RANDOM VARIABLE

Intuitively by a random variable we mean a real number X connected with the outcome of a random experiment E. For example, if E consists of three tosses of a coin, one can consider the random variable which is the number of heads (0, 1, 2 or 3).

Outcome	HHH	HHT	HTH	THH	TTH	THT	HTT	TTT
Value of X	3	2	2	2	1	1	1	0

Let S denote the sample space of a random experiment. A random variable means it is a rule which assigns a numerical value to each and every outcome of the experiment. Thus, random variable is a function X (ω) with domain S and range (-∞, ∞) such that for every real number *a*, the event [ω: X (ω) ≤ *a*] ϵ B the field of subsets in S. It is denoted as f: S → R.

Note that all the outcomes of the experiment are associated with a unique number. Therefore, f is an example of a random variable. Usually a random variable is denoted by letters such as X, Y, Z etc. The image set of the random variable may be written as f(S) = {0, 1, 2, 3}.

There are two types of random variables. They are;

1. Discrete Random Variable (DRV)
2. Continuous Random Variable (CRV).

Discrete Random Variable: A discrete random variable is one which may take on only a countable number of distinct values such as 0, 1, 2, 3, … . Discrete random variables are usually (but not necessarily) counts. If a random variable takes at most a countable number of values, it is called a **discrete random variable**. In other words, a real valued function defined on a discrete sample space is called a discrete random variable.

Examples of Discrete Random Variable:

i. In the experiment of throwing a die, define X as the number that is obtained. Then X takes any of the values 1 – 6. Thus, X(S) = {1, 2, 3…6} which is a finite set and hence X is a DRV.

ii. If X be the random variable denoting the number of marks scored by a student in a subject of an examination, then X(S) = {0, 1, 2, 3,…100}. Then, X is a DRV.

iii. The number of children in a family is a DRV.

iv. The number of defective light bulbs in a box of ten is a DRV.

Probability Mass Function: Suppose X is a one-dimensional discrete random variable taking at most a countably infinite number of values x_1, x_2, …. With each possible outcome x_i, one can associate a number $p_i = P(X = x_i) = p(x_i)$, called the probability of x_i.

The numbers $p(x_i)$; $i = 1, 2, \ldots$ must satisfy the following conditions:

(i) $p(x_i) \geq 0 \ \forall\ i$,

(ii) $\sum_{i=1}^{\infty} p(x_i) = 1$.

This function p is called the **probability mass function** of the random variable X and the set $\{x_i, p(x_i)\}$ is called the probability distribution of the random variable X.

Remarks:

1. The set of values which X takes is called the spectrum of the random variable.
2. For discrete random variable, knowledge of the probability mass function enables us to compute probabilities of arbitrary events. In fact, if E is a set of real numbers, we have $(X \in E) = \sum_{x \in E \cap S} p(x)$, where S is the sample space.

Discrete Distribution Function: In this case there is a countable number of points $x_1, x_2, x_3, \ldots$ and numbers $p_i \geq 0, \sum_{i=1}^{\infty} p_i = 1$ such that $(X) = \sum_{(i: x_i \leq x)} p_i$.

Mean/Expected Value, Variance and Standard Deviation of DRV:

The **mean or expected value** of a DRV X is defined as

$E(X) = \mu = \sum P(X = x_i) * x_i = \sum P_i X_i$.

The **variance** of a DRV X is defined as

$Var\ (X) = \sigma^2 = \sum P(X = x_i) * (x_i - \mu)^2 = \sum P_i (x_i - \mu)^2 = \sum P_i x_i^2 - \mu^2$.

The **standard deviation** of DRV X is defined as

$SD\ (X) = \sigma = \sqrt{\sigma^2} = \sqrt{Var\ (X)}$.

Continuous Random Variable: A continuous random variable is not defined at specific values. Instead, it is defined over an interval of values, and is represented by the area under a curve. Thus, a random variable X is said to be continuous if it can take all possible values between certain limits. In other words, a random variable is said to be continuous when its different values cannot be put in 1-1 correspondence with a set of positive integers. Here, the probability of observing any single value is equal to zero, since the number of values which may be assumed by the random variable is infinite.

A continuous random variable is a random variable that (at least conceptually) can be measured to any desired degree of accuracy.

Examples of Continuous Random Variable:

i. Rainfall in a particular area can be treated as CRV.
ii. Age, height and weight related problems can be included under CRV.
iii. The amount of sugar in an orange is a CRV.
iv. The time required to run a mile is a CRV.

Important Remark: In case of DRV, the probability at a point i.e., P ($x = c$) is not zero for some fixed c. However, in case of CRV the probability at a point is always zero, i.e., P($x = c$) = 0 for all possible values of c.

Probability Density Function: The probability density function (p.d.f) of a random variable X usually denoted by $f_X(x)$ or simply by $f(x)$ has the following obvious properties:

i) $f(x) \geq 0, -\infty < x < \infty$

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

iii) The probability $P(E)$ given by $P(E) = \int f(x)dx$ is well defined for any event E.

If $f(x)$ is the p.d.f of x, then the probability that x belongs to A, where A is some interval (a, b) is given by the integral of $f(x)$ over that interval.

i.e., $P(X \in A) = \int_a^b f(x)dx$

Cumulative Density Function: Cumulative density function of a continuous random variable is defined as $F(x) = \int_{-\infty}^{x} f(t)dt$ for $-\infty < x < \infty$.

Mean/Expectation, Variance and Standard deviation of CRV:

The mean or expected value of a CRV X is defined as $\mu = E(X) = \int_{-\infty}^{\infty} x\, f(x)dx$

The variance of a CRV X is defined as $Var(X) = \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$

The standard deviation of a CRV X is given by = $\sqrt{Var(X)}$.

EXAMPLES:

1. The probability density function of a discrete random variable X is given below:

x	0	1	2	3	4	5	6
$P(X=x) = f(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

Find (i) k; (ii) F (4); (iii) $P(X \geq 5)$; (iv) $P(2 \leq X < 5)$; (v) E(X) and (vi) Var (X).

Solution: To find the value of k, consider the sum of all the probabilities which equals to $49k$. Equating this to 1, we obtain $k = 1/49$. Therefore, distribution of X may now be written as

x	0	1	2	3	4	5	6
P(*X*=*x*) = *f*(*x*)	1/49	3/49	5/49	7/49	9/49	11/49	13/49

Using this, we may solve the other problems in hand.

$F(4) = P[X \leq 4] = P[X = 0] + P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4] = \frac{25}{49}$.

$P[X \geq 5] = P[X = 5] + P[X = 6] = \dfrac{24}{49}$.

$P[2 \leq X < 5] = P[X = 2] + P[X = 3] + P[X = 4] = \dfrac{21}{49}$.

Next to find *E*(*X*), consider

$$E(X) = \sum_i x_i * f(x_i) = \frac{203}{49}.$$

To obtain Variance, it is necessary to compute

$E(X^2) = \sum_i x_i^2 * f(x_i) = \frac{973}{49}$.

Thus, Variance of X is obtained by using the relation,

$Var(X) = E(X^2) - [E(X)]^2 = \frac{973}{49} - \left(\frac{203}{49}\right)^2$.

2. A random variable, X, has the following distribution function.

X	-2	-1	0	1	2	3
$f(x_i)$	0.1	k	0.2	$2k$	0.3	k

Find (i) k; (ii) F (2); (iii) $P(-2 < X < 2)$; (iv) $P(-1 < X \leq 2)$; (v) *E*(*X*) and (vi) Variance.

Solution: Consider the result, namely, sum of all the probabilities equals 1,

$0.1 + k + 0.2 + 2k + 0.3 + k = 1$ yields $k = 0.1$. In view of this, distribution function of X may be formulated as

X	-2	-1	0	1	2	3
$f(x_i)$	0.1	0.1	0.2	0.2	0.3	0.1

Note that

$F(2) = P[X \leq 2] = P[X = -2] + P[X = -1] + P[X = 0] + P[X = 1] + P[X = 2]$

$= 0.9$. The same also be obtained using the result,

$\mathrm{F}(2) = \mathrm{P}[\mathrm{X} \leq 2] = 1 - \mathrm{P}[\mathrm{X} < 1] = 1 - \{\mathrm{P}[\mathrm{X} = -2] + \mathrm{P}[\mathrm{X} = -1] + \mathrm{P}[\mathrm{X} = 0]\} = 0.6$.

Next, $P(-2 < X < 2) = P[X = -1] + P[X = 0] + P[X = 1] = 0.5$.

Clearly, $P(-1 < X \leq 2) = 0.7$.

Now, consider $(X) = \sum_i x_i * f(x_i) = 0.8$.

Then $E(X^2) = \sum_i x_i^2 * f(x_i) = 2.8$. $Var(X) = E(X^2) - \{E(X)\}^2 = 2.8 - 0.64 = 2.16$.

3. A shipment of 20 similar laptop computers to a retail outlet contains 3 that are defective. If a school makes a random purchase of 2 of these computers, find the probability distribution for the number of defectives.

Solution: Let *X* be a random variable whose values *x* are the possible numbers of defective computers purchased by the school. Then *x* can only take the numbers 0, 1, and 2. Now

$f(0) = P(X = 0) = \frac{\binom{3}{0}\binom{17}{2}}{\binom{20}{2}} = \frac{68}{95}$, $f(1) = P(X = 1) = \frac{\binom{3}{1}\binom{17}{1}}{\binom{20}{2}} = \frac{51}{190}$

$f(2) = P(X = 2) = \frac{\binom{3}{2}\binom{17}{0}}{\binom{20}{2}} = \frac{3}{190}$.

Thus, the probability distribution of *X* is

x	0	1	2
$f(x)$	68/95	51/190	3/190

4. If a car agency sells 50% of its inventory of a certain foreign car equipped with side airbags, find a formula for the probability distribution of the number of cars with side airbags among the next 4 cars sold by the agency.

Solution: Since the probability of selling an automobile with side airbags is 0.5, the $2^4 = 16$ points in the sample space are equally likely to occur. Therefore, the denominator for all probabilities, and also for our function, is 16. To obtain the number of ways of selling 3 cars with side airbags, we need to consider the number of ways of partitioning 4 outcomes into two cells, with 3 cars with side airbags assigned to one cell and the model without side airbags assigned to the other. This can be done in $\binom{4}{3} = 4$ ways. In general, the event of selling *x* models with side airbags and 4 - x models without side airbags can occur in $\binom{4}{x}$ ways, where *x* can be 0, 1, 2, 3, or 4. Thus, the probability distribution *f*(*x*) = *P*(*X* = *x*) is

$$f(x) = \binom{1}{16}\binom{4}{x} \; for \;\; x = 0,1,2,3,4.$$

5. The diameter of an electric cable, say X, is assumed to be a continuous random variable with p.d.f $f(x) = \begin{cases} 6x(1-x) & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

(i) Check that above is p.d.f.

(ii) Find $P\left(\frac{2}{3} < x < 1\right)$

(iii) Determine a number *b* such that $P(X < b) = P(X > b)$.

Solution: (i) $f(x) \ge 0$ in the given interval.

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{0} f(x)dx + \int_{0}^{1} f(x)dx + \int_{1}^{\infty} f(x)dx$$

$= 0 + \int_0^1 6x(1-x)\,dx + 0$

$= \left\{\frac{6x^2}{2} - \frac{6x^3}{3}\right\}$ by putting limits x = 0 to 1 we get

$= 1$

(ii) $P\left(\frac{2}{3} < x < 1\right) = \int_{2/3}^{1} f(x)dx = \int_{2/3}^{1} (6x - 6x^2)dx = \frac{7}{27}$.

(iii) $P(X < b) = P(X > b)$

$\int_0^b f(x)dx \; = \int_b^1 f(x)dx$

$6\int_0^b x(1-x)dx = 6\int_b^1 x(1-x)dx$

$\left(\frac{b^2}{2} - \frac{b^3}{3}\right) = \left[\left(\frac{1}{2} - \frac{1}{3}\right) - \left(\frac{b^2}{2} - \frac{b^3}{3}\right)\right]$

$3b^2 - 2b^3 = [1 - 3b^2 + 2b^3]$

$4b^3 - 6b^2 + 1 = 0$

$(2b-1)(2b^2 - 2b - 1) = 0$

From this b = ½ is the only real value lying between 0 and 1 and satisfying the given condition.

6. Suppose that the error in the reaction temperature, in ◦C, for a controlled laboratory experiment is a continuous random variable X having the probability density function

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{elsewhere} \end{cases}$$

(i) Verify that $f(x)$ is a probability density function.

(ii) Find P(0 < X ≤ 1).

Solution: a) $\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{2} \frac{x^2}{3} dx = 1$. Hence the given function is a p.d.f.

b) P(0 < X ≤ 1) = $\int_0^1 \frac{x^2}{3} dx = \frac{1}{9}$.

7. The length of time (in minutes) that a certain lady speaks on telephone is found to be a random variable with probability function $f(x) = \begin{cases} Ae^{\frac{-x}{5}} & for\ x \geq 0 \\ 0 & \text{otherwise} \end{cases}$

(i) Find A

(ii) Find the probability that she will speak on the phone

(a) more than 10 min (b) less than 5 min (c) between 5 & 10 min.

Solution: (i) Given $f(x)$ is p.d.f. i.e., $\int_{-\infty}^{\infty} f(x)dx = 1$

$$\int_{-\infty}^{0} f(x)dx + \int_{0}^{\infty} f(x)dx = 1$$

$$\xrightarrow{yields} 0 + \int_{0}^{\infty} Ae^{\frac{-x}{5}}dx = 1$$

$$\xrightarrow{yields} A = \frac{1}{5}$$

(ii) (a) $P(x > 10) = \int_{10}^{\infty} f(x)dx = \int_{10}^{\infty} \frac{1}{5}e^{\frac{-x}{5}}\,dx = e^{-2} = 0.1353$

(b) $P(x < 5) = \int_{-\infty}^{5} f(x)dx = \int_{0}^{5} \frac{1}{5}e^{\frac{-x}{5}}\,dx = -e^{-1} + 1 = 0.6322$

(c) $P(5 < x < 10) = \int_{5}^{10} f(x)dx = \int_{5}^{10} \frac{1}{5}e^{\frac{-x}{5}}\,dx = -e^{-2} + e^{-1} = 0.2325$.

8. Suppose X is a continuous random variable with the following probability density function $f(x) = 3x^2$ for $0 < x < 1$. Find the mean and variance of X.

Solution: Mean $= \mu = \int_{-\infty}^{\infty} xf(x)dx$

$$= \int_{-\infty}^{0} xf(x)dx + \int_{0}^{1} xf(x)dx + \int_{1}^{\infty} xf(x)dx$$

$$= 0 + \int_{0}^{1} x * 3x^2dx + 0 = \int_{0}^{1} 3x^3dx = \frac{3}{4} .$$

Variance $= \sigma^2 = \int_{-\infty}^{\infty} x^2f(x)dx - \mu^2$

$$= \int_{0}^{1} x^2f(x)dx - \mu^2$$

$$= \int_{0}^{1} x^2 * 3x^2dx - \left(\frac{3}{4}\right)^2$$

$$= \int_{0}^{1} 3x^4dx - \left(\frac{3}{4}\right)^2 = \frac{3}{80}.$$

Exercise:

1. Two cards are drawn randomly, simultaneously from a well shuffled deck of 52 cards. Find the variance for the number of aces. Ans: 0.1392
2. If X is a discrete random variable taking values 1,2,3,… with $P(x) = \frac{1}{2}\left(\frac{2}{3}\right)^x$. Find P(X being an odd number) by first establishing that P(x) is a probability function. Ans: 3/5.
3. The probability mass function of a random variable X is zero except the points x = 0,1,2. At these points it has the values $p(0) = 3c^3, p(1) = 4c - 10c^2$ and $p(2) = 5c - 1$ for some $c > 0$.
   a) Determine the value of c. Ans: 1/3
   b) Compute the probabilities $P(X < 2)$ and $P(1 < X \leq 2)$. Ans: 1/3, 2/3
   c) Find the largest x such that $F(x) < \frac{1}{2}$. Ans: 1
   d) Find the smallest x such that $F(x) \geq \frac{1}{3}$. Ans: 1

4. If X is a random variable with $P(X = x) = \frac{1}{2^x}$, where $x = 1,2,3, \dots \infty$.

   Find i) P(X) (ii) P(X = even) (iii) P(X = divisible by 3). Ans: 1, 1/3, 1/7.

5. A continuous random variable has the density function $f(x) = \begin{cases} kx^2 & -3 < x < 3 \\ 0 & \text{otherwise} \end{cases}$

   Find k and hence find $P(x < 3)$, $P(x > 1)$.

6. Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} ax , & 0 \le x \le 1 \\ a , & 1 \le x \le 2 \\ -ax + 3a , & 2 \le x \le 3 \\ 0 , & \text{otherwise} \end{cases}$$

   (i) Determine the constant. (ii) Compute $P(X \le 1.5)$. Ans: 1/2, 1/2.

7. Find the mean and variance of the probability density function $f(x) = \frac{1}{2} e^{-|x|}$

   Ans: Mean = 0 and Variance = 2.

8. A continuous distribution of a variable X in the range (-3, 3) is defined by

$$f(x) = \begin{cases} \frac{1}{16}(3 + x)^2 , & -3 \le x \le -1 \\ \frac{1}{16}(6 - 2x^2), & -1 \le x \le 1 \\ \frac{1}{16}(3 - x)^2 , & 1 \le x \le 3 \end{cases}$$

   (i) Verify that the area under the curve is unity. Ans: Yes, area is unity.

(ii) Find the mean and variance of the above distribution. Ans: 0, 1.

PROBABILITY DISTRIBUTIONS

Introduction:

In this section, we will discuss discrete probability distributions and continuous probability distributions.

Discrete probability distribution is used when the sample space is discrete but not countable, whereas continuous probability distribution is used when the sample space is continuous or sample space is defined in a continuous interval.

In discrete distributions, the variables are distributed according to some definite probability law which can be expressed mathematically. The present study will also enable us to fit a mathematical' model or a function of the form *y* = *p(x)* to the observed data. In discrete distributions we discuss Binomial distribution, Poisson distribution and Geometric distributions. In continuous distributions we discuss Normal distribution and Exponential distribution.

Bernoulli distribution: A random variable X which takes two values 0 and 1, with probabilities q and p respectively, i.e ., P (X = 1) = p, P(X = 0) = q, q = 1 - p is called a Bernoulli variate and is said to have a Bernoulli distribution.

The probability of getting a head or a tail on tossing a coin is 1/2. If a coin is tossed thrice, the sample space S = {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}.

The probability of getting one head and two tails = 3/8. i.e., {HTT, TTH, THT}.

The probability of each one (one head, one tail, one tail) of these being (1/2) * (1/2) * (1/2) i.e., $(1/2)^3$, their total probability shall be 3 * $(1/2)^3$.

Similarly if a trial is repeated 'n' times and if 'p' is the probability of a success and 'q' that of a failure, then the probability of 'r' successes and 'n – r' failures is given by '$p^r q^{n-r}$'. But these 'r' successes and 'n – r' failures can occur in any of n_{C_r} ways in each of which the probability is same. Thus the probability of 'r' successes in $n_{C_r} p^r q^{n-r}$. The probability of at least 'r' successes in 'n' trials = Sum of probabilities of 'r, r + 1, ..., n' successes.

$$= n_{C_r} p^r q^{n-r} + n_{C_{r+1}} p^{r+1} q^{n-r-1} + \cdots + n_{C_n} p^n.$$

Binomial distribution: Binomial distribution was discovered by James Bernoulli (1654-1705) in the year 1700. It is concerned with trials of a repetitive nature in which only the occurrence or non-occurrence, success or failure, acceptance or rejection, yes or no of a particular event is of interest.

If we perform a series of independent trials such that for each trial 'p' is the probability of success and 'q' that of a failure, then the probability of 'r' successes in a series of 'n' trials is given by $n_{C_r} p^r q^{n-r}$, where 'r' takes any integral value from 0 to n. The probabilities of 0, 1, 2, ..., r,..., n successes are, therefore, given by

$$q^n, n_{C_1} p q^{n-1}, n_{C_2} p^2 q^{n-2}, \dots, n_{C_r} p^r q^{n-r}, \dots, p^n$$

The probability of the number of successes so obtained is called the **Binomial distribution**.

The sum of the probabilities $= q^n + n_{C_1} p q^{n-1} + n_{C_2} p^2 q^{n-2} + \cdots + p^n = (q+p)^n = 1.$

The most obvious application deals with the testing of items as they come off an assembly line, where each trial may indicate a defective or a non defective item. We may choose to define either outcome as a success. The process is referred to as a **Bernoulli process**. Each trial is called a **Bernoulli trial**. The number *X* of successes in *n* Bernoulli trials is called a

binomial random variable. The probability distribution of this discrete random variable is called the **binomial distribution**, and its values will be denoted by $b(x;\ n,\ p)$ since they depend on the number of trials and the probability of a success ‘p’ on a given trial.

To prove that for a binomial distribution $\sum_{x=0}^{n} p(x) = 1$

Proof:

$$\sum_{x=0}^{n} p(x) = \sum_{x=0}^{n} n_{C_x} p^x q^{n-x} = n_{C_0} p^0 q^{n-0} + n_{C_1} p^1 q^{n-1} + \cdots + n_{C_n} p^n q^{n-n}$$
$$= q^n + n_{C_1} p^1 q^{n-1} + n_{C_2} p^2 q^{n-2} + \cdots + p^n$$
$$= (p+q)^n = 1$$

Note:

In a binomial distribution

1. n, the number of trials is finite.
2. each trial has two possible outcomes called success & failure.
3. all the trials are independent
4. p & q are constants for all the trials.

Mean and Variance of a Binomial distribution:

$$\textbf{Mean} = \mu = \frac{\sum x\, p(x)}{\sum p(x)}$$

But for Binomial distribution $\sum p(x) = 1$

Mean $= \ \mu = \sum x\, p(x)$

$$= \sum x\, n_{C_x} p^x q^{n-x}$$
$$= 0.\, n_{C_0} p^0 q^{n-0} + 1.\, n_{C_1} p^1 q^{n-1} + 2.\, n_{C_2} p^2 q^{n-2} + \cdots + n.\, n_{C_n} p^n q^{n-n}$$
$$= n.\, p^1 q^{n-1} + n(n-1) p^2 q^{n-2} + \cdots + n.\, p^n$$
$$= np(q^{n-1} + (n-1) p q^{n-2} + \cdots + p^{n-1})$$
$$= np(q+p)^{n-1}$$
$$\boldsymbol{\mu = np}$$

$$\textbf{Variance} = \sigma^2 = \frac{\sum x^2\, p(x)}{\sum p(x)} - \mu^2$$

But for Binomial distribution $\sum p(x) = 1$

$$\sigma^2 = \sum x^2\, p(x) - \mu^2$$
$$= 0.\, n_{C_0} p^0 q^{n-0} + 1.\, n_{C_1} p^1 q^{n-1} + 4.\, n_{C_2} p^2 q^{n-2} + \cdots + n^2.\, n_{C_n} p^n q^{n-n} - (np)^2$$
$$= n.\, p^1 q^{n-1} + 2n(n-1) p^2 q^{n-2} + \tfrac{3}{2} n(n-1)(n-2) p^3 q^{n-3} + \cdots + n^2.\, p^n - (np)^2$$
$$= np(q^{n-1} + 2(n-1) p q^{n-2} + \tfrac{3}{2}(n-1)(n-2) p^2 q^{n-3} \ldots + n p^{n-1}) - (np)^2$$

$$= np[(q^{n-1} + (n-1)pq^{n-2} + \frac{1}{2}(n-1)(n-2)p^2q^{n-3} \ldots + p^{n-1}) +$$

$$((n-1)pq^{n-2} + 1.(n-1)(n-2)p^2q^{n-3} \ldots + (n-1)p^{n-1})] - (np)^2$$

$$= np[(q+p)^{n-1} + (n-1)p(q^{n-2} + (n-2)pq^{n-3} + \cdots + p^{n-2})] - (np)^2$$

$$= np[(q+p)^{n-1} + (n-1)p(q+p)^{n-2}] - (np)^2$$

$$= np[(1)^{n-1} + (n-1)p(1)^{n-2}] - (np)^2$$

$$= np[1 + (n-1)p] - n^2p^2$$

$$= np + n^2p^2 - np^2 - n^2p^2$$

$$= np(1-p)$$

$$\boldsymbol{\sigma^2 = npq}$$

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

Problems:

1. The mean and variance of a binomial variate are respectively 16 & 8. Find (i) $\mathrm{P}(\mathrm{X} = 0)$ (ii) $\mathrm{P}(\mathrm{X} \geq 2)$

   Solution: Given, Mean = $\mu = np = 16$ and Variance = $\sigma^2 = npq = 8$

   $\frac{npq}{np} = \frac{8}{16} = \frac{1}{2}$

   i.e., $q = \frac{1}{2}$. Therefore, $p = 1 - q = \frac{1}{2}$. Also, $np = 16$ i.e., $n = 32$.

   (i) $\mathrm{P}(\mathrm{X} = 0) = n_{C_0}p^0q^{n-0} = \left(\frac{1}{2}\right)^0\left(\frac{1}{2}\right)^{32} = \left(\frac{1}{2}\right)^{32}$

   (ii) $\mathrm{P}(\mathrm{X} \geq 2) = 1 - \mathrm{P}(\mathrm{X} < 2) = 1 - \mathrm{P}(\mathrm{X} = 0) - \mathrm{P}(\mathrm{X} = 1)$

   $= 1 - \left(\frac{1}{2}\right)^{32} - 32\left(\frac{1}{2}\right)^{32} = 1 - 33\left(\frac{1}{2}\right)^{32}$.

2. Six dice are thrown 729 times. How many times do you expect at least 3 dice to shown a 5 or 6?

   Solution: Here n = 6, N = 729

   $\mathrm{P}(x \geq 3) = 6_{C_x}p^xq^{6-x}$

   Let p be the probability of getting 5 or 6 with 1 dice

   i.e., p = 2/6 = 1/3. Thus, q = 1 – 1/3 = 2/3

   $\mathrm{p}(x \geq 3) = \mathrm{p}(x = 3, 4, 5, 6)$

   $= \mathrm{p}(x = 3) + \mathrm{p}(x = 4) + \mathrm{p}(x = 5) + \mathrm{p}(x = 6)$

   = 0 .3196

   Therefore, number of times = 729*0.3196 = 233

3. A basket contains 20 good oranges and 80 bad oranges. 3 oranges are drawn at random from this basket. Find the probability that out of 3 (i) exactly 2 (ii) at least 2 (iii) at most 2 are good oranges.

   Solution: Let p be the probability of getting a good orange i.e., p $= \frac{80C_1}{100C_1}$

   p = 0.8 and q = 1- 0.8 = 0.2

   (i) p (x = 2) = $3C_2(0.8)^2(0.2)^1$ = 0.384

   (ii) p (x ≥ 2) = p(2) + p(3) = 0.896

   (iii) p (x ≤ 2) = p(0) + p(1) + p(2) = 0.488

4. In a sampling a large number of parts manufactured by a machine, the mean number of defective in a sample of 20 is 2. Out of 1000 such samples how many would expected to contain at least 3 defective parts.

   Solution: Given; n = 20, n p = 2

   i.e., p = 1/10 and q = 1-p = 9/10

   p (x ≥ 3) = 1 – p (x < 3)

   = 1 – p (x = 0,1,2) = 0.323

   Number of samples having at least 3 defective parts = 0.323 * 1000 = 323

5. The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that

   (i) at least 10 survive, (ii) from 3 to 8 survive.

   Solution: Let *X* be the number of people who survive.

   (i) $p(X \geq 10) = 1 - p(X < 10) = 1 - \sum_{x=0}^{9} b(x; 15,0.4) = 1 - 0.9662 = 0.0338$

   (ii) $p(3 \leq X \leq 8) = \sum_{x=3}^{8} b(x; 15,0.4) = \sum_{x=0}^{8} b(x; 15,0.4) - \sum_{x=0}^{2} b(x; 15,0.4)$

   = 0.9050 - 0.0271 = 0.8779

Exercise:

1. Ten coins are thrown simultaneously. Find the probability of getting at least seven heads.
2. In 256 sets of 12 tosses of a coin, in how many cases one can expect 8 heads and 4 tails.
3. Let ‘x’ be a binomial variate with mean 6 and variance 4. Find the distribution of ‘x’,
4. In a binomial distribution consisting of 5 independent trials, probability of 1 & 2 successes are 0.4096 & 0.2048 respectively. Find the parameter ‘p’ of the distributive function.

5. A and B play a game in which their chances of winning are in the ratio 3:2. Find A's chance of winning at least three games out of five games played.

Answers: 1. $1.\frac{176}{1024}$ $2.\,31(\text{approx})$ $3.\,18_{C_x}\left(\frac{1}{3}\right)^x\left(\frac{2}{3}\right)^{18-x}$ $4.\,p = 0.2$ $5.\,0.68$

Poisson distribution:

Poisson distribution was discovered by the French mathematician and physicist Simeon Denis Poisson (1781-1840) who published it in 1837. Poisson distribution is a limiting case of the binomial distribution under the following conditions:

i) *n,* the number of trials is indefinitely large, *i.e.,* $n \to \infty$.

ii) *p,* the constant probability of success for each trial is indefinitely small, *i.e., $p \to 0$.*

iii) $np = \lambda$, (say), is finite. Thus $p = \lambda$ /n, $q = 1 - \lambda /n$, where λ is a positive real number.

The probability function of the Poisson distribution is given by

$\boldsymbol{p(x,\lambda) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}}$ where $\boldsymbol{\lambda}$ is known as the parameter of poisson distribution.

Definition: A random variable X is said to follow a Poisson distribution if it assumes only non-negative values and its probability mass function is given by

$$p(\mathrm{x},\lambda) = P(X = \mathrm{x}) = \frac{e^{-\lambda}\lambda^{\mathrm{x}}}{\mathrm{x}!}\ ;\ \ x = 0, 1, 2, \ldots;\ \lambda > 0$$

$$= 0, \text{ otherwise}$$

Remarks:

1. It should be noted that $\sum_{x=0}^{\infty} P(X = \mathrm{x}) = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^{\mathrm{x}}}{\mathrm{x}!} = e^{-\lambda}e^{\lambda} = 1$
2. The corresponding distribution function is:

   $F(x) = P(X \le \mathrm{x}) = \sum_{r=0}^{x} P(r) = e^{-\lambda} \sum_{r=0}^{\mathrm{x}} \frac{\lambda^{\mathrm{r}}}{\mathrm{r}!}; x = 0, 1, 2, \ldots$
3. Poisson distribution occurs when there are events which do not occur as outcomes of a definite number of trials (unlike that in binomial) of an experiment but which occur at random points of time and space wherein our interest lies only in the number of occurrences of the event, not in its non-occurrences.
4. Following are some instances where Poisson distribution may be successfully employed.

   i) Number of deaths from a disease (not in the form of an epidemic) such as heart attack or cancer or due to snake bite.

   ii) Number of suicides reported in a particular city.

   iii)The number of defective material in a packing manufactured by a good concern.

iv) Number of faulty blades in a packet of 100.

v) Number of air accidents in some unit of time.

vi) Number of printing mistakes at each page of the book.

vii) Number of telephone calls received at a particular telephone exchange in some unit of time or connections to wrong numbers in a telephone exchange.

viii) Number of cars passing a crossing per minute during the busy hours of a day.

ix) The number of fragments received by a surface area *'t'* from a fragment atom bomb.

x) The emission of radioactive (alpha) particles.

Mean and Variance of a Poisson distribution

$$\textbf{Mean} = \mu = \frac{\sum x\, p(x)}{\sum p(x)}$$

But for Poisson distribution $\sum p(x) = 1$

Mean $= \mu = \sum x\, p(x)$

$$= \sum x \frac{e^{-\lambda}\lambda^{x}}{x!} = \sum \frac{e^{-\lambda}\lambda^{x}}{(x-1)!}$$

$$= e^{-\lambda} \sum \frac{\lambda^{x}}{(x-1)!}$$

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

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

$$= \lambda\, e^{-\lambda}\, e^{\lambda}$$

$$\textbf{Mean} = \boldsymbol{\mu} = \boldsymbol{\lambda}$$

$$\textbf{Variance} = \sigma^2 = \frac{\sum x^2\, p(x)}{\sum p(x)} - \mu^2$$

But for Poisson distribution $\sum p(x) = 1$

$$\sigma^2 = \sum x^2\, p(x) - \mu^2$$

$$= \sum[x(x-1) + x]p(x) - \mu^2 = \sum[x(x-1) + x]\, \frac{e^{-\lambda}\lambda^{x}}{x!} - \mu^2$$

$$= \sum x(x-1)\frac{e^{-\lambda}\lambda^{x}}{x!} + \sum x\, \frac{e^{-\lambda}\lambda^{x}}{x!} - \mu^2$$

$$= \sum \frac{e^{-\lambda}\lambda^{x}}{(x-2)!} + \sum x\, \frac{e^{-\lambda}\lambda^{x}}{x!} - \mu^2$$

$$= e^{-\lambda}\left(\lambda^2 + \frac{\lambda^3}{1!} + \frac{\lambda^4}{2!} + \cdots\right) + \lambda - \lambda^2$$

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

Variance $= \boldsymbol{\sigma^2} = \boldsymbol{\lambda}$

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

Problems

1. It is known that the chance of an error in the transmission of a message through a communication channel is 0.002. 1000 messages are sent through the channel; find the probability that at least 3 messages will be received incorrectly.

   Solution: Here, the random experiment consists of finding an error in the transmission of a message. It is given that n = 1000 messages are sent, a very large number, if p denote the probability of error in the transmission, we have p = 0.002, relatively a small number, therefore, this problem may be viewed as Poisson oriented. Thus, average number of messages with an error is $\lambda = np = 2$.

   Therefore, required probability function is

   $$\mathrm{p}(\lambda, x) = \frac{e^{-\lambda}\lambda^{\mathrm{x}}}{\mathrm{x}!}\ , x = 0,1,2,3,\ldots,\infty.$$

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

   Here, the problem is about finding the probability of the event, namely,

   $$P(X \geq 3) = 1 - P(X < 3) = 1 - \{P[X = 0] + P[X = 1] + P[X = 2]\}$$

   $$= 1 - \left[\sum_{x=0}^{2} \frac{e^{-2}2^{x}}{x!}\right]$$

   $$= 1 - e^{-2}(1 + 2 + 2) = 1 - 5e^{-2}$$

2. A Car-hire firm has two cars it hires out daily. The number of demands for a car on each day is distributed as poisson variate with mean 1.5. Obtain the proportion of days on which i) there was no demand ii) demand is refused.

   Solution: Here λ = 1.5

   i) $\mathrm{p}(x, 0) = \frac{e^{-1.5}(1.5)^{0}}{0!} = 0.2231$

   ii) $\mathrm{p}(x > 2) = 1 - p(x \leq 2) = 1 - p(x = 0, 1, 2)$

   $$= 1 - \mathrm{p}(x = 0) - p(x = 1) - p(x = 2)$$

   $$= 1 - \frac{e^{-1.5}(1.5)^{0}}{0!} - \frac{e^{-1.5}(1.5)^{1}}{1!} - \frac{e^{-1.5}(1.5)^{2}}{2!}$$

   $$= 0.1913$$

3. Assuming that the probability of an individual being killed in a mine accident during a year is 1/2400. Use poisson distribution to calculate the probability that in a mine employing 200 miners there will be at least one fatal accident in a year?

   Solution: Here p = 1/2400, n = 200, λ= np = 0.083

   $\mathrm{p}(x \geq 1) = 1 - p(x < 1) = 1 - p(x = 0)$

   $= 1 - \mathrm{e}^{-0.083} = 0.0796$

4. In a poisson distribution if $\mathrm{P}(2) = \frac{2}{3}\mathrm{P}(1)$, find $\mathrm{P}(0)$. Find also its mean and standard deviation.

   Solution: $\mathrm{p}(\lambda, x) = \frac{e^{-\lambda}\lambda^{\mathrm{x}}}{\mathrm{x}!}$

   Given $\mathrm{P}(2) = \frac{2}{3}\mathrm{P}(1)$

   i.e.,$\mathrm{p}(\lambda, 2) = \frac{2}{3}\mathrm{p}(\lambda, 1)$

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

   $\lambda = \frac{4}{3}$

   Thus, $\mathrm{p}(\lambda, 0) = \frac{e^{-\lambda}\lambda^0}{0!} = e^{-\lambda} = e^{-4/3} = 0.2636$

   Mean $= \mu = \lambda = \frac{4}{3}$ and standard deviation = $\sigma = \sqrt{\lambda} = \sqrt{\frac{4}{3}}$

5. 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 in a group of seven, 5 or more will suffer from it.

   Solution: p = 10% = 0.1, n = 7

   $\mu = np = 0.1 * 7 = 0.7$

   $\mathrm{P}(\mathrm{x} \geq 5) = \mathrm{P}(5) + \mathrm{P}(6) + \mathrm{P}(7)$

   $= \frac{e^{-\lambda}\lambda^5}{5!} + \frac{e^{-\lambda}\lambda^6}{6!} + \frac{e^{-\lambda}\lambda^7}{7!}$

   $= \frac{e^{-0.7}(0.7)^5}{5!} + \frac{e^{-0.7}(0.7)^6}{6!} + \frac{e^{-0.7}(0.7)^7}{7!} = 0.0008$

Exercise:

1. For a poisson variable $3\mathrm{P}(2) = \mathrm{P}(4)$, find standard deviation.
2. If the probability of a bad reaction from a certain injection is 0.001, determine the chance that out of 2000 individuals more than two will get a bad reaction.
3. Fir a poisson distribution to the set of observations given below.

x	0	1	2	3	4
f(x)	122	60	15	2	1

4. In a certain factory turning out razor blades there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10. Use poisson distribution to calculate the approximate number of packets containing no defective, one defective two blades defective respectively ina consignment of 10,000 packets.
5. A manufacturer of cotter pins knows that 5% of his product is defective. If he sells cotter pins in boxes of 100 and guarantees that not more than 10 pins will be defective, what is the approximate probability that a box will fail to meet the guaranteed quality?

Answers: 1. 2.45 2. 0.32 3. $\mathrm{f}(x) = \frac{e^{-0.5}(0.5)^x}{x!}$, for $\mathrm{N} = 200$, it is $\mathrm{N} * \mathrm{f}(x)$.

4. 9802, 196, 2 5. $1 - e^{-5} \sum_{x=0}^{10} \frac{(5)^x}{x!}$

Exponential distribution

Many experiments involve the measurement of time X between an initial point of time and the occurrence of some phenomenon of interest. Exponential distribution deals with such type of continuous random variable X.

A continuous random variable X assuming non-negative values is said to have an exponential distribution with parameter $\lambda > 0$, if its probability density function is given by

$$\mathrm{f}(x) = \begin{cases} \lambda \mathrm{e}^{-\lambda x}, & x \geq 0 \\ 0, & \text{otherwise} \end{cases}$$

Examples such as time between two successive job arrivals, duration of telephone calls, life time of a component or a product, server time at a server in a queue can be taken under Exponential distribution.

Mean and variance of Exponential distribution

$$\textbf{Mean} = \boldsymbol{\mu} = \int_0^\infty x \, \mathrm{f}(x) dx$$

$$= \int_0^\infty x \, \lambda \mathrm{e}^{-\lambda x} dx$$

$$= \lambda \int_0^\infty x \, \mathrm{e}^{-\lambda x} dx$$

$$= \lambda \left[x * \frac{\mathrm{e}^{-\lambda x}}{(-\lambda)} - 1 * \frac{\mathrm{e}^{-\lambda x}}{(-\lambda)^2} \right]_{x=0 \text{ to } \infty}$$

$$\boldsymbol{\mu} = \frac{\mathbf{1}}{\boldsymbol{\lambda}}$$

Variance $= \boldsymbol{\sigma^2} = \displaystyle\int_0^{\infty} x^2\, \mathrm{f}(x)dx - \mu^2$

$$= \int_0^{\infty} x^2\, \lambda e^{-\lambda x} dx - \left(\frac{1}{\lambda}\right)^2$$

$$= \lambda\left[x^2 * \frac{e^{-\lambda x}}{(-\lambda)} - 2x * \frac{e^{-\lambda x}}{(-\lambda)^2} + 2 * \frac{e^{-\lambda x}}{(-\lambda)^3}\right]_{x=0 \text{ to } \infty} - \left(\frac{1}{\lambda}\right)^2$$

$$\boldsymbol{\sigma^2} = \left(\frac{1}{\lambda}\right)^2$$

Standard deviation = $\boldsymbol{\sigma} = \frac{1}{\lambda}$

Problems:

1. Let the mileage (in thousands of miles) of a particular tyre be a random variable X having the probability density $(x) = \begin{cases} \frac{1}{20} e^{\frac{-x}{20}} & , x > 0 \\ 0, x \le 0 \end{cases}$.

   Find the probability that one of these tyres will last (i) at most 10,000 miles (ii) anywhere between 16,000 to 24,000 miles (iii) at least 30,000 miles. Also, find The mean and variance of the given probability density function.

   Solution: (i) $\mathrm{P}(x \le 10) = \int_0^{10} f(x)dx = \int_0^{10} \frac{1}{20} e^{\frac{-x}{20}} dx = \frac{1}{20}\int_0^{10} e^{\frac{-x}{20}} dx = 0.3934$

   (ii) $\mathrm{P}(16 \le x \le 24) = \int_{16}^{24} f(x)dx = \int_{16}^{24} \frac{1}{20} e^{\frac{-x}{20}} dx = \frac{1}{20}\int_{16}^{24} e^{\frac{-x}{20}} dx = 0.148$

   (iii) $\mathrm{P}(x \ge 30) = \int_{30}^{\infty} f(x)dx = \int_{30}^{\infty} \frac{1}{20} e^{\frac{-x}{20}} dx = \frac{1}{20}\int_{30}^{\infty} e^{\frac{-x}{20}} dx = 0.223$

   Mean $= \mu = \frac{1}{\lambda} = \frac{1}{\frac{1}{20}} = 20$ and Variance $= \sigma^2 = \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\left(\frac{1}{20}\right)^2} = 400.$

2. The length of time for one person to be served at a cafeteria is a random variable X having an exponential distribution with a mean of 4 minutes. Find the probability that a person is served in less than 3 minutes on at least 4 of the next 6 days.

   Solution: Given, Mean = 4. i.e., Mean $= 4 = \frac{1}{\lambda} \rightarrow \lambda = \frac{1}{4}$

   The probability density function is $f(x) = \lambda e^{-\lambda x} = \frac{1}{4} e^{\frac{-x}{4}}$

   $$\mathrm{P}(x < 3) = 1 - \mathrm{P}(x \ge 3) = 1 - \int_3^{\infty} f(x)dx$$

   $$= 1 - \int_3^{\infty} \frac{1}{4} e^{\frac{-x}{4}} dx = 1 - e^{\frac{-3}{4}} = 0.9875$$

   Let D represents the number of days on which a person is served in less than 3 minutes. Then using the binomial distribution, the probability that a person is served in less than 3 minutes on at least 4 of next 6 days is;

$P(D \geq 4) = P(D = 4) + P(D = 5) + P(D = 6)$

$= 6_{C_4}\left(1 - e^{-\frac{3}{4}}\right)^4 (e^{-\frac{3}{4}})^2 + 6_{C_5}(1 - e^{-3/4})^5(e^{-\frac{3}{4}})^1 + 6_{C_6}(1 - e^{-3/4})^6(e^{-\frac{3}{4}})^0$

$= 0.3968$

3. The increase in sales per day in a shop is exponentially distributed with Rs 800 as the average. If sales tax is paid at the rate of 6%, find the probability that increase in sales tax return from that shop will exceed Rs 30 per day.

   Solution: Given, Mean = 800

   i.e., $\text{Mean} = 800 = \frac{1}{\lambda} \rightarrow \lambda = \frac{1}{800}$

   The probability density function is $f(x) = \lambda e^{-\lambda x} = \frac{1}{800} e^{\frac{-x}{800}}$

   Let X denotes the sales per day. Total sales tax on X items = $\frac{6}{100}X$

   Given total sales tax exceeds Rs 30 per day. i.e., $\frac{6}{100}X > 30.\ \text{i.e.,} X > 500$

   Probability of sales tax exceeding Rs 30 = Probability of sales per day exceeding 500

   $= P(X > 500) = 1 - P(X \leq 500)$

   $= 1 - \int_0^{500} f(x)dx$

   $= 1 - \int_0^{500} \frac{1}{800} e^{\frac{-x}{800}} dx = 0.5353$

4. After the appointment of a new sales manager the sales in a 2 wheeler showroom is exponentially distributed with mean 4. If 2 days are selected at random what is the probability that (i) on both days, the sales is over 5 units (ii) the sales is over 5 times at least 1 of 2 days.

   Solution: Given, Mean = 4. i.e., $\text{Mean} = 4 = \frac{1}{\lambda} \rightarrow \lambda = \frac{1}{4}$

   The probability density function is $f(x) = \lambda e^{-\lambda x} = \frac{1}{4} e^{\frac{-x}{4}}$

   Let X represents the sales per day

   $P(x > 5) = \int_5^{\infty} f(x)dx = \int_5^{\infty} \frac{1}{4} e^{\frac{-x}{4}} dx = \frac{1}{4}\int_5^{\infty} e^{\frac{-x}{4}} dx = 0.2865$

   Let D = number of days on which sales is over 5 units

   (i) $P(D = 2) = n_{C_x} p^n q^{n-x} = 2_{C_2}(e^{-\frac{5}{4}})^2 \left(1 - e^{-\frac{5}{4}}\right)^{2-2} = 0.082$

   (ii) $P(D = \text{at least 1 of 2 days}) = P(D = 1) + P(D = 2)$

   $2_{C_1}(e^{-\frac{5}{4}})^1 \left(1 - e^{-\frac{5}{4}}\right)^{2-1} + 2_{C_2}(e^{-\frac{5}{4}})^2 \left(1 - e^{-\frac{5}{4}}\right)^{2-2} = 0.4908.$

Exercise:

1. The sales per day in a shop are exponentially distributed with average sale amounting to Rs 100 and net profit is 8%. Find the probability that net profit exceed Rs 30 on 2 consecutive days.
2. Let X and Y have common p.d.f $\alpha e^{-\alpha x}, 0 < x < \infty, \alpha > 0$. Find the p.d.f of

   (i) 3 + 2X (ii) X – Y.
3. If X has exponential distribution with mean 2, find $P(X < 1|X < 2)$.
4. The life (in years) of a certain electrical switch has an exponential distribution with an average life of 2 years. If 100 of these switches are installed in different systems, find the probability that at most 30 fail during the first year.

Answers: 1. $(e^{-3.75})^2$ 2. $\frac{\alpha}{2}\exp\left(-\frac{\alpha(x-3)}{2}\right), x > 3$, $\frac{\alpha}{2}\exp(-\alpha|x|)$, $\forall\, x$

3. $\frac{(1-e^{-\lambda})}{(1-e^{-2\lambda})}$, where $\lambda = \frac{1}{2}$. 4. $P(X \le 30) = \sum_{x=0}^{30} 100_{C_x}(0.606)^x(0.394)^{100-x}$

Normal distribution

The normal distribution was first discovered in 1733 by English mathematician De-Moivre, who obtained this continuous distribution as a limiting case of the binomial distribution and applied it to problems arising in the game of chance.

Among all the distribution of a continuous random variable, the most popular and widely used one is normal distribution function. Most of the work in correlation and regression analysis, testing of hypothesis, has been done based on the assumption that problem follows a normal distribution function or just everything normal. Also, this distribution is extremely important in statistical applications because of the central limit theorem, which states that "under very general assumptions, the mean of a sample of n mutually Independent random variables (having finite mean and variance) are normally distributed in the limit $n \rightarrow \infty$". It has been observed that errors of measurement often possess this distribution.

Definition: A random variable X is said to have a normal distribution with parameters μ (called "mean") and σ^2 (called "variance") if its density function is given by the probability law:

$$f(x;\ \mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left[\frac{-1}{2}\left\{\frac{x-\mu}{\sigma}\right\}^2\right] \text{ for } -\infty < x < \infty, -\infty < \mu < \infty \text{ and } 0 < \sigma < \infty .$$

Examples such as marks scored by students and life span of a product can be included under normal distribution.

Remarks:

1. A random variable X with mean μ and variance σ^2 and following the normal law is expressed by $X \sim N(\mu, \sigma^2)$.
2. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$, is a standard normal variate with E(Z) = 0 and Var (Z) = 1 and we write $Z \sim N(0, 1)$.
3. The p.d.f of standard normal variate Z is given by $\varphi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}, -\infty < z < \infty$ and the corresponding distribution function, denoted by $F(z)$is given by $F(z) = P(Z \leq z) = \int_{-\infty}^{z} \varphi(u)du = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-u^2/2} du$.
   Also, $F(-z) = \int_{-\infty}^{-z} \varphi(u)du = 1 - F(z)$.
4. The graph of f(x) is a famous 'bell-shaped' curve. The top of the bell is directly above the mean μ. For large values of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak.

Note: The limiting form of the binomial distribution for large values of n with neither p nor q is very small, is the normal distribution.

Properties of Normal Distribution:

1. All normal curves are bell-shaped.
2. All normal curves are symmetric about the mean μ.
3. The area under an entire normal curve is 1.
4. All normal curves are positive for all x. i.e., f(x) > 0 for all x.
5. The shape of any normal curve depends on its mean and the standard deviation.

The probabilities are computed numerically and recorded in a special table called the normal distribution table (the probabilities can also be computed using a standard calculator). We can use the following results for the calculation of probabilities.

(i) $P(a \leq X \leq b) = F(b) - F(a)$

(ii) $P(a < X < b) = F(b) - F(a)$

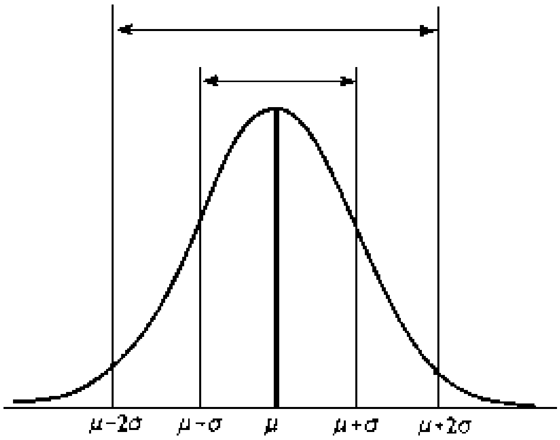
(iii) $P(a < X) = 1 - P(X \leq a) = 1 - F(a)$

(iv) $F(-b) = 1 - F(b)$, where b is positive.

The distributions of some variables including aptitude-tests scores, heights of women/men, have roughly the shape of a normal curve (bell shaped curve)

Normally Distributed Variable

A variable is said to be **normally distributed** or to have a **normal distribution** if its distribution has the shape of a normal curve.



Problems:

2. A sample of 100 battery cells is tested to find the length of life, gave the following results. Mean = 12 hrs. Standard Deviation = 3 hrs. Assuming the data to be normally distributed what % of battery cells are expected to have life (i) more than 15 hrs. (ii) less than 6hrs. (iii) between 10 & 14 hrs .

   Solution: (i) when x = 15 for given mean = 12 hrs and standard deviation = 3 hrs;

   $$\mathrm{P}(\mathrm{x} > 15) = \mathrm{P}\left(\frac{\mathrm{X}-\mu}{\sigma} > \frac{15-\mu}{\sigma}\right) = \mathrm{P}\left(\frac{\mathrm{X}-\mu}{\sigma} > \frac{15-12}{3}\right)$$

   $$= \mathrm{P}(\mathrm{z} > 1)$$

   = 0.5 - 0.3413

   = 0.1587 = 16%

   (ii) When x = 6

   $$\mathrm{P}(\mathrm{x} < 6) = \mathrm{P}\left(\frac{\mathrm{X}-\mu}{\sigma} < \frac{6-\mu}{\sigma}\right) = \mathrm{P}\left(\frac{\mathrm{X}-\mu}{\sigma} < \frac{6-12}{3}\right)$$

   $$= \mathrm{P}(\mathrm{z} < -2)$$

   = 0.5 - 0.4772

   = 0.0228 = 2.28%

   (iii) $\mathrm{P}(10 < x < 14) = \mathrm{P}\left(\frac{10-\mu}{\sigma} < \frac{\mathrm{X}-\mu}{\sigma} < \frac{6-\mu}{\sigma}\right)$

   $$= \mathrm{P}(-0.6667 < z < 0.6667)$$

   $$= 2 * \mathrm{P}(0 < z < 0.6667)$$

   = 2*0.2485 = 0.497 = 50%

3. Find the mean and standard deviation of an examination in which grades 70 and 88 corresponds to standard scores of -0.6 and 1.4 respectively.

   Solution: Standard variable $z = \frac{\mathrm{X}-\mu}{\sigma}$

Here, $-0.6 = \frac{70-\mu}{\sigma}$ gives $\mu - 0.6\sigma = 70$

and $1.4 = \frac{88-\mu}{\sigma}$ gives $\mu + 1.4\sigma = 88$

by solving the above equations, we get $\mu = 75.4$ and $\sigma = 9$.

4. The marks X obtained in mathematics by 1000 students is normally distributed with mean 78% and standard deviation 11%. Determine how many students got marks above 90%.

Solution: Here, mean = 78% = 0.78 and standard deviation = 11% = 0.11.

Thus, $z = \frac{X-\mu}{\sigma} = \frac{X-0.78}{0.11}$

For X = 0.9, we write $z = \frac{0.9-0.78}{0.11} = 1.09$

$P(x > 0.9) = 1 - P(X \leq 0.9) = 1 - P(z \leq 1.09)$

$= 1 - 0.86214 = 0.13786$

5. X is a normal variate with mean 30 and standard deviation 5. Find the probabilities that (i) $26 \leq X \leq 40$ (ii) $X \geq 45$ (iii) $|X - 30| > 5$.

Solution: Given, mean = 30 and standard deviation = 5

Thus, $z = \frac{X-\mu}{\sigma} = \frac{X-30}{5}$

(i) For X = 26, we get $z = \frac{26-30}{5} = -0.8$ and

For X = 40, we get $z = \frac{40-30}{5} = 2$

Therefore, $P(26 \leq X \leq 40) = P(-0.8 \leq z \leq 2)$

$= F(2) - F(-0.8) = 0.97725 - 0.21186$

$= 0.76539$

(ii) For X = 45, we get $z = \frac{45-30}{5} = 3$

$P(X \geq 45) = 1 - P(X \leq 45) = 1 - P(z \leq 3)$

$= 1 - F(3) = 1 - 0.99865 = 0.00135$

(iii) $P(|X - 30| > 5) = 1 - P(|X - 30| \leq 5)$

$= 1 - P(-5 \leq X - 30 \leq 5)$

$= 1 - P(25 \leq X \leq 35)$

$$= 1 - P(-1 \leq z \leq 1)$$

$$= 1 - (F(1) - F(-1))$$

= 1- (0.84134 - 0.15866) = 0.31732

Exercise:

1. In a test of 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average life of 2040 hours and standard deviation of 60 hours. Estimate the number of bulbs likely to burn for (i) more than 2150 hours (ii) less than 1950 hours (iii) more than 1920 hours but less than 2060 hours.
2. Assume that the reduction of a person's oxygen consumption during a period of transcendental meditation (T M) is a continuous random variable X normally distributed with mean 37.6 cc/min and standard deviation 4.6 cc/min. Determine the probability that during a period of T M a person's oxygen consumption will be reduced by (i) at least 44.5 cc/min (ii) at most 35 cc/min (iii) anywhere from 30 cc/min to 40 cc/min/
3. An analog signal received at a detector (measured in micro volts) may be modeled as a Gaussian random variable N (200, 256) at a fixed point in time. What is the probability that the signal will exceed 240 micro volts? What is the probability that the signal is larger than 240 micro volts, given that it is larger than 210 micro volts.
4. In an examination it is laid down that a student passes if he secures 30 percent or more marks. He is placed in the first, second or third division according as he secures 60% or more marks, between 45% to 60% marks and marks between 30% and 45% respectively. He gets distinction in case he secures 80% or more marks. It is noticed from the result that 10% of the students failed in the examination, whereas 5% of them obtained distinction. Calculate the percentage of students placed in the second division. (Assume normal distribution of marks).

Answers: 1. 0.0336 (67 bulbs), 0.0668 (134 bulbs), 0.6065 (1213). 2. 0.0668, 0.2877, 0.649
3. 0.0062, 0.02335 4. 34%.