Page No.
Date

Session-40

- Probability

1) R.V Defination ✓

Coin toss, $X = \{0, 1\}$

Dice $Y = \{1, 2, 3, 4, 5, 6\}$

R.V

- Discrete: $X = \{H, T\}$
- Continuous: $X = [0, 10]$

- Probability Distribution

dice roll -

1	2	3	4	5	6
1/6	1/6	1/6	1/6	1/6	1/6

- Prob. distribution function - (PDF)
  - how prob. are distributed over possible value of R.V
  - It is of two types
    1) Discrete PDF (PMF)
    2) Continuous PDF (Prob. Density fn)

→ Cumulative Dist. function

- A single dice is rolled

$$PMF = \begin{cases} 1/6 & x \in \{1, 2, 3, 4, 5, 6\} \\ 0 & \text{else} \end{cases}$$

Page No.
Date

- 2 dice are rolled and their sum is calculated.

$$\text{PMF} = \begin{cases} 1/36 & x \in [2, 12] \\ 2/36 & x = [3, 11] \end{cases}$$

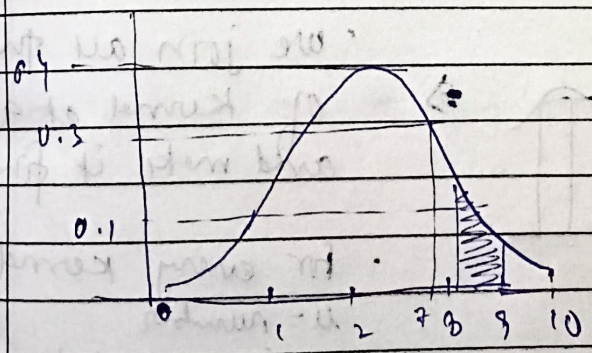
- Cumulative Dist fn for PMF –

$$f(x) = P(X \le x)$$

→ $P(X \le 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$$= \frac{4}{6}$$

- Probability Density function –



- Y axis → tells probability density
- area under graph tells probability

$$\int_0^{10} f(x)\,dx = 1 = P(0 \le X \le 10)$$

- At $y = 0.3$ ⇒ it tells prob. at $x = [7., 7.01]$ basically very small interval..

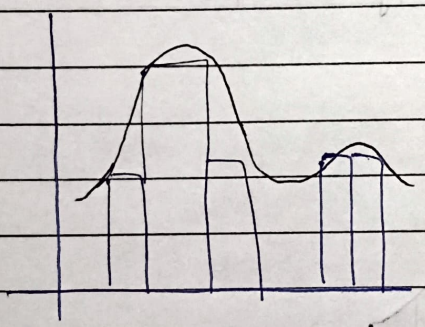
Page No.
Date

• Density Estimation
  ↳ It is process of estimating [PDF] of a R.V based on observed data
  It is of 2 types-

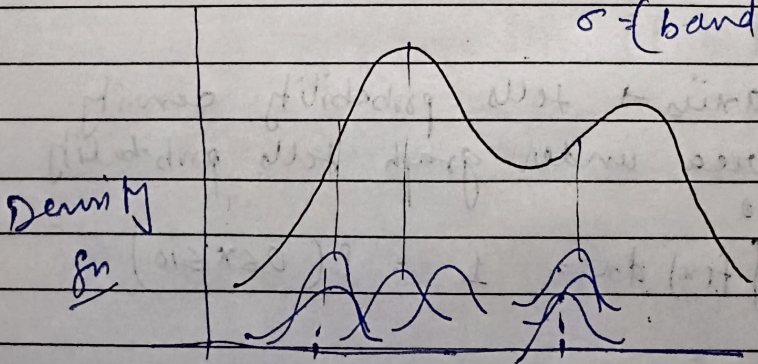
- Parametric Density Estimation
  ↳ we assume data follows a known distribution.

- Non parametric Density estimation (KDE)
  ↳ No assumption about distribution shape

Kernel Density Estimation -
  ↳ At each point we make kernel (mostly Normal dist)

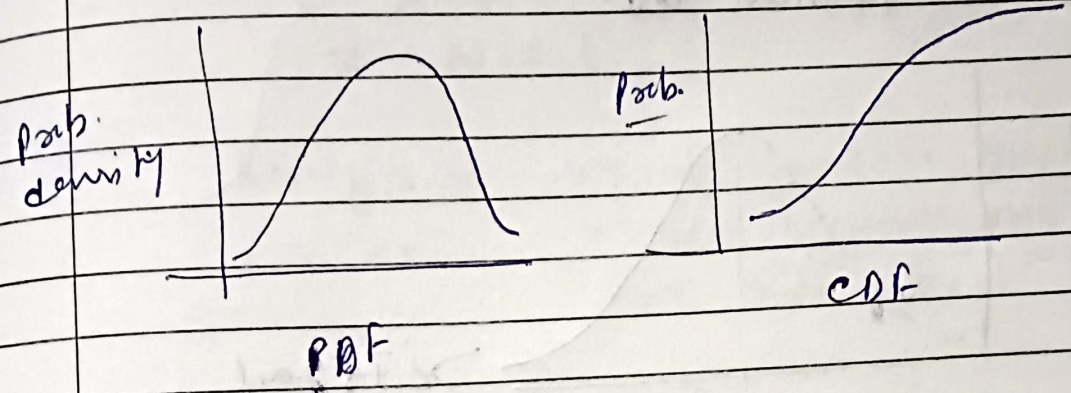


⇒
- we join all the 'y' of kernel at a point and make it final
- for every kernel
  μ = number
  σ = (bandwidth) = we can decide



Page No.
Date

CDF of PDF



→ from Area of PDF (integ") we get CDF

→ from slope of CDF (diff.) we get PDF