Probability Distribution

Probability is a measure of the likelihood or chance that a particular event will occur.

A distribution shows the possible values a random variable can take and how frequently they occur.

Random variable

$$X = \begin{cases} 0, & \text{if Head} \\ 1, & \text{if Tail} \end{cases}$$

Tossing a coin $\rightarrow$ H, T

$P(H) = \frac{1}{2} = \frac{\text{Favourable}}{\text{Sample space}}$

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

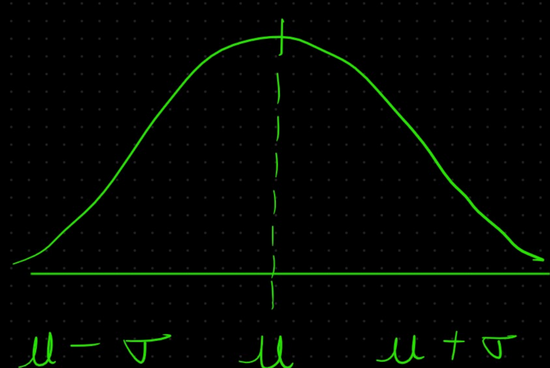
Rolling a dice $= \{1, 2, 3, 4, 5, 6\}$

$P(3) = \frac{1}{6}$, $P(4) = \frac{1}{6}$

$$0 \leq P \leq 1$$

impossibility

certainity



→ Probability Distribution describe the likelihood of different outcomes in a Random process or experiment.

Type of Distribution

(1) Discrete → finite number of outcomes

(2) Continuous → Infinite many consecutive possible values

Probability Distribution Func^n

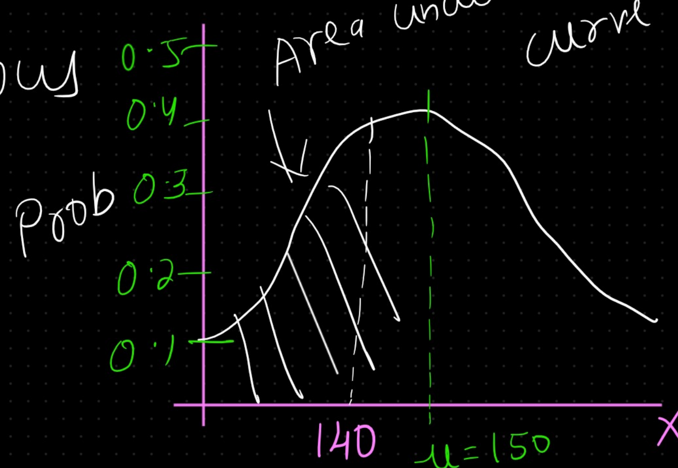
(1) Prob. Density Func^n (Pdf)

(2) Prob. Mass Func^n (pmf)

(3) Cummulative Distribution Func^n (cdf)

Height of students

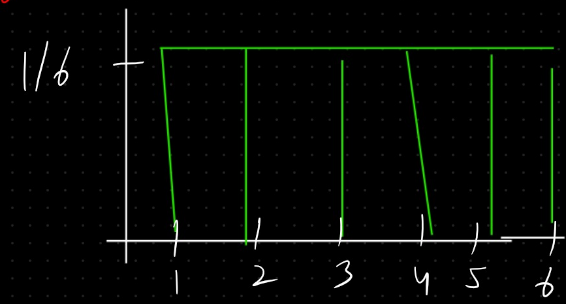
a) Pdf → continuous



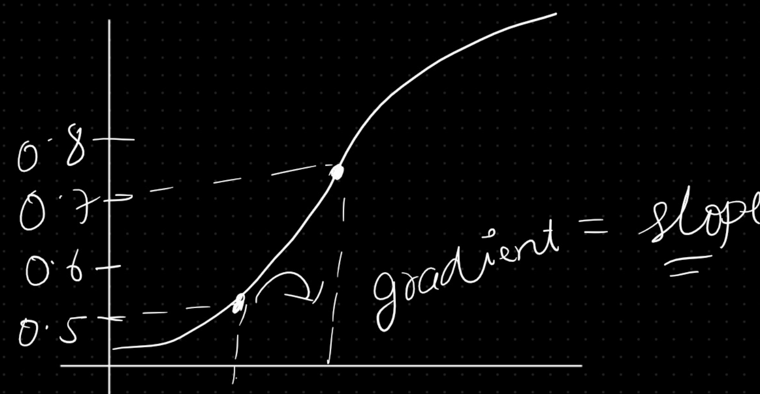
b) Pmf → Discrete →

Rolling a dice = $\{1, 2, 3, 4, 5, 6\}$

$P(1) = \frac{1}{6}$, $P(2) = \frac{1}{6}$, $P(3) = \frac{1}{6}$



3) cdf



→ CDF at a given point X is the sum of the prob. Values less than or equal to Random variable

Discrete Prob. Distribution

1. Bernoulli Distribution (pmf)

$$X = \begin{cases} 0 \rightarrow q = 1-p \\ 1 \rightarrow p \end{cases}$$

$\downarrow$

single experiement / trail

A single trail with possible outcomes

$p \rightarrow$ success

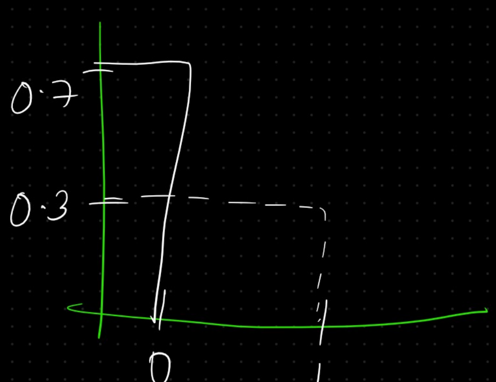
$q \rightarrow (1-p) \rightarrow$ failure

Exam: T / F

T

F

Notation: $\boxed{X \sim Bern(p)}$



$$p = \frac{1}{7} \qquad = \frac{7-1}{7}$$

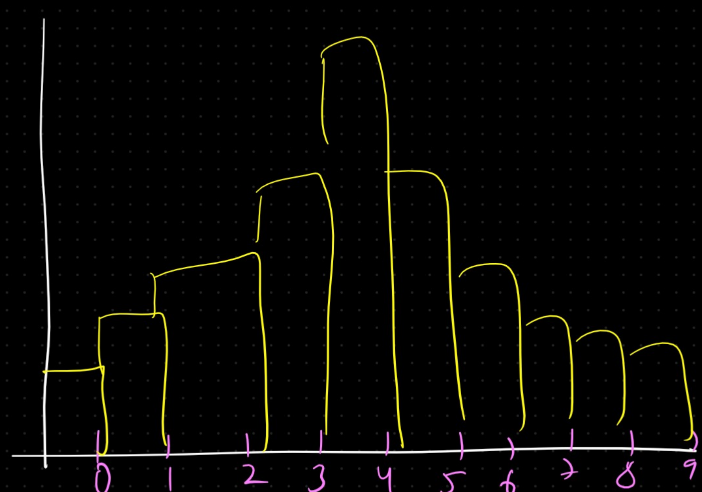
$$q = 1 - p = 1 - \frac{1}{7} = \frac{6}{7}$$

II) Binomial Distribution

no. of trial = n idependent experiement

→ The number of success in a fixed number of independent Bernoulli trial

$$\therefore \boxed{X \sim B(n, p)}$$



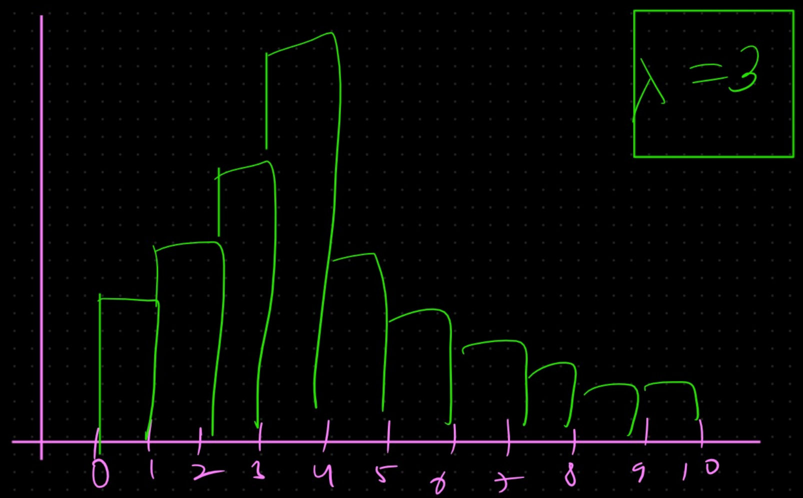
3. Poisson Distribution

→ It describe the events occurring in a fixed time Interval.

Eg: No of people visiting every hour Bank / Hospital

λ = Average Rate of events
Expected No of events to occur at every time interval

$$\boxed{X \sim P_0(\lambda)}$$

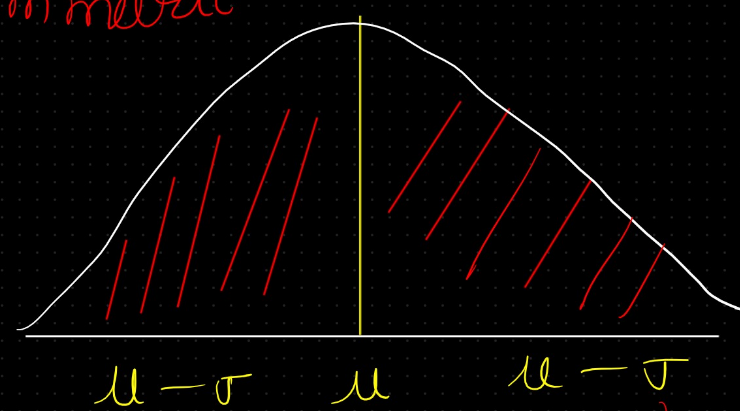


Continuous Probability Distribution

(1) Normal (Gaussian) Distribution

a → Bell - shaped

b) symmetric



$$\boxed{X \sim N(\mu, \sigma)}$$

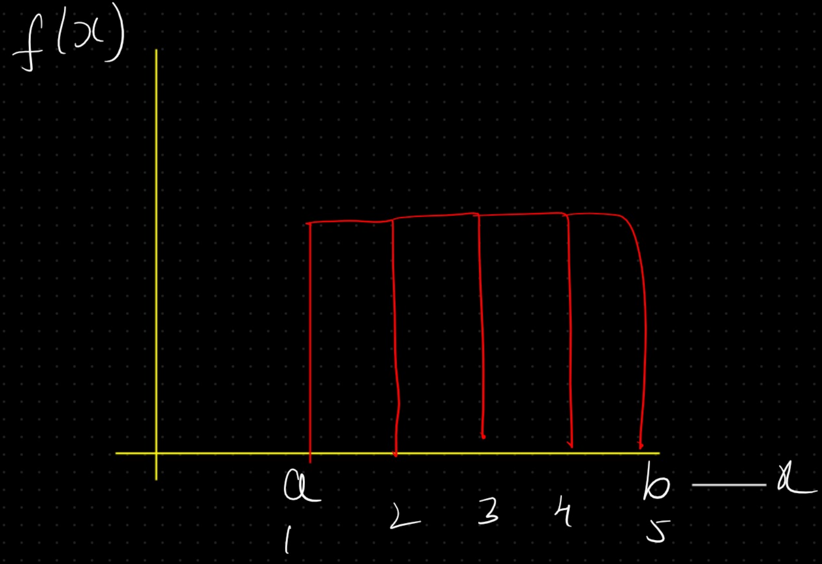
mean

standard deviation

II) Uniform Distribution

All outcomes are equally likely within a given range.

Parameters → Minimum (a) and Maximum (b)



Standard Normal Distribution

$$\boxed{\mu = 0, \quad \sigma = 1}$$

z-score

$$\boxed{Z = \frac{x - \mu}{\sigma}}$$

$N \sim (0, 1)$

→ detect outliers
→ test hypothesis testing
→ create confidence Interval
→ Regression Analysis

Central Limit Theorem

Objective: Simulate the Central Limit Theorem using a simple example of rolling a fair six-sided die.

Steps:

Roll a Die: We simulate rolling a fair six-sided die, which means it can land on any number from 1 to 6 with equal probability.

Calculate Sample Mean: We roll the die multiple times (let's say 30 times) to create a sample. Then, we calculate the average (mean) of those 30 rolls.

$S_{30} =$

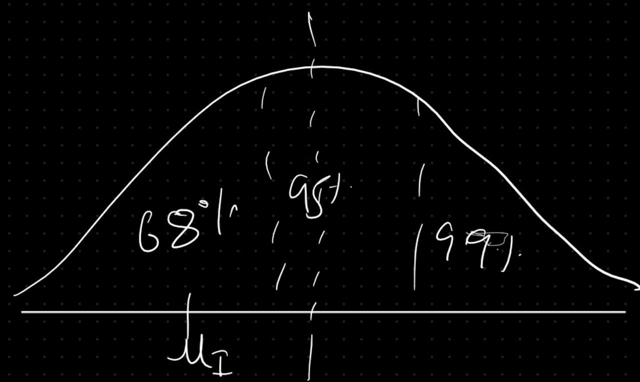
Repeat Many Times: We repeat this process many times (let's say 1000 times) to create a collection of sample means.

Observe Distribution: We plot a histogram of these sample means. According to the Central Limit Theorem, as we increase the number of samples, the distribution of sample means should start to resemble a normal (bell-shaped) distribution.

The Central Limit Theorem (CLT) is one of the greatest statistical insights. It states that no matter the underlying distribution of the dataset, the sampling distribution of the means would approximate a normal distribution. Moreover, the mean of the sampling distribution would be equal to the mean of the original distribution and the variance would be n times smaller, where n is the size of the samples. The CLT applies whenever we have a sum or an average of many variables (e.g. sum of rolled numbers when rolling dice)

Emperical Rule

68%, 95%. 99%



68% $(\mu - \sigma, \ \mu + \sigma)$

95% $(\mu - 2\sigma, \ \mu + 2\sigma)$

99% $(\mu - 3\sigma, \ \mu + 3\sigma)$