Probability Distributions

1. Binomial Distribution
2. Bernoulli Distribution
3. Uniform Distribution
4. Normal Distribution
5. Poisson Distibution

1. Binomial Distribution :

- The binomial distribution used when there are only two possible outcomes. success or failure. for no. of trials (n)

- Probability of both outcomes same for all trials.

Properties :

1) Each trial is independant
2) Only two possible outcomes.
3) Probability of success and failure is same for all trials.

P - binomial probability
n - no. of trials
p - probability (Success)
q - probability (Failure)

$$P(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$$

$$P(x) = \frac{n!}{(n-x)!\,x!}\, p^{x}\, q^{n-x}$$

x - no. of times for specific outcome within n trials.

Ex. Hospital records show that of patients suffering from a specific disease, 75% die of it. What is the probability that of six randomly selected patients; four will recover?

→

$n = 6$, $x = 4$, $q = 0.75$, $p = 1 - 0.75 = 0.25$

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

$$= {}^6C_4 \, (0.25)^4 \, (0.75)^{6-4}$$

$$= \frac{6!}{(6-4)! \, 4!} (0.0039)(0.5625)$$

$$= \frac{720}{2 \times 24} \times (0.0021)$$

$$P(x) = \frac{1.5795}{48} = 0.03295$$

∴ probability that four will recover is 0.03295

Ex. A marksman finds that on the average, he hits the target '4' times out of '5'. If he fires '4' shots, what is probability of

a) more than '2' hits

b) at least 3 misses.

→

he hits target 4 times out of 5 times on an avg.

∴ $p = 4/5 = 0.8$

$q = 1 - p = 0.2$

x - be the no. of hits.

a> more than 2 Hits, means it can be 3 or 4 hits.

$\therefore P(x) = P(x_3) + P(x_4)$

$= {}^4C_3\, p^3 q^1 + {}^4C_4\, p^4 q^0$

$= \frac{4!}{(4-3)!\,3!}(0.8)^3(0.2)^1 + \frac{4!}{0!\,4!}(0.8)^4(0.2)^0$

$= 4(0.8)^3(0.2) + (0.8)^4$

$\therefore P(x) = 0.8192$

b> at least 3 misses. $\therefore$ only 1 hit or 0 hit

$P(x) = P(x1) + P(x0)$

$= {}^4C_1\,(0.8)^1(0.2)^3 + {}^4C_0\,(0.8)^0(0.2)^4$

$= 4(0.8)(0.2)^3 + (0.2)^4$

$\therefore P(x) = 0.0272$

② Bernoulli Distribution:

- Bernoulli distribution has only two outcomes, success (1) or Failure (0), and only single trial

- In tossing a coin, head denotes success/failure and tail denotes failure/success.

- In this case, probability of head = probability of tail = 0.5

- Bernoulli distribution is similar to Binomial distribution. Only difference is in Bernoulli $n = 1$ always. and x will take only two value 0 or 1

So, probability mass Function will be,

$$P(x) = {}^{n}C_{x}\, p^{x}\, q^{n-x}$$ -- pmf of Binomial distribution.

In Bernoulli, $n = 1$, $x = 0$ or 1

$$\therefore P(x) = {}^{1}C_{x}\, p^{x}\, q^{1-x} \quad \text{where } x \in (0,1) \quad \text{--- (1)}$$

x will be 0 or 1.

$\therefore$ When $x = 0$, ${}^{1}C_{x} = {}^{1}C_{0} = \frac{1!}{(1-0)!\,0!} = 1$

$\therefore$ When $x = 1$, ${}^{1}C_{1} = 1$

$\therefore$ ${}^{1}C_{x}$ will always 1.

from (1),

$$\therefore P(x) = 1 \cdot p^{x}\, q^{1-x}$$

$$\therefore \boxed{P(x) = p^{x}\, q^{1-x} \quad \text{where } x \in (0,1)}$$

probability mass Function for Bernoulli Distribution.

- Probabilities of success and failure need not be equal always.

example: Result of fight between me and Undertaker, here probability of my

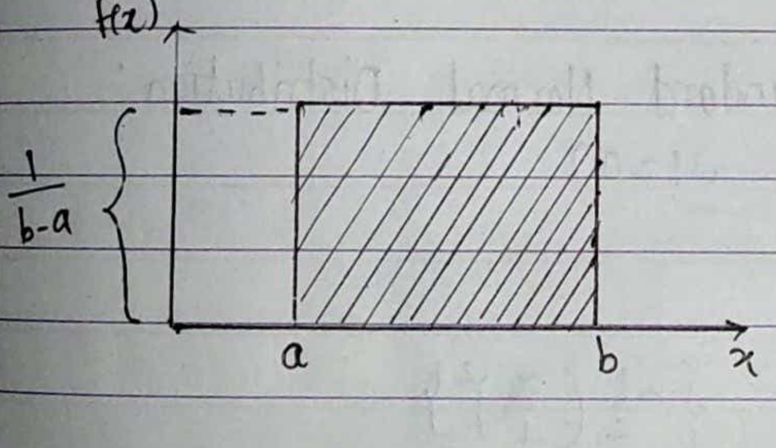
win/ success is 0.01 and Failure is 0.99
here, probability of success (p) is not same as probability of failure (q).

③ Uniform Distribution : (Rectangular Distribution)

- The probabilities of getting an outcome, when you roll a die is always equal and that is the basis of uniform distribution.

- A variable x is said to be uniformly distributed if density function is,

$$F(x) = \frac{1}{b-a} \quad \text{for } (-\infty < a \leq x \leq b < \infty)$$



- Area under the curve is always. 1.

- $(b-a)\left(1/(b-a)\right) = 1$

-

Normal / Gaussian Distribution & Emperical Formula:

Normal Distribution:

mean = median = mode

- In Normal Distribution, the Data is symmetrically distributed around it's mean
- Mean = Median = Mode
- No skewness
- No Outliers

- Many groups follow this type of pattern.
  - Heights of people.
  - Students marks
  - salaries.

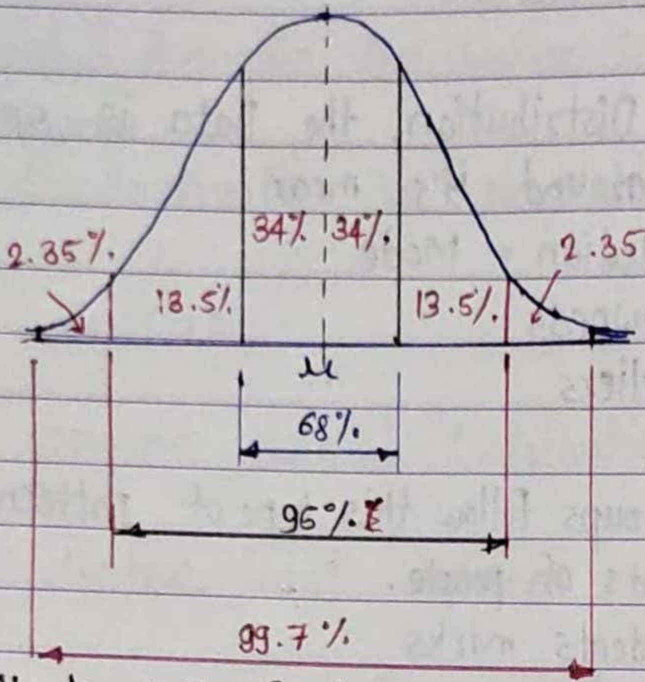
Why is Normal Distribution is important ?

1. Statistical hypothesis test assumes that the data follows a normal distribution.

2. Both linear and non-linear regression assumes that the residual follows the normal distribution.

3. Central Limit theorem states that, as the sample size increases the distribution of mean follows Normal Distribution. irrespective of distribution of original variable.

Emperical formula :-



+ About 68% of data fall within 1 std dev. of the mean.

+ About 95% of data fall within 2 std.dev from mean.

+ Almost all of the values - about 99.7% - fall within 3 std. dev. from mean.

+ These facts are the 68 95 99.7 Rule. also called as Emperical Rule, because the rule originally came from observations. (emperical means " based on observations".

When to Use :

+ · Data is normal, nearly normal or if you have a unimodal distribution (i.e. one with single peak) that is symmetric.

Problem: The weights of dogs at a particular pound average 70 lbs with standard deviation of 2.5 lbs. Assuming the weights follow Guassian Distribution.

1. What weight is 2 std. deviation below the mean?
2. What weight is 1 std. deviation above the mean?
3. The middle 68% of dogs weigh how much?

Mean(μ) = 70 lbs
std. deviation (σ) = 2.5 lbs.

① 2 std. dev below mean = $70 - 2 \times 2.5 = 65$ lbs.

So, dog is 2.5 std. dev. below the mean they weigh 65 lbs.

② 1 std. dev. above the mean = $70 + 1 \times 2.5 = 72.5$ lbs

So, if dog is 1 std. deviatn above the mean they weigh 72.5 lbs.

③ Midde. 68% of dogs weigh = $(\mu - \sigma, \mu + \sigma)$

$= (70 - 2.5, 70 + 2.5)$

$= 67.5, 72.5$

∴ Therefore, the dogs weigh between 67.5 lbs & 72.5 lbs

④ Normal Distribution :

characteristics :

① mean, median, mode of distribution coincide.

② curve of distribution is Bell- shaped and symmetrical about line $x = \mu$.

③ Total area under the curve is 1.

④

PDF of random variable x following a normal distribution is given by:

$$\boxed{f(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\}} \quad \text{for } -\infty < x < \infty} \quad —①$$

PDF for Standard Normal Distribution :
$(\sigma = 1, \ \mu = 0)$

from —①

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{\left\{-\frac{1}{2}\left(\frac{x}{1}\right)^2\right\}}$$

$$\therefore \boxed{f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}} \quad \text{for } -\infty < x < \infty}$$

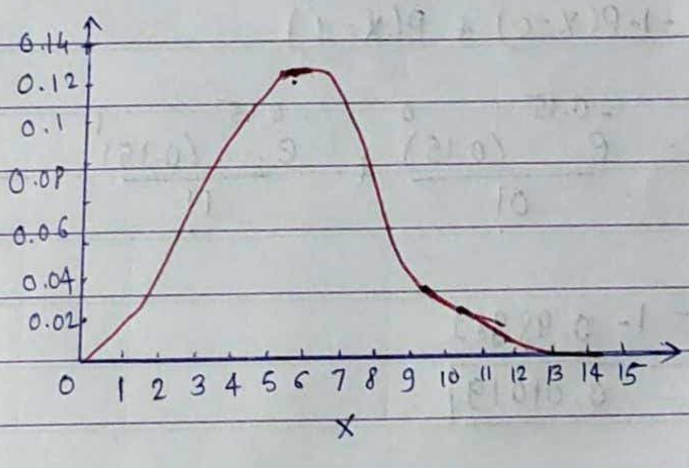
⑤ Poisson Distribution:

Poisson Distribution can be used to find probability of several events in a time period.

conditions:- 1) Events can occur independantly
2) An event can occur any number of times.

The PMF of X - Poisson Random Variable . Following poisson distribution.

$$\boxed{P(X=x) = \frac{e^{-\mu} \mu^{x}}{x!} \quad \mu\text{- avg. of success}}$$



Ex. ① A life insurance salesman sells on an average 3 life insurance policies per week. Use poissons law to calculate probability
a) He will sell 2 or more but not less than five.

→

$\mu = 3$

$$P(X=x) = \frac{e^{-3} 3^{x}}{x!}$$

a) $P(2 \leq x < 5) = P(X=2) + P(X=3) + P(X=4)$

$$= \frac{e^{-3} 3^2}{2!} + \frac{e^{-3} 3^3}{3!} + \frac{e^{-3} 3^4}{4!}$$

$$= 0.61611$$

② If electric power failure occur according to the poisson distribution with an average of 3 failure every twenty weeks. Calculate probability of more than 1 failure during particular week.

→

$$\mu = 3/20 = \cancel{0.00}\ 0.15$$

$$P(X>1) = ?$$

$$P(X>1) = 1 - P(X=0) + P(X=1)$$

$$= \frac{e^{-0.15} (0.15)^0}{0!} + \frac{e^{-0.15} (0.15)^1}{1!}$$

$$= 1 - 0.98981$$

$$\therefore \boxed{P(X>1) = 0.01019}$$