

# SESSION 40 PROBABILITY-DISTRIBUTION

Random variable : →

\* What are algebraic Variables?

↳ In Algebra, like  $x$ , is (unknown value)

$$x + 10 = 5, \quad x = 5 - 10, \quad \boxed{x = -5}$$

\* Random variable in stats & probability  
→ (A random variable is set of possible value from a random experiment)

like A coin have  $\{H, T\}$  } sample  
dice =  $\{1, 2, 3, 4, 5, 6\}$  } space

Types

Discrete Random variable

Continuous Random variable

{ exact random value }  
 $\{H, T\}$   
 $\{1, 2, 3, 4, 5, 6\}$

→ value in a range  
 $\{0, 10\}$

\* Probability Distribution : →

↳ is a list of all of the possible outcome of a random variable along with their corresponding probability values.

Coin toss	1(H)	0(T)
Probability	$\frac{1}{2}$	$\frac{1}{2}$

Dice

1	2	3	4	5	6
$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

\* Lets take example of 2 Dice.

	1	2	3	4	5	6
1	3	4	5	6	7	8
2	4	5	6	7	8	9
3	5	6	7	8	9	10
4	6	7	8	9	10	11
5	7	8	9	10	11	12
6	8	9	10	11	12	

Possible outcome

value

P

2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	5/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

Problem with distribution →

↳ in many scenario the No. of outcome is much larger and hence Table would be impossible to write down, or outcome are infinite  
→ so, it impossible to write down table.

Example मानलो 1 Dice = {1, 2, 3, 4, 5, 6} possible  
लेकिन वही 11 Dice है possible outcome

$6^{11} = (6 \text{ Billions})$  impossible

Solution → instead of Table banana Hum formula or function use karate

① Probability Density function (PDF) →  
for continuous data

② Probability Mass function (PMF)  
for discrete data.

Probability distribution function



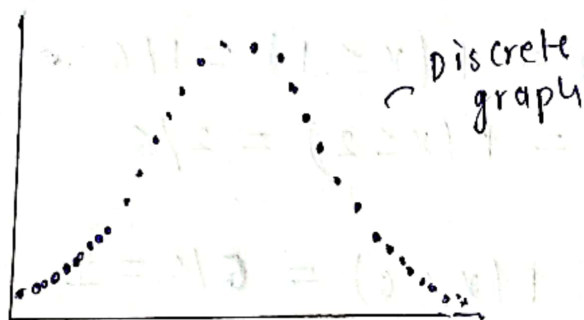
example : एक part, हम चलेगी किसी भी No. में गिरने का chance uniform hai  $(0, 1)$  by using formula

$$P(x) = 1 \quad (\text{for } 0 \leq x \leq 1)$$

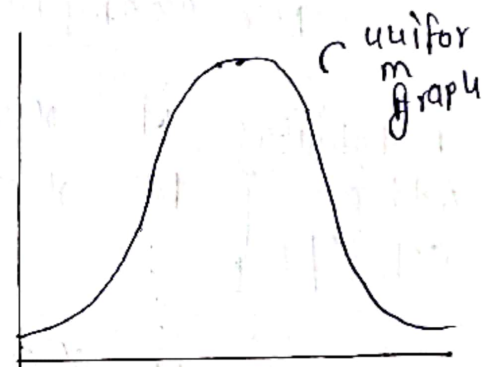
\* { A lot of time P-distribution & P-distribution function used interchangeably.

### Type of probability distribution (PDF)

↓  
Random  
Discrete variable



↓  
continuous  
Random variable



PMF  $\Rightarrow$  describe the probability distribution of discrete random variable.  $p \geq 0$  it cannot be -ve. and the sum of Probability equal to 1.

$$y = f(x) \quad y = \begin{cases} 1/6 & \text{if } x \in \{1, 2, 3, 4, 5\} \\ 0 & \text{otherwise} \end{cases}$$

PMF

for 2 Dice  $y = \begin{cases} 1/36 & \text{if } x \in \{2, 12\} \\ 2/36 & \text{if } x \in \{3, 11\} \\ \vdots & \vdots \end{cases}$

otherwise

\* Cumulative distribution function (CDF) or PMF

→ The CDF  $f(x)$  describes the probability that random variable  $x$  with a given probability distribution will be found at a value less than or equal to  $x$

$$f(x \leq x)$$

मतलब किसी random variable  $x$  किसी particular value  $x$  से छोटी या उसके बराबर होने की probability क्या है।

for example: → Dice =  $\{1, 2, 3, 4, 5, 6\}$

Probabilities Add हो रही हैं।

if  $x = 1$ ,  $f(1) = P(x \leq 1) = 1/6$   
if  $x = 2$ ,  $f(2) = P(x \leq 2) = 2/6$   
:  
if  $x = 6$ ,  $f(6) = P(x \leq 6) = 6/6 = 1$

\* Probability Density Function (PDF) :-

→ mathematical function which describes the probability distribution of continuous random variable.

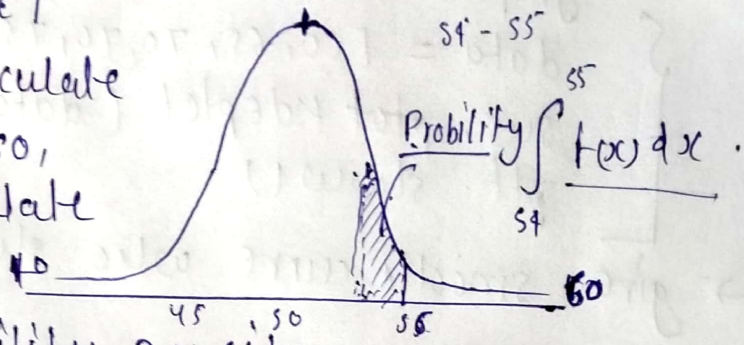
\* exact value की probability zero for continuous random variable (because their infinite possibilities होती हैं। इसीलिए एक Range के अन्दर probabilities बताता है।)

$$\int_a^b f(x) dx$$



Ex. एक exam में student के marks का distribution student के marks 40 से 60 के बीच में है।

→ is difficult to calculate the probability so, instead of calculate Probability we calculate probability density :  $\rightarrow$



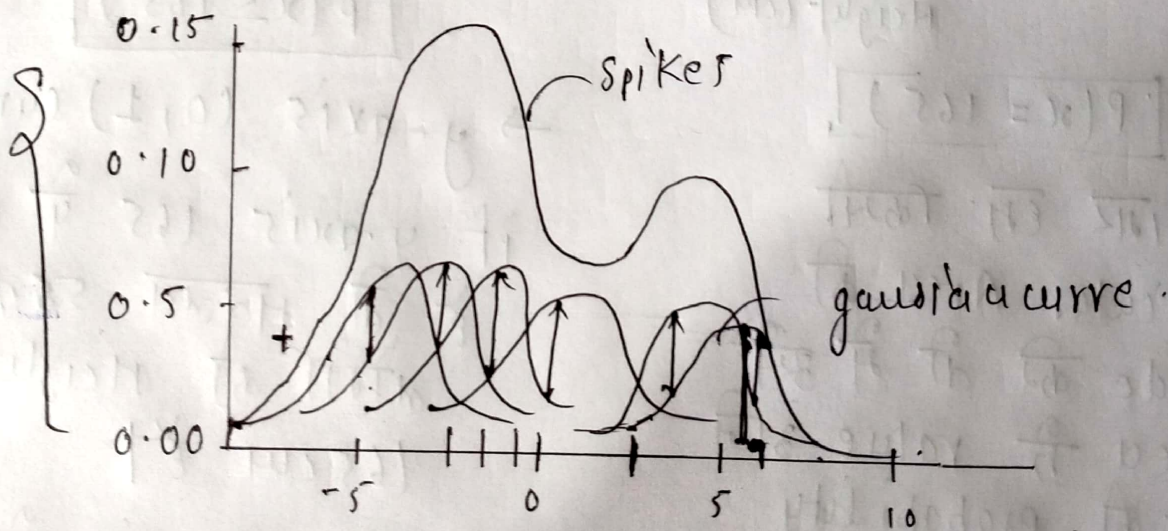
\* How the graph calculate :  $\rightarrow$

- ① collect points - ex. [60, 65, 70, 70, 75]
- ② Kernel  $\rightarrow$  Kernel is a mathematical function में हर data point को mean मान के gaussian curve बनाता है।

③ Decide smoothness :  $\rightarrow$  (Bandwidth) :

$\downarrow$  Bandwidth = Detail curve  
 $\uparrow$  Bandwidth = smooth curve.

$\rightarrow$  decide curve  
 / मा \  
 narrow wide



$\rightarrow$  हर data point के लिए Kernel function का curve बनाता है।

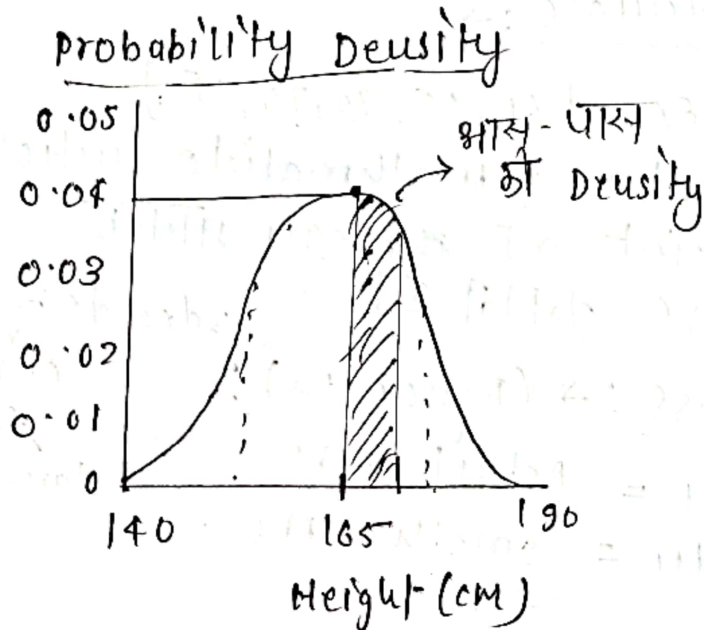
$\rightarrow$  फिर इन्हें Add करते हैं ताकि overall smooth curve मिले जो distribution दिखाता है।

Python : KDE plot

```
import seaborn as sns  
data = [60, 65, 70, 70, 75]  
sns.kdeplot(data, bw_adjust=0.5)  
plt.show()
```

it → give smooth curve who show the distribution

CDF of PDF

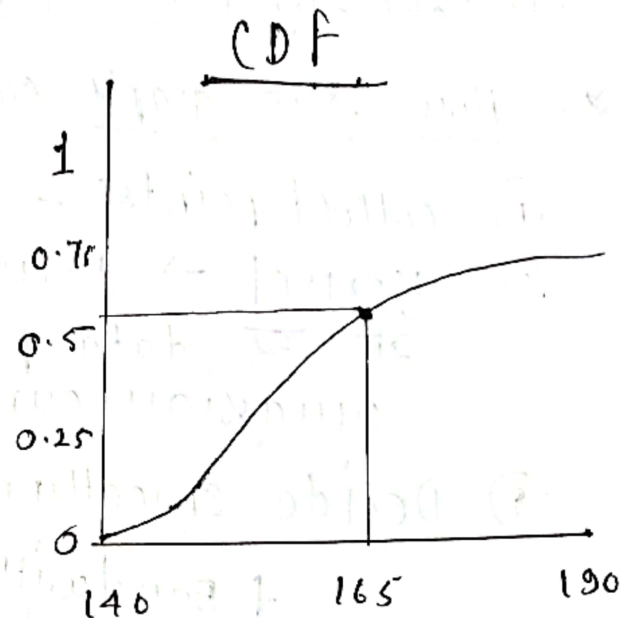


$$P(x = 165)$$

→ अगर हम किसी particular Area की shade करे तो मे उस Area में value आने की probability बताता है।

→ direct y-axis is not Probabilities

→ but if Area select then it is probability



$$P(x \leq 165)$$

→ y-axis (0, 1) cumm  
if y-axis 165 पे 0.5  
है तो मतलब 50% लोगो की height 165cm है।

cdf ka slope pdf के proportional होता है।