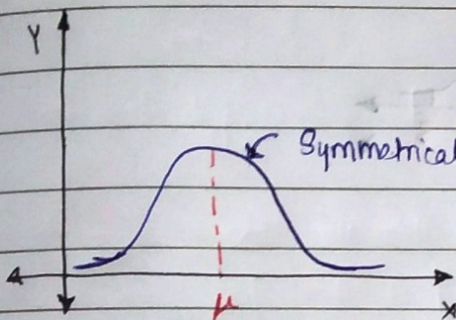


Normal distribution \rightarrow Gaussian Distribution

Key Feature: -

It is completely described by its mean (μ), and variance (σ^2) stated as $X \sim N(\mu, \sigma^2)$.
In other words, "X is normally distributed with mean (μ) and variance (σ^2)".



Mathematically,

$$Y = f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- Always symmetrical
- long tail
- Bell-shaped
- mean/median/mode are same.
- Two factors define the shape of curve are σ & μ .

② The normal distribution is the most important probability distribution stats, it fits in many natural phenomena. Eg:- Height, blood type, IQ, etc.

③ Symmetry - about σ and μ .

④ Measure of central tendency are equal - mean = median = mode.

⑤ Empirical Rule - (68-95-99.7 Rule)

(i) what % of data falls within a certain number of std. dev from mean.

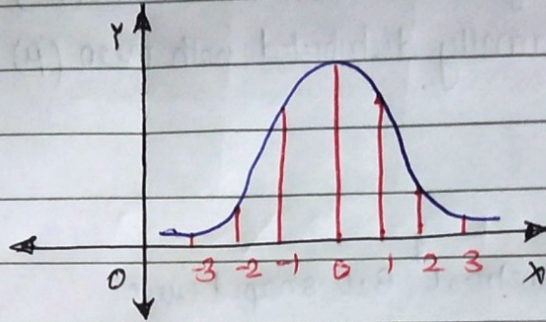
(ii) 68% of data falls within one std of the mean

(iii) 95.45% of data falls within two std. of the mean

(iv) 99.73% of data falls within three std. of the mean

- Standard Normal Variate :-

- standard Normal Variate (z) is a standardized form of the normal distribution with mean = 0 & Std. = 1



⑤ Skewness :- (asymmetry of Probability distribution)

- In a symmetrical distribution, the mean, median & mode are equal. In contrast, in a skewed distribution (Asymmetrical dist.), the mean, median & mode are not equal.
- Skewness can be positive, negative or zero.

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x - \bar{x}}{\sigma} \right)^3$$

Moment

→ Application in Data science :-

- ① Outlier detection
- ② Assumption on data for ML algorithm - Linear Regression, & GMM
- ③ Hypothesis testing
- ④ Central Limit theorem