Repeated trials

If a coin is tossed thrice, the probability of getting one head and two tails can be combined as HTT, THT and TTH. The probability of each one of these being $\left(\frac{1}{2}\right)^3$, their total probability shall be $3\left(\frac{1}{2}\right)^3$.

Similarly, if trial is repeated 'n' times and if p is the probability of success and q is that of failure, then probability of 'r' success and '$n-r$' failure is given by $p^r q^{n-r}$. But these 'r' success and '$n-r$' failure can occur in any of these n_{Cr} ways in each of which the probability is same.

Thus the probability of r success is $n_{Cr}\, p^r q^{n-r}$.

The probability of at least r success in n trials= sum of probabilities $r, r+1, \ldots, n$ success = $n_{C_r} p^r q^{n-r} + n_{C_{r+1}} p^{r+1} q^{n-r-1} + \ldots + n_{Cn} p^n$.

Binomial distribution:

It is concerned with trials of repetitive nature in which only the occurrence or non-occurrence i.e., success or failure of particular event is of interest.

A random experiment with only two types of outcomes, success or failure of a particular event is called a Bernoulli trial. A random variable X that takes the value either 0 or 1 is known as Bernoulli variable.

x_i	0	1
$P(x_i)$	$1-p$	p

Mean $\mu = \sum x_i P(x_i) = p$.

Variance $\sigma^2 = \sum (x_i - \mu)^2 P(x_i)$

$$¿(0-p)^2(1-p)+(1-p)^2 p = pq.$$

Standard deviation $\sigma = \sqrt{pq}$.

$f(x) = P(X = x) = {}^nC_x\, p^x (1-p)^{n-x}$, where the random variable X denotes the number of successes in n trials and x = 0, 1, . . . , n.

The Random variable X given by $x=0,1,2,\ldots,n$ would have probability as

Number of success	Probability $P(x)$
0	q^n
1	$n_{C_1} p q^{n-1}$
2	$n_{C_2} p^2 q^{n-2}$
$\vdots$	$\vdots$
n	$n_{C_n} p_n q^{n-n}$

This table is in the form of frequency distribution. Where in it should be observed that the values of $P(x)$ for different values of $x=0,1,2,\ldots$ are the various terms expansion of $(p+q)^n$ and accordingly the frequency distribution is called as the Binomial distribution.

Mean and variance of Binomial distribution (B.D):

For various values of $x=0,1,2,\ldots,n$, we get the corresponding values of $P(x)=n_{C_x} p^x q^{n-x}$.

$\therefore$ Sum of the probability $=q^n+n_{C_1} p q^{n-1}+\ldots+p^n=(q+p)^n=1$.

$$\therefore \sum P(x_i)=1.$$

Mean $\mu=\sum x_i P(x_i)=\sum_{x=0}^{n} x\, n_{C_x} p^x q^{n-x}$

$$\text{¿}\sum_{x=0}^{n} x \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

$$\text{¿}\sum_{x=1}^{n} \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x}$$

$$\text{¿}n\sum_{x=1}^{n} \frac{(n-1)!}{(x-1)!\text{¿¿}}\text{¿}$$

$$\text{¿}np(q+p)^{n-1}=np.$$

Variance $\sigma^2=\sum (x_i-\mu)^2 P(x_i)$

$$\text{¿}\sum_{x=0}^{n} (x-\mu)^2 P(x)$$

$$\text{¿}\sum_{x=0}^{n} (x^2-2x\mu+\mu^2) P(x)$$

$$\text{¿}\sum_{x=0}^{n} x^2 P(x)-2\sum_{x=0}^{n} x\mu P(x)+\sum_{x=0}^{n} \mu^2 P(x)$$

$$¿\sum_{x=0}^{n} x^2 n_{C_x} p^x q^{n-x} - 2\mu^2 + \mu^2$$

$$¿\sum_{x=0}^{n} [x(x-1)+x] n_{C_x} p^x q^{n-x} - \mu^2$$

$$¿\sum_{x=0}^{n} x(x-1)\frac{n!}{(n-x)!\,x!} p^x q^{n-x} + \mu - \mu^2$$

$$¿n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{[n-2-(x-2)]!\,(x-2)!} p^{x-2} q^{n-2-(x-2)} + \mu - \mu^2$$

$$¿n(n-1)p^2 (q+p)^{n-2} + \mu - \mu^2$$

$$¿n(n-1)p^2 + \mu - \mu^2$$

$$¿n^2 p^2 - n p^2 + np - n^2 p^2 = np(1-p) = npq.$$

Standard deviation $\sigma = \sqrt{npq}$.

<u>Problems on Binomial distribution:</u>

1. The probability that a pen manufactured by a company will be defective is $\frac{1}{10}$. If 12 such pens are manufactured, find the probability that a) exactly 2 will be defective b) at least 2 will be defective c) none will be defective.

<u>Solution:</u> The probability of a defective pen=0.1

$\therefore$ Probability of a non- defective pen $q = 1 - 0.1 = 0.9$

a) Exactly 2 will be defective $¿12_{C2} p^2 q^{n-2}$

$$¿12_{C_2} p^2 q^{12-2}$$

$$¿12_{C_2} p^2 q^{10} = 12_{C_2} (0.1)^2 (0.9)^{10} = 0.2301.$$

b) The probability that at least 2 will be defective=1-pobability that less than 2 are defective

=1-[$12_{C0} p^0 q^{12} + 12_{C_1} p^1 q^{11}¿$

=1-$[(0.9)^{12} + 12(0.1)(0.9)^{11}]$

=$0.3412.$

c) The probability that none will be defective =$12_{C_0} p^0 q^{12} = (0.9)^{12} = 0.2824.$

2. 4 coins are tossed 100 times and the following results were obtained. Fit a binomial distribution for the data and calculate the theoretical frequency (corresponding probability frequencies).

No. of heads	0	1	2	3	4
Frequency	5	29	36	25	5

Solution: Let X be random variable denoting no. of heads. If p is the probability of head, then q represents the probability of tail.

From the given frequency distribution, mean $\mu=\frac{\sum f_i x_i}{\sum f_i}=\frac{196}{100}=1.96.$

But $\mu=np=1.96$

$$p=\frac{1.96}{n}=\frac{1.96}{4}=0.49$$

$$\therefore q=1-p=1-0.49=0.51.$$

$\therefore$ The binomial distribution fit for this data is $100(0.49+0.51)^4$.

Hence, the corresponding probability frequencies are obtained as $100\,4_{C_0}(0.49)^0(0.51)^4=7$

$$100\,4_{C_1}(0.49)^1(0.51)^3=26$$

$$100\,4_{C_2}(0.49)^2(0.51)^2=37$$

$$100\,4_{C_3}(0.49)^3(0.51)^1=24$$

$$100\,4_{C_4}(0.49)^4(0.51)^0=6$$

Thus the theoretical frequencies are 7, 26, 37, 24, 6.

3. Out of 800 families with 5 children each, how many would have expect to have a) 3 boys b) 5 girls c) either 2 or 3 boys? Assume equal probability for boys and girls.

Solution: Let $p=¿$ probability of a boy$=\frac{1}{2}$ and $q=¿$probability of girl$¿\frac{1}{2}$.

a) Probability of a family having 3 boys$=5_{C_3}(0.5)^3(0.5)^2=0.3125.$
   $\therefore$Expected no. of families=800 $\times 0.3125$=250.

b) Probability of a family having 5 girls$=5_{C0}(0.5)^0(0.5)^5=0.03125.$
   $\therefore$ Expected no. of families$=800\times 0.03125$=25.

c) Probability of a family having either 2 or 3 boys= $5_{C_2}(0.5)^2(0.5)^3+5_{C_3}(0.5)^3(0.5)^2$
   $¿0.625.$

$\therefore$ Expected no. of families = 800 $\times 0.625$=500.

4. The number of telephone lines busy at a particular time is a binomial variable with probability 0.1 that a line is busy. If 10 lines are selected at random, what is the probability that i) no line is busy ii) at least one line is busy iii) at most 2 lines are busy.

Solution: By data we have $p=0.1, q=1-p=0.9$. Also n=10.

$\therefore$ Out of 10 lines, the probability that r lines are busy is given by $10_{C_r}(0.1)^r(0.9)^{10-r}$.

i) No. line is busy $= 10_{C_0}(0.1)^0(0.9)^{10} = 0.3487$.

ii) Probability that at least one line is busy =1- probability that no. line is busy.

$$¿1-0.3487=0.6513.$$

iii) At most 2 lines are busy = $10_{C_0}(0.1)^0(0.9)^{10}+10_{C_1}(0.1)(0.9)^9+10_{C_2}(0.1)^2(0.9)^8$ =0.929830.

5. In a bombing action, there is 50% chance that any bomb will strike the target. Two direct hits are needed to destroy the target completely. How many bombs are required to be dropped to give 99% chance or better of destroying the target?

<u>Solution:</u> From data, $p=0.5$ and $q=0.5$.

Where p is the probability that bomb strike the target.

$P(x)=n_{C_x}(0.5)^x(0.5)^{n-x}=n_{C_x}(0.5)^n$ represents the probability that x bombs out of n strikes the target.

We need to find minimum value of n such that $2\le k\le n$ bombs destroy the target completely.

i.e, $P(2\le k\le n)\ge 0.99$

$$\sum n_{C_x}(0.5)^n \ge 0.99$$

$$n_{C_0}(0.5)^n+n_{C_1}(0.5)^n\le 1-0.99$$

$$\frac{1+n}{2^n}\le 0.01$$

i.e., $100(1+n)\le 2^n$

till $n=10$, we get $100(1+n)>2^n$

i.e., for $n=10, 100(1+10)=1100>2^{10}$

for n=11, we get $100(1+n)=1200<2^{11}$

$$\therefore n=11$$

Minimum 11 bombs are required.

6. An airline knows that 5% of the people making reservation on a certain flight will not turn up. Consequently their policy is to sell 52 tickets for a flight that can hold only 50 passengers. What is the probability that there will be a seat available for every passengers who turns up.

<u>Solution</u>: Probability that passenger will not turn up=0.05.

i.e., $¿0.05$,$q=1-0.05=0.95$.

x=no. of passenger will not turn up, $n=52$.

$$P(x)=52_{C_x}(0.05)^x(0.95)^{52-x}$$

Required probability $¿P(x\ge 50)=1-P(x>50)$

$$¿1-\{P(x=51)+P(x=52)\}$$

$$¿1-\{52_{C_{51}}(0.05)^{51}(0.95)^1+52_{C_{52}}(0.05)^{52}(0.95)^0\}.$$

7. If 20% of the bolts produced by a machine are defective, determine the probability that out of 4 bolts chosen at random, (a) 1 (b) 0 (c) less than 2, bolts will be defective.

Solution: The probability of a defective bolt is p = 0.2, of a non - defective bolt is $q = 1 - p = 0.8$.

Let the random variable X be the number of defective bolts. Then

(a) $P(X = 1) = {}^4C_1(0.2)^1(0.8)^3 = 0.4096$

(b) $P(X = 0) = {}^4C_0(0.2)^0(0.8)^4 = 0.4096$

(c) $P(X < 2) = P(X = 0) + P(X = 1) = 0.4096 + 0.4096 = 0.8192$.

8. Let the probability that the birth weight (in grams) of babies in America is less than 2547 grams be 0.1. If X equals the number of babies that weigh less than 2547 grams at birth among 20 of these babies selected at random, then what is $P(X \le 3)$?

Solution: If a baby weighs less than 2547, then it is a success; otherwise it is a failure. Thus X is a binomial random variable with probability of success p and n = 20. We are given that p = 0.1. Hence $P(X \le 3) = \sum_{k=0}^{3} {}^{20}C_k(0.1)^k(0.9)^{20-k} = 0.867$

9. A gambler plays roulette at Monte Carlo and continues gambling, wagering the same amount each time on "Red", until he wins for the first time. If the probability of "Red" is $\frac{18}{38}$ and the gambler has only enough money for 5 trials, then (a) what is the probability that he will win before he exhausts his funds; (b) what is the probability that he wins on the second trial?

Solution: $p = P(\text{Red}) = \frac{18}{38}$.

(a) Hence the probability that he will win before he exhausts his funds is given by

$P(X \le 5) = 1 - P(X \ge 6) = 1 - (1-p)^5 = 1 - \left(1-\frac{18}{38}\right)^5 = 0.956$.

(b) Similarly, the probability that he wins on the second trial is given by

$P(X = 2) = (1 - p)\, p = (1 - \frac{18}{38})\frac{18}{38} = 0.2493$.

10. A continuous random variable has pdf $f(x) = \begin{cases} \frac{x}{2}, & 0 < x < 2 \\ 0, & elsewhere \end{cases}$.

(a) If two independent determination of X are made, then what is the probability that both these determinations will be greater than 1.

(b) If three independent determinations are made, what is the probability that atleast 2 of these are greater than 1.

Solution:

$$f(x) = \begin{cases} \frac{x}{2}, & 0 < x < 2 \\ 0, & elsewhere \end{cases}$$

Z : Number of independent determination greater than 1.

$$p = P(X > 1) = \int_1^2 \frac{x}{2} dx = 0.75$$

(a) $P(Z = z) = {}^nC_z\, p^z (1 - p)^{n-z}$

$P(Z = 2) = \frac{9}{16}$.

(b) $n = 3,\ P(Z \geq 2) = P(Z = 2) + P(Z = 3)$

$= {}^3C_2 (0.75)^2 (0.25)^{3-2} + {}^3C_3 (0.75)^3 (0.25)^0$

$= \frac{27}{32}$.

The Poisson Distribution

It is a distribution related to the probabilities of events which are extremely rare, but which have large number of independent opportunities for occurrence.

Eg: Number of persons born blind per year in a large city, how many hits will websites get in a particular minute.

Poisson distribution is regarded as the limiting form of the binomial distribution when $n$is very large i.e., $n \to \infty$ and p, the probability of success is very small i.e., $p=0$. So that the mean nptends to a fixed finite constant.

$\therefore$ Probability of x success in Poisson distribution is given by $P(x)=\frac{\lambda^x e^{-\lambda}}{x!}$.

Mean and standard deviation of Poisson distribution:

Mean $\mu=\sum x_i P(x_i)$

¿ $\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} x$

¿ $e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{(x-1)!}$

¿ $e^{-\lambda} \lambda \sum_{x=0}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$

¿ $e^{-\lambda} \lambda \left[1+\frac{\lambda}{1!}+\frac{\lambda^2}{2!}+\ldots\right]=e^{-\lambda} \lambda e^{\lambda}=\lambda.$

Variance $\sigma^2=\sum (x-\mu)^2 P(x)$

¿ $\sum (x^2-2\mu x+\mu^2) P(x)$

$$\text{¿}\sum_{x=0}^{\infty} x^2 P(x) - 2\mu \sum_{x=0}^{\infty} x\,P(x) + \mu^2$$

$$\text{¿}\sum_{x=0}^{\infty} [x(x-1)+x] P(x) - \lambda^2$$

$$\text{¿}\sum_{x=0}^{\infty} x(x-1)P(x) + \sum xP(x) - \lambda^2$$

$$\text{¿}\sum_{x=0}^{\infty} \frac{x(x-1)\lambda^x e^{-\lambda}}{x!} + \lambda - \lambda^2$$

$$\text{¿}\sum_{x=2}^{\infty} \frac{\lambda^x e^{-\lambda}}{(x-2)!} + \lambda - \lambda^2$$

$$\sigma^2 = \lambda^2 \sum_{x=2}^{\infty} \frac{\lambda^{x-2}}{(x-2)!} e^{-\lambda} + \lambda - \lambda^2$$

$$\text{¿}\lambda^2 e^{-\lambda} e^{\lambda} + \lambda - \lambda^2 = \lambda$$

Standard deviation $\sigma = \sqrt{\lambda}$.

Note: Mean and Variance are equal in Poisson distribution.

Theorem : Prove that Poisson distribution is the limiting case of binomial distribution.
Proof: If X is binomial distributed, then

$$f(x) = P(X = x) = {}^{n}C_x\, p^x (1-p)^{n-x} \quad \text{……………………..(1)}$$

where E(x) = np.

Let λ = np so that $p = \dfrac{\lambda}{n}$.

$$P(X = x) = {}^{n}C_x \left(\frac{\lambda}{n}\right)^x \left(1-\frac{\lambda}{n}\right)^{n-x} = \frac{n(n-1)(n-2)\ldots(n-k+1)}{x!\,n^x} \lambda^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

$$= \frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x-1}{n}\right)}{x!} \lambda^x \left(1-\frac{\lambda}{n}\right)^{n-x}$$

As n $\to \infty$

$$\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{x-1}{n}\right) \to 1$$

While $\left(1-\dfrac{\lambda}{n}\right)^{n-x} = \left(1-\dfrac{\lambda}{n}\right)^{n}\left(1-\dfrac{\lambda}{n}\right)^{-x} = e^{-\lambda}$

using the well-known result from calculus that $\lim_{n\to\infty}\left(1+\dfrac{u}{n}\right)^n = e^u$

It follows that when n $\to \infty$, $P(X = x) = \dfrac{\lambda^x e^{-\lambda}}{x!}$
which is the Poisson distribution.

Problems on Poisson distribution:

1. X is a Poisson variable and it is found that the probability that $X = 2$ is two third of probability that $X = 1$. Find the probability that $X = 0$ and $X = 3$. What is the probability that X exceeds 3?

   Solution: $P(x) = \dfrac{\lambda^x e^{-\lambda}}{x!}, x = 0, 1, 2, \ldots$

$$\therefore P(X=2) = \frac{2}{3} P\{X=1\}$$

   i.e., $\dfrac{3\lambda^2 e^{-\lambda}}{2!} = \dfrac{2\lambda e^{-\lambda}}{1!}$

   $3\lambda^2 - 4\lambda = 0$

   $\lambda(3\lambda - 4) = 0$

   $\lambda = 0$ or $\lambda = \dfrac{4}{3}$

   If $\lambda = 0$, then $P(x) = 0$.

$$\therefore \lambda \neq 0 \wedge \lambda = \frac{4}{3} = ¿ P(x) = \frac{\left(\frac{4}{3}\right)^x e^{\frac{-4}{3}}}{x!}$$

$$P(X=0) = \left(\frac{4}{3}\right)^0 e^{\frac{-4}{3}} = e^{\frac{-4}{3}} = 0.26359$$

$$P(X=3) = \left(\frac{4}{3}\right)^0 e^{\frac{-4}{3}} = 0.10413714$$

$$P(X>3) = 1 - P(X \leq 3) = 1 - ¿$$

$$¿ 1 - \left[\frac{\left(\frac{4}{3}\right)^0 e^{\frac{-4}{3}}}{0!} + \frac{\left(\frac{4}{3}\right)^1 e^{\frac{-4}{3}}}{1!} + \frac{\left(\frac{4}{3}\right)^2 e^{\frac{-4}{3}}}{2!} + \frac{\left(\frac{4}{3}\right)^3 e^{\frac{-4}{3}}}{3!}\right]$$

$$¿ 1 - 0.953505 = 0.046494.$$

2. A manufacturer knows that the condensers that he makes contain on the average 1% defective. He packs them in box of 100. What is the probability that a box picked at random will contain 3 or more faulty condensers?

   Solution: Let p=probability of a defective=0.01

$$n = 100, mean\, \lambda = np = 100 \times 0.01 = 1.$$

$\therefore$ Probability that a box will contain 3 or more faulty condensers

¿ $1-P\{X<3\}$

¿ $1-\left[\frac{\lambda^0 e^{-\lambda}}{0!}+\frac{\lambda^1 e^{-\lambda}}{1!}+\frac{\lambda^2 e^{-\lambda}}{2!}\right]$

¿ $1\left[e^{-1}+e^{-1}+\frac{e^{-1}}{2!}\right]$

$= 1-\frac{5e^{-1}}{2}=0.0803013$

3. The number of accidents in a year to taxi drivers in a city follows a Poisson distribution with mean 3 out of 1000 taxi drivers. Find approximately the number of the drivers with i) no accident in a year ii) more than 3 accidents in a year.

Solution: $P(x)=\frac{3^x e^{-3}}{x!}$ gives the probability of accidents to taxi drivers.

Approximate number of drivers out of 1000 with x accidents¿ $1000\,P(x)$.

i) No. of drivers with no accidents ¿ $1000\times P(0)$

¿ $1000\frac{3^0 e^{-3}}{0!}=49.7870\cong 50.$

ii) No. of drivers with more than 3 accidents in a year$=1000-$¿no. of drivers with less than or equal to 3 accidents in a year.

¿ $1000-1000[P(0)+P(1)+P(2)+P(3)]$

¿ $1000-1000\left[\frac{3^0 e^{-3}}{0!}+\frac{3e^{-3}}{1!}+\frac{3^2 e^{-3}}{2!}+\frac{3^3 2^{-3}}{3!}\right]$

¿ $353.$

4. A bag contains 1 red and 7 white marbles. A marble is drawn from the bag and its color is observed. Then the marble is put back into the bag and the contents are thoroughly mixed. Find the probability that in 8 such drawings, a red ball is selected exactly 3 times?

Solution: Let X be a random variable denoting the number of times red ball is selected in 8 drawings.

By data, mean $\lambda=np=8\times\frac{1}{8}=1.$

$P(x)=\frac{1^x e^{-1}}{x!},x=0,1,2,3,\ldots$

$\therefore$Probability of selecting a red ball exactly 3 times ¿ $P(X=3)=\frac{1^3 e^{-1}}{3!}=0.06131.$

5. Suppose that 0.01% of the population of the city with population 10,000 suffers from certain disease. Find the probability that there is at least two persons, who suffer from the disease. If there are 10 such cities in state, what is the probability that at least one city will have at least one person who suffer from the disease.

Solution: $n=10{,}000\,,\,p=\dfrac{0.01}{100}=0.0001\,,\,\lambda=np=10{,}000\times 0.0001=1$

$$\therefore P(X\geq 2)=1-P(X<2)$$

$$¿\,1-\left[P(X=0)+P(X=1)\right]$$

$$¿\,1-\left(\frac{e^{-1}\lambda^{0}}{0!}+\frac{e^{-1}\lambda^{1}}{1!}\right)=1-\frac{2}{e}=0.2642.$$

No. of cities, $n=10$.

For each city P(Person suffer from disease)$=P(X\geq 1)=1-P(X<1)$

$¿\,1-P(X=0)=0.6321.$

Probability of at least one suffer from each city = 0.6321.

$$p=0.6321,\,n=10\,,\,q=1-p=1-0.6321=0.3679.$$

Using binomial distribution,

Probability that at least one city will have at least one person who suffer from disease $¿\,P(X\geq 1)=1-P(X=0)=1-10_{C_0}p^{0}q^{10}=0.999.$

6. An insurance company has discovered that only about 0.1% of the population is involved in a certain type of accident each year. If its 10,000 policy holders were randomly selected from the population, what is the probability that not more than 5 of its clients are involved in such an accident next year?

Solution: Let X be a number of clients involved in accidents.

$$n=10{,}000\,,\,p=0{,}0001\,,\,\lambda=np=10$$

$$P(X=x)=\frac{e^{-\lambda}\lambda^{x}}{x!}\,,\,\lambda=10$$

$$P(X\leq 5)=e^{-10}\,¿$$

$¿\,0.0671.$

7. The number of traffic accidents per week in a small city has a Poisson distribution with mean equal to 3. What is the probability of exactly 2 accidents occur in 2 weeks?

Solution: The mean traffic accident is 3. Thus, the mean accidents in two weeks are λ = (3) (2) = 6.

Since $f(x)=\dfrac{\lambda^{x}e^{-\lambda}}{x!}$

$$f(2)=\frac{6^{2}e^{6}}{2!}=18e^{6}.$$

8. The distributor of been seeds determines from extensive test that 5% of large batch of seeds won't germinates. He sells seeds in packets of 50 and guarantees 90% germination. Determine the probability that particular packet violate the guaranty.

Solution: p = 0.05, n = 50, λ = 50 × 0.05 = 2.5.

X = Number of seeds that do not germinate

P(X > 10% of 50) = P(X > 5) = 1 – P(X ≤ 5) = $1 - \sum_{0}^{5} \frac{2.5^x e^{-2.5}}{x!}$ = 0.042

The Normal Distribution

One of the most important examples of a continuous probability distribution is the normal distribution, sometimes called the Gaussian distribution. The density function for this distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty \quad -\infty<\mu<\infty, \sigma>0.$$

,

Such a variable X following the normal law is expressed as $X \ \ N(\mu,\sigma^2)$.

If $z = \frac{x-\mu}{\sigma}$, then $f(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}, \ -\infty < z < \infty$. This is often referred to as the standard normal density function.

This $Z = \frac{X-\mu}{\sigma}$ is the standard normal variate with $E(Z)=0$ and $var(Z)=1$and we write $Z \ \ N(0,1)$.

The probability of X lying between x_1 and x_2 is given by the area under normal curve from x_1 to x_2.

$$P(x_1 \le X \le x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

¿ $\frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{\frac{-z^2}{2}} dz$ when $z = \frac{x-\mu}{\sigma}$

¿ $\frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^{z_2} e^{\frac{-z^2}{2}} dz - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_1} e^{\frac{-z^2}{2}} dz \right]$

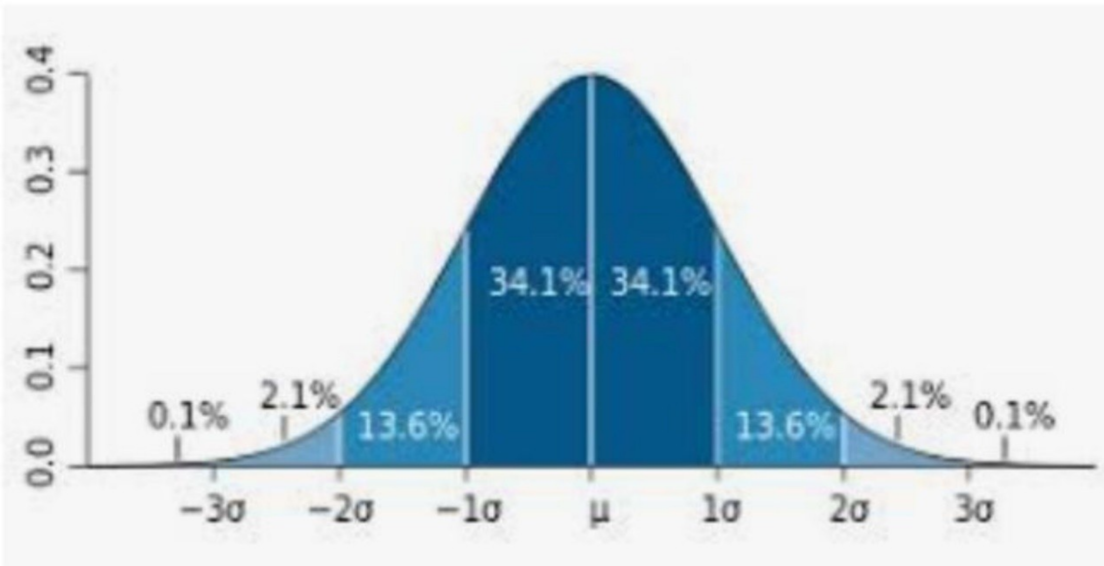
¿ $\phi(z_2) - \phi(z_1)$

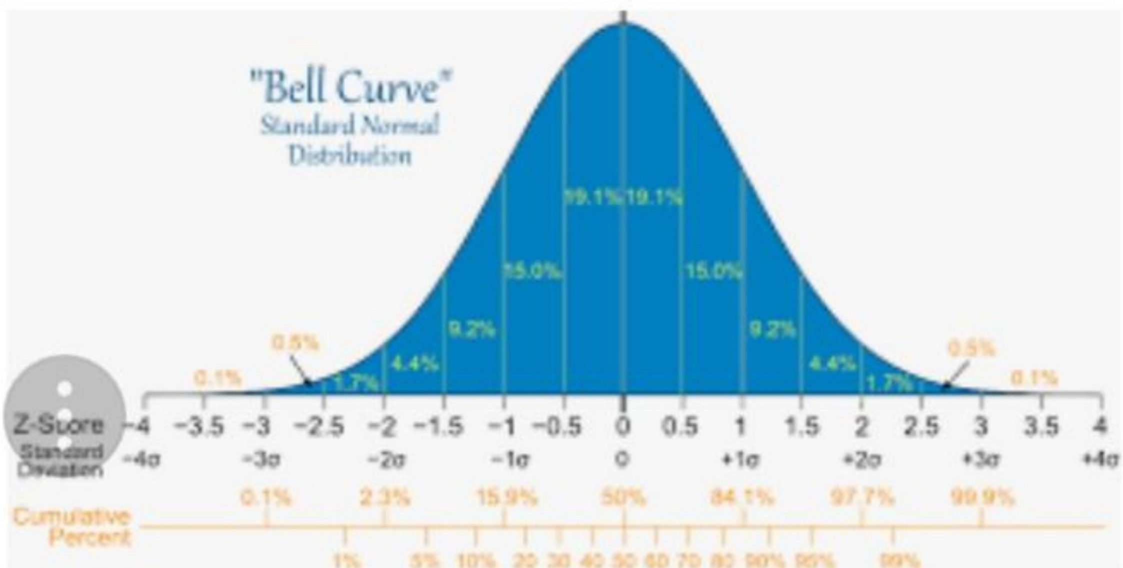
Where $\phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{\frac{-\mu^2}{2}} dx$for various values of $z$gives the corresponding standard normal function.

The cumulative function for z is given by

$$\phi(z)=P(Z\le z)=\int_{-\infty}^{z}\phi(\mu)\,dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{\frac{-\mu^2}{2}}\,d\mu.$$

A graph of the density function f(z) sometimes called the standard normal curve, is shown in Figure.





In this graph $P(-1\le z\le 1)=0.6827$, $P(-2\le z\le 2)=0.9545$ and $P(-3\le z\le 3)=0.9973$.
Two important result on $\phi(z)$ are given by

1) $\phi(-z) = P\{Z \le -z\} = P\{Z \ge z\}$ $(by\ symmetry)$

¿ $1 - P\{Z \le z\}$

¿ $1 - \phi(z)$

2) $P\{a \le X \le b\} = P\left\{\frac{a-\mu}{\sigma} \le z \le \frac{b-\mu}{\sigma}\right\} = P\left\{z \le \frac{b-\mu}{\sigma}\right\} \quad P\left\{z - \frac{a-\mu}{\sigma}\right\}$

¿ $\phi\left(\frac{b-\mu}{\sigma}\right) - \phi\left(\frac{a-\mu}{\sigma}\right)$

Expectation of normal distribution $E(X) = \mu$.

Variance of normal distribution $var(X) = \sigma^2$.

<u>Note</u>: 1) If $X \quad N(\mu, \sigma^2)$ and $Z \quad N(0,1)$, then X is called a normal variation with parameters μ and σ^2 while Z is the standard normal variate with the parameters 0 and 1.

1) To find $P\{a \le X \le b\}$,we change the variable X to the standard normal variable Z and hence find the area under the standard normal curve $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$.

i.e, $\phi(z) = P\{Z \le z\} = \int_{-\infty}^{z} \frac{e^{\frac{-x^2}{2}}}{\sqrt{2\pi}} dx$.

Examples 1:

$P(z \ge 1.66) = 1 - P(z < 1.66) = 1 - 0.9515 = 0.0485.$ (Using standard distribution table).

Examples 2:

$P(-1.96 \le z \le 1.96) = F(1.96) - F(-1.96) = F(1.96) - [1 - F(1.96)] = 0.9750 - (1 - 0.9750) = 0.95$ (Using standard distribution table).

<u>Problems on Normal distribution:</u>

1. In a normal distribution, 31% of the items are under 45 and 8% are over 64. Find the mean and variance of distribution.

<u>Solution:</u> Given $P\{X < 45\} = 31\%$ and $P\{X > 64\} = 8\%$

i.e., $P\left\{\frac{X-\mu}{\sigma} < \frac{45-\mu}{\sigma}\right\} = 0.31$

$$P\left\{Z<\frac{45-\mu}{\sigma}\right\}=0.31$$

$$\phi\left(\frac{45-\mu}{\sigma}\right)=0.31$$

$$1-P\left(\frac{-45+\mu}{\sigma}\right)=0.31$$

$$\text{i.e., } \phi\left(\frac{\mu-45}{\sigma}\right)=0.69.$$

$$\frac{\mu-45}{\sigma}=0.5$$

$\mu-0.5\sigma=45$--------------------(1)

$P\{X>64\}=8\%$

$$\text{i.e., } P>\left\{\frac{X-\mu}{\sigma}>\frac{64-\mu}{\sigma}\right\}$$

$$P\left\{Z>\frac{64-\mu}{\sigma}\right\}=0.08$$

$$\text{i.e., } 1-\phi\left(\frac{64-\mu}{\sigma}\right)=0.08$$

$$\phi\left(\frac{64-\mu}{\sigma}\right)=0.92$$

$$\frac{64-\mu}{\sigma}=1.4$$

$\mu+1.4\sigma=64$-------------------(2)

From (1) and (2), we get $\mu=50$ and $\sigma=10$.

2. Suppose X has the distribution $N(\mu,\sigma^2)$. Determine c as function of μ and σ such that $P\{X\le c\}=2P\{X>C\}$.

<u>Solution:</u> $P\{X\le c\}=2P\{X>C\}$

i.e., $P\{X\le c\}=2(1-P\{X\le c\})$

$P\{X\le c\}=2-2P\{X\le c\}$

$3P\{X\le c\}=2$

$$P\{X\le c\}=\frac{2}{3}=0.666$$

$$\text{i.e., } P\left\{\frac{X-\mu}{\sigma} \le \frac{c-\mu}{\sigma}\right\} = 0.666$$

$$P\left\{Z \le \frac{c-\mu}{\sigma}\right\} = 0.666$$

$$\phi\left(\frac{c-\mu}{\sigma}\right) = 0.666$$

$$\frac{c-\mu}{\sigma} = 0.43$$

$$c = \mu + 0.43.$$

3. Suppose that life length of two electronic device say D_1 and D_2 have distributions $N(40,36)$ and $N(45,9)$ respectively. If the electronic device is to be used for 45hours period, which device is to be preferred? If it is to be used for 48 hours period, which device is to be preferred?

Solution: Let X represents the life lengths of the electronic device.

For device D_1,

(i) For a period of 45 hours,

$$P\{X \ge 45\} = 1 - P\{X < 45\} = 1 - P\{Z - 0.8333\} = 1 - \phi(0.8333)$$

¿ $1 - 0.7967 = 0.2053$.

(ii) For a period of 48 hours,

$$P\{X \ge 48\} = 1 - P\{X < 48\} = 1 - P\{Z < 1.333\} = 1 - \phi(1.333)$$

¿ $1 - 0.9082 = 0.0918.$

For device D_2,

(i) For a period of 45 hours,

$$P\{X \ge 45\} = 1 - P\{X < 45\} = 1 - P\{Z < 0\} = 1 - \phi(0) = 1 - 0.5$$

¿ $0.5.$

(ii) For a period of 48 hours,

$$P\{X \ge 48\} = 1 - P\{X < 48\} = 1 - P\{Z < 1\} = 1 - \phi(1)$$

¿ $1 - 0.8413 = 0.1587.$

For a period of 45 hours, device D_2 is to be preferred whereas for a period of 48 hours, device D_2 is preferred.

4. For a normally distributed variate with mean 1 and standard deviation 3. Find the probabilities that i) $3.43 \le X \le 6.19$ ii) $-1.43 \le X \le 6.19$

Solution:

(i) $P\{3.43 \le X \le 6.19\} = P\left\{\frac{3.43-1}{3} \le Z \le \frac{6.19-1}{3}\right\}$

¿ $P\{0.81 \le Z \le 1.73\}$

¿ $\phi(1.73) - \phi(0.81)$

¿ $0.9582 - 0.7910 = 0.1672.$

(ii) $P\{-1.43 \le X \le 6.19\} = P\left\{\frac{-1.43-1}{3} \le Z \le \frac{6.19-1}{3}\right\}$

¿ $P\{-0.81 \le Z \le 1.73\}$

¿ $\phi(1.73) - \phi(-0.81)$

¿ $\phi(1.73) - (1 - \phi(0.81))$

¿ $0.9582 + 0.7910 - 1 = 0.7492$.

5. A fair coin is tossed 500 times. Find the probabilities that the number of heads will not differ from 250 by (a) more than 10 (b) more than 30.

<u>**Solution**</u>: Let X represents the number of heads in 500 tosses.

$\therefore p =$ ¿probability of head turning up¿ $\frac{1}{2}$ and $n = 500$.

Mean¿ $np = 500 \times \frac{1}{2} = 250$ and $\sigma = \sqrt{npq} = \sqrt{500 \times \frac{1}{2} \times \frac{1}{2}} = 11.18033$.

Since n is very large and p is not small. We can use normal approximation to binomial.

a) Probability that number of heads will not differ from 250 by more than 1

¿ $P\{240 \le X \le 260\} = P\left\{\frac{240-250}{11.18033} \le Z \le \frac{260-250}{11.18033}\right\}$

¿ $P\{-0.894427191 \le Z \le 0.894427191\}$

¿ $2\phi(0.894427191) - 1$

¿ $2(0.8133) - 1 = 0.6266$.

b) Probability that number of heads will not differ from 250 by more than 30=

$P\{220 \le X \le 280\} = P\left\{\frac{220-250}{11.180339} \le Z \le \frac{280-250}{11.180339}\right\}$

¿ $P\{-2.68328172 \le Z \le 2.68328172\}$

¿ $2\phi(2.68328172) - 1$

¿ $2 \times 0.9963 - 1 = 0.9920$.

6. The weekly wages of workers in a certain factory was found to be normally distributed with mean $Rs.500$ and standard deviation $Rs.50$. There are 228 persons getting at least $Rs.600$. Find the number of workers in the factory.

<u>**Solution**</u>: For $P\{X \le 600\} = P\left\{Z \le \frac{600-500}{50}\right\} = P\{Z \le 2\} = 1 - P\{Z \le 2\}$

$P\{X \ge 600\} = 1 - \phi(Z) = 0.0228$.

If n is total number of workers then $nP\{X \ge 600\} = 228$.

i.e., $n \times 0.0228 = 228$.

$$n = \frac{228}{0.0228} = 10{,}000.$$

There are 10,000 workers in the factory.

7. If X is normally distributed with N(1, 4). Find P(|X| > 4).

Solution: If $z = \dfrac{x - \mu}{\sigma}$,
then,

$$P(|X|>4) = 1 - P(|X| \le 4) = 1 - P(-4 \le X \le 4)$$
$$= 1 - P\left(\frac{-4-1}{2} \le Z \le \frac{4-1}{2}\right) = 1 - \left[F(1.5) - F(-2.5)\right]$$
$$= 2 - 0.9332 - 0.9938 = 0.073$$

8. In a normal distribution 31% of items rare under 45 and 48% are over 64. Find mean and standard deviation.

Solution:

X = 45, area = 31%

$$\Rightarrow Z = -0.5$$
$$Z = \frac{X - \mu}{\sigma} \Rightarrow -0.5 = \frac{45 - \mu}{\sigma}$$
$$\Rightarrow -0.5\sigma = 45 - \mu \quad \dots. \quad (1)$$

X = 64 , area = 48%

$$\Rightarrow Z = 0.05$$
$$Z = \frac{X - \mu}{\sigma} \Rightarrow 0.05 = \frac{64 - \mu}{\sigma}$$
$$\Rightarrow 0.05\sigma = 64 - \mu \quad \dots. \quad (2)$$

Solving (1) and (2),
$\mu = 62.27,\ \sigma = 34.54.$

9. If mean marks is 60 and standard deviation is 10, 70% failed in examination. What is the grace marks given to obtain 70% pass the examination?

Solution:

$\mu = 60,\ \sigma = 10.$

$P(Z \le a) = 0.7 \Rightarrow\ Z = 0.525$

$Z = \dfrac{X - \mu}{\sigma} \Rightarrow 0.525 = \dfrac{X - 60}{10}$

$\Rightarrow\ X = 65.25$

$P(Z \le a) = 0.3 \Rightarrow\ Z = -0.525$

$Z = \dfrac{X - \mu}{\sigma} \Rightarrow -\ 0.525 = \dfrac{X - 60}{10}$

$\Rightarrow\ X = 54.75$

Therefore, grace mark = 65.25 – 54.75 = 10.5

10. Local authorities in a certain city install 10000 electric lamps in streets of city. If these lamps have average life of 1000 burning hours with standard deviation of 200 hours. What is the number of lamps might be expected to fail (i) in first 800 hours (ii) between 800 and 1200 hours.

Solution: $\mu = 1000,\ \sigma = 200.$

(a)

$$P(Z \le 800) = P\left(Z \le \frac{800 - 1000}{200}\right)$$
$$= P(Z \le -1)$$
$$= 1 - P(Z \le 1)$$
$$= 0.1587$$

Number of lamps = 0.1587 X 10000 = 1587.

(b)

$$P(800 \le Z \le 1200) = P\left(\frac{800 - 1000}{200} \le Z \le \frac{1200 - 1000}{200}\right)$$
$$= P(-1 \le Z \le -1)$$
$$= 0.6826$$

Number of lamps = 0.6826 X 10000 = 6826

11. Annual rainfall at a place is known to be normal distributed with $\mu = 29.5$ inches and $\sigma = 2.5$ inches. How many inches of rain is expected to exceeded about 5% of the time?

Solution:

$$\mu = 29.5, \sigma = 2.5.$$
$$P(Z > a) = 0.05$$
$$\Rightarrow P(Z \leq a) = 0.95$$
$$\Rightarrow Z = 1.645$$
$$X = \mu + \sigma Z = 33.61$$

Therefore, amount of rain exceeded = 33.61 - 29.5 = 4.11 inches.

Exponential distributions

A continuous random variable X is said to have exponential distribution with parameter $\alpha > 0$ if its pdf is given by

$$f(x) = \begin{cases} \alpha e^{-\alpha x}, x > 0 \\ 0, elsewhere \end{cases}$$

and we write $X \quad \exp(\alpha)$

Mean:

$$E(X) = \int_0^\infty x \alpha e^{-\alpha x} dx$$

Put $\alpha x = z$

Then $$E(X) = \frac{1}{\alpha} \int_0^\infty z e^{-z} dz = \frac{1}{\alpha} \int_0^\infty e^{-z} z^{2-1} dz = \frac{1}{\alpha} \Gamma(2) = \frac{1}{\alpha}$$

Variance:

$$V(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_0^\infty x^2 \alpha e^{-\alpha x} dx$$

Put $\alpha x = z$

$$E\left(X^2\right)=\frac{1}{\alpha^2}\int_0^\infty z^2 e^{-z}\,dz=\frac{1}{\alpha^2}\int_0^\infty e^{-z} z^{3-1}\,dz=\frac{1}{\alpha}\Gamma(3)=\frac{2!}{\alpha^2}=\frac{2}{\alpha^2}$$

$$V(X)=\frac{1}{\alpha^2}$$

Gamma Distribution

A continuous random variable X is said to have Gamma distribution with parameter $r>0 \wedge \alpha>0$ if its pdf is given by

$$f(x)=\begin{cases}\frac{\alpha}{\Gamma(r)}e^{-\alpha x}(\alpha x)^{r-1}, x>0\\ 0, elsewhere\end{cases}$$

and we write $X \ G(r,\alpha)$

Note: In Gamma distribution if r = 1 we get exponential distribution. Therefore, exponential distribution is a special case of Gamma distribution.

Mean:

$$E(X)=\int_0^\infty x\frac{\alpha}{\Gamma(r)}e^{-\alpha x}(\alpha x)^{r-1}dx$$

Put $\alpha x=z$

Then $$E(X)=\frac{1}{\alpha}\frac{1}{\Gamma(r)}\int_0^\infty e^{-z}z^r\,dz=\frac{1}{\alpha}\frac{1}{\Gamma(r)}\Gamma(r+1)=\frac{1}{\alpha}\frac{1}{\Gamma(r)}r\Gamma(r)=\frac{r}{\alpha}$$

Variance:

$$V(X)=E\left(X^2\right)-[E(X)]^2$$

$$E\left(X^2\right)=\int_0^\infty x^2\frac{\alpha}{\Gamma(r)}e^{-\alpha x}(\alpha x)^{r-1}dx=\frac{r^2+r}{\alpha^2}$$

$$V(X)=\frac{r}{\alpha^2}$$

Chi-square distribution

A continuous random variable X is said to have Chi-square distribution if its pdf is given by

$$f(x)=\begin{cases}\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}e^{\frac{-x}{2}}x^{\frac{n}{2}-1}, x>0, n>0\\ 0, elsewhere\end{cases}$$

We write $X \ \ \lambda^2(n)$ where n is called the number of degrees of freedom.

Mean: $$E(X)=\int_0^\infty x\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}e^{\frac{-x}{2}}x^{\frac{n}{2}-1}dx$$

Put $\frac{x}{2}=z$

Then $$E(X)=\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}\int_0^\infty e^{-z} ¿¿$$

Variance:

$$V(X)=E(X^2)-[E(X)]^2$$

$$E(X^2)=\int_0^\infty x^2\frac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}e^{\frac{-x}{2}}x^{\frac{n}{2}-1}dx=n^2+2n$$

$$V(X)=2n$$