Theoretical Probability Distributions

CA FDS

A theoretical probability distribution is a model that describes the probabilities of different outcomes in a random experiment. It's based on mathematical principles and assumptions, not on actual experimental results. These distributions help us understand and predict the likelihood of various outcomes

Key Concepts:

Mathematical Model: Theoretical distributions are mathematical functions that assign probabilities to each possible value of a random variable.

Based on Assumptions: They are built upon specific assumptions about the nature of the random process being modeled.

Not Experimental: Unlike experimental probability, which is determined from real-world data, theoretical probability is calculated using theoretical models.

Common Types: Some common examples include the binomial distribution, normal distribution, and Poisson distribution.

Examples of Theoretical Distributions:

Binomial Distribution: Models the probability of a certain number of successes in a fixed number of independent trials, like coin flips or multiple-choice questions.

Normal Distribution: A bell-shaped curve that describes the distribution of many natural phenomena, like height, weight, or test scores.

Poisson Distribution: Models the probability of a certain number of events occurring within a given time or space interval, like the number of customers arriving at a store in an hour.

Uniform Distribution: Assigns equal probability to each possible value within a given range.

Exponential Distribution: Models the time until an event occurs, like the time between customer arrivals or the time a piece of equipment operates before failure.

Uses of Theoretical Distributions:

Probability Theory: They provide a framework for understanding and calculating probabilities in various scenarios.

Statistics: They are used to analyze data, test hypotheses, and make inferences about populations.

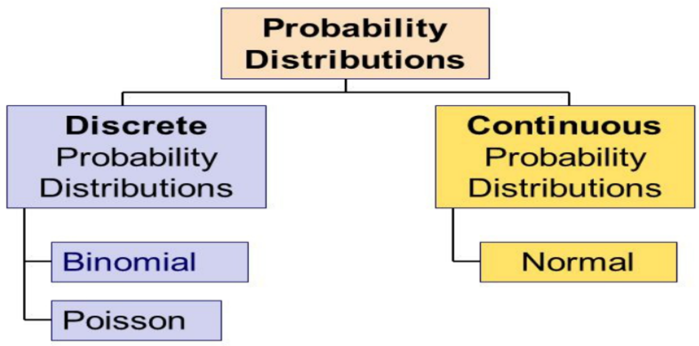
Data Science: They are essential for modeling and understanding random phenomena in various fields like forecasting, machine learning, and finance.

Physics and Engineering: They are used to model and analyze systems with random components

Note :-

1. Theorectical prob. dist^n may be profitably employed to make short term projections for the future.
2. Statistical analysis based on the basis of theoretical prob. dist^n.

Probability Distributions



* BINOMIAL DISTRIBUTION :-

Definition

The binomial distribution is a discrete probability distribution that describes the number of successes in a fixed number of independent Bernoulli trials, each with the same probability of success. It is an essential concept in AP Statistics and can be summarized with the following properties:

- **Bernoulli Trial:** An experiment or process that results in a binary outcome (success or failure).
- **Number of Trials (n):** The fixed number of independent Bernoulli trials.
- **Probability of Success (p):** The probability of success on a single trial.
- **Probability of Failure (q):** The probability of failure on a single trial, where q=1-p → $\boxed{p+q=1}$

- Each trial is associated with two mutually exclusive and exhaustive outcomes.
- No. of trials is a finite positive integer and trials are independent.

$B \sim (n, p)$ then

$$P(X=x) = {}^{n}C_{x} \cdot p^{x} \cdot q^{n-x} \quad \text{for } x = 0, 1, 2, \ldots n$$

Note :-

① Binomial distribution is known as biparametric distribution as it is characterised by two parameters n and p (written as $X \sim B(n, p)$).

No. of trials   Prob. of Success

This means if the values of n and p are known, then the dist^n is known completely.

(2) Mean & Variance of Binomial distribution $X \sim B(n, P)$

Mean $\mu = nP$

Variance $\sigma^2 = nPq$

Standard Deviation $\sigma = \sqrt{nPq}$

Mode $= [(n+1)P]$; $[\cdot] \to$ GINT (greatest integer function)

$= (n+1)P, \; (n+1)P-1$ if Not integer

$\because$ P and q are neumerically less than or equal to 1 i.e $nPq < nP$

$\Rightarrow$ Variance < mean

Variance of a Binomial distⁿ is always less than its mean.

& Variance of X attains its max value only when $P = q = 1/2$ & max value of Variance is $\frac{n}{4}$.

(3) If X and Y are two independent Variables such that

$X \sim B(n_1, P)$ and $Y \sim B(n_2, P)$

then $(X+Y) \sim B(n_1 + n_2, P)$

Q-1) If mean and SD of a Binomial distribution is 10 and $\sqrt{5}$ respectively. find mode.

Ans.

$nP = 10$

$\sqrt{nPq} = \sqrt{5}$

$nPq = 5$

$q = \frac{5}{10}$

$q = 1/2$

$\because P + q = 1$

$\therefore P = 1/2$

$\therefore n\left(\frac{1}{2}\right) = 10$

$n = 20$

Mode $= [(n+1)P]$

$= \left[(20+1)\frac{1}{2}\right]$

$= [10.5]$

$= 10$ Ans

Q-2) A coin is tossed 10 times, Assuming the coin to be unbiased, what is the probability of getting

1) 4 Heads
2) Atleast 4 Heads
3) Atmost 3 Heads

Ans

P: getting Head $P = 1/2$

$q = 1/2$, $n = 10$

$$P(4 \text{ Heads}) = {}^{10}C_4 \cdot \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^6 = \frac{210}{2^{10}} = \frac{210}{1024} = \frac{105}{512}$$

$$P(\text{At least 4 Heads}) = P(X \geq 4)$$

$$= 1 - \left(P(X=0) + P(X=1) + P(X=2) + P(X=3)\right)$$

$$= 1 - \left({}^{10}C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10} + {}^{10}C_1 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)^9 + {}^{10}C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^8 + {}^{10}C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^8\right)$$

$$= 1 - \left(\frac{1 + 10 + 45 + 120}{1024}\right)$$

$$= 1 - \frac{176}{1024} = \frac{848}{1024}$$

$P(\text{at most 3 Heads}) = P(X \le 3)$

$$= P(X=0) + P(X=1) + P(X=2) + P(X=3) = \frac{176}{1024}$$

Q-3) If 15 dates are selected at random, what is the prob. of getting two Sundays?

A. p : Prob. of a Sunday in a week $= 1/7$

$q = 1 - 1/7 = 6/7$

$$P(\text{Two Sundays}) = {}^{15}C_2 \cdot \left(\tfrac{1}{7}\right)^2 \left(\tfrac{6}{7}\right)^{13} = 0.29$$

Q-4) The incidence of occupational disease in an industry is such that the workmen have a 10% chance of suffering from it. What is the probability that out of 5 workmen, 3 or more will contract the disease?

A. p : Prob of Person having disease $= \frac{10}{100} = \frac{1}{10}$

$q = 1 - \frac{1}{10} = \frac{9}{10}$

$n = 5$

X : No. of Persons having disease

$$P(X \ge 3) = P(X=3) + P(X=4) + P(X=5)$$

$$= {}^5C_3 \cdot \left(\tfrac{1}{10}\right)^3 \left(\tfrac{9}{10}\right)^2 + {}^5C_4 \left(\tfrac{1}{10}\right)^4 \left(\tfrac{9}{10}\right)^1 + {}^5C_5 \left(\tfrac{1}{10}\right)^5 \left(\tfrac{9}{10}\right)^0$$

$$= \frac{(10)(81) + (5)(9) + 1}{10^5} = \frac{856}{100000} = 0.00856 \quad \text{A.}$$

Q Find the probability of a success for the binomial distribution satisfying the following relation 4 P (x = 4) = P (x = 2) and having the parameter n as six.

A. $4P(X=4) = P(X=2)$; $n = 6$; p = Prob. of success, q = Prob. of failure

$$4 \cdot \cancel{{}^6C_4} \cdot (p)^4 (q)^2 = \cancel{{}^6C_2} (p)^2 (q)^4$$

$$4p^2 = q^2$$

$$4p^2 = (1-p)^2$$

$$4p^2 = 1 + p^2 - 2p$$

$$3p^2 + 2p - 1 = 0 \Rightarrow p = -1 \; ✗, \; \boxed{p = 1/3} \; ✓$$

Q Find the binomial distribution for which mean and standard deviation are 6 and 2 respectively.

A. $np = 6$

$\sqrt{npq} = 2$

$npq = 4$

$6q = 4 \Rightarrow q = 2/3$

$\therefore \quad p = 1/3$

$\therefore n\left(\frac{1}{3}\right) = 6 \Rightarrow n = 18$

$\therefore \quad X \sim B(n, p)$

$\Rightarrow X \sim B(18, 1/3)$

$P(X = x) = {}^{18}C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{18-x}$ Ans.

Q. An experiment succeeds thrice as after it fails. If the experiment is repeated 5 times, what is the probability of having no success at all ?

Ans. p: Prob. of success
q: Prob. of failure

$p = 3q$

$\because p + q = 1 \Rightarrow 3q + q = 1 \Rightarrow q = 1/4$

$p = 3/4$

$n = 5$

X: having success

$P(X=0) = P(\text{No success})$

$= {}^5C_0 \left(\frac{3}{4}\right)^0 \left(\frac{1}{4}\right)^5 = \frac{1}{1024}$

Q. 6 coins are tossed 512 times. Find the expected frequencies of heads. Also, compute the mean and SD of the number of heads.

Ans. p: Prob. of a Head $= 1/2$
q: Prob. of Not a head $= 1/2$

$n = 6$

X: No. of Heads

$P(X = x) = {}^nC_x \cdot p^x q^{n-x}$

$= {}^6C_x \cdot \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{n-x}$

Mean $= np$

$= 6 \cdot \frac{1}{2} = 3$

Variance $= npq$

$= 6 \cdot \frac{1}{2} \cdot \frac{1}{2} = 1.5$

$SD = \sqrt{npq} = \sqrt{1.5} = 1.22$

Mode $= [(n+1)p]$

$= [7 \cdot \frac{1}{2}] = [3.5] = 3$ Ans.

Q. If x and y are 2 independent binomial variables with parameters 6 and 1/2 and 4 and 1/2 respectively, what is P (x + y ≥ 1)?

Ans. $X \sim B(6, 1/2)$

$Y \sim B(4, 1/2)$

$X + Y \sim B(10, 1/2)$ $\quad \underline{n=10} \quad \underline{p = 1/2} \quad \underline{q = 1/2}$

So, $P(X+Y \geq 1) = 1 - P(X+Y=0)$

$= 1 - {}^{10}C_0 \cdot \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{10} = 1 - \frac{1}{2^{10}} = \frac{1023}{1024}$

Q. Fit a binomial distribution to the following data:

x:	0	1	2	3	4	5
f:	3	6	10	8	3	2

Solution: In order to fit a theoretical probability distribution to an observed frequency distribution it is necessary to estimate the parameters of the probability distribution. There are several methods of estimating population parameters. One rather, convenient method is 'Method of Moments'. This comprises equating p moments of a probability distribution to p moments of the observed frequency distribution, where p is the number of parameters to be estimated. Since n = 5 is given, we need estimate only one parameter p. We equate the first moment about origin i.e. AM of the probability distribution to the AM of the given distribution and estimate p.

Mean $\mu = \frac{\sum f_i x_i}{\sum f_i}$ $\quad x \to 0$ to 5 $\quad \therefore \underline{n=5}$ ✓

$$\mu = \frac{0+6+20+24+12+10}{32}$$

$$= \frac{72}{32} = 2.25$$

$\because$ Mean $= np$

$2.25 = 5 \cdot p$

$p = \frac{2.25}{5} = 0.45$

$q = 0.55$

So, $X \sim B(5, 0.45)$

$$P(X=x) = {}^5C_x \cdot (0.45)^x (0.55)^{5-x}$$

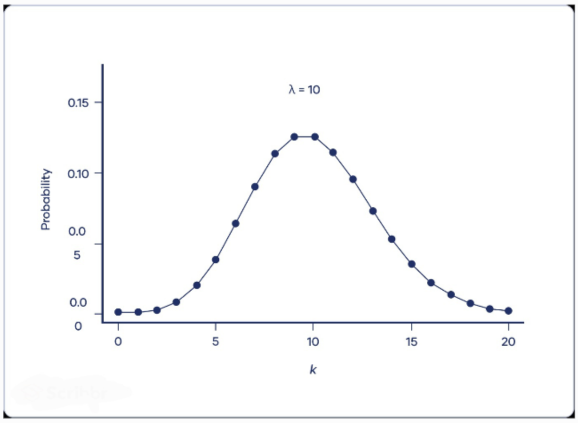
Ans

*

* Poisson Distribution :-

In probability theory and statistics, the Poisson distribution is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known constant mean rate and independently of the time since the last event.

A Poisson distribution is a discrete probability distribution. It gives the probability of an event happening a certain number of times (k) within a given interval of time or space. The Poisson distribution has only one parameter, λ (lambda), which is the mean number of events.



Binomial distribution describes the distribution of binary data from a finite sample. Thus it gives the probability of getting r events out of n trials. Poisson distribution describes the distribution of binary data from an infinite sample. Thus it gives the probability of getting r events in a population.

For the Poisson distribution to be accurate, all events must be independent of each other, the rate of events over time must be constant, and events cannot occur simultaneously. Moreover, the mean and the variance will be equal.

The Poisson distribution is a limiting case of the binomial distribution which arises when the number of trials n increases indefinitely whilst the product μ = np, which is the expected value of the number of successes from the trials, remains constant.

A Random Variable x is define to follow Poisson distribution with parameter λ, to be denoted by $X \sim P(\lambda)$ if the prob. mass function of x is given by

$$\boxed{\begin{aligned} f(x) = P(X=x) &= \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} \quad \text{for } x = 0, 1, 2, \ldots, \infty \\ &= 0 \quad \text{Elsewhere} \end{aligned}}$$

e = Transcendental Quantity

$e = 2.71828$

Note :-

1. $P(X=x) \geq 0 \ \forall x$ for every value of λ. ($\lambda > 0$)
2. Poisson distⁿ is uniparametric distⁿ (only one parameter λ)
3. for poisson distⁿ

   mean $\mu = \lambda$

   Variance $\sigma^2 = \lambda$
4. Like Binomial distⁿ, poisson distⁿ Could be also unimodel or bimodel depending upon the value of the parameter λ.

   If λ is an integer then mode $= \lambda$ and $\lambda - 1$

   If λ is not an integer then mode $= [\lambda]$; $[\cdot] \to$ GINT

5. **Poisson approximation to Binomial distribution**

   If n, the number of independent trials of a binomial distribution, tends to infinity and p, the probability of a success, tends to zero, so that m = np remains finite, then a binomial distribution with parameters n and p can be approximated by a Poisson distribution with parameter m (= np).

   In other words when n is rather large and p is rather small so that m = np is moderate then

   β (n, p) $\cong$ P (m). .... (16.14)

6. **Additive property of Poisson distribution**

   If X and y are two independent variables following Poisson distribution with parameters m_1 and m_2 respectively, then Z = X + Y also follows Poisson distribution with parameter $(m_1 + m_2)$.

   i.e. if $X \sim P(m_1)$

   and $Y \sim P(m_2)$

   and X and Y are independent, then

   $Z = X + Y \sim P(m_1 + m_2)$ .... (16.15)

Examples of Poisson distⁿ :-

1. The distⁿ of the no. of printing mistakes per page of a large book
2. The distⁿ of the no. of road accidents on a busy road per minute
3. The distⁿ of the no. of radio-active elements per minute in a fusion process
4. The distⁿ of the no. of demands per minute for health centre.

Q. find the mean and SD of x, where x is a Poisson variate function satisfying the Cond^n $P(X=2) = P(X=3)$

Ans $X \sim P(m)$

$$P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$$

here, $P(X=2) = P(X=3)$

$$\frac{e^{-\lambda} \cdot \lambda^2}{2!} = \frac{e^{-\lambda} \cdot \lambda^3}{3!} \Rightarrow \lambda = 3$$

So, mean $= 3$

Variance $= 3$

$SD = \sqrt{3} = 1.732$

Q The Prob. that a random variable X follows Poisson dist^n would assume a positive value is $(1-e^{-2.7})$. find mode of the dist^n

Ans $X \sim P(\lambda)$

$$P(X>0) = 1-e^{-2.7}$$

$$\Rightarrow 1 - P(X=0) = 1-e^{-2.7}$$

$$\Rightarrow 1 - \frac{e^{-\lambda} \cdot \lambda^0}{0!} = 1-e^{-2.7}$$

$$\Rightarrow 1-e^{-\lambda} = 1-e^{-2.7}$$

$$\lambda = 2.7$$

Mode $= [\lambda]$

$= [2.7] = 2$ Ans.

Q The standard deviation of a poisson variate is 1.732. what is the prob. that the variate lies b/w -2.3 to 3.68

Ans $\sqrt{\lambda} = 1.732 = \sqrt{3}$; $X \sim P(\lambda)$

$\lambda = 3$

$$P(-2.3 < x < 3.68) = \cancel{P(x=-2)}^{0} + \cancel{P(x=-1)}^{0} + P(x=0) + P(x=1) + P(x=2) + P(x=3)$$

$$= \sum_{x=0}^{3} \frac{e^{-\lambda} \cdot \lambda^x}{x!} = e^{-\lambda}\left(\frac{\lambda^0}{0!} + \frac{\lambda^1}{1!} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!}\right)$$

$$= e^{-3}\left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6}\right)$$

$$= e^{-3}\left(1 + 3 + \frac{9}{2} + \frac{27}{6}\right)$$

$$= 13 \cdot e^{-3}$$

$$= \frac{13}{e^3} = \frac{13}{(2.71)^3} = 0.65$$ Ans.

Q. X is a Poisson variate satisfying

$P(X=2) = 9P(X=4) + 90P(X=6)$, find SD of X.

Ans.

$$\frac{e^{-\lambda} \cdot \lambda^2}{2!} = \frac{9e^{-\lambda} \cdot \lambda^4}{4!} + \frac{90 \cdot e^{-\lambda} \cdot \lambda^6}{6!}$$

$$\Rightarrow \frac{\lambda^2}{2} = \frac{9\lambda^4}{24} + \frac{90\lambda^6}{720}$$

$\Rightarrow \quad \frac{1}{2} = \frac{3}{8}\lambda^2 + \frac{\lambda^6}{8}$

$\Rightarrow \quad 4 = 3\lambda^2 + \lambda^6$

$\lambda = 1 \qquad$ so, $SD = \sqrt{\lambda} = \sqrt{1} = 1$ Ans.

Q. A discrete random variable x follows Poisson distⁿ. find
i) $P(X = \text{at least } 1)$ $\qquad$ ii) $P(X \le 2 \mid X \ge 1)$
given that $E(X) = 2.20$ and $e^{-2.2} = 0.1108$

Ans. $X \sim P(\lambda)$ $\qquad P(X=x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}$, $\quad E(X) = 2.20$, $\lambda = 2.2$

1) $P(X \ge 1) = 1 - P(X=0)$

$$= 1 - \frac{e^{-2.2} \cdot (2.2)^0}{0!}$$

$$= 1 - (0.1108) = 0.8892$$

2) $P(X \le 2 \mid X \ge 1) = \frac{P(X \le 2 \cap X \ge 1)}{P(X \ge 1)}$

$$= \frac{P(1 \le X \le 2)}{P(X \ge 1)} = \frac{P(X=1) + P(X=2)}{1 - P(X=0)}$$

$$= \frac{\frac{e^{-\lambda} \cdot \lambda^1}{1!} + \frac{e^{-\lambda} \cdot \lambda^2}{2!}}{1 - \frac{e^{-\lambda} \cdot \lambda^0}{0!}}$$

$$= \frac{\lambda e^{-\lambda} + \frac{\lambda^2 e^{-\lambda}}{2}}{1 - e^{-\lambda}} = \frac{e^{-\lambda}\left(\lambda + \frac{\lambda^2}{2}\right)}{1 - e^{-\lambda}}$$

$\lambda = 2.2$

$$= \frac{e^{-2.2}\left(2.2 + \frac{4.84}{2}\right)}{1 - e^{-2.2}}$$

$$= \frac{(0.1108)(4.62)}{1 - 0.1108} = 0.5756$$

Q. Between 9 and 10 AM, the average number of phone calls per minute coming into the switchboard of a company is 4. Find the probability that during one particular minute, there will be,

1. no phone calls
2. at most 3 phone calls (given $e^{-4} = 0.018316$)

Ans. $\lambda = 4$ $\qquad$ Poisson distⁿ as average given & small time interval, X: Phone calls

i) $P(\text{No Phone Calls}) = \frac{e^{-\lambda} \cdot \lambda^0}{0!} = \frac{e^{-4} \cdot 1}{1} = e^{-4} = 0.018316$
$= P(X=0)$

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

$$= \frac{e^{-\lambda} \cdot \lambda^0}{0!} + \frac{e^{-\lambda} \cdot \lambda^1}{1!} + \frac{e^{-\lambda} \cdot \lambda^2}{2!} + \frac{e^{-\lambda} \cdot \lambda^3}{3!}$$

$$= e^{-d}\left(1+d+\frac{d^2}{2}+\frac{d^3}{6}\right)$$

$$= e^{-4}\left(1+4+8+\frac{32}{3}\right) = 0.4334$$

Q. If 2 per cent of electric bulbs manufactured by a company are known to be defectives, what is the probability that a sample of 150 electric bulbs taken from the production process of that company would contain

1. exactly one defective bulb?
2. more than 2 defective bulbs?

Ans. 2 Percent out of 150

(Why Poisson ↓ bcz large sample से small sample को use कर रहे है)

Mean $d = \frac{2}{100}\cdot 150 = 3$

X: defective bulbs

i) $P(X=1) = \frac{e^{-d}\, d^1}{1!} = 3\cdot e^{-3} = 0.149$

ii) $P(X>2) = 1-\left(P(X=0)+P(X=1)+P(X=2)\right)$

$$= 1-\left(\frac{e^{-d}.d^0}{0!}+\frac{e^{-d}.d^1}{1!}+\frac{e^{-d}.d^2}{2!}\right)$$

$$= 1-e^{-d}\left(1+d+\frac{d^2}{2}\right)$$

$$= 1-e^{-3}(8.5) = 0.57$$ Ans.

Q. The manufacturer of a certain electronic component is certain that two per cent of his product is defective. He sells the components in boxes of 120 and guarantees that not more than two per cent in any box will be defective. Find the probability that a box, selected at random, would fail to meet the guarantee? Given that $e^{-2.40} = 0.0907$.

Ans. 2 % defective

Mean of defective items $= \frac{2}{100}\times 120 = 2.4$

$$P(X>2.4) = 1-\left(P(X=0)+P(X=1)+P(X=2)\right)$$

$$= 1-\left(\frac{e^{-d}.d^0}{0!}+\frac{e^{-d}.d^1}{1!}+\frac{e^{-d}.d^2}{2!}\right)$$

$$= 1-e^{-d}\left(1+d+\frac{d^2}{2}\right)$$

$$= 1-e^{-2.4}\left(1+2.4+\frac{5.76}{2}\right)$$

$$= 1-(0.0907)(6.28)$$

$$= 0.43$$ Ans

* NORMAL DISTRIBUTION :-

A normal distribution, also known as a Gaussian distribution, is a probability distribution that is symmetrical around its mean, forming a bell-shaped curve when graphed. In this distribution, data points cluster most frequently around the mean and become less frequent as they move further away from the mean in either direction

Symmetry: The left and right sides of the bell curve are mirror images of each other.

Mean, Median, and Mode: In a normal distribution, the mean, median, and mode are all equal and located at the center of the curve.

Bell Curve: The characteristic shape of the normal distribution is a bell curve, with the highest point representing the mean and the curve tapering off towards the tails.

Frequency: Data points are most frequently found near the mean, with the frequency decreasing as you move away from the mean in either direction.

Standard Deviation: The standard deviation of a normal distribution indicates how spread out the data is. A larger standard deviation means the data is more spread out from the mean, while a smaller standard deviation means the data is more clustered around the mean.

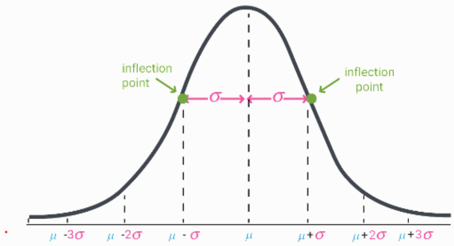
Empirical Rule:

The empirical rule (or 68-95-99.7 rule) states that in a normal distribution:

Approximately 68% of the data falls within one standard deviation of the mean.

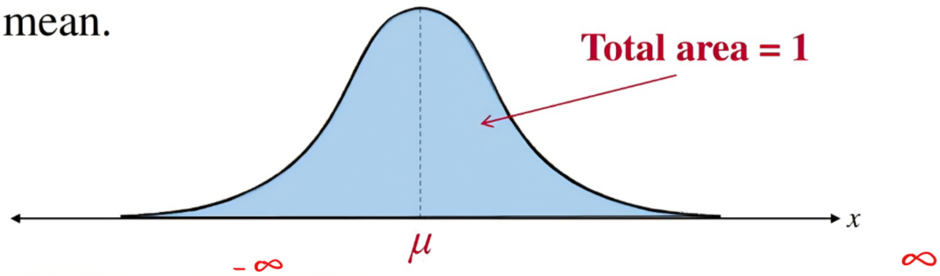
Approximately 95% of the data falls within two standard deviations of the mean.

Approximately 99.7% of the data falls within three standard deviations of the mean.



Properties of Normal Distributions

1. The mean, median, and mode are equal. (unique)
2. The normal curve is bell-shaped and is symmetric about the mean.
3. The total area under the normal curve is equal to 1.
4. The normal curve approaches, but never touches, the x-axis as it extends farther and farther away from the mean.



Normal distⁿ is known as biparametric distribution as it is characterized by two parameters μ & σ^2

Area b/w $-\infty$ to $\mu = 1/2 =$ Area b/w μ to ∞

A Continuous random variable X is defined to follow normal distribution with parameters μ and σ^2, to be denoted by

$$X \sim N(\mu, \sigma^2)$$

$\mu \to$ mean
$\sigma^2 \to$ variance

If the Probability density funcn of the random variable X is given by

$$\boxed{f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(\bar{x}-\mu)^2}{2\sigma^2}}} \quad \text{for } x \in (-\infty, \infty)$$

$\mu, \sigma \to$ constants
$\sigma > 0$

If we take $\mu = 0, \sigma = 1$

then $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$; $z \to$ Standard normal variate/ Standard normal deviate

→ The Prob. that a standard normal variate X would take a value less than or equal to a particular value say $X = x$ is given by

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

↑
Cumulative distribution function

$\phi(0) = P(X \le 0)$
= Area of the Standard normal Curve b/w $-\infty$ and 0
$= 0.5$

* Properties of Normal distribution :-

i) $\because \pi = \frac{22}{7}$, $e^{-\theta} = \frac{1}{e^{\theta}} > 0 \;\; \forall \; \theta \in \mathbb{R}$

So, $f(x) \geq 0 \;\; \forall x$

also $\int_{-\infty}^{\infty} f(x) = 1$

ii) The SD of the Normal distn is given by σ

and mean deviation of normal distn is $\sigma\sqrt{\frac{2}{\pi}} = 0.8\sigma$

also, the first & third Quartile is given by

$Q_1 = \mu - 0.675\sigma$

$Q_3 = \mu + 0.675\sigma$

So, Quartile deviation $= 0.675\sigma$

iii) The Normal distn is Symmetrical about $x = \mu$.

Its Skewness is zero i.e. the normal curve is neither inclined more towards the right (negatively skewed) nor towards the left (positively skewed)

iv) The Normal Curve $y = f(x)$ has two points of inflexion to be given by $x = \mu - \sigma$ and $x = \mu + \sigma$ i.e, at these two points, the normal curve changes its curvature from concave to convex and from convex to concave.

v) If $X \sim N(\mu, \sigma^2)$

then $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$

Z is known as standardised Normal Variate or Normal deviate

$P(Z \le K) = \phi(K)$

$\uparrow$

The values of $\phi(K)$ for different 'K' is known as "Biometrika"

also $\phi(-K) = 1 - \phi(K)$

$P(X \le a) = P(X < a)$ as X is continuous

and $P(X<a) = P\left(\frac{X-\mu}{\sigma} < \frac{a-\mu}{\sigma}\right)$

$= P(Z<K)$, $K = \frac{a-\mu}{\sigma}$

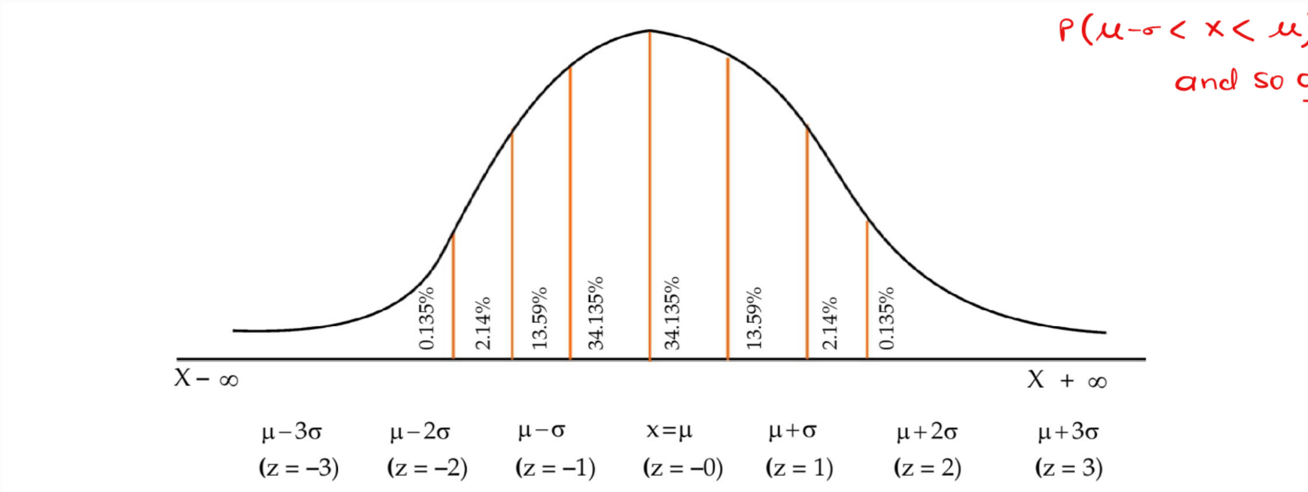
$= \phi(K)$

$P(X>b) = 1 - P(X \le b)$

$= 1 - \phi\left(\frac{b-\mu}{\sigma}\right)$

and $P(a<X<b) = \phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$

vi) Area under the Normal Curve :-



$P(\mu - \sigma < X < \mu) = 0.34135$ and so on.

99.73 % of the values of a normal variable lies b/w $(\mu - 3\sigma)$ to $(\mu + 3\sigma)$ thus, the prob. that a value of X lies outside the limit is as low as 0.0027.

vii) If X and Y are independent normal variables with means and S.D's as μ_1 and μ_2 and σ_1 and σ_2 respectively then $Z = X + Y$ also follows Normal dist^n with mean $(\mu_1 + \mu_2)$ and $SD = \sqrt{\sigma_1^2 + \sigma_2^2}$ respectively i.e If $X \sim N(\mu_1, \sigma_1^2)$

$Y \sim N(\mu_2, \sigma_2^2)$

then $X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$

* Applications of Normal Distribution :-

The applications of normal distribution is not restricted to statistics only. Many science subjects, social science subjects, management, commerce etc. find many applications of normal distributions. Most of the continuous variables like height, weight, wage, profit etc. follow normal distribution. If the variable under study does not follow normal distribution, a simple transformation of the variable, in many a case, would lead to the normal distribution of the changed variable. When n, the number of trials of a binomial distribution, is large and p, the probability of a success, is moderate i.e. neither too large nor too small then the binomial distribution, also, tends to normal distribution. Poisson distribution, also for large value of m approaches normal distribution. Such transformations become necessary as it is easier to compute probabilities under the assumption of a normal distribution. Not only the distribution of discrete random variable, the probability distributions of t, chi-square and F also tend to normal distribution under certain specific conditions. In order to infer about the unknown universe, we take recourse to sampling and inferences regarding the universe is made possible only on the basis of normality assumption. Also the distributions of many a sample statistic approach normal distribution for large sample size.

Q. for a random variable x, the prob. density func^n is given by $f(x) = \dfrac{e^{-(x-4)^2}}{\sqrt{\pi}}$ for $-\infty < x < \infty$. Identify the dist^n & also find its mean & variance

Ans $f(x) = \dfrac{e^{-(x-4)^2}}{\sqrt{\pi}}$ Compare with $\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

$\mu = 4$, $2\sigma^2 = 1$, $\sigma\sqrt{2} = 1$

$\mu = 4$ $\sigma^2 = 1/2$ $\sigma = 1/\sqrt{2}$

Normal dist^n $N \sim (\mu, \sigma^2)$

$= N \sim (4, 1/2)$ Mean $= 4$

Variance $= 1/2$

Q. If the two quartiles of a Normal dist^n are 47.30 and 52.70 respectively, what is the mode of the dist^n ?

Also find Mean deviation about the median.

Ans $Q_1 = 47.30$ $Q_3 = 52.70$ $X \sim N(\mu, \sigma^2)$

$\mu - 0.675\sigma = 47.30$ $\mu + 0.675\sigma = 52.70$ $\sim N(50, 16)$

$2\mu = 100$ $1.35\sigma = 5.4$ Mode = Median = Mean

$\mu = 50$ $\sigma = 4$ $= \mu = 50$ Ans.

→ Mean deviation about the Median/Mode/Mean $= 0.8\sigma$

$= (0.8)(4) = 3.2$ Ans.

Q. X follows Normal distⁿ with mean as 50 & variance as 100. What is $P(X \geq 60)$; given $\phi(1) = 0.8413$

Ans $\mu = 50$ $\quad P(X<a) = P(Z<K) = \phi(K)$

$\sigma^2 = 100$ $\quad$ where $K = \frac{a-\mu}{\sigma}$, $Z = \frac{X-\mu}{\sigma}$

$\sigma = 10$

Now,

$$P(X \geq 60) = 1 - P(X<60)$$
$$= 1 - P(Z<K)$$
$$= 1 - P\left(Z < \frac{60-50}{10}\right)$$
$$= 1 - P(Z<1)$$
$$= 1 - \phi(1)$$
$$= 1 - 0.8413 = 0.16 \quad \text{Ans.}$$

Q X is a Normal variable with mean 25 and SD = 10 find the value of b such that the prob. of the interval [25, b] is 0.4772, given $\phi(2) = 0.9772$

Ans. $X \sim N(\mu, \sigma^2)$

$X \sim N(25, 100)$

given $\quad P(25 < X < b) = 0.4772$

$$P\left(\frac{25-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right) = 0.4772$$

$$P\left(0 < \frac{X-25}{10} < \frac{b-25}{10}\right) = 0.4772$$

$$\phi\left(\frac{b-25}{10}\right) - \phi(0) = 0.4772 \qquad \phi(0) = 0.5$$

$$\phi\left(\frac{b-25}{10}\right) = 0.4772 + 0.5$$
$$= 0.9772$$
$$= \phi(2)$$

$$\frac{b-25}{10} = 2 \Rightarrow b = 25 \quad \text{Ans}$$

Q X and Y are independent normal variables with mean 100 & 80 respectively & standard deviation as 4 & 3 respectively. What is the distⁿ of X+Y.

Ans $X \sim N(\mu_1, \sigma_1^2) = X \sim N(100, 16)$

$Y \sim N(\mu_2, \sigma_2^2) = Y \sim N(80, 9)$

So, $X+Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$

$X+Y \sim N(180, 16+9) = N(180, 25)$ Ans.

Some Imp. Problems :-

Q. If the 1st Quartile & Mean deviation about median of a Normal distn are 13.25 & 8 respectively. find mode of the distn.

A $X \sim N(\mu, \sigma^2)$

$Q_1 = \mu - 0.675\sigma = 13.25$, Mean deviation about Median/Mode/Mean $= 0.8\sigma$

$0.8\sigma = 8$

$\boxed{\sigma = 10}$

$\mu - (0.675)(10) = 13.25$

$\mu = 13.25 + 6.75$

$\mu = 20 \rightarrow$ Mean/Mode/Median.

Q If the Quartile deviation of a Normal Curve is 4.05 then find its Mean deviation.

A Quartile deviation $= 4.05$

$0.675\sigma = 4.05$

$\sigma = 6$

Mean deviation $= 0.8\sigma$
$= (0.8)(6)$
$= 4.8$

Q. If the points of inflexion of a normal Curve are 40 & 60 respectively then find its mean deviation.

A. $\mu - \sigma = 40$
$\mu + \sigma = 60$

$2\sigma = 20$

$\sigma = 10$

& Mean $= 50$

Meandeviation $= 0.8\sigma$
$= (0.8)(10)$
$= 8$

Q. If x having the following prob. density function $f(x) = \frac{1}{\sqrt{72\pi}} e^{\frac{-(x-10)^2}{72}}$

then find Mean, Median, Mode, first Quartile (Q_1), third Quartile (Q_3), Mean deviation, Quartile deviation, SD, Variance.

A
$$\frac{1}{\sqrt{72\pi}} e^{\frac{-(x-10)^2}{72}} = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Compare, $\sigma\sqrt{2\pi} = \sqrt{72\pi}$ $2\sigma^2 = 72$

$\sigma^2 \cdot 2\pi = 72\pi$ $\sigma^2 = 36$

$\sigma^2 = 36$, $\mu = 10$

$X \sim N(\mu, \sigma^2)$
$X \sim N(10, 36)$

So, Mean = Median = Mode $= \mu = 10$

Variance $= \sigma^2 = 36$

SD $= \sigma = 6$

first Quartile $Q_1 = \mu - 0.675\sigma = 10 - (0.675)(6) = 5.95$

Third Quartile $Q_3 = \mu + 0.675\sigma = 10 + (0.675)(6) = 14.05$

Mean deviation $= 0.8\sigma$
$= (0.8)(6) = 4.8$

Quartile deviation $= 0.675\sigma = (0.675)(6) = 4.05$

Q If the two Quartiles of $N(\mu, \sigma^2)$ are 14.6 and 25.4 respectively, what is the SD, Mean deviation, Quartile deviation of the distn.

A) $Q_1 = 14.6$ $\qquad$ $Q_3 = 25.4$

$\mu - 0.675\sigma = 14.6$ $\qquad$ $\mu + 0.675\sigma = 25.4$

on solving $2\mu = 40.0$ & $1.35\sigma = 10.8 \Rightarrow \sigma = \frac{10.8}{1.35} = 8$

$\mu = 20$

i) $SD = \sigma = 8$

ii) Mean deviation $= 0.8\sigma = (0.8)(8) = 6.4$

iii) Quartile deviation $= 0.675\sigma$
$= (0.675)(8) = 5.4$

iv) Mean = Median = Mode $= \mu = 20$

* Standard Normal distribution :-

If a continuous random variable z follows standard normal distribution, to be denoted by $z \sim N(0, 1)$, then the probability density function of z is given by

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{for } -\infty < z < \infty$$

Some important properties of z are listed below :

(i) z has mean, median and mode all equal to zero.

(ii) The standard deviation of z is 1. Also the approximate values of mean deviation and quartile deviation are 0.8 and 0.675 respectively.

(iii) The standard normal distribution is symmetrical about $z = 0$.

(iv) The two points of inflexion of the probability curve of the standard normal distribution are –1 and 1.

(v) The two tails of the standard normal curve never touch the horizontal axis.

(vi) The upper and lower p per cent points of the standard normal variable z are given by

$$P(Z > z_p) = p$$

And $P(Z < z_{1-p}) = p$

i.e. $P(Z < -z_p) = p$ respectively

(since for a standard normal distribution $z_{1-p} = -z_p$)

Selecting P = 0.005, 0.025, 0.01 and 0.05 respectively,

We have $\quad$ $z_{0.005} = 2.58$

$z_{0.025} = 1.96$

$z_{0.01} = 2.33$

$z_{0.05} = 1.645$

(vii) **If $\overline{x}$ denotes the arithmetic mean of a random sample of size n drawn from a normal population then,**

$$Z = \frac{\sqrt{n}\,(\overline{x} - \mu)}{\sigma} \sim N(0, 1)$$