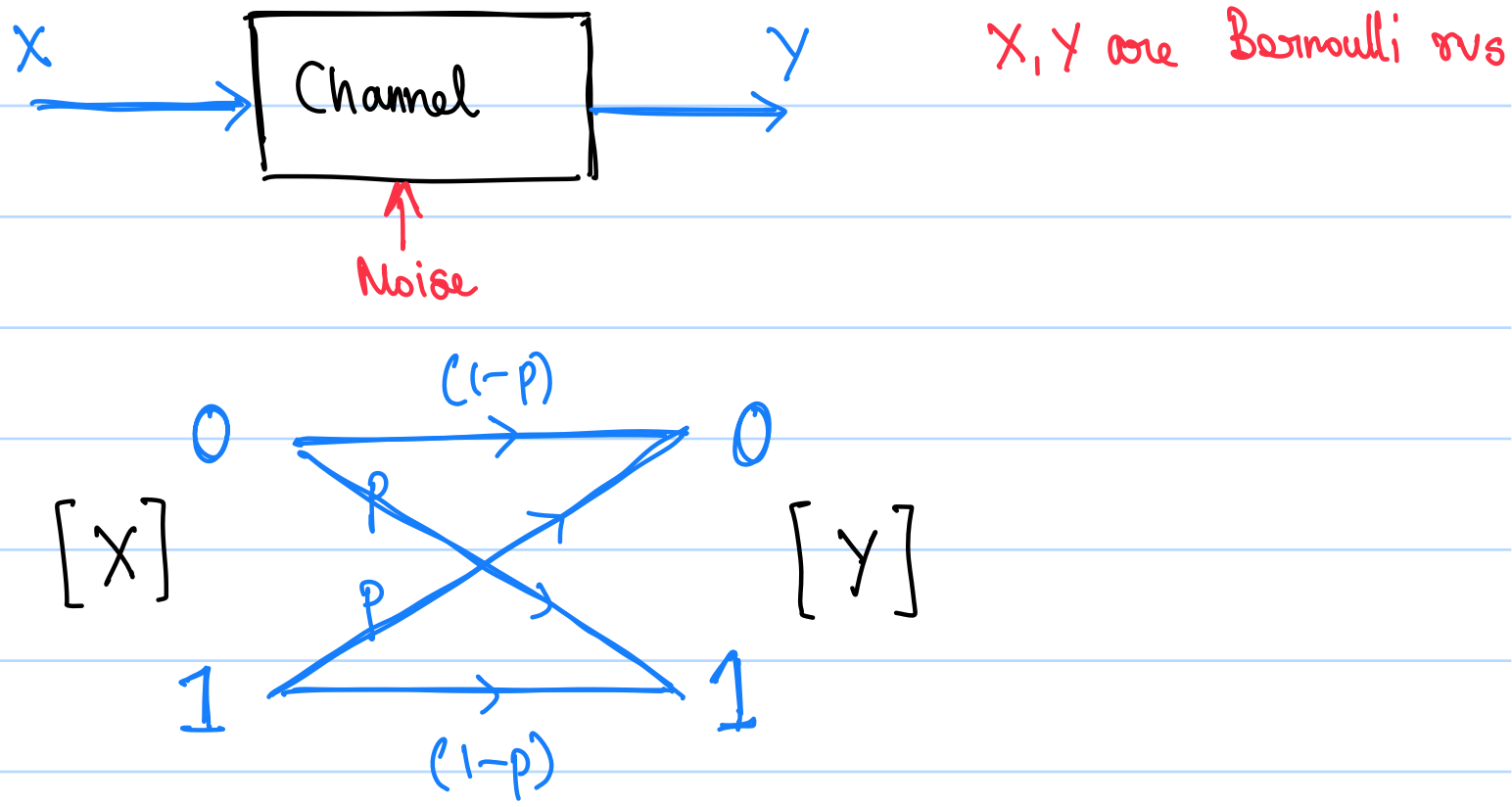
$$= 6/8 + 12/8 + 5/8 + 4/8$$

$$= 27/8$$

→ Mutual Information :-

- Entropy of a r.v is the measure of uncertainity of the r.v.

- A problem in communication is,



The given channel is said to be binary symmetric if

$$P_{Y|X}(0,0) = P_{Y|X}(1,1) = (1-p)$$

↳ p = Probability of flipping

The channel is therefore, conditioned by a pmf.

The problem is to figure out what is the maximum amount of data that can be transferred across this channel, if given $P_{X,Y}(x,y)$.

- Mutual Information is defined as,

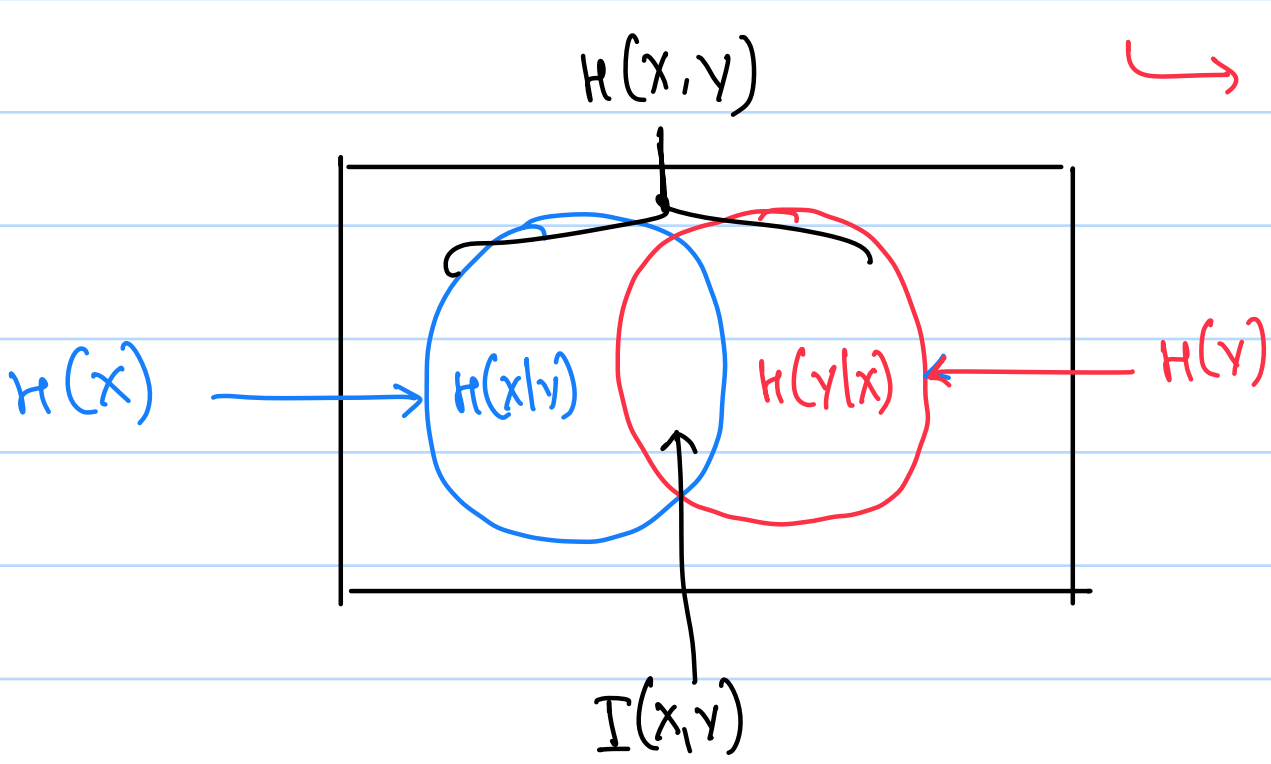
$$I(X,Y) = H(X) - H(X|Y) \rightarrow$$ Amount of info Y is telling about X.

$H(X)$ ↑ Amount of uncertainity in X

$H(X|Y)$ ↑ Amount of uncertainity in X even after observing Y.

$$I(Y,X) = H(Y) - H(Y|X) \rightarrow$$ Amount of info X is telling about Y.

↳ Can be proved using above formulae.



$$I(X_1, X_2, X_3, X_4) = H(X_1) - H(X_1 | X_2, X_3, X_4)$$

- If X, Y are independent, $I(X,Y) = 0$

→ Relative Entropy :-

why is $P \neq Q$, isn't it the same r.v ???

- Measure of "distance" b/w 2 pmfs.

- If P and Q are 2 Pmfs over the same r.v X.

$$D(P||Q) = \sum_{x \in X} P(x) \log_2 \frac{P(x)}{Q(x)}$$

- Properties of a 'proper' distance metric:

  - $d(x,y) \geq 0$ & $d(x,y) = 0 \Rightarrow x = y$
  - $d(x,y) = d(y,x)$
  - $d(x,z) \leq d(x,y) + d(y,z)$

- But Relative entropy does not satisfy properties 2, 3.

- Relation b/w Relative Entropy & Mutual Information :-

$$I(X,Y) = D(P_{X,Y} || P_X P_Y)$$

$$D(P_{X,Y} || P_X P_Y) = \sum_{x \in X} \sum_{y \in Y} P_{X,Y}(x,y) \log_a \frac{P_{XY}(x,y)}{P_X(x) P_Y(y)}$$

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

$\Rightarrow$ $I(X,Y) \geq 0$ & $I(X,Y) = 0 \Rightarrow X, Y$ is indp.

↳ $H(X,Y) = H(X) + H(X)$

$\Rightarrow$ $I(X,Y) = H(X) - H(X|Y) \geq 0$

$\Rightarrow$ $H(X|Y) \leq H(X)$ → Conditioning a r.v decreases the uncertainity

- Entropy is maximum in a uniformly distributed prnf.
  ↳ $V(X)\uparrow$, $H(X)\uparrow$
  ↳ Cannot guess which event is more probable.

3/3/25

Textbook: Elements of information theory.

→ Intro to IT Summary :-

- Entropy <=> Uncertainity <=> Information (Unit is bits)
- If the entropy of a r.v is x, then it implies that we need a minimum of x bits to represent the information associated w/ the r.v.
- w.k.t

$$H(X|Y) = \sum_{y \in Y} H(X|Y=y)\, P_Y(y)$$

$$H(X,Y) = \sum_{x \in X} \sum_{y \in Y} P_{X,Y}(x,y) \log_2 \frac{1}{P_{X,Y}(x,y)}$$

- Given 2 r.v's X and Y,

$$\text{Product pmf} = P(X=x)\,P(Y=y)$$ is also a valid pmf

$$\text{Relative Entropy} = D(P_1||P_2) = \sum_{x\in X} P_1(x)\log_2\frac{P_1(x)}{P_2(x)}$$ one of multiple ways to define a "distance"

- The relation b/w Relative Entropy and Mutual Information is,

$$I(X,Y) = D(\text{joint pmf}(X,Y)\,||\,\text{prod. pmf}(X,Y))$$

When X, Y are independent,

$$I(X,Y) = D(\text{prod. pmf}(X,Y)\,||\,\text{prod. pmf}(X,Y)) = \underline{0}$$

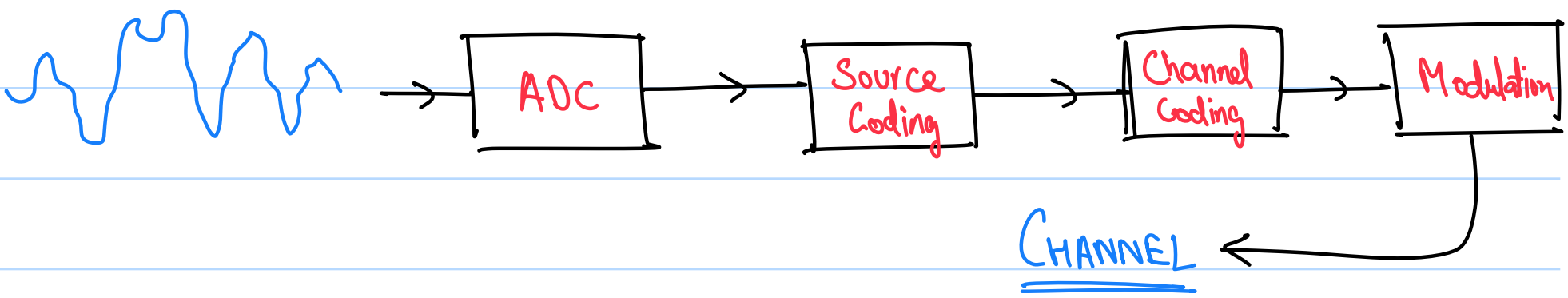
i.e., they are not sharing any information.

When X, Y are dependent,

$I(X,Y) \neq 0$, i.e., they share some information

→ Source Coding :-

- Block diagram of a Transmitter.



- Source coding is responsible for data compression.

○ Data Compression :-

- Suppose the file is

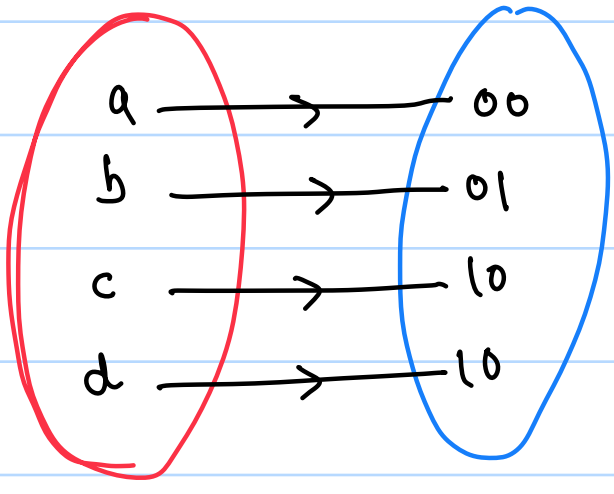
  acbaaadaabbaabcd (Ln : 16)

- One way of encoding the file is

  a = 00 b = 01 c = 10 d = 11 (Naive Approach)

  This uses 32 bits

- However this can be represented by less than 32 bits, by compressing the binary string.



00, 01, 10, 11 are the codewords of a, b, c, d respectively

- Defn of a source code,
  - Define a r.v X with support set $\mathcal{X}$.
  - Let D^* be the set of finite length strings of symbols from a D-ary alphabet. Assume the D-ary alphabet is to be $\{0, 1, \ldots, D-1\}$
  - A source code is defined as a mapping from $\mathcal{X}$ to D^*.
  - $C(x)$: Codeword of $x \in \mathcal{X}$.
  - $l(x)$: Length of $C(x)$.

Same as a D-base number system

$C : \mathcal{X} \rightarrow D^*$

- **Parameters of a Source Code:**

a c b a a a d a a b b a a b c d

- Let the file be rep.d by an r.v X.

- $P(X=a) = \frac{1}{2}$
  $P(X=b) = \frac{1}{4}$
  $P(X=c) = \frac{1}{8}$
  $P(X=d) = \frac{1}{8}$

- The expectation of a general function is given by,

$$E(f(X)) = \sum_{x \in X} f(x)\, p_X(x)$$

- The expected length of a source code is given by,

$$E(l(X)) = \sum_{x \in X} l(x)\, p_X(x)$$

$$\Rightarrow L(C) = \sum_{x \in X} l(x)\, p_X(x)$$

- Rate = Total No. of Bits after Source Coding / No. of symbols

- In a good source coding we want $L(C)$ to be as low as possible

- Note: A source code that maps 2 symbols to the same codeword can also be useful, if the other symbol appears very rarely.

- Defn of a Good Source Code,

1) $L(C)$ should be low for good compression
2) Should not loose the info contained in the r.v X.

$\Rightarrow$ $L(C) \geq H(X)$ $\longrightarrow$ Source Coding Theorem

- A source code is said to be optimal if $L(C) = H(X)$

- Source Coding Theorem holds only for lossless source coding.

Example: Pmf of X is given by,

$$P(X=a) = \frac{1}{2},\ P(X=b) = \frac{1}{4},\ P(X=c) = \frac{1}{8},\ P(X=d) = \frac{1}{8}$$

Construct optimal source code.

A:
$$H(X) = \frac{1}{2}\log 2 + \frac{1}{4}\log 4 + 2\left(\frac{1}{8}\log 8\right)$$
$$= \frac{1}{2}(1) + \frac{1}{4}(2) + \frac{1}{4}(3)$$
$$= \frac{1}{2} + \frac{1}{2} + \frac{3}{4}$$
$$= \frac{7}{4} = 1.75$$

Let $C(A) = 0$, $C(B) = 1$, $C(C) = 00$, $C(D) = 11$

$$L(C) = (1)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{4}\right) + 2\left(\frac{1}{8}\right) + 2\left(\frac{1}{8}\right)$$

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

$$= \underline{1.25}$$

- $x \xrightarrow{\text{encoding}} C(x) \xrightarrow{\text{decoding}} \dot{x}$

- A source code is said to be non singular if

$$x \neq x' \Rightarrow C(x) \neq C(x') \quad \forall\ x, x' \in \mathcal{X}$$

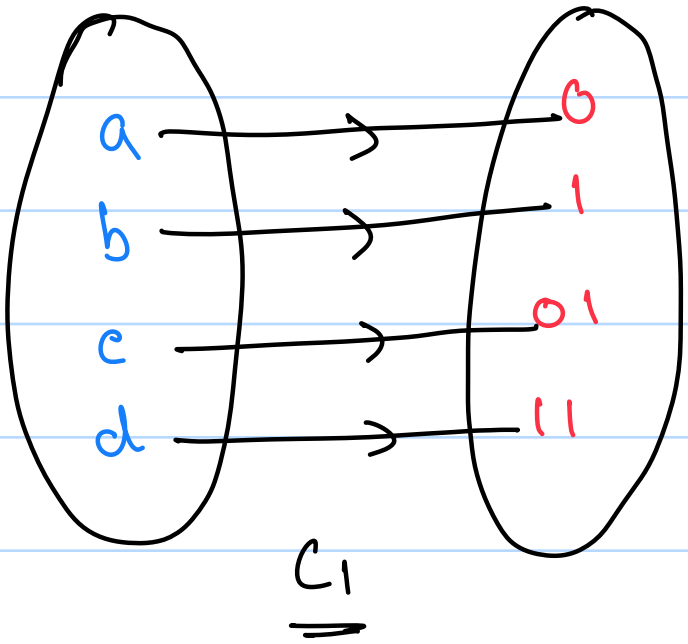
ie, every element of $\mathcal{X}$ maps to a different string in $\mathcal{D}^*$

$C(a) = 0$, $C(b) = 1$, $C(c) = 01$, $C(d) = 11$ → A singular code.

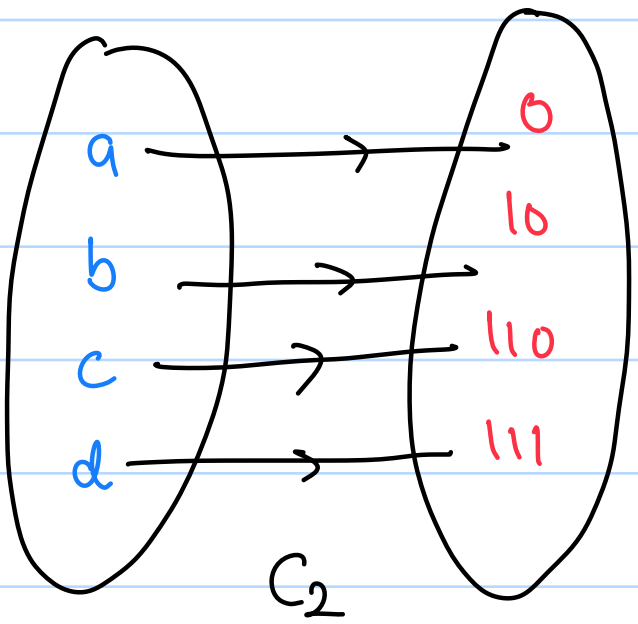
Contradictory to common sense, singular codes also have a lot of use cases. (Ex: In above, c and d have very low probability of occuring, ie. $P(a) >> P(c)$, $P(b) >> P(d)$)

- Unique Decodability of a Source Code:

- Define 2 source codes as,



C_1



C_2

- abca

C_1 : 01010

↓

ababa
caba
abca
⋮

} C_1 is not uniquely decodable

C_2 : 0101100

↓

abca

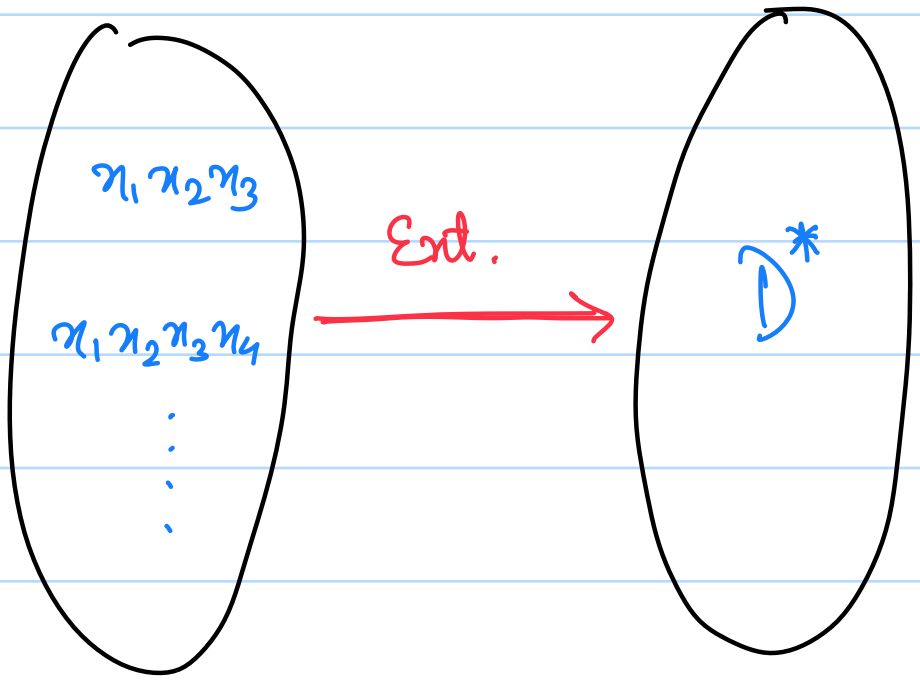
C_2 is uniquely decodable

- In C_2, each codeword is not a prefix of any of the other codewords.

- $c(x_1 x_2) = c(x_1) \cdot c(x_2)$ ← Concatenation

- The "extension" of a source code C is defined as the mapping from finite length strings of X to finite length strings of D^*.

Only finite length strings are considered since the unique decodability of ∞ strings can't be checked



- A source code is singular iff its extension is non singular.

Example: For a file containing a, b, c's, check if the following source codes are uniquely decodable.

1) $\{0, 10, 11\}$ ⟶ Yes since it follows the prefix property.

2) $\{0, 01, 11\}$ ⇒ Yes, since we can point out decoding errors.

bcc ⟶ 01111

3) $\{0, 01, 10\}$

ba ⟶ 010 → ba, ac ⇒ No

4) $\{0, 01\}$ ⇒ Yes since if 1 appears, we can say its a b.

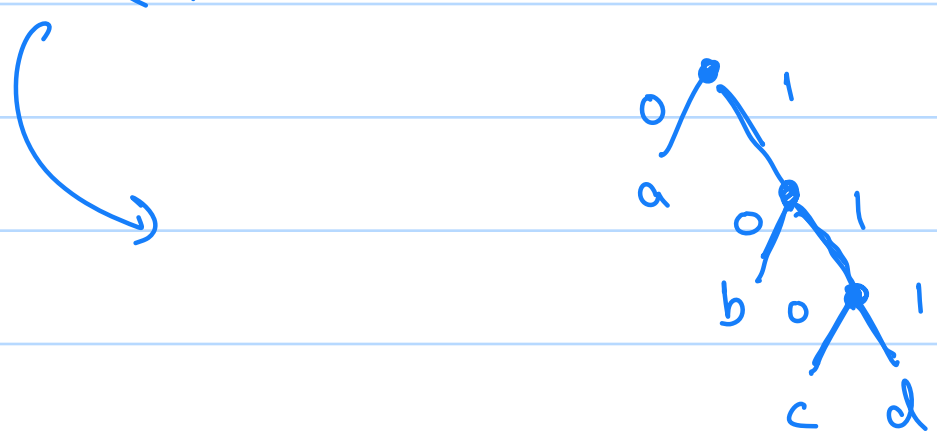
5) $\{0, 01, 10, 11\}$ ⇒ No since $\{0, 01\}$ is not.

6) $\{110, 11, 10\}$ 1110 ⇒ 110|1110 1110|10 ⇒ Yes

1011 ⇒ 110|011 10|11|10

- A source code that follows the prefix property is termed as a prefix code / instantaneous source code / Self punctuating code.

$C(a) = 0, \quad C(b) = 10, \quad C(c) = 110, \quad C(d) = 111$



- Prefix Code ⇒ Uniquely Decodable
  Uniquely Decodable ⇏ Prefix Code.

→ Optimal Code Design :-

1. Must be uniquely decodable.
2. L(C) should be minimum
3. Encoding / Decoding should be efficient.

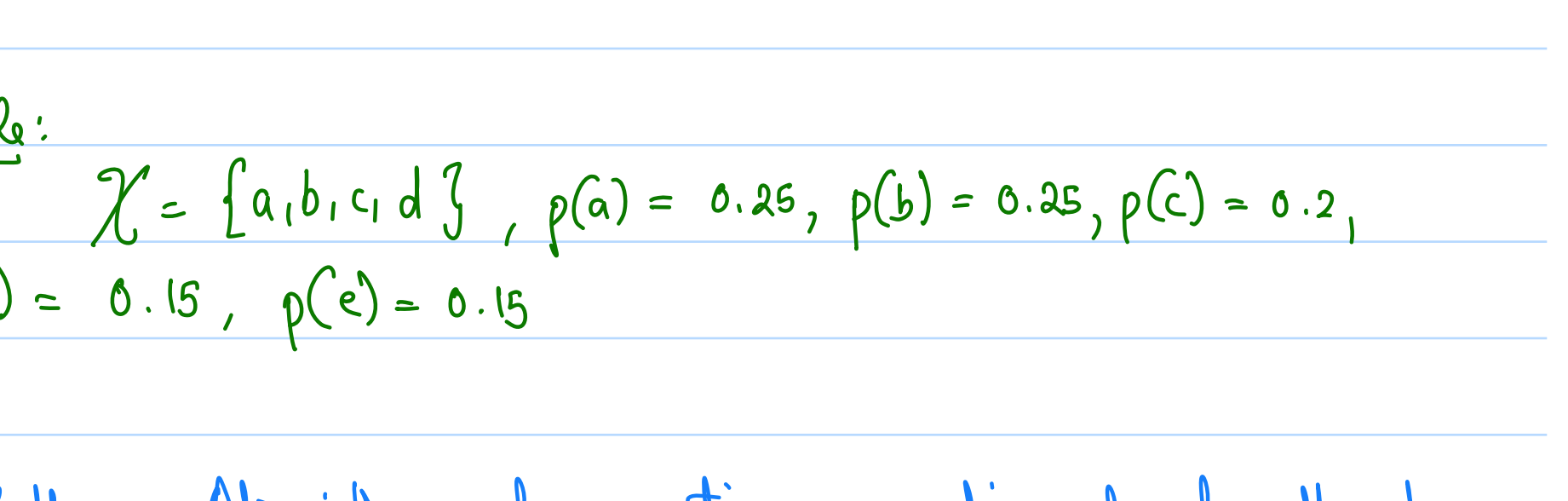
→ Huffman Algorithm :-

Example:

$X = \{a, b, c, d\}$, $p(a) = 0.25$, $p(b) = 0.25$, $p(c) = 0.2$,
$p(d) = 0.15$, $p(e) = 0.15$

- A Huffman Algorithm for creating a prefix code for the above scenario is as follows.

In decreasing order of $p(x)$,



C(a) = 00, C(b) = 10, C(c) = 11, C(d) = 010, C(e) = 011

Note: ASCII is a source code.

- Huffman encoding is used in JPEG compression.
- Lempel-Ziv Algorithm is used in Zip files.

To find Entropy, Expected length of the prev. scenario.

$$H(X) = 2\left(\frac{1}{4}\log 4\right) + \frac{1}{5}\log 5 + \left(\frac{3}{20}\log\frac{20}{3}\right)2$$

$$= \frac{1}{\ln 2}\left(\frac{1}{2}\ln 4 + \frac{1}{5}\ln 5 + \frac{3}{10}\ln\frac{20}{3}\right)$$

$$\approx 2.285$$

$$L(C) = \frac{1}{4}(2) + \frac{1}{4}(2) + \frac{1}{5}(2) + \frac{3}{20}(3) + \frac{3}{20}(3) = \underline{2.3}$$

$$\Rightarrow \text{Efficiency} = \frac{2.285}{2.3} = 0.9934$$

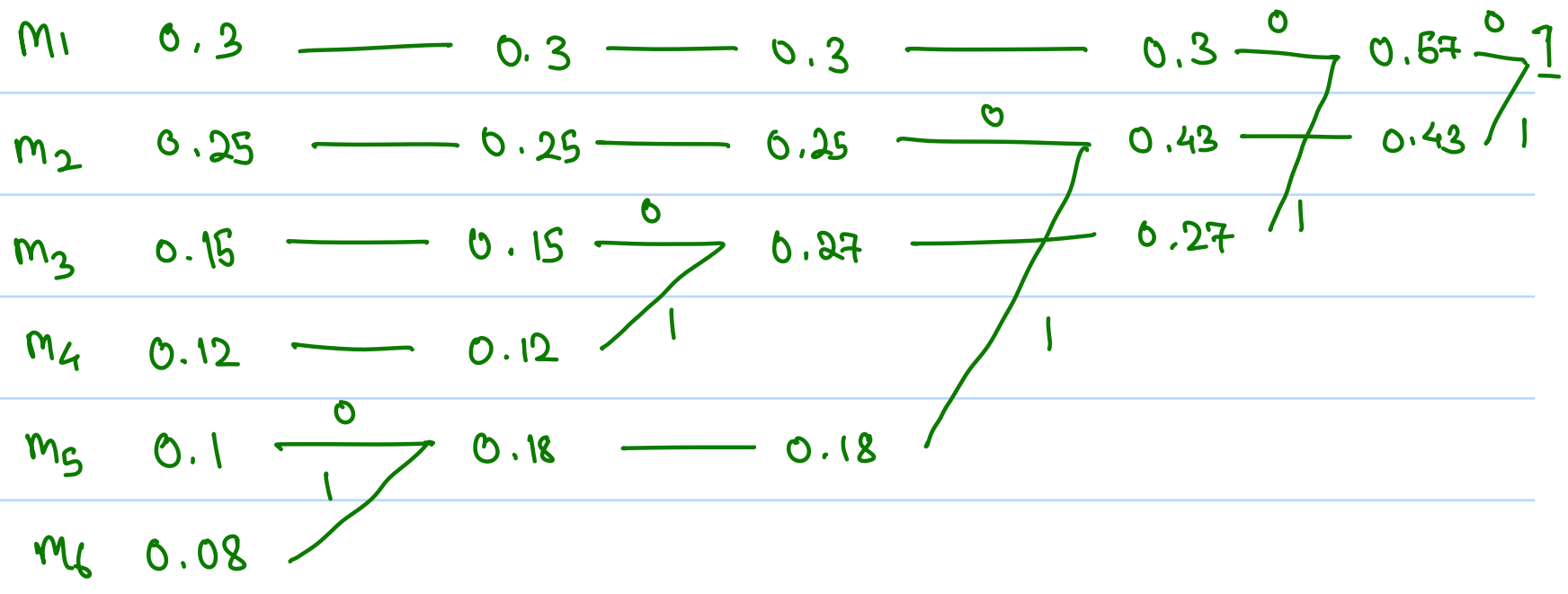
Example: Construct a ternary code for.

$$X = \{m_i \mid i \in [1,6], i \in \mathbb{N}\}$$

$p(m_1) = 0.3$ $p(m_2) = 0.25$ $p(m_3) = 0.15$ $p(m_4) = 0.12$ $p(m_5) = 0.1$

$p(m_6) = 0.08$

Binary code,



$C(m_1) = 00 \quad C(m_2) = 10 \quad C(m_3) = 010 \quad C(m_4) = 011$

$C(m_5) = 110 \quad C(m_6) = 111$

$$L(C) = 0.3(2) + 0.25(2) + 0.15(3) + 0.12(3) + 0.1(3) + 0.08(3)$$

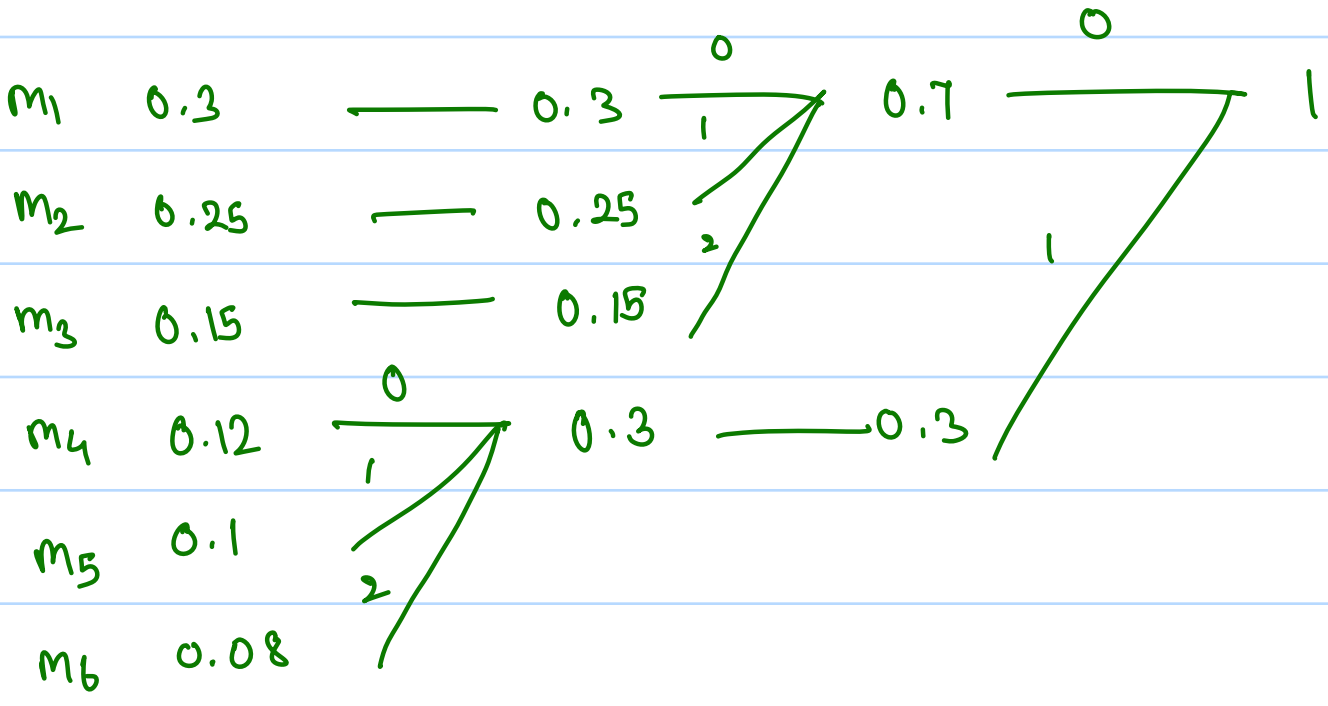
$$= 2.45$$

$$H(X) = \frac{1}{\ln 2}\left(0.3 \ln \frac{10}{3} + 0.25 \ln 4 + 0.15 \ln \frac{20}{3} + 0.12 \ln \frac{25}{3} + 0.1 \ln 10 + 0.08 \ln \frac{25}{2}\right)$$

$$= 2.422$$

$$\text{Eff} = 0.988$$

Ternary Code:



$C(m_1) = 00 \quad C(m_2) = 01 \quad C(m_3) = 02 \quad C(m_4) = 10$

$C(m_5) = 11 \quad C(m_6) = 12$

$$L(C) = \underline{2}$$

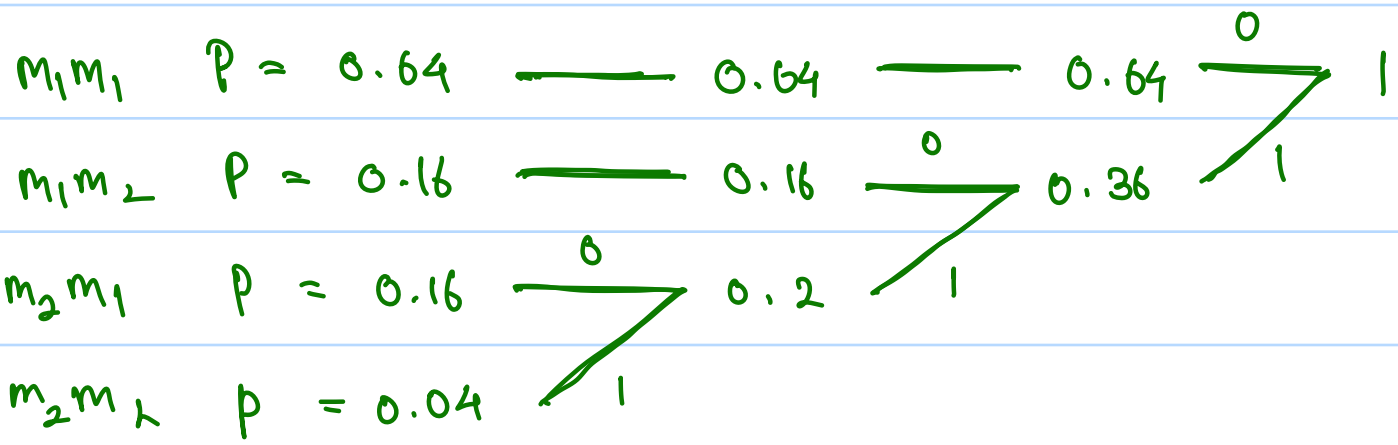
$$H(X) = H_2(X) \times \frac{\ln 2}{\ln 3} = 1.528$$

$$\text{Eff} = \frac{1.528}{2} = 0.764$$

- The key idea of the Huffman algorithm is to give more frequently occuring symbols, shorter strings, which gives us good coding efficiency.
- For a given pmf, Huffman codes will have the lowest expected length, among the uniquely decodable codes.
- Huffman codes are also prefix codes.

- To find for extensions,

Example: Find the Huffman code for the second order extension of $\{m_1, m_2\}$. $P(m_1) = 0.8$, $P(m_2) = 0.2$



$C(m_1m_1) = 0 \quad C(m_1m_2) = 10 \quad C(m_2m_1) = 110 \quad C(m_2m_2) = 111$

$$L(C) = 0.64(1) + 0.16(2) + 0.16(3) + 0.04(3)$$

$$= 1.56 \longrightarrow \text{For 2nd order extension}$$

$$\Rightarrow L(C) = 0.78 \longrightarrow \text{Per symbol}$$

$$H(X) = \frac{1}{\ln 2}\left(0.8 \ln \frac{10}{8} + 0.2 \ln \frac{10}{2}\right)$$

$$= 0.7219$$

$$H(M_1M_2) = H(M_1 | M_2) + H(M_2)$$

$$=$$

$$\text{Efficiency} = 0.9255$$

- $H(X) \leq L(C) \leq H(X) + 1$

- For n^{th} order extension, $H(X^n) \leq L(C) \leq H(X^n) + 1$

- If $H(X^n) = nH(X)$ holds, then X is said to be memoryless
  Each symbol in X is independent of the other symbols. ↓ Also $L(C) \geq \frac{L(C^n)}{n}$

→ Lempel Ziv Coding :-

Consider the sequence 1010110110101010

- Parsing: Identify phrases of the smallest length that haven't appeared before.

1010110110 1 0 1 0 1 0

1010110110 → 1, 0, 10, 11, 01, 0 − ϕ
1 2 3 4 5

- Once parsing has been done, the phrases are compiled into a dictionary

Dictionary Location	Phrase	Encoding
0	ϕ	
1	1	(0,1)
2	0	(0,0)
3	10	(1,0)
4	11	(1,1)
5	01	(2,1)
6	101	(3,1)
7	010	(5,0)
8	1010	(6,0)

⇒ 1 0 1 0 1 1 0 1 1 0 1 0 1 0 1 0 = 0 (0,1)(0,0)(1,0)(1,1)(2,1)(3,1)(5,0)(6,0)

- This algorithm is inefficient for small files but very efficient for large files and also for decoding, ie, its a practical coding algorithm.

10|3|25

- Decoding Lempel-Ziv Coding :-

Example: (0,1), (0,0), (1,0), (1,1), (2,1), (3,1), (5,0)
1 2 3 4 5 6 7

location	Content
0	ϕ
1	1
2	0
3	10
4	11
5	01
6	101
7	010

(Recursive approach)

=) 1010110101010

→ Kraft Inequality :-

(prefix code)

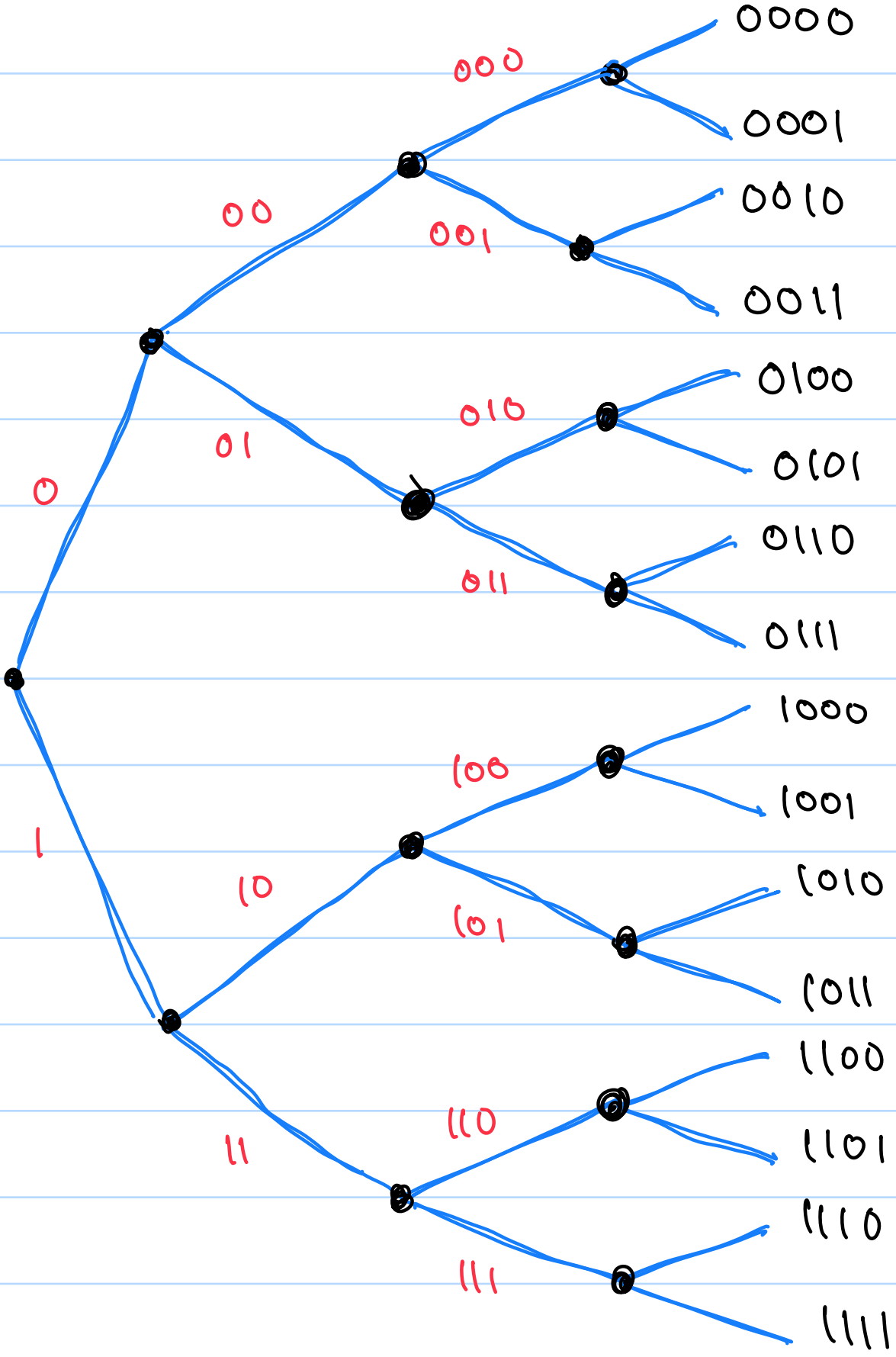
- Suppose we wish to construct an instantaneous code
- Let $l_1, l_2 \ldots l_m$ be the codeword lengths of the source code.

- Kraft Inequality,

$$\sum_{i=1}^{m} 2^{-l_i} \leq 1$$

Proof:

Suppose $l_1 \leq l_2 \leq l_3 \leq l_4 \ldots \ldots \leq l_m$



If $l_m = 4$, this binary tree can be used in the proof.

- In a prefix code, if one codeword is valid, then all of its descendants in the above binary tree are invalid., ie,

  if 01 is valid 0100, 0101, 0110, 0111 are all invalid.

- For a general case, assume that the binary tree is of depth l_m.

- The binary strings at level l_m are either codewords, descendants of a codeword or neither.

- If we have a binary string of level l_i, it will have $2^{l_m - l_i}$ descendants at l_m.

$$\sum_{i=1}^{n} 2^{l_m - l_i} \leq 2^{l_m}$$

since 2^{l_m} is the total. no. of descendants.

$$= \frac{1}{2^{l_m}} \sum_{i=1}^{n} 2^{l_m - l_i} \leq 1$$

$$\Rightarrow \sum_{i=1}^{n} 2^{-l_i} \leq 1$$

- Theorem:

  The average length of any prefix-free code is lower bounded by the entropy of the source, ie.

  $$L(C) \geq H(X)$$

Proof:

$$L(C) = \sum_{i=1}^{m} l_i\, p(x_i) \qquad H(X) = \sum_{i=1}^{m} p(x_i) \log_2 \frac{1}{p(x_i)}$$

$$L(C) - H(X) = \sum_{i=1}^{m} l_i\, p(x_i) + \sum_{i=1}^{m} p(x_i) \log_2 p(x_i)$$

We know that,

1) $l_i = \log_2 \frac{1}{2^{-l_i}}$ 2) $\sum_{i=1}^{m} \frac{1}{2^{l_i}} \leq 1$

3) $D(p\|q) = \sum_{i=1}^{m} p(x_i) \log \frac{p(x_i)}{q(x_i)}$

$$\Rightarrow \sum_{i=1}^{m} l_i\, p(x_i) = \sum_{i=1}^{m} \log_2 \frac{1}{2^{-l_i}}\, p(x_i) \geqslant \sum_{i=1}^{m} \log \frac{\sum_{j=1}^{m} 2^{-l_j}}{2^{-l_i}}\, p(x_i)$$

$$\sum_{i=1}^{m} \log \frac{\sum_{j=1}^{m} 2^{-l_j}}{2^{-l_i}}\, p(x_i) = \sum_{i=1}^{m} p(x_i) \log \frac{\frac{1}{2^{l_i}}}{\sum_{j=1}^{m} 2^{-l_j}}$$

$$\Rightarrow L(C) - H(X) \geqslant \sum_{i=1}^{m} p(x_i) \log \left(\frac{\frac{p(x_i)}{2^{-l_i}}}{\sum_{j=1}^{m} 2^{-l_j}} \right)$$

$$\geqslant D\left(p(x_i) \,\|\, \frac{2^{-l_i}}{\sum_{j=1}^{m} 2^{-l_j}}\right) \geqslant 0$$

→ Source Coding Theorem :-

- As we use Huffmann encoding over larger and larger sets of symbols, the efficiency approaches 1, ie, $L(C) \longrightarrow H(X)$.

- If the source X has support set $\mathcal{X}$, then the extension X^n has the support set $\mathcal{X}^n$.

  $\forall \; x^n \in \mathcal{X}^n$, $f^n(x^n)$ is its codeword.

- Codeword $f^n(x^n)$ is decoded as $g^n(f^n(x^n)) = \hat{x}^n$.

- If $\hat{x}^n \neq x^n$, we say that decoding error has occured.

- Define a source code as (f^n, g^n).

- Source Coding Theorem states that, (Achievability Statement)

  If $L(C) > H(X)$, then $\exists$ a sequence of codes $\{f^n, g^n\}$ of average length $L(C)$ such that,

  $$P(\hat{X}^n \neq X^n) \longrightarrow 0 \text{ as } n \longrightarrow \infty$$

- $P_e^n := P(\hat{x}^n \neq x^n) \longrightarrow$ Probability of decoding error.

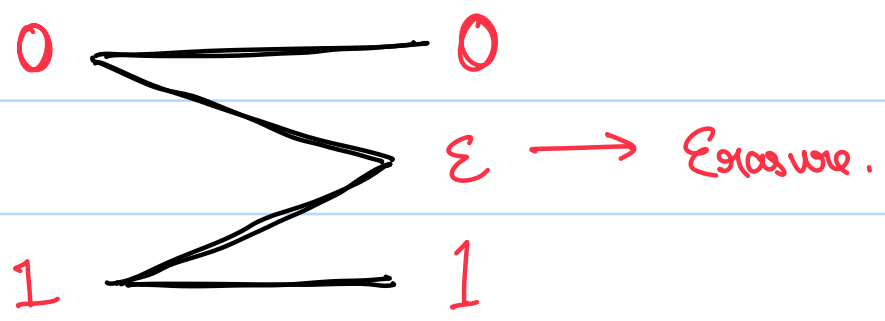
- Converse statement,

  If $L(C) < H(X)$, then $P_e^n > 0 \; \forall \; n$

Average length is also termed as data compression factor

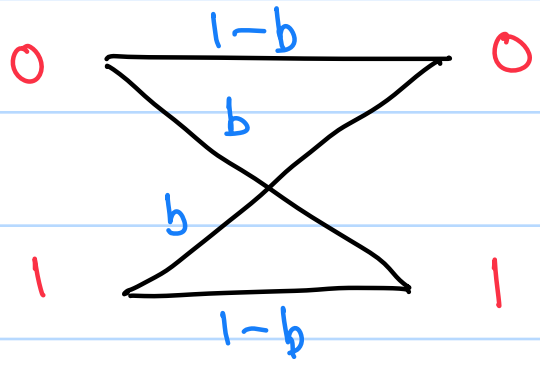
→ Channel Encoding :-

- Channel encoding is also termed as error correction code.
- Messages now are just sequences of bits, thanks to source coding.
- Binary Erasure Channel :-



- A channel such that each bit is either transmitted correctly or erased.
- $P(Y=0|X=1) = P(Y=1|X=0) = 0$ (ie, no bit flip)
- $P(Y=\varepsilon|X=0) = P(Y=\varepsilon|X=1) = \varepsilon$ (Probability of Erasure)
- To reduce the loss of information, the same signal can be sent multiple times.

- Binary Symmetric Channel :



- The probability of a bit flip does not depend on the value of the bit.

- $P(Y=0|X=1) = P(Y=1|X=0) = b$

- We can use Parity codes to find bit flips.

- If $m_1, m_2, m_3 \ldots\ldots m_n$ is our string of bits, whose each $m_i \in \{0,1\}$ $\forall i$, then,

$$\underbrace{m_1 \oplus m_2 \oplus m_3 \oplus \cdots m_n = 1 \Rightarrow \text{No. of 1's is odd}}_{\text{Parity Check Equation}}$$
$$= 0 \Rightarrow \text{No. of 1's is even}$$

- Single Parity Check Codes :-

The parity of the whole message is taken and the parity information is encoded into a "parity bit"

- **Purpose of Channel Codes:**

  To prevent the corruption of transmitted information by channel noise.

- **Types of Channel Codes:**

  1) Cyclic Codes 2) Convolutional Codes 3) LDPC Codes 4) Turbo Code
  5) Polar Codes 6) Reed Solomon Codes

→ **Orthogonal Subspaces :-**

V and W are said to be orthogonal to each other if $\forall\ v \in V$ & $w \in W$, $v \cdot w^T = 0$

(↑ inner product)

- **Our Focus:-** Vector space spanned by vectors over the field $F = \{0, 1\}$
  - ↳ For n vectors, there will be 2^n possible linear combinations.
  - ↳ Follows binary addition (XOR) and multiplication (AND)

→ **Binary Linear Block Codes :-**

- **REP-3 Codes:** Each bit is repeated 3 times

  $C(0) = 000$

  $C(1) = 111$

$$\chi_{REP-3} = \{000, 111\}, \quad \chi_{SPC-3} = \{000, 011, 101, 110\}$$

→ verify later if correct

- SPC-3: Every 3rd bit is the XOR of the first 2 bits

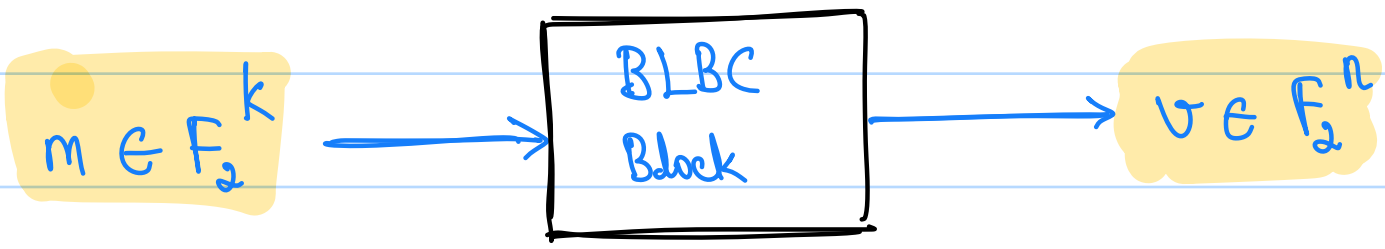
$$C(00) = 000, \quad C(01) = 011, \quad C(10) = 101, \quad C(11) = 110$$

- A binary linear block, denoted by $C(n,k)$, with parameters n and k is defined as a subspace of F_2^n with dimensions k.

- REP-3 and SPC-3 are subspaces of F_n^3 with dimensions 1 and 2 respectively.

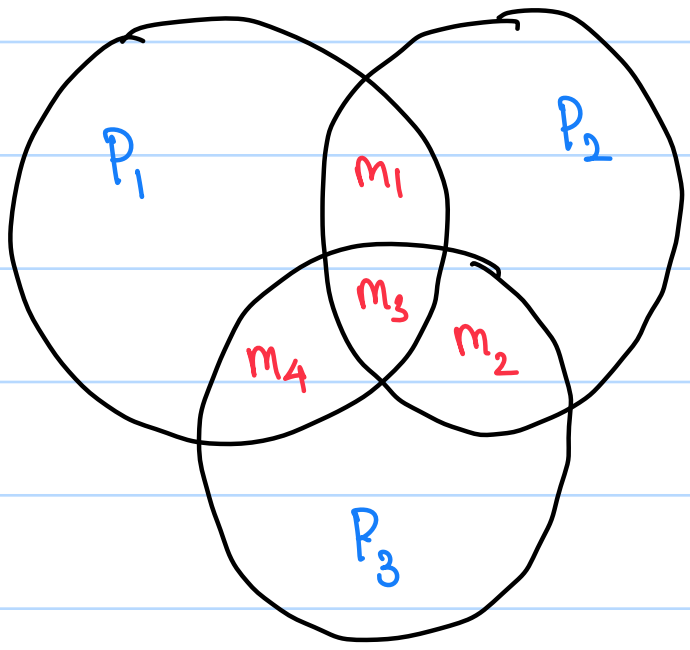
- The size of a code $C(n,k)$ is the cardinality of the subspace, ie, the number of codewords.

- Length of codeword = n since the coding, is a subspace of the vector space F_2^n.



- Rate = Initial Size / Final Size = k/n

∘ Hamming Codes :-



P - parity bits

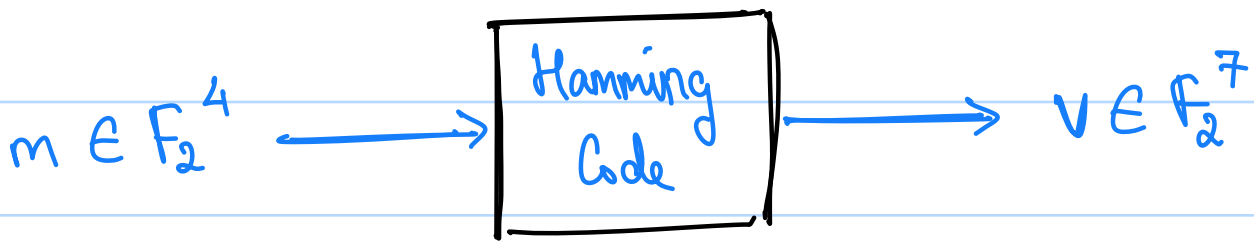
m - message bits

- Message vector $m = \{m_1, m_2, m_3, m_4\}$

- We define 3 parity bits P_1, P_2, P_3 as,

$$P_1 = m_1 \oplus m_3 \oplus m_4$$
$$P_2 = m_1 \oplus m_2 \oplus m_3 \Rightarrow v = \{m_1, m_2, m_3, m_4, P_1, P_2, P_3\}$$
$$P_3 = m_2 \oplus m_3 \oplus m_4 \qquad v \in F_2^7$$

- Hamming Codes can correct single bit errors. The parity bits can be located anywhere in the encoded sequence.



∘ Any subspace of F_2^n will give us a BLBC. The dimension of that subspace will be k.

↳ Only the message bits are indp.

BLBC ⇒ Binary ⟶ Since F is a binary field

Linear ⟶ $C(m_1+m_2) = C(m_1) + C(m_2)$

Block ⟶ Block of k bits ⟼ Block of n bits

Code

Note: Collection of codewords = Codebook

- For each n, k, the no. of BLBC's possible is,

The no. of subspaces of $\mathbb{F}_2^n$ of dim = k, is ${}^nC_k$, since the any basis vector of the BLBC should have k indp. variables among n total variables.

- Generator Matrix of a BLBC :-

- Let m be the message vector and v be the corresponding codeword, st dim(m) = k, dim(v) = n, then the generator matrix G is the matrix such that,

$$m_{1\times k}\, G_{k\times n} = v_{1\times n}$$

- G is a $k \times n$ matrix over the field $(F_2, \oplus, \cdot)$.

Example:

Define a BLBC as,

$00 \longrightarrow 000$

$10 \longrightarrow 101$ (SPC-3)

$01 \longrightarrow 011$

$11 \longrightarrow 110$

Find the Generator matrix.

The basis of the i/p vector space is 10, 01 and they are mapped as,

$10 \longrightarrow 101$ & $01 \longrightarrow 011$

$\Rightarrow G = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ (wkt from LA, that a LT is defined by the way it acts on the basis of the i/p Vector space)

Example: Find the Gen. matrix of

$00 \longrightarrow 00000$

$10 \longrightarrow 11000$

$01 \longrightarrow 11101$

$11 \longrightarrow 00101$

Using the same logic as before,

$G = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}$ (As per how it acts over the basis $\{01, 10\}$)

Example: Find the Gen. matrix of the Hamming code of $C(5,3)$ (The parity bits are added at the end).

The parity bits will be given by :

$$P_1 = m_1 \oplus m_2$$

$$P_2 = m_2 \oplus m_3$$

The basis of the i/p space will be mapped as,

$$100 \longrightarrow 10010$$

$$010 \longrightarrow 01010$$

$$001 \longrightarrow 00101$$

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

- In general, to find the Generator matrix of a LBC, we need to find out how it maps the basis vectors of the input vector space.

- The rows of the Generator Matrix is the basis of $C(n,k)$.

Example: Given a generator matrix G,

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

Find the code words.

From the Gen. matrix, we can say that $n = 6, k = 3$;

and also,

$100 \rightarrow 011100$

$010 \rightarrow 101010$

$001 \rightarrow 110001$

From the basis vector, we can construct the codebook as

000	⟶	000000
001	⟶	110001
010	⟶	101010
011	⟶	011011
100	⟶	011100
101	⟶	101101
110	⟶	110110
111	⟶	000111

- **Systematic Generator Matrix :-**

  - The codewords can be written as a concatenation of the original message and the coded bits (parity bits).

  - The generator matrix will be of the form,

$$G = \left[\begin{array}{c|c} \text{original message} & \text{Coded bits} \end{array}\right]$$

↑ Only the basis

- **Cyclic BLDC :-**

  - If $\{m_1, m_2, m_3 \ldots\ldots m_n\}$ is a codeword, then $\{m_n, m_1, m_2 m_3 \ldots\ldots, m_{n-1}\}$ is also a codeword.

  - They are represented using polynomials.

- **Parity Check Matrices :-**

Consider the Gen. matrix,

$$G = \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$ of C(n, k)

  - We can see that the codebook formed by the generator matrix

will satisfy the parity equations,

$$\text{eqn (1)} \quad v_0 + v_4 + v_5 = 0$$
$$\text{eqn (2)} \quad v_1 + v_3 + v_5 = 0$$
$$\text{eqn (3)} \quad v_2 + v_3 + v_4 = 0$$

How ? → The indices in each eqn corresponds to each 1 in the corresponding row of G.

Define rows of a matrix H as

$$H_0 = [1\,0\,0\,0\,1\,1]$$
$$H_1 = [0\,1\,0\,1\,0\,1]$$
$$H_2 = [0\,0\,1\,1\,1\,0]$$

These are just the complements of the rows of G, by the way we wrote the parity equations.

We see that,

$$[v_0\ v_1\ v_2\ v_3\ v_4\ v_5]\, H_0^T = \text{eqn (1)} \quad \forall \text{ codewords} \in C(n,k)$$

Why for eqn (2) & eqn (3)

→ dim: $(n-k) \times n$

This matrix is termed as the Parity Check Matrix of $C(n,k)$

- We know that $C(n,k)$ is the span of rows of G. Let S be the span of rows of H. We see that,

  $C(n,k)$ and S are orthogonal subspaces

- S is termed as the dual of $C(n,k)$

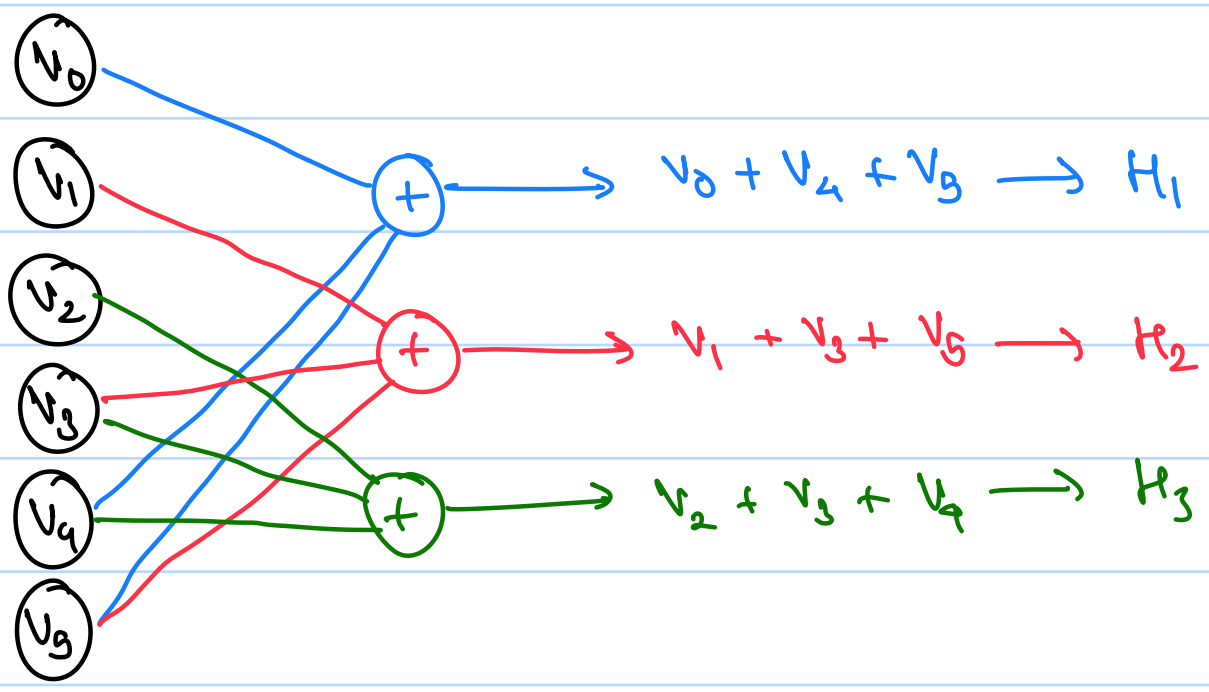
- **Alternate Definition of A BLDC :-**

A set of codewords such that given a matrix H, $\forall\, c \in C(n,k)$, $cH_i^T = 0 \;\forall\, i \in (0, n-k)$

- **Parity Check Matrix of a Systematic LBC :-**

- If the Generator matrix is given as $G = [P \mid I_n]$, then the Parity Check matrix is $H = [I_{n'} \mid P^T]$, for a systematic LBC.

- **Tanner Graph Representation :-**



Example:

Consider a systematic code of $l = 8$ and $dim = 4$ whose parity check equations are,

$$p_0 = v_1 + v_2 + v_3$$
$$p_1 = v_0 + v_1 + v_2$$
$$p_2 = v_0 + v_1 + v_3$$
$$p_3 = v_1 + v_2 + v_3$$

, where v_0, v_1, v_2, v_3 are the msg. bits and p_0, p_1, p_2, p_3 are the parity bits.

Find the Gen matrix and PC matrix.

G,

Message				Parity			
v_0	v_1	v_2	v_3	P_0	P_1	P_2	P_3
1	0	0	0	0	1	1	0
0	1	0	0	1	1	1	1
0	0	1	0	1	1	0	1
0	0	0	1	1	0	1	1

$$\Rightarrow \quad G = \left[\begin{array}{cccc:cccc} 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \end{array}\right]$$

$$\Rightarrow \quad H = \left[\begin{array}{cccc:cccc} 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 1 & 1 \end{array}\right]$$

- **Dual of a LBC :-**

The dual of a code C, $C^{\perp}$, consists of all the vectors $w \in F_2^n$, st $wv^T = 0 \ \forall \ v \in C$.

Example: Find the dual of REP-3.

$$\text{REP-3} = \{000, 111\} \qquad G_{REP3} = [1 \ 1 \ 1]$$

(0 ↑ 1 ↑)

Parity eqn : $v_0 + v_1 + v_2 = 1$

- **Hamming Weight** :-

  For a $w \in \mathbb{F}_2^n$, the Hamming weight of w is the no. of 1's in w, denoted $d_H(w)$.

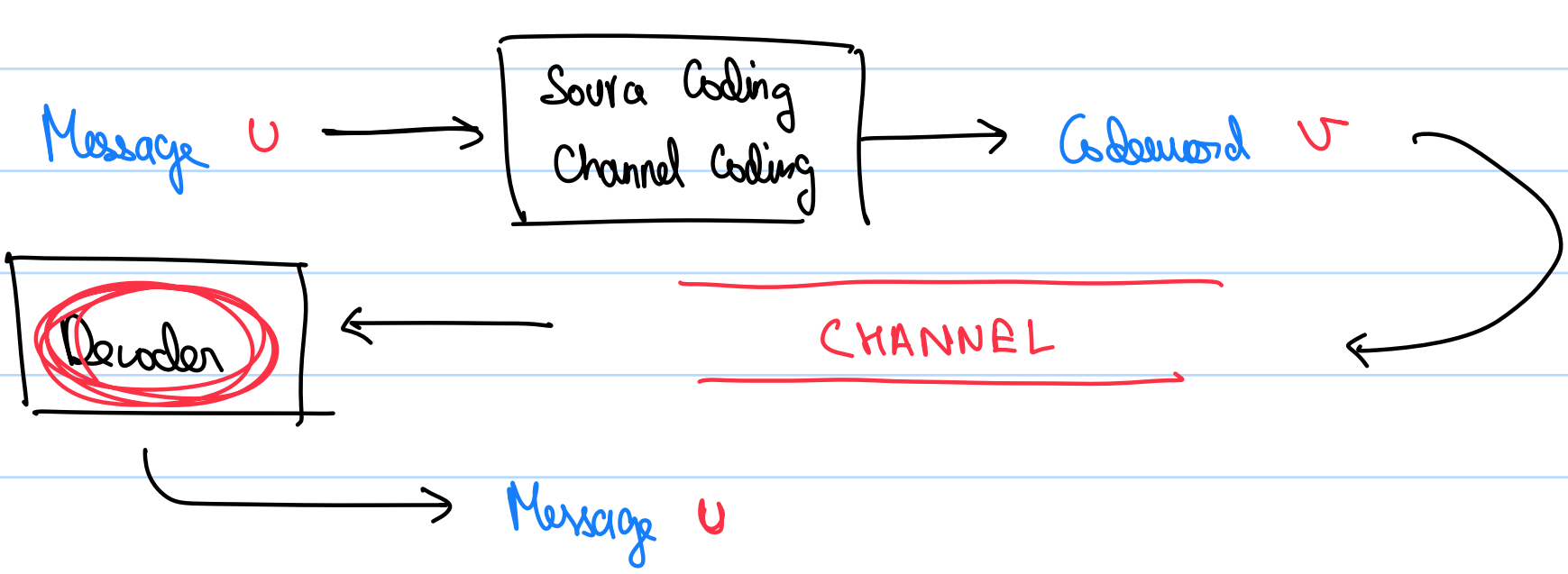
- The min. distance of a binary LBC C, denoted by $d_{min}(C)$, is the minimum weight among all $w \in C(n,k)$.

The linearity of a block code allows us to have a Generator matrix. (Each codeword is the linear combination of the rows of G).

7/4/25

→ **Decoding**:

- **Decoder** :-



- A decoder is a map b/w codewords and their corresponding message.

- ML Decoding:

- Suppose $\bar{v}_0, \bar{v}_1, \bar{v}_2, \ldots, \bar{v}_{2^k-1}$ are the possible messages.

- Given $Y = y$, ML decoder maps y to the "Most Likely" transmitted message.

  Y – Possible inputs to the decoder.

- To find the image of y in the decoder, find the likelihood of each original message, i.e.

  $P(U = \bar{v}_0 | Y = y), P(U = \bar{v}_1 | Y = y), P(U = \bar{v}_2 | Y = y) \ldots$

- Suppose $P_i = P(U = \bar{v}_i | Y = y)$ is maximum, the decoder will return $\bar{v}_i$ as the decoded output of y.

Example: REP-3 Code. Possible messages $\bar{v}_0 = 0$, $\bar{v}_1 = 1$. Probability of bit flip is BSC(p)

$$0, 1 \xrightarrow{\text{Encode}} 000, 111 \xrightarrow{\text{Bit flip}} 000, 001, 010 \ldots, 111$$

$$P(U = 0 | Y = 000) = \frac{P(Y = 000 | U = 0)\, P(U = 0)}{P(Y = 000)}$$

$$= \frac{(1-p)^3\, p(0)}{P(Y = 000)}$$

$\therefore P(Y = 000)$ need not be computed since the likelihoods will have the same denominator.

$$P(U=1 \mid Y=000) = \frac{P(Y=000 \mid U=1)\, P(U=1)}{P(Y=000)}$$

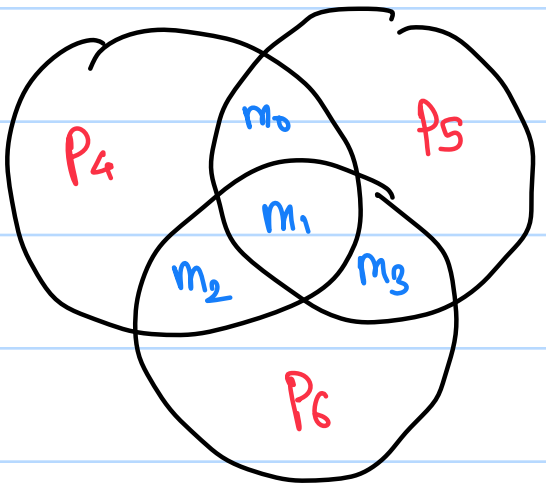
$$= \frac{p^3 \cdot p(1)}{P(Y=000)}$$

Example: Decoding Hamming Code of length 7.

(Error Correction in Hamming Code)

Message bits = m_0, m_1, m_2, m_3

Parity = p_4, p_5, p_6



$$m_0 \oplus m_1 \oplus m_2 \oplus p_4 = 0$$
$$m_0 \oplus m_1 \oplus m_3 \oplus p_5 = 0$$
$$m_1 \oplus m_2 \oplus m_3 \oplus p_6 = 0$$

After the decoding step, we get,

$$y = \{ y_0, y_1, y_2, y_3, y_4, y_5, y_6 \}$$

Let, $\oplus \{ y_0, y_1, y_2, y_4 \} = S_1$

$\oplus \{ y_0, y_1 y_3, y_5 \} = S_2$

$\oplus \{ y_1, y_2, y_3, y_6 \} = S_3$

S_1, S_2, S_3 are the results of the parity checks.

Possible values of S_1, S_2, S_3

S_1	S_2	S_3	Flipped Bit
0	0	0	ϕ
0	0	1	$y_6 \Rightarrow P_6$
0	1	0	$y_5 \Rightarrow P_5$
0	1	1	$y_4 \Rightarrow P_4$
1	0	0	$y_3 \Rightarrow m_3$
1	0	1	$y_2 \Rightarrow m_2$
1	1	0	$y_1 \Rightarrow m_1$
1	1	1	$y_0 \Rightarrow m_0$

Example: $y = \{ \underset{y_0}{1} \; \underset{y_1}{0} \; \underset{y_2}{1} \; \underset{y_3}{1} \; \underset{y_4}{1} \; \underset{y_5}{0} \; \underset{y_6}{0} \}$

$$S_1 = \oplus \{0\,1\,2\,4\} = 1 \oplus 0 \oplus 1 \oplus 1 = 1$$

$$S_2 = \oplus \{0\,1\,3\,5\} = 1 \oplus 0 \oplus 1 \oplus 0 = 0$$

$$S_3 = \oplus \{1\,2\,3\,6\} = 0 \oplus 1 \oplus 1 \oplus 0 = 0$$

$S_1 = 1, \; S_2 = 0, S_3 = 0 \Rightarrow y_3$ (aka m_3) is flipped

$\Rightarrow \bar{v} = \{1\,0\,1\,0\,1\,0\,0\}$

→ Standard Array :-

- All the possible 2^n vectors are arranged in a table.

Example: $G = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$ ⇒ Message length = 2

Codebook = { 0000, 1100, 0110, 1010 }

- The column headers have the codewords.

0000	1100	0110	1010

- Fill the first column with Coset leaders, vector of (not always. min weight) that hasn't appeared before. Add the header value of each column to the Coset value of each row, for the other cells of the table.

→	0000	1100	0110	1010
Coset Leaders ←	0001	1101	0111	1011
	1000	0100	1110	0010
	1001	0101	1111	0011

- If our received word is $\bar{y}$, the decoder output will be the header value of the column of $\bar{y}$.

The Coset Leaders represent the correctable error patterns, ie only these bit flips which match a Coset Leader, can be corrected.

Note: Each vector appears only once (By construction).

Example: Construct a Standard table for REP 3

Ans:

000	111
001	110
010	101
100	011

↑ Coset leader.

(X) To minimize decoding error, choose the Coset leader of it covers the most probable noise patterns.

Example: We have a Gen. matrix G_1 for $C(6,3)$

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

000000	011100	101010	110001	110110	101101	011011	000111
000001	011101	101011	110000	110111	101100	011010	000110
000010	011110	101000	110011	110100	101111	011001	000101
000100	011000	101110	110101	110010	101001	011111	000011
001000	010100	100010	111001	111110	100101	010011	001111
010000	001100	111010	100001	100110	111101	001011	010111
100000	111100	001010	010001	010110	001101	111011	100111

- If BSC(p) < 1/2, then the choice of Coset leaders should prefer lower weights. (Bit flip less likely to happen)

  If BSC(p) > 1/2, then the choice of Coset leaders should prefer higher weights (Bit flip more likely to happen)

- For a Code C(n,k) of d_{min}, then the no. of correctable error patterns is $t = \lfloor (d_{min} - 1)/2 \rfloor$. (t-error correcting code.)

10/4/25

Modulation

- Modulation is the last process done before the message goes through the channel.

- In an Analog Communication System,

Source → Modulation ⇝ (Channel) ⇝ Demodulation → Target

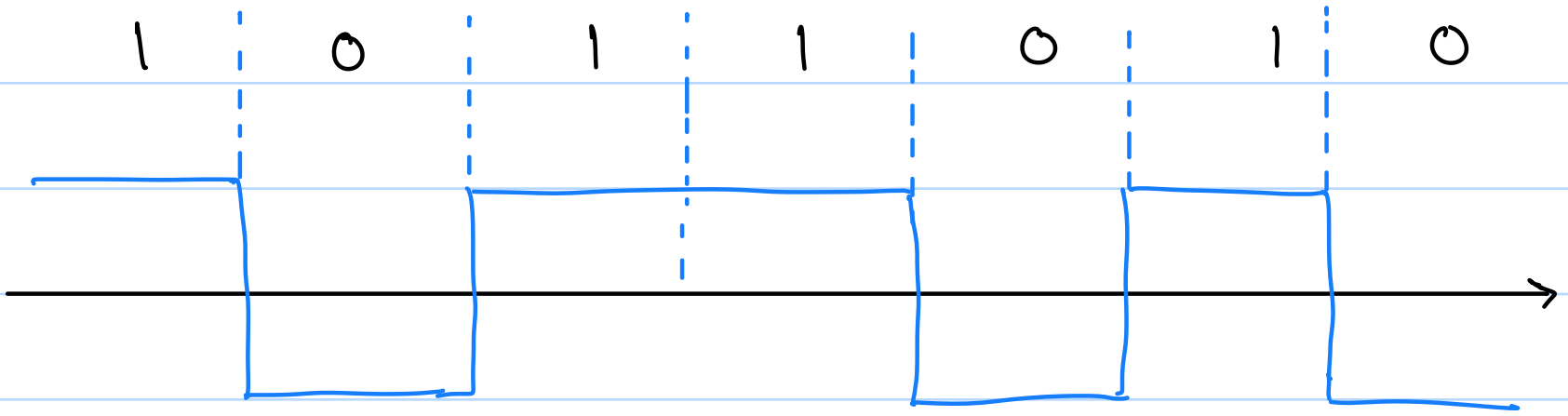
→ **Digital Modulation:-**

Conversion of a bit sequence into an Analog waveform, to be sent over a physical channel.

ex: Bit sequence to EM Waves.

- We can use a square wave to model the bit sequence;

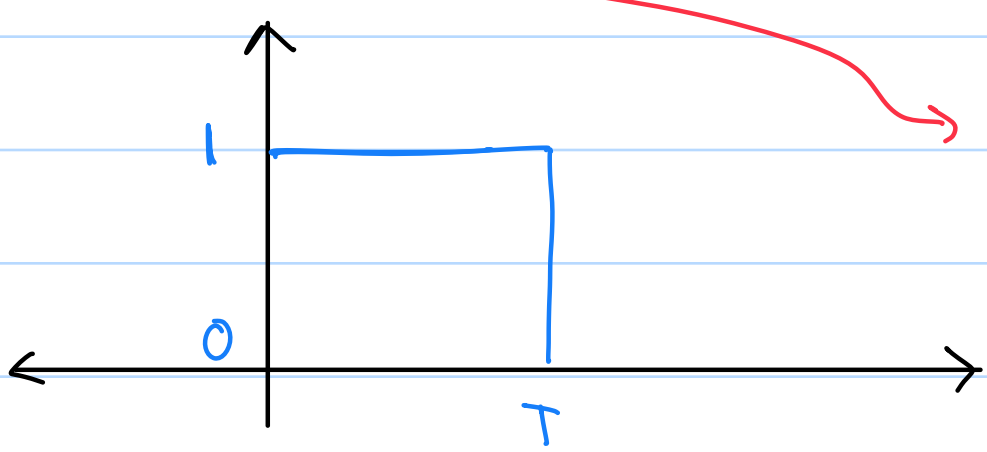
$$y(t) = \begin{cases} 1, & \text{bit} = 1 \\ -1, & \text{bit} = 0 \end{cases} \quad \forall \text{ bits} \in \text{sequence}.$$



$$\Rightarrow \quad x(t) = \sum_{k \in \mathbb{Z}} a_k \, p(t - kT)$$

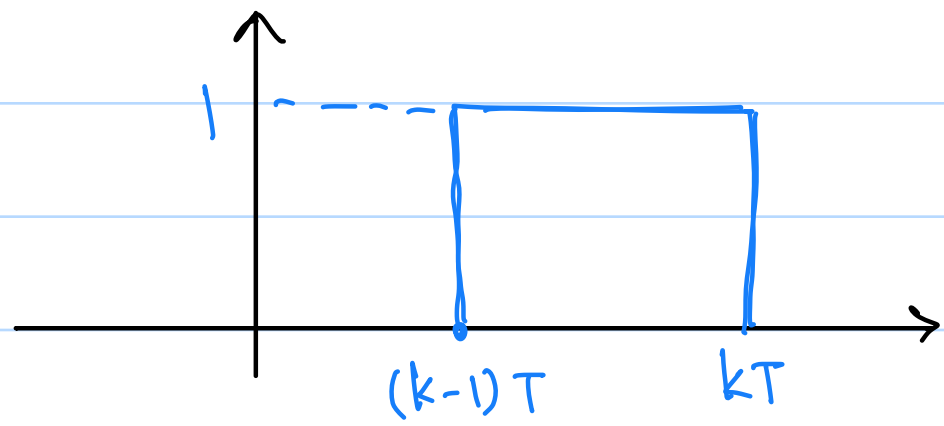
where p is each pulse of duration T.

$p(t)$:



$$a_k = \begin{cases} -1, & k^{th} \text{ bit} = 0 \\ 1, & k^{th} \text{ bit} = 1 \end{cases}$$

$p(t - kT)$



Note:

Our pulse can be any signal (like a bell curve / sinusoid)

- Another requirement of modulation is increasing the frequency of the signal → Reduces wavelength ⇒ Smaller antennas
  → Reduces interference and Power usage.

- Modulating signal : Signal that has the info. (Low freq) $m(t)$.
  Carrier signal : High freq signal used in modulation. $c(t)$.

Recap:

Fourier transform of a signal $x(t)$

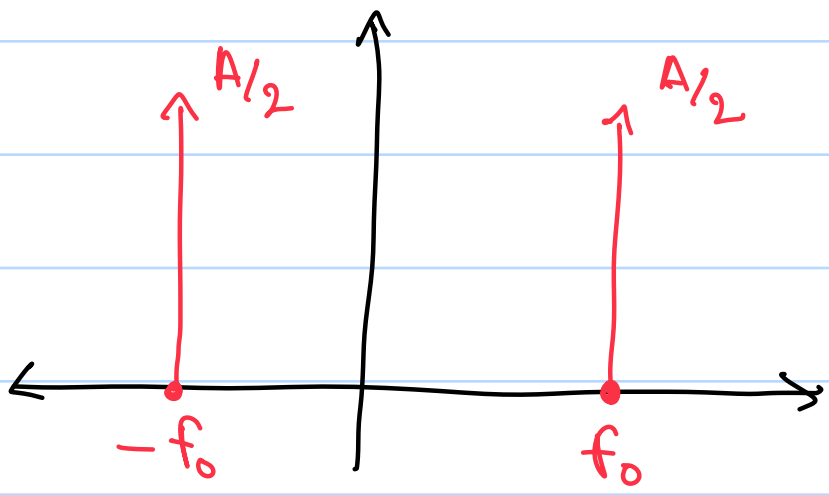
$$X(f) = \int_{-\infty}^{\infty} x(t)\, e^{-j2\pi f t}\, dt$$

We know that if $x(t) = A\cos(2\pi f_0 t)$

$$x(t) = A\cos(2\pi f_0 t) = A\,\frac{e^{j2\pi f_0} - e^{-j2\pi f_0}}{2}$$

$$\Rightarrow \quad X(s) = \infty\int \frac{A}{2} e^{j2\pi f_0 t - j2\pi f t}\, dt + \infty\int \frac{A}{2} e^{-j2\pi f_0 t - j2\pi f t}\, dt$$

$\Rightarrow$ $X(s)$:



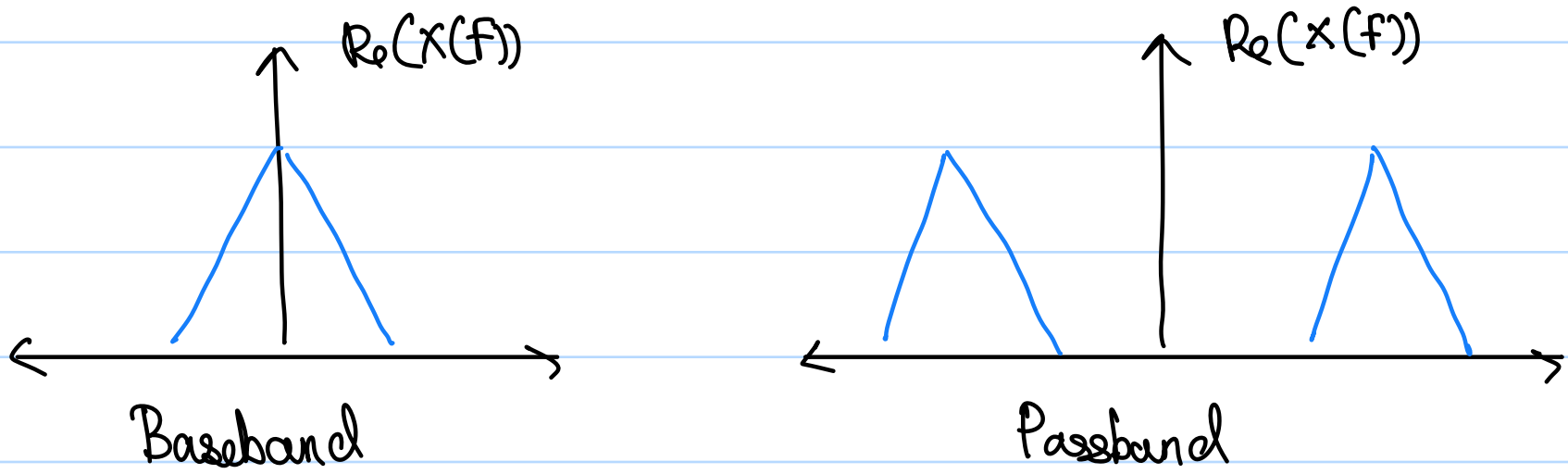
- An important property of FT is,

Given $x(t) \xrightarrow{FT} X(f)$, } Modulation Propety

Then $x(t)\, e^{2\pi f_0 t} \longrightarrow X(f - f_0)$ (Shifting of the Spectrum)

- Also, the FT of a real signal is **conjugate - symmetric**

- **Baseband and Passband :-**



frequencies centered at zero (DC)

frequencies centered at a non zero frequency.

Note: Re (X(f)) - Magnitude of X(f) at each sinusoid
Im (X(f)) - Phase of X(f) at each sinusoid.

- **Objective of Modulation :-**

To design a passband signal that is to be transmitted, to convey information in a baseband signal.

$\therefore \quad U(f) \longrightarrow U(f - f_c)$

- We know that, $u(t)\, e^{2\pi f_c t} \longrightarrow U(f - f_c)$

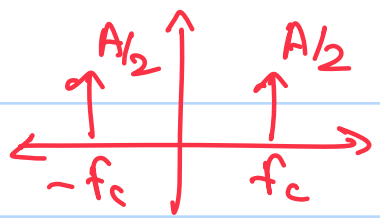
$\therefore$ $e^{2\pi f_c t}$ is the **carrier signal**.

∴ To convert a LF to HF signal

1. Multiply with $e^{2\pi f_c t}$

(or)

2. Multiply with $\cos(2\pi f_c t)$ / $\sin(2\pi f_c t)$



↳ Amplitude modulation ↳ Phase modulation

- The passband signal will be,

$$v_p(t) = v_c(t)\cos(2\pi f_c t) - v_s(t)\sin(2\pi f_c t)$$

Where $v_c(t)$: In phase component (I - Component)

$v_s(t)$: Quadrature component (Q - Component)

The message $u(t)$ must be encoded into $v_c(t)$ and $v_s(t)$.

$v_c(t)$ = Pure cosine components → Control the amplitude

$v_s(t)$ = Pure sine components → Control the phase shift

Since $v(t) = \sum a_w \sin(\omega t + \phi)$

$$= \sum_{\infty} a_w (\sin\omega t \cos\phi + \cos\omega t \sin\phi)$$

$$= \sum a_{sw} \sin\omega t + a_{cw} \cos\omega t$$

$$= \underbrace{\sum a_{sw} \sin\omega t}_{v_s(t)} + \underbrace{\sum a_{cw} \cos\omega t}_{v_c(t)}$$

→ Amplitude Modulation :-

- A passband signal is of the form,

$$u_p(t) = v_c(t)\cos(2\pi f_c t) - v_s(t)\sin(2\pi f_c t)$$

- In Amplitude Modulation, only the I-Component is modulated.

- Double Side Band Suppressed Carrier :- → The carrier wave does not appear in the spectrum

$$y(t) = \underbrace{A\,x(t)}_{\text{Amplitude of Carrier wave}}\ \underbrace{\cos(2\pi f_c t)}_{\text{Carrier wave}}$$

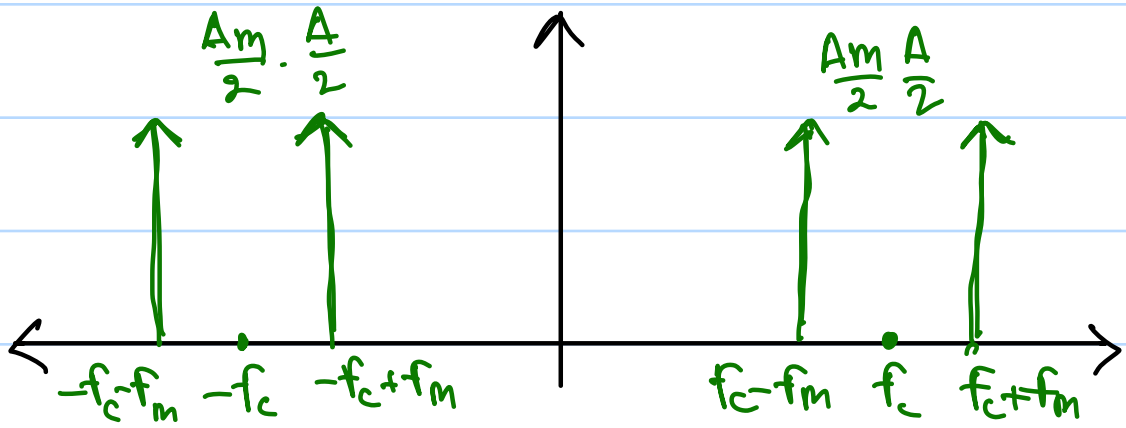
→ Modifies only the cosine components since they control the magnitude

$$\Rightarrow \quad Y(f) = \frac{A}{2}\left(X(f-f_c) + X(f+f_c)\right)$$

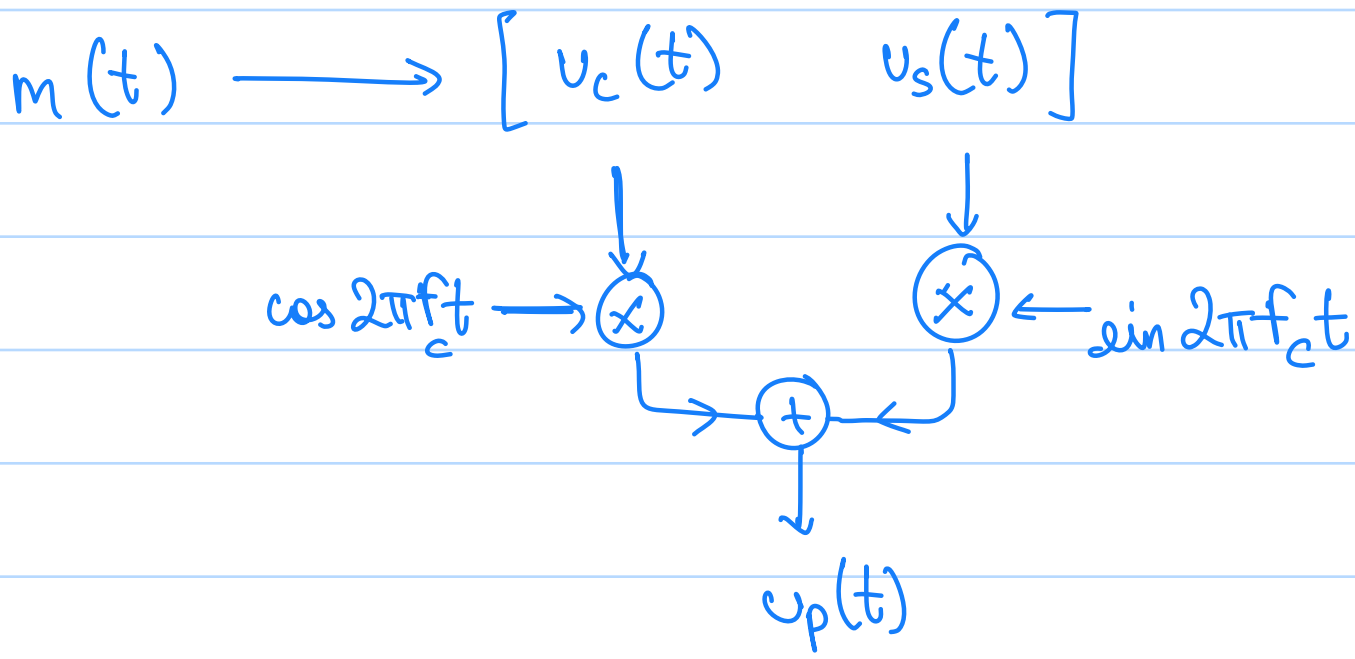
- The amplitude of the carrier wave is varied according to the amplitude of the message.

Example: If $x(t) = A_m\cos(2\pi f_m t)$, Carrier $= A\cos(2\pi f_c t)$

$Y(f)$:



- Inphase and Quadrature Components :



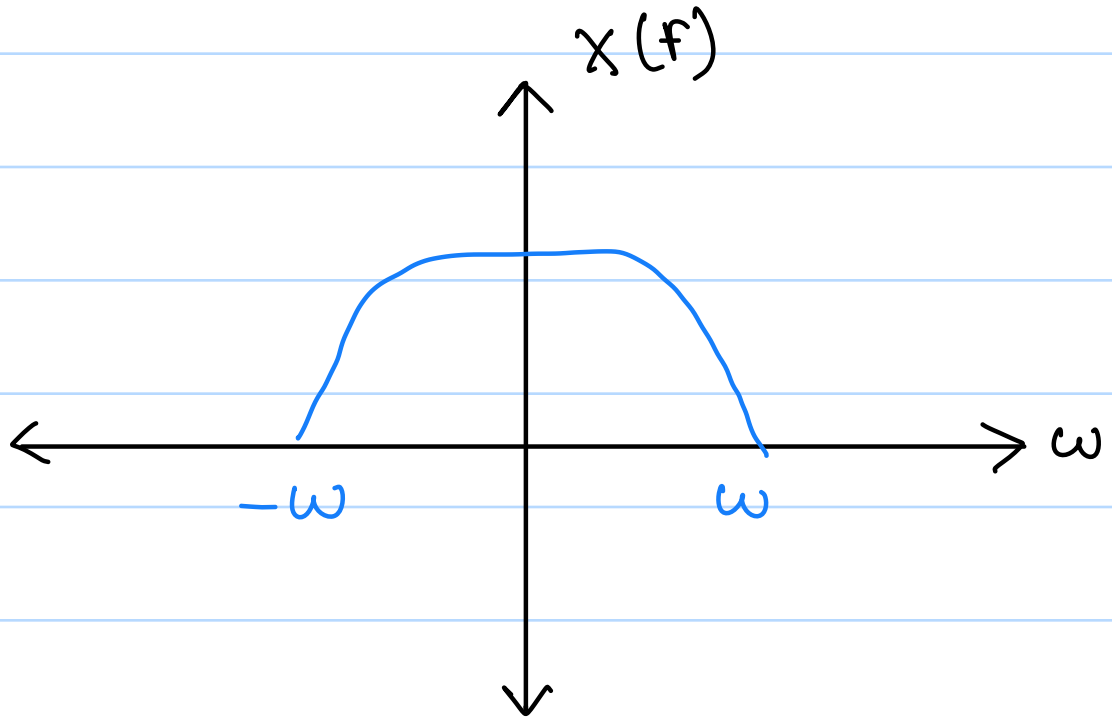
$v_c(t)$ = In phase component

$v_s(t)$ = Quadrature component

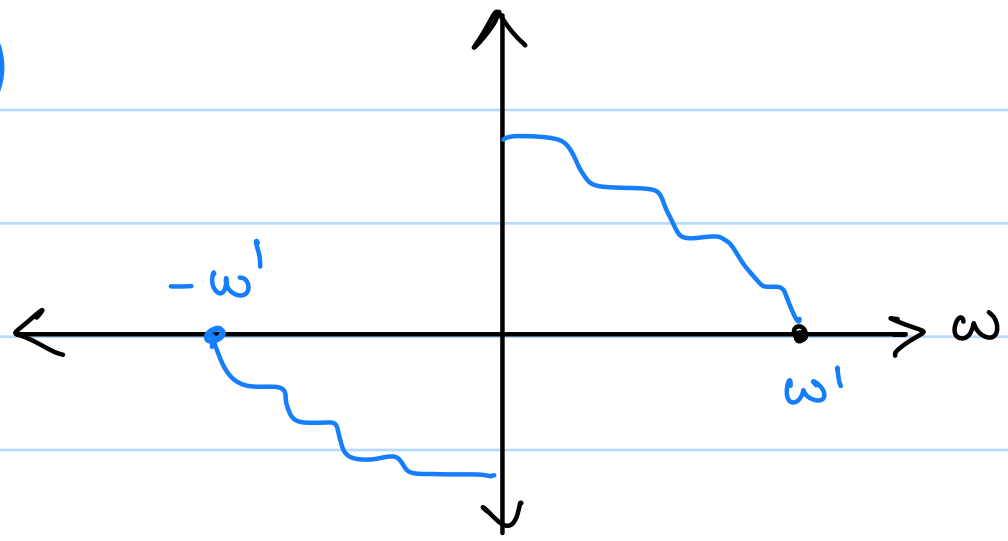
- $m(t) \xrightarrow{DSD \cdot SC} [Am(t) \quad 0]$

- Bandwidth: Let $X(f)$ be the Fourier Transform of any signal $x(t)$ such that

$Re(X(f))$

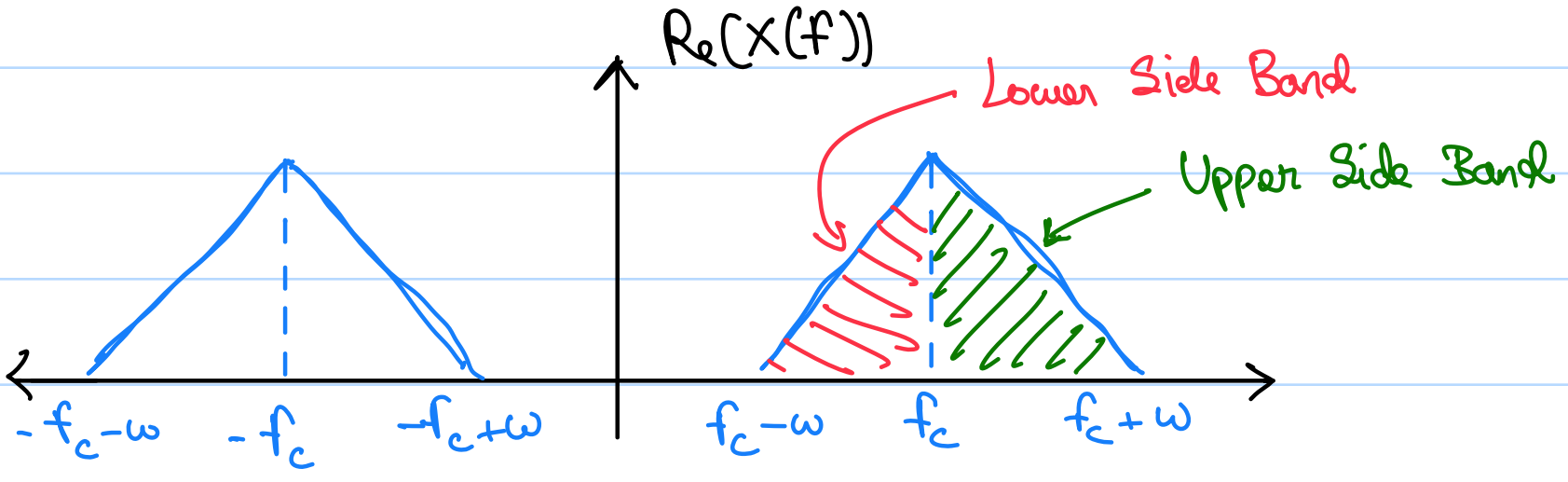


$Im(X(f))$



Bandwidth = max $\{\omega, \omega'\}$

If $X(f)$ undergoes DSB-SC modulation,



The entire information of the signal is contained in just one of the sidebands. (So, transmitting just a single sideband also works)

↓ (SSB-SC)

Bandwidth = $f_c + \omega - (f_c - \omega)$

= 2ω

- Carrier Wave Expression:

In DSB-SC, the spectrum of the carrier signal does not appear anywhere in the final spectrum. (Hence DSB - Suppressed Carrier)

→ Digital Modulation :-

- We have a baseband signal,

$$v(t) = \sum_n b[n]\, p(n - nT)$$

- To convert this into a passband signal, one approach would be to,

$$v_p(t) = v(t) \cos 2\pi f_c t$$

$$\Rightarrow v_p(t) = \begin{cases} \cos 2\pi f_c t, & \text{when } b[n] = 1 \rightarrow \text{bit} = 1 \\ -\cos 2\pi f_c t, & \text{when } b[n] = -1 \rightarrow \text{bit} = 0 \end{cases}$$

- Here, we see that the phase of the carrier signal ($\cos 2\pi f_c t$) switches abruptly between 0 and π, corresponding to the bit value at that point.

- This is called BPSK (Binary Phase Shift Keying) Modulation.

- In analog modulation,

$$v_p(t) = v_c(t) \cos(2\pi f_c t) + v_s(t) \sin(2\pi f_c t)$$

where we define the complex envelope of $v_p(t)$ as $v_c(t) + j v_s(t)$

Similarly in digital modulation we can define the complex envelope of $b[n]$ as $b_c[n] + j b_s[n]$

$$\Rightarrow \quad [v_c(t) \quad v_s(t)] \longleftrightarrow [b_c[t] \quad b_s[t]]$$

In BPSK

$$b[n] \xrightarrow{\text{BPSK}} [\pm 1 \quad 0] \rightarrow \text{Only cos is there}$$

But if we take $b[n] \xrightarrow{\text{BPSK}} [\pm 1 \quad \pm 1]$

$$\Rightarrow v_p(t) = b_c[n]\cos 2\pi f_c t + b_s[n]\sin 2\pi f_c t \quad \text{(For each period)}$$

$$= \sqrt{2}\left(\frac{1}{\sqrt{2}} b_c[n]\cos 2\pi f_c t + \frac{1}{\sqrt{2}} b_s[n]\sin 2\pi f_c t\right)$$

$$= \sqrt{2}\left(b_c[n]\cos\frac{\pi}{4}\cos 2\pi f_c t + b_s[n]\sin\frac{\pi}{4}\sin 2\pi f_c t\right)$$

$b_c[n]$	$b_s[n]$	$v_p(t)$	
1	1	$\sqrt{2}\cos(2\pi f_c t + \frac{\pi}{4})$	①
1	-1	$\sqrt{2}\cos(2\pi f_c t - \frac{\pi}{4})$	②
-1	1	$\sqrt{2}\cos(2\pi f_c t + \frac{3\pi}{4})$	③
-1	-1	$\sqrt{2}\cos(2\pi f_c t - \frac{3\pi}{4})$	④

The above values are used in QPSK, which combines bits into 2 bit blocks and modulates them as a whole.

2 Bit Seq	Comp. Envelope	$[b_c[n] \quad b_s[n]]$	$v_p(t)$
0 0	$1+j$	1 1	①
0 1	$1-j$	1 -1	②
1 0	$-1+j$	-1 1	③
1 1	$-1-j$	-1 -1	④

Since the phase shifts between 4 different phase values. Therefore it is termed as Quadrature Phase Shift Keying.

Illy we have 8 - PSK, 16 -PSK, 32-PSK .....

For n -PSK, we need n different complex numbers and assign each n bit sequence a complex number.

→ Fourier Transform :-

$$X(\omega) = \int_{-\infty}^{\infty} x(t)\, e^{-j\omega t}\, dt$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)\, e^{j\omega t}\, d\omega$$

- Sinc function :- $\operatorname{sinc}(x) = \frac{\sin x}{x}$

- The FT of a triangular pulse is sinc(ω).

- Properties :-

1. Linearity (Is a linear canonical transform)

2. Time Shift Property : $x(t) \xrightarrow{F} X(\omega)$

$$x(t - t_0) \xrightarrow{F} X(\omega)\, e^{-j\omega t_0}$$

3. Freq - Shift Property : $x(t)\, e^{j\omega_0 t} \longrightarrow X(\omega - \omega_0)$

4. Convolution : $x(t) * y(t) \xrightarrow{f} X(\omega)\, Y(\omega)$

$x(t)\, y(t) \xrightarrow{f} X(\omega) * Y(\omega)$

Proof of $x(t) * y(t) \xrightarrow{f} X(\omega) Y(\omega)$:

$$x(t) * y(t) = \int_{-\infty}^{\infty} x(\tau)\, y(t-\tau)\, d\tau$$

$$\downarrow f$$

$$\int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} x(\tau)\, y(t-\tau)\, d\tau \right) e^{-j\omega t}\, dt$$

$$= \int_{-\infty}^{\infty} x(\tau) \left(\int y(t-\tau)\, e^{-j\omega t}\, dt \right) d\tau$$

$$= \int_{-\infty}^{\infty} x(\tau)\, Y(s)\, e^{-j\omega\tau}\, d\tau$$

$$= X(s) \cdot Y(s)$$

∎

- FT of a Pulse Train :

$$p(n) = \sum_{n \in \mathbb{N}} \delta(t - nT_s)$$

$$P(\omega) = \int_{-\infty}^{\infty} \sum \delta(t - nT_s)\, e^{-j\omega t}\, dt$$

$$= \sum_{n \in \mathbb{N}} \int_{\infty} \delta(t - nT_s)\, e^{-j\omega t}\, dt$$

$$= \sum_{n \in N} e^{-j\omega n T_s}$$

$$= \sum_{n=0}^{\infty} e^{-j\omega n T_s} + \sum_{n=-\infty}^{-1} e^{-j\omega n T_s}$$

$$= \frac{1}{1 - e^{-j\omega T_s}} + \sum_{n=1}^{\infty} e^{j\omega n T_s}$$

$$= \frac{1}{1 - e^{-j\omega T_s}} + \frac{e^{j\omega T_s}}{1 - e^{j\omega T_s}}$$

$$= \frac{(1 - e^{j\omega T_s}) + (1 - e^{-j\omega T_s}) e^{j\omega T_s}}{1 - e^{j\omega T_s} - e^{-j\omega T_s} + 1}$$

$$= \frac{1 - e^{j\omega T_s} + e^{j\omega T_s} - 1}{2 - e^{j\omega T_s} - e^{-j\omega T_s}}$$

- The collection of the complex envelopes of 2^M-ary PSK is termed as the constellation of the 2^M-ary PSK on the Argand plane.

One way of selecting Complex Envelopes is using the n^{th} roots of unity concept.

- In BPSK, we have the result that,

$$v_p(t) = \begin{cases} \cos(2\pi f_c t), & b[n] = +1 \quad \text{phase} = 0 \\ -\cos(2\pi f_c t), & b[n] = -1 \quad \text{phase} = \pi \end{cases}$$

Illy in QPSK

$$v_p(t) = \begin{cases} \sqrt{2}\cos\left(2\pi f_c t - \frac{\pi}{4}\right) & b_2[n] = 00 \\ \sqrt{2}\cos\left(2\pi f_c t + \frac{\pi}{4}\right) & " = 01 \\ \sqrt{2}\cos\left(2\pi f_c t - \frac{3\pi}{4}\right) & " = 10 \\ \sqrt{2}\cos\left(2\pi f_c t + \frac{3\pi}{4}\right) & " = 11 \end{cases}$$

$$\Rightarrow \quad v_p(t) = \begin{cases} \cos(2\pi f_c t) + \sin(2\pi f_c t) & b_2[n] = 00 \\ \cos(2\pi f_c t) - \sin(2\pi f_c t) & " = 01 \\ -\cos(2\pi f_c t) + \sin(2\pi f_c t) & " = 10 \\ -\cos(2\pi f_c t) - \sin(2\pi f_c t) & " = 11 \end{cases}$$

why we need both sin and cos components

→ Discrete Memoryless Channel :-

- A system consisting of i/p alphabet $\mathcal{X}$ and o/p alphabet $\mathcal{Y}$ and a probability transition matrix $p(y|x)$ that expresses the probability of observing y given i/p x, such that it is independent of any previous input or output.

Disc-rete

Memoryless

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

- BSC, BEC, AWGN are Discrete Memoryless Channels.

- AWGN (Additive White Gaussian Noise) :-

- For an AWGN channel, $Y = X + Z$, where $Z \sim N(0, \sigma^2)$. (Z → Nois)

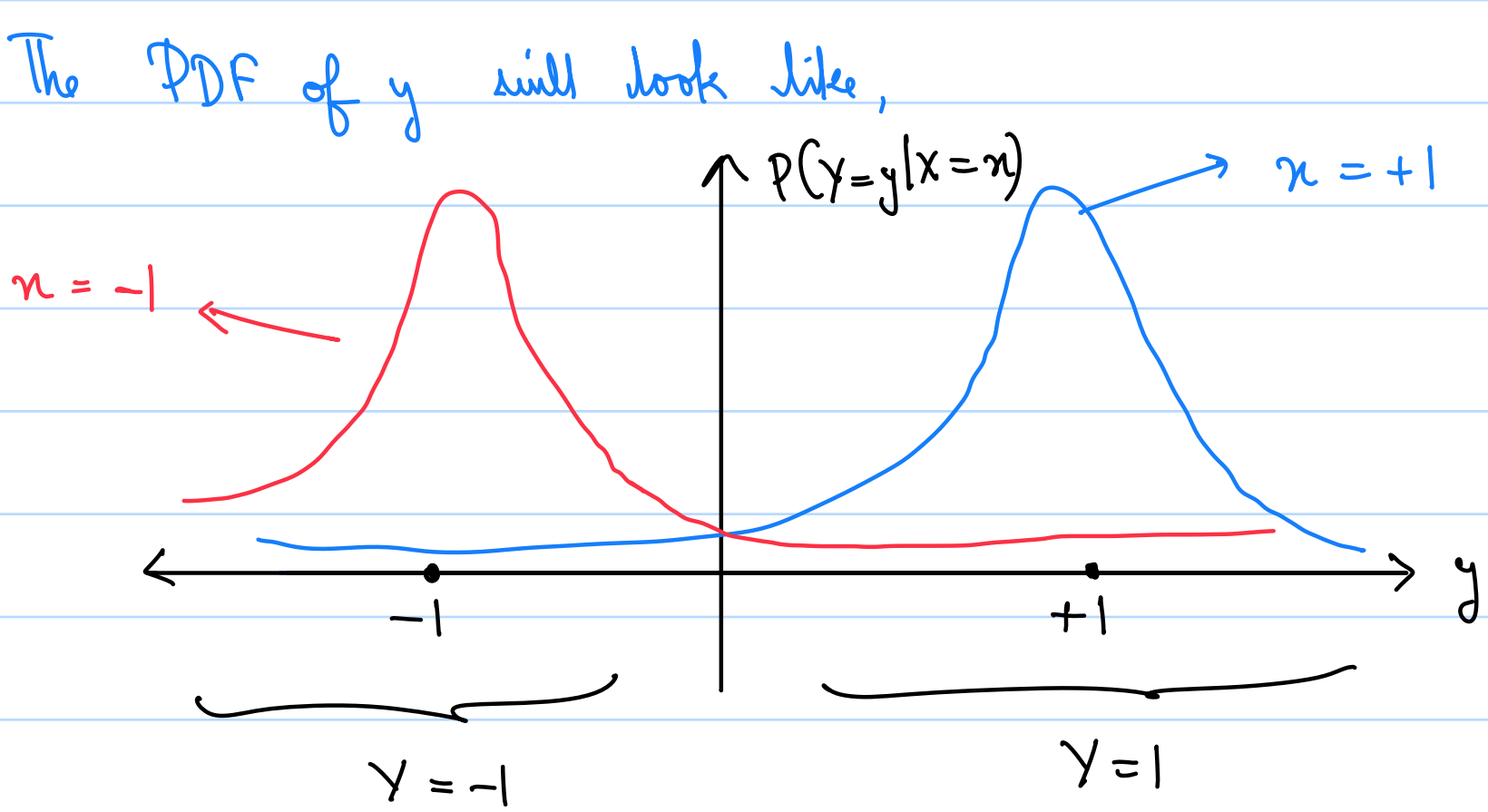
- The support set of X depends on the modulation scheme.

$$f_Z(z) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(-z-\mu)^2}{2\sigma^2}}$$ → Gaussian Bell Curve

- BPSK Demodulation :-

- In a BPSK modulator, $0 \rightarrow +1V$, $1 \rightarrow -1V$. $\therefore X = \{+1, -1\}$

- The PDF of y will look like,



- Seeing the PDF, we see that

$$P(Y=-1|X=-1) > P(Y=-1|X=1) \quad \forall \quad y<0$$
$$P(Y=1|X=1) > P(Y=1|X=-1) \quad \forall \quad y>0$$

- Therefore, we can define our demodulator as,

$$\hat{X} = \begin{cases} +1, & y \geq 0 \\ -1, & y<0 \end{cases} \quad \text{(ML Decoding)}$$

Decoding error,

$$P(\hat{X} \neq X) = P(y>0|X=-1) + P(y<0|X=1)$$
$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z-1)^2}{2\sigma^2}} dz + \int_{0}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(z+1)^2}{2\sigma^2}} dz$$
$$= \int_{-\infty}^{0} N(1,\sigma^2,z)\,dz + \int_{0}^{\infty} N(-1,\sigma^2,z)\,dz$$

Cannot be computed using simplification

→ Channel Capacity :-

$$\text{Capacity } C = \max_{p(x)} \{ I(X,Y) \}$$

It is the maximum mutual information between X and Y over different different probability distributions of X.

- Properties:

1. $C \geq 0$ (Mutual information is always positive)

2. $C \leq |\log X|$

↳

$$I(X,Y) = H(X) - H(X|Y)$$
$$\leq H(X)$$
$$\leq |\log X|$$

- Noiseless Binary Channel :-

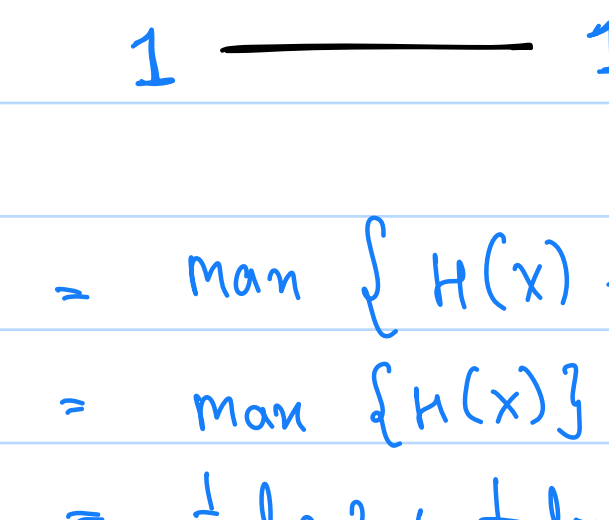
0 ——— 0 $\quad P(Y=0|X=0) = 1$

1 ——— 1 $\quad P(Y=1|X=1) = 1$

$$C = \max \{ H(X) - \overset{0}{H(X|Y)} \}$$
$$= \max \{H(X)\} \longrightarrow$$ Entropy is maximum in uniform distribution
$$= \frac{1}{2}\log 2 + \frac{1}{2}\log 2 = \underline{1}$$

∴ Channel Capacity of a noiseless channel is 1 bit.

- Noiseless Channel with Overlapping Outputs :-

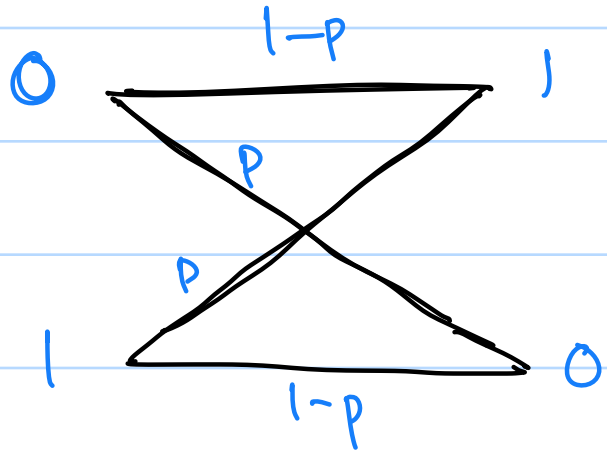


Since the channel is actually not noisy, ie.

$$P(X=0|Y=1) = 1, \quad P(X=0|Y=2) = 1$$
$$P(X=1|Y=3) = 1, \quad P(X=1|Y=4) = 1$$

$$\underline{H(X|Y) = 0} \longrightarrow \text{If } Y \text{ is given, } X \text{ can be determined}$$

- **BSC(p)**



$$H(Y|X) = H(Y|X=0)\,P(X=0) + H(Y|X=1)\,P(X=1)$$

Pmf of $Y|X=0$, $P(Y=1|X=0) = p$

$P(Y=0|X=0) = 1-p$

" $Y|X=1$, $P(Y=1|X=1) = 1-p$

$P(Y=0|X=1) = p$

$$H(Y|X=0) = p\log_2\frac{1}{p} + (1-p)\log_2\frac{1}{1-p}$$
$$H(Y|X=1) = p\log_2\frac{1}{p} + (1-p)\log_2\frac{1}{1-p}$$

$$\Rightarrow H(Y|X) = (P(X=0) + P(X=1))\left(p\log_2\frac{1}{p} + (1-p)\log_2\frac{1}{1-p}\right)$$

$$\Rightarrow H(Y|X) = p\log_2\frac{1}{p} + (1-p)\log_2\frac{1}{1-p}$$

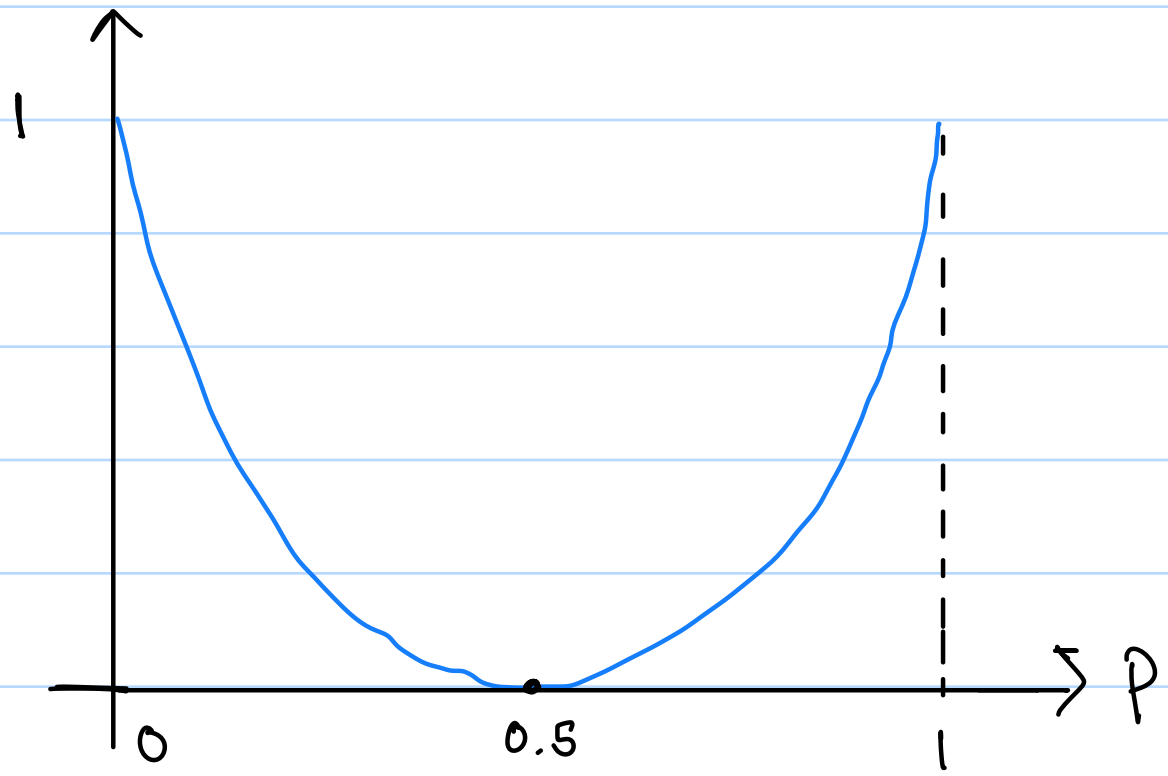
Max of $H(Y) = 1$

$$\Rightarrow \quad \max(I(X,Y)) = \max\left(H(Y) - H(Y|X)\right)$$

$$= 1 - p\log_2\frac{1}{p} - (1-p)\log_2\frac{1}{1-p}$$

$$\Rightarrow \quad C = 1 - p\log_2\frac{1}{p} - (1-p)\log_2\frac{1}{1-p}$$

$\Rightarrow$



The above graph implies that Channel capacity is maximum when $p = 0$ or 1, which makes intuitive sense.