Basic Concepts of Probability

Events and Outcomes
The result of an experiment is called an **outcome**. An **event** is any particular outcome or group of out comes.
A **simple event** is an event that cannot be broken down further. The **sample space** is the set of all possible simple events.
Basic Probability
Given that all outcomes are equally likely, we can compute the probability of an event *E* using this formula:

$$P(E)=\frac{\text{Number of outcomes corresponding to the event E}}{\text{Total number of equally - likely outcomes}}$$

Cards: A standard deck of 52 playing cards consists of four suits (hearts, spades, diamonds and clubs). Spades and clubs are black while hearts and diamonds are red. Each suit contains 13 cards, each of a different rank: an Ace (which in many games functions as both a low card and a high card), cards numbered 2 through 10, a Jack, a Queen and a King.
Complement of an Event: The complement of an event is the event "*E* doesn't happen".
The notation $\bar{E}$ is used for the complement of event E we can compute the probability of the complement using $P\left(\bar{E}\right)=1-P(E)$

Independent Events : Events A and B are independent events if the probability of Event B occurring is the same whether or not Event A occurs *P*(*A* and *B*) for independent events
If events *A* and *B* are independent, then the probability of both *A* and *B* occurring is

$P(A \text{ and } B) = P(A) \cdot P(B)$

where *P*(*A* and *B*) is the probability of events *A* and *B* both occurring, *P*(*A*) is the probability of event *A* occurring, and *P*(*B*) is the probability of event *B* occurring *P*(*A* or *B*).
The probability of either *A* or *B* occurring (or both) is **$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$**
Conditional Probability:
The probability the event *B* occurs, given that event *A* has happened, is represented as $P(B\,|\,A)$
This is read as "the probability of *B* given *A*" .
If Events *A* and *B* are not independent, then **$P(A \text{ and } B) = P(A) \cdot P(B\,|\,A)$**
Bayes' Theorem :

$$P(A\,|\,B)=\frac{P(A)P(B\,|\,A)}{P(A)P(B\,|\,A)+P(\bar{A})P(B\,|\,\bar{A})}$$

Experiment : An experiment is a test or series of tests in which purposeful changes are made to the input variables of a process or system so that we may observe and identify reasons for changes in the output response.
Random Experiment : A random experiment is one whose outcome cannot be predicted with certainty.
Random Variable : A random variable is a numerical description of the outcome of an experiment.
In other words : A random variable is a numerical variable whose measured value is determined by chance.
OR A random variable is a real valued function having domain as the sample space associated with a given random experiment.
Note: We will denote a random variable with an uppercase letter, such as X, and a measured value of the random variable with a lowercase letter, such as x.
Types of Random Variable:
There are two common types of random variables. They are:
(a) **Discrete random variable:** a quantity assumes either a finite number of values or an infinite sequence of values, such as 0, 1, 2, ...
or a variable, when real valued function defined on a discrete sample space is called a discrete random variable.
Example: The marks obtained in a paper , number of telephone calls per unit time , number of success in n-trails
(b) **Continuous random variable:** a quantity assumes any numerical value in an interval or collection of intervals,

Or A random variable X is said to be continuous if it takes aa possible values between certain limits. **Example**: time, weight, distance, and temperature.

PROBABILITY MASS FUNCTION: Let X is a discrete random variable. A probability mass function (*p.m.f.*) is given by

(a) $P\left(X=a_i\right)=f(a_i)\geq 0,$ for every *i*.

(b) $\sum_{i=1}^{\infty} f\left(a_i\right)=f\left(a_1\right)+f\left(a_2\right)+\cdots+f\left(a_n\right)+\cdots=1$

Or $f(x)=P\{x:X(x_i)=x\}$

Example : Let X be the number of heads , Then P.m.f.

Number of heads X={0,1}	**Elementary events E**	**probability mass function (*p.m.f.*) *f(x)=P(X=x)***
x=0	T	1/2
x=1	H	1/2

Example: Toss a balanced coin twice. Let X be the number of heads . Find the probability mass function of X.

Solution.: Random variable $X=\{x=0,1,2\}$, Sample Space $S=\{HH,TT,HT,TH\}$ $n(S)=4$

Number of heads X={0,1,2}	**Elementary events E**	**probability mass function (*p.m.f.*) *f(x)=P(X=x)***
x=0	TT	¼
x=1	HT,TH	2/4=1/2
X=2	HH	1/4

PROBABILITY DISTRIBUTION (OR DENSITY) FUNCTION (P.D.F)): [RGPV 2012]

A function which describes, how probabilities are distributed over the values of the random variable is called distributive function.

i.e Let *f(x)* is probability function then the probability density / probability distribution function is the function which represent the probabilities which lies in the given interval [*a*,*b*], and defined by $P\left(a\leq x\leq b\right)=\int_a^b f\left(x\right)dx$.

Types of Distribution function:

(1) **Discrete probability distributions**: Discrete probability distributions are used when the sampling space is discrete but not countable. Following is a list of discrete probability distributions:

- discrete uniform
- binomial and multinomial
- hypergeometric
- negative binomial
- geometric
- Poisson

➢ **Required conditions for a discrete probability distribution function:**

Let $a_1,a_2,\ldots,a_n,\ldots$ be all the possible values of the *discrete* random variable *X*. Then, the required conditions for $f(x)$ to be the discrete probability distribution for *X* is

(a) $P\left(X=a_i\right)=f(a_i)\geq 0,$ for every *i*. (b) $\sum_{i=1}^{\infty} f\left(a_i\right)=f\left(a_1\right)+f\left(a_2\right)+\cdots+f\left(a_n\right)+\cdots=1$

(2) Continuous probability distribution: Continuous probability distribution is used when the sample space is continuous. Following is a list of continuous probability distributions:

- Uniform
- Normal (or Guassian)
- Gamma
- Beta
- t distribution
- F distribution

- χ^2 distribution

- **Required conditions for a continuous probability density:**

Let the *continuous* random variable *Z* taking values in subsets of $(-\infty,\infty)$. Then, the required conditions for $f(x)$ to be the continuous probability density function for *Z* are

(a) $f(x)\geq 0, \quad -\infty < x < \infty.$ (b) $\int_{-\infty}^{\infty} f(x)dx = 1$

Example : Whether check the following function is p.d.f? $f(x)=6x(1-x), 0\leq x\leq 1$

Solution : Since $f(x)=6x(1-x)\geq 0$ in $0\leq x\leq 1$ and $\int_0^1 f(x)dx=\int_0^1 6x(1-x)dx=1$, hence function isp.d.f..

Remark : for given distribution function p.d.f = $\boxed{f(x)=\frac{d}{dx}F(x)}$, where F(x)= distribution function.

- **Mean / Arithimic mean (or Expected Value) :**

If *X* is discrete, $E(X)=\sum_{i=1}^{\infty} a_i f(a_i) = a_1 f(a_1)+a_2 f(a_2)+\cdots+a_n f(a_n)+\cdots$

If *X* is continuous, $E(X)=\int_{-\infty}^{\infty} xf(x)dx$

- **Variance:**

If *X* is discrete, $\boxed{Var(X)=\sigma^2=E\left[X-E(X)\right]^2=\sum_i (a_i-\mu)^2 f(a_i)}$

$$=(a_1-\mu)^2 f(a_1)+(a_2-\mu)^2 f(a_2)+\cdots+(a_n-\mu)^2 f(a_n)+\cdots$$

If *X* is continuous $\boxed{Var(X)=\sigma^2=E\left[X-E(X)\right]^2=\int_{-\infty}^{\infty}(x-u)^2 f(x)dx}$

Note: In practice, it is easier to use the computational formula for the variance, rather than the defining formula:

$$\sigma^2=E\left[X^2\right]-\mu^2=\int_{-\infty}^{+\infty} x^2 f(x)dx-\mu^2 .$$

Moments about origin:

First moment about origin = Mean : $\mu_1'=\int_{-\infty}^{\infty} xf(x)dx$

- **r^th^ moment about origin :** $\mu_r'=\int_{-\infty}^{\infty} x^r f(x)dx$
- **First moment about mean =** $\mu_1=0$
- **Variance= Second moment about mean =** $\mu_2=\mu_2'-(\mu_1')^2$
- **r^th^ moment about mean :** $\int_a^1 3x^2 dx=0.05$
- **Median:** Median is the line which divide the whole area under the curve in to two equal parts .

  If M_d is median then $\boxed{\int_a^{m_d} f(x)dx=\frac{1}{2} \text{ or } \int_{m_d}^{b} f(x)dx=\frac{1}{2}}$
- **Mode : Mode is the value of x for which *f(x)* is maximum.**
- **Mean deviation from mean:** $M.D.=\int_a^b |x-\bar{x}|\, f(x)dx$

Remark: For symmetric distribution mean, mode and median coincides at origin. i.e. mean=mode=median

- **DISTRIBUTION FUNCTION or CUMULATIVE DISTRIBUTION FUNCTION (OR C.D.F.)** :

1. Let X be a discrete random variable , then distributive function of X is

$$\boxed{F(x)=P(X\le x)=\sum_{x_i\le x} p_i} \text{,such that } p_i\ge 0 \text{ and } \sum_{i=1} p_i=1\text{, where } p=p(x_i)$$

2. Let X be continuous random variable then Cumulative distribution function is given by

$$\boxed{F(x)=P(X\le x)=\int_{-\infty}^{x} f(x)dx}\text{, for all } x\in\Re\ .$$

If the distribution does not have a *p.d.f.*, we may still define the *c.d.f.* for any x as the probability that X takes on a value no greater than x.

Note: The *c.d.f.* for the distribution of a random variable is unique, and completely describes the distribution.

Discrete Probability distribution :

Example: A random variable X has the following probability function, find the distributive function

X=x	0	1	2	3
p(x)	0	1/5	2/5	2/5

Solution: $F(x)=P(X=x)=\sum_{x_i\le x} p_i$, $F(0)=P(X\le 0)=p(0)=0$,

$F(1)=P(X\le 1)=p(0)+p(1)=0+1/5=1/5$, $F(2)=P(X\le 2)=p(0)+p(1)+p(2)=0+1/5+2/5=3/5$,

$F(3)=P(X\le 3)=p(0)+p(1)+p(2)+p(3)=0+1/5+2/5+2/5=5/5=1$

Q 1. What do you mean by Probability density function ? **[May 2019]**

Q 2. The **P.D.F.** of a random variable X is given by $p(x)=\begin{cases} k, & if\ x=0 \\ 2k, & if\ x=1 \\ 3k, & if\ x=2 \\ 0 & Otherwise \end{cases}$

(a) Find the value of k (b) Evaluate P(X<2) , P(X≤2), P(0<X<2) **[Ans: (a)** $\sum_{x_i\le x} p_i=1\Rightarrow k=1/6$ (b)1/2,11/3,

---

Continuous Probability distribution (value of function is given in certain intervals):

Q 3. If $f(x)=Kx^2$, $0<x<1$ is probability density function (p.d.f) ,(i) determine K (ii) find $P(\frac{1}{3}<x<\frac{1}{2})$ (iii) $P(X>a)$.[**Ans** (i) K=3, (ii) $P(\frac{1}{3}<x<\frac{1}{2})=\int_{1/3}^{1/2}3x^2dx=\frac{19}{216}$ (iii) given $P(X>a)=0.05$, $\int_a^1 3x^2dx=0.05$ then a=$(0.95)^{1/3}$

Q 4. The probability density function $p(x)$ of a random variable is given by $p(x)=Ce^{-x}$,$0\le x<\infty$. Find the value of C. [**Ans** $\int_{-\infty}^{\infty} f(x)dx=1\Rightarrow\int_0^{\infty} Ce^{-x}dx=1\Rightarrow C=1$

Q 5. A random variable X has the **P.D.F.** $f(x)=\begin{cases} kx^2, & 0\le x\le 3 \\ 0, & otherwise \end{cases}$ .Find **the value of *k***. Also compute P(1≤x≤2).

[**May 2018 EC]**

Q 6. The **P.D.F.** of a random variable X is given by $p(x)=\begin{cases} ax, & 0\le x\le 1 \\ a, & 1\le x\le 2 \\ -ax+3a, & 2\le x\le 3 \\ 0 & Otherwise \end{cases}$

(a) Find the value of "*a*" (b) Evaluate P(X<1.5)

[**Ans:** (a) $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_{-\infty}^{0} 0dx + \int_{0}^{1} ax\,dx + \int_{1}^{2} a\,dx + \int_{2}^{3}(-ax+3a)\,dx = 1 \Rightarrow a = \frac{1}{2}$

(b) $P(X \le 1.5) = \int_{-\infty}^{1.5} f(x)dx = \int_{0}^{1} ax\,dx + \int_{1}^{1.5} a\,dx = a = \frac{1}{2}$

Q 7. A random variable X has the following **probability function**. Determine the **distributive function.**

X=x	0	1	2	3
P(x)	0	1/5	2/5	2/5

Q 8. A random variable X has the following probability function/ density function $f(x) = \begin{cases} \frac{1}{9}x^2, & 0 \le x \le 3 \\ 0, & Otherwise \end{cases}$

find distributive function.

[**Hint**: $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow \int_{0}^{\infty} Ce^{-x}dx = 1 \Rightarrow C = 1$ Hence D.F. $F(x) = \begin{cases} 0; & x < 0 \\ \frac{x^3}{27}; & 0 \le x \le 3 \\ 1; & x \ge 3 \end{cases}$

Q 9. If $f(x) = \frac{c}{1+x^2}, -\infty < x < \infty$, then find "*c*" and show that its corresponding **distribution function** is $F(x) = (\frac{1}{\pi}\tan^{-1}x) + \frac{1}{2}$.

Q 10. The **distributive function** of a random variable X is given by $F(x) = \begin{cases} 0, & x \le 1 \\ \frac{(x-1)^4}{16}, & 1 \le x \le 3 \\ 1, & x > 3 \end{cases}$

then (i) Find the corresponding **Probability density function** of variable *X, (ii) Compute P(2<x≤3).*

[**Hint**: (i) p.d.f = $\boxed{f(x) = \frac{d}{dx}F(x)}$, where F(x)= distribution function then

pdf= $F(x) = \begin{cases} \frac{d}{dx}0, & x \le 1 \\ \frac{d}{dx}\frac{(x-1)^4}{16}, & 1 \le x \le 3 \\ \frac{d}{dx}1, & x > 3 \end{cases} = \begin{cases} 0, & x \le 1 \\ 1/4(x-1)^3, & 1 \le x \le 3 \\ 0, & x > 3 \end{cases}$ (ii) $P(2 < x \le 3) = \int_{2}^{3} f(x)dx = \int_{2}^{3} 1/4(x-1)^3 dx = 15/16$

Q 11. A random variable X has the **P.D.F.** $f(x) = \begin{cases} x, & 0 \le x \le 1 \\ 2-x, & 1 \le x \le 2 \\ 0, & x \ge 2 \end{cases}$ .Find **cumulative distribution function** of *X*.

Mean, Median and Mode:

Q 12. Define Mean, Median, Mode, Variance and Standard deviation for P.D.F..

Q 13. A frequency distribution is defined by the following function, Prove that f(x) is P.D.F. . Also find Mean and Standard Deviation. $f(x) = \begin{cases} x^3, & 0 \le x \le 1 \\ 3(2-x)^3, & 1 \le x \le 2 \end{cases}$

Q 14. Determine the value of k so that the following function represents the P.D.F. Also find Median.

$$f(x)=\begin{cases}0, & x\le -1\\ k(x+1), & -1\le x\le 3\\ 4k, & 3\le x\le 4\\ 0 & Otherwise\end{cases}$$

Q 15. For the distribution function $dF = y_0 e^{-|x|}\,dx\,,\ -\infty < x < \infty$. Prove that $y_0=1/2$, Mean=0, S.D.=$\sqrt{2}$, Variance =2, and mean deviation about mean is 1. **[June 14]**

Q 16. For the distribution function $dF = y_0 e^{-x/\sigma}dx\,,\ 0\le x<\infty$. Find Mean, S.D., Variance and r^{th} moment about the origin.

Q 17. For the frequency of probability curve $y=\frac{1}{2}\sin x,\ 0\le x\le\pi$, Find **median and mode**.

Q 18. For the distribution $dF=\sin x,\ 0\le x\le\frac{\pi}{2}$, Find **Mode , Mean and Variance..**

Q 19. For the distribution function $dF=6(x-x^2)dx\,,\ 0\le x\le 1$, find Arithmetic men, Harmonic mean, Median, Mode, Mean deviation and S.D.. Is it a symmetrical distribution?

Q 20. For the Beta distribution $dF=\frac{1}{\beta(m,n)}x^{n-1}(1-x)^{m-1}dx\,, 0\le x\le 1, m>0$. Find the Mean, S.D., Harmonic Mean and μ'_r.

Q 21. Define Theoretical distribution. Write the types of theoretical distribution..

3.5 PROBABILITY DISTRIBUTIONS

An example will make clear the relationship between random variables and probability distributions. Suppose you throw a coin two times. This simple statistical experiment can have four possible outcomes: HH, HT, TH, and TT. Now, let the variable *X* represent the number of Heads that result from this experiment. The variable *X* can take on the values 0, 1, or 2. In this example, *X* is **a random variable;** because its value is determined by the outcome of a statistical experiment.

A **probability distribution** is a table or an equation that links each outcome of a statistical experiment with its probability of occurrence. Consider the coin throw experiment described above. The table below, which associates each outcome with its probability, is an example of a probability distribution.

Number of heads	Probability
0	0.25
1	0.50
2	0.25

The above table represents the probability distribution of the random variable X.

- **Cumulative Probability Distributions:**

A **cumulative probability** refers to the probability that the value of a random variable falls within a specified range.

Let us return to the coin throw experiment. If we throw a coin two times, we might ask: What is the probability that the coin throws would result in one or fewer heads? The answer would be a cumulative probability. It would be the probability that the coin throw experiment results in zero heads plus the probability that the experiment results in one head.

$$P(X < 1) = P(X = 0) + P(X = 1) = 0.25 + 0.50 = 0.75$$

Like a probability distribution, a cumulative probability distribution can be represented by a table or an equation. In the table below, the cumulative probability refers to the probability than the random variable X is less than or equal to x.

Number of heads: x	Probability: $P(X = x)$	Cumulative Probability: $P(X < x)$
0	0.25	0.25
1	0.50	0.75
2	0.25	1

- **Uniform Probability Distribution:**

The simplest probability distribution occurs when all of the values of a random variable occur with equal probability. This probability distribution is called the **uniform distribution**.

Uniform Distribution. Suppose the random variable X can assume k different values. Suppose also that the $P(X = x_k)$ is constant. Then,

$$P(X = x_k) = 1/k$$

Example: Suppose a die is tossed. What is the probability that the die will land on 6 ?

Solution: When a die is tossed, there are 6 possible outcomes represented by: $S = \{1, 2, 3, 4, 5, 6\}$. Each possible outcome is a random variable (X), and each outcome is equally likely to occur. Thus, we have a uniform distribution. Therefore, the $P(X = 6) = 1/6$.

Example: Suppose we repeat the dice tossing experiment described in Example 1. This time, we ask what is the probability that the die will land on a number that is smaller than 5?

Solution: When a die is tossed, there are 6 possible outcomes represented by: $S = \{1, 2, 3, 4, 5, 6\}$. Each possible outcome is equally likely to occur. Thus, we have a uniform distribution.

This problem involves a cumulative probability. The probability that the die will land on a number smaller than 5 is equal to:

$$P(X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) = 1/6 + 1/6 + 1/6 + 1/6 = 2/3$$

Three main types of probability distributions are discussed in next section.

➢ BINOMIAL DISTRIBUTION

To understand binomial distributions and binomial probability, it helps to understand binomial experiments and some associated notation; so we cover those topics first.

Binomial Experiment: A **binomial experiment** (also known as a **Bernoulli trial**) is a statistical experiment that has the following properties:

- The experiment consists of n repeated trials.
- Each trial can result in just two possible outcomes. We call one of these outcomes a success and the other, a failure.
- The probability of success, denoted by P, is the same on every trial.
- The trials are independent; that is, the outcome on one trial does not affect the outcome on other trials.

Consider the following statistical experiment. You throw a coin 2 times and count the number of times the coin lands on heads. This is a binomial experiment because:

- The experiment consists of repeated trials. We throw a coin 2 times.
- Each trial can result in just two possible outcomes - heads or tails.
- The probability of success is constant - 0.5 on every trial.
- The trials are independent; that is, getting heads on one trial does not affect whether we get heads on other trials.

Notations:

The following notation is helpful, when we talk about binomial probability.

- x: The number of successes that result from the binomial experiment.

- *n*: The number of trials in the binomial experiment.
- *p*: The probability of success on an individual trial.
- *q*: The probability of failure on an individual trial. (This is equal to 1 - *p*.)
- b(*x*; *n*, *p*): Binomial probability - the probability that an *n*-trial binomial experiment results in exactly *x* successes, when the probability of success on an individual trial is *p*.
- ${}^nC_r$: The number of combinations of *n* things, taken *r* at a time.

Binomial Distribution: Binomial distribution is a discrete probability distribution. A random variable is said to follow binomial distribution if it takes non negative values and its probability mass function is given by

$$P(x=r)={}^nC_r p^r q^{n-r} = \frac{n!}{r!\ n-r!} p^r q^{n-r}\ ,\ r=0,1,2,3.....$$

If an experiment is conducted in N- sets then , No. of *r*-Success in *n*- trails (or Frequency of success)=

$$N.P(x=r)={}^nC_r p^r q^{n-r} = \frac{n!}{r!\ n-r!} p^r q^{n-r}\ ,\ r=0,1,2,3.....$$

Suppose we throw a coin two times and count the number of heads (successes). The binomial random variable is the number of heads, which can take on values of 0, 1, or 2. The binomial distribution is presented below.

Number of heads	Probability
0	0.25
1	0.50
2	0.25

The binomial distribution has the following properties:

The mean of the binomial distribution $\mu_{x=}\ np$.
The variance $\sigma_{2x=}\ n\,pq$.
The standard deviation $\sigma_x=\sqrt{n\,pq}$.

Binomial Probability: Suppose a binomial experiment consists of *n* trials and results in *x* successes. If the probability of success on an individual trial is P, then the binomial probability is:

$$b(x;\ n,\ P)\ =\ {}^nC_x\ p^x\ q^{n-x} = \frac{n!}{(n-x)!x!} p^x q^{n-x}\ \ for\ x=0,1,2,...,n$$

Or Probability of r – success in n-trails $P(X=r)=\ {}^nC_r\ p^r\ q^{n-r}$

Hypothesis of Binomial distribution :

1. The procedure has a **fixed number of trials**. [n trials]
2. The trials must be **independent**.
3. Each trial is in **one of two mutually exclusive categories**.
4. The **probabilities remain constant** for each trial.

Q 22. Define Binomial Distribution/ Binomial Theorem **[RGPV. June 2004 / May 2019]**

Q 23. Find the mean of binomial distribution.
[June 2004,07,09,10,11,,Dec.04,07,10,.Feb. 08,April 09, June 2012,Dec.2014,2015, Nov.18, May 2018 EC]

Q 24. Find the variance of Binomial Distribution. **[RGPV. Feb. 08, June 09, 10,11,12,Dec. 2015, May 2018 EC]**

Q 25. Find the Standard Deviation of Binomial Distribution. **[RGPV. June 04,Dec.04,10,Feb. 08,April 09]**

Q 26. A coin is tossed 4 times , what is the probability to getting(i) Two heads(ii) at least two heads.
Ans: (i)3/8, (ii) 11/16 **[May 2018 EC]**

Q 27. Six dice are thrown 729 times. How many times do you expect **at least three** dice to show five or six?
Ans: *p= 233/729 , No. of Dice= 233*

Q 28. A perfectly cubical dice is thrown a large number of times in sets of 8. The occurrence of 5 or 6 is called a success. In what proportion of the sets do you expect 3 successes? **Ans: 27.31%**

Q 29. The incidence of occupational disease in an industry is such that the workers have a 20% chance of suffering from it. Find the chance that out of 6 workers 4 or more will be catch by the disease. **Ans: 0.01696**

Q 30. If the 10% of the bolts produces by a machine are defective, find the probability that out of 5 bolts chosen at

Random at least two will be defective. **[June2014, May 2019]**

Q 31. In sampling a large number of parts manufactured by a machine the mean number of defectives in a sample of 20 is 2. Out of 1000 such samples how many would be expected to contain (i) At least 3 defective parts (ii) none defective. **Ans.** (i) **323 (ii) 122**

Q 32. Assuming that half of the population is consumer of chocolate. Show that the chance of an individual being a consumer is ½ and assuming that 100 investigators each takes 10 individuals to see whether they are consumers, how many investigators would you expect to report that three people or less were consumers? **Ans: 17**

Q 33. A bag contains 3 red and 4 black balls, one ball is drawn and replaced in the bag and the process is repeated. Getting a red ball in a draw is considered a success. Find the distribution of X. Where X denotes the number of success in 3 drawn, assuming that in each draw each ball is exactly likely to be selected.
Ans: $(4/7)^3$, $9/7(4/7)^2$, $12/7(3/7)^2$, $(3/7)^3$

Q 34. Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) at least one boy (iii) no girl(iv) at most 2 girls? Assume equal probability for boys and girls.
Ans: (i) 300(ii) 750 (iii) 50 (iv) 550

Q 35. Out of 800 families with 5 children each, how many families would be expected to have (i) 3 boys (ii) at least one boy (iii) 5 girls (iv) either 2 or 3 boys ? Assume equal probability for boys and girls.
Ans: (i) 250 (ii) 25 (iii) 500

Q 36. The probability that a bomb dropped from a place will strike the target is 1/5 . If six bombs are dropped, find the probability that (i) exactly two will strike the target (ii) at least 2 will strike the target.
Ans: 0.246, 0.345 **[June 13]**

Q 37. In a binomial distribution the mean is 12 and S.D. is 2 find *n* and *p*. **[Nov.18 EC, May 2019 Ex.]**

Q 38. Criticize the following statement: "For binomial distribution Mean=5 and S.D.=3". Ans: Wrong

Q 39. Find the Binomial Distribution, whose mean=4 and variance is 3. Also find Mode. Ans: $(3/4+1/4)^{16}$, 4

Q 40. The mean and variance of binomial distribution are 4 and 4/3 respectively. Find $P(X\geq 2)$ and the probability of 2 successes. Also find $P(X>2)$. **Ans: 20/243 , 716/729 , 656/729, 656/729**

Q 41. Fit a Binomial Distribution for the following data and compare the theoretical with the actual ones.

x	0	1	2	3	4	5	**[RGPV. June 2002,06]**
f	2	14	20	34	22	8	***Ans: f: 2,10,26,34,22,6***

Q 42. The following data are the number of seeds germinating out of 10 on a damp filter paper for 80 sets of seeds. Fit a binomial distribution.

x	0	1	2	3	4	5	6	7	8	9	10	Total
y	6	20	28	12	8	6	0	0	0	0	0	80

Ans: f: 6.9, 19.1, 24 , 17.8 ,8.6, 2.9, 0.7, 0.1, 0,,0,0,0

Poisson distribution:

Definition: The **Poisson distribution** is a discrete probability distribution of a random variable *x* that satisfies the following conditions.

1. The experiment consists of counting the number of times, *x*, an event occurs in a given interval. The interval can be an interval of time, area, or volume.
2. The probability of two or more success in any sufficiently small subinterval is 0. For example, the fixed interval might be any time between 0 and 5 minutes. A subinterval could be any time between 1 and 2 minutes.
3. The probability of the event occurring is the same for any two intervals of equal length.
4. The number of occurrences (success) in any interval is independent of the number of occurrences in any other interval provided the intervals are not overlapping.

NECESSARY CONDITIONS FOR POISSON DISTRIBUTION:

Poisson distribution is a discrete probability distribution, which is the limiting case of the binomial distribution under certain conditions.

1. When n is very indefinitely very large
2. Probability of success is very small.
3. $np = \lambda$ is finite, $\lambda \in R^+$

Def: A discrete random variable X is said to be follow a Poisson distribution if the probability mass function is given by

$$p(X = x) = P(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2,3.........\infty \quad \text{Where e = 2.7183 and } \lambda > 0$$

Here λ is called the ***parameter*** of the Poisson distribution.

➢ **Examples where the Poisson distribution is used (or) Applications of Poisson distribution:**
This distribution is used to describe the behavior of the rare events like

1. The number of blind born per year in a large city.
2. The number of printing mistakes per page in a large volume of a book.
3. The number of air pockets in a glass sheet.
4. The number of accidents occurred annually at a busy crossing of city.
5. The number of defective articles produced by a quality machine.
6. This is widely used in waiting lines or queuing problems in management studies.
7. It has wide applications in industrial quality control.
8. In determining the number of deaths in a given period by a rare disease.

For a Poisson distribution the probability mass function is given by

$$\boxed{p(X = x) = P(x;\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0,1,2,3.........\infty}$$

Example: 1) Number of printing mistakes on each page of a book published by a good publisher
2) Number of telephone calls arriving at a telephone switch board per minute.

Q 1. Define Poisson distribution. Write the conditions for Poisson distribution.

Q 2. Find the Mean, Variance and S.D. of Poisson distribution. **[Nov.18EC, May 2019(Ex.)]**

Q 3. Show that in a poisons distribution with unit mean, the mean deviation about the mean is 2/e times of the S.D..

Q 4. For a Poisson distribution prove that $M\sigma\gamma_1\gamma_2 = 1$

Q 5. For a Poisson distribution prove that $\sqrt{\beta_1}\,(\beta_2 - 3)m\sigma = 1$.

Q 6. Fit a Poisson distribution for the following data.

x	0	1	2	3	4	**[June 2007, Dec. 2007]**
f	122	60	15	2	1	***Ans: f:122,61,15,2,0***

Q 7. Fit a Poisson Distribution for the following data .

x	0	1	2	3	4	**[June 2008, Feb. 2010, June 2013]**
f	46	38	22	9	1	***Ans: f:44,43,21,7,2***

Q 8. Fit a Poisson distribution for the following data.

x	0	1	2	3	4	5	6	7	8	9	10	***Ans:*****104,141,93,41,14,4,1,0,0,0,0**
f	103	143	98	42	8	4	2	0	0	0	0	**[RGPV. Sept. 2009]**

Q 9. A manufacturer, who produces medicine bottles, finds that 0.1% of the bottles are defective. The bottles are packed in boxes containing 500 bottles. A drug manufacturer buys 100 boxes from the producer of bottles. Using Poisson distribution find how many boxes will contain no defectives. **[Dec.2015]**

Q 10. Find the probability that at most 5 defective fuses will be found in a box of 200 fuses, if experiences shows that 2 % of such fuses are defective. **Ans: 0.7845** **[Dec.2012, Nov. 2018 , May 2018 EC]**

Q 11. If 3% of the electric bulbs manufactured by a company are defective, find the probability that in a sample of 100 bulbs exactly 5 bulbs are defective. **[Nov. 18 EC]**

Q 12. If there are 3 misprint in a book of 1000 pages , find the probability that a given page will contain

(i) No mis print (ii) more than 2 mis print. **[May 2018 EC]**

Q 13. In a certain factory turning razor blades there is a small chance of 0.002 for any blade to be defective. The blades are in packets of 10. Use poisons distribution to calculate the approximately number of packets containing (i) No defective (ii) One defective (iii) Two defective, blades in a box of 50,000 packets.
Ans: (i)49010(ii) 980(iii)10 [June 2006]

Q 14. The probability that an evening college will graduate is 0.4. Determine the probability that out of 5 students. (i) None (ii) One and (iii) At least one will graduate. **[Dec.2015]**

Q 15. The mean height of 500 students in 151 cm and the standard deviation is 15 cm. assuming that the heights are normally distributed. Find how many students have heights between 120 and 155 cm. **[June2015]**

Q 16. If the probability of a bad reaction from certain injection is 0.001. Determine the chance that out of 2000 individuals (i) 3 (ii) more than 2 (iii) none (iv) more than 1 (v) more than two, individuals will get a bad reaction. **Ans: (i)0.180(ii)0.323 (iii)0.135 (iv)0.594** **[Dec.14]**

Q 17. A car hire firm has two cars, which it hires out day by day. The number of demands for a car on each day is distributed as a Poisson distribution with mean 1.5 . Calculate the population of days on which neither car is used and the proportion of days on which some demand is refused**.**

Ans: (i)0.2231 (ii)0.20[June 2003, Dec. 2006]

Q 18. A telephone switch handles 600 calls on the average during a rush hour. The board can make a maximum 20 connections per minute. Use poissons distribution to estimate the probability that the board will be over-taxed during any given minute. **Ans :** $P(r>20)=1-P(r\le 20)=1-e^{-10}\sum_{r=0}^{20}\frac{m^r}{r!}$

Q 19. The randam variable X has a Poisson distribution if P (X=1)=0.01487 , P (X=2)= 0.04461 then find P(X)=3. [**May 2019]**

Normal Distribution: Defn: A random variable X is said to be <u>normally distributed</u> or to have a <u>normal distribution</u> if its p.d.f has the form

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

Here μ and σ are the parameters of the distribution; μ = the mean of the random variable X (or of the probability distribution); and σ = the standard deviation of X.

<u>Note</u>: The normal distribution is not just a single distribution, but rather a family of distributions; each member of the family is characterized by a particular pair of values of μ and σ.

The graph of the p.d.f. has the following characteristics:

1) It is a bell-shaped curve;
2) It is symmetric about μ;
3) The inflection points are at $\mu - \sigma$ and $\mu + \sigma$.

Importance of Normal Distribution:

The normal distribution is very important in statistics for the following reasons:

1) Many phenomena occurring in nature or in industry have normal, or approximately normal, distributions.
<u>Examples</u>: a) heights of people in the general population of adults;

b) for a particular species of pine tree in a forest, the trunk diameter at a point 3 feet above the ground;

c) fill weights of 12-oz. cans of Pepsi-Cola; d) IQ scores in the general population of adults;

e) diameters of metal shafts used in disk drive units.

2) Under general conditions (independence of members of a sample), the possible values of the sample mean for samples of a given (large) size have an approximate normal distribution (Central Limit Theorem).

<u>The Empirical Rule:</u>

For the normal distribution,

1) The probability that X will be found to have a value in the interval (μ - σ, μ + σ) is approximately 0.6827;

2) The probability that X will be found to have a value in the interval (μ - 2σ, μ + 2σ) is approximately 0.9545;

3) The probability that X will be found to have a value in the interval (μ - 3σ, μ + 3σ) is approximately 0.9973.

Unfortunately, the p.d.f. of the normal distribution does not have a closed-form anti-derivative. Probabilities must be calculated using numerical integration methods. This difficulty is the reason for the importance of a particular member of the family of normal distributions, the standard normal distribution, which has *p.d.f.*

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

Note: For shorthand, we will write X ~ Normal(μ, σ) to mean that the continuous r.v. X has a normal distribution with mean μ and standard deviation σ.
The c.d.f. of the standard normal distribution will be denoted by

$$\Phi(z) = P(Z \le z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{\frac{-w^2}{2}} dw .$$

Values of this function have been tabulated in Table 1 of Appendix A.

Q 20. Define Normal Distribution. [**RGPV. June 2006]**
Q 21. Find the mean and standard deviation of the normal distribution. [**RGPV June 2005]**
Q 22. Write the basic properties and standard form of normal distribution. **[RGPV. June 2012]**
Q 23. Prove Mean, Mode and Median of normal distribution coincides at origin.
Q 24. Define normal curve .Find the point of inflexion of normal curve.
Q 25. Prove that the mean deviation from mean of a normal distribution is approximately 4/5 times of standard deviation. [**RGPV. June 2006]**
Q 26. For the normal curve $y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-m)^2/2\sigma^2}$ .Find (i) Mean and S.D. (ii) Point of inflexion**[June2001]**
Q 27. The life of army shoes is normally distributed with mean 8 month and S.D. 2 month. If 5000 pairs are issued , howmanny pairs would be expected to need repladcement after 12 month? Given that P(z≥2)=0.0228.
Ans: 4886 **[Jan. 2007]**
Q 28. The mean height of 500 students in 151 cm and the standard deviation is 15 cm. assuming that the heights are normally distributed. Find how many students have heights between 120 and 155 cm.
Ans : 294 **[June2015]**
Q 29. A sample of 100 dry battery cells tested to find the length of life product the following results: Mean =12 hours, S.D.=3 hours . Assuming that the data to be normally distributed, what percentage of battery cells are expected to have life (i) more than 15 hours (ii) Less than 6 hours (iii) between 10 and 14 Hours?
Ans: (i) 15.87% (ii) 2.28% (iii)49.74%
Q 30. Ten individuals are chosen at random from a population and their heights are found to be in inches 63, 63, 64,65,66,69,70,70,71. Discuss the suggestion that the Mean heights of universe are 65. **[June 2014]**

Beta distribution: It is a continuous distribution.

- It is bounded on both sides. In this respect it resembles the binomial distribution. The standard beta distribution is constrained so that its domain is the interval (0, 1).
- The beta distribution has two parameters *a* and *b* both referred to as shape parameters.
- The formula for the beta density is the following. $f(x;\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma\alpha\Gamma\beta} x^{\alpha-1}(1-x)^{\beta-1} = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$

The reciprocal of the ratio of gamma functions that appears in front as the normalizing constant is generally called the beta function and is denoted *B(α, β).*

- The beta distribution is often used in conjunction with the binomial distribution particularly in Bayesian models where it plays the role of a prior distribution for *p*.

- It also can be used to give rise to a beta-binomial model. Here the probability of success *p* is assumed to arise from a beta distribution and then, given the value of *p*, the observed number of successes has a binomial distribution with parameters *n* and this value of *p*. The significance of this approach is that it allows *p* to vary randomly between subjects and is a way of modeling what's called binomial over dispersion.

Gamma distribution **Gamma Distribution**

Defn: The gamma function is defined by the integral $\Gamma(r) = \int_0^{+\infty} t^{r-1}e^{-t}dt$, for r > 0.

It may be shown using integration by parts that $\Gamma(r) = (r-1)\Gamma(r-1)$. Hence, in particular, if r is a positive integer, $\Gamma(r) = (r-1)!$. We also have $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

Defn: A continuous r.v. X is said to have a gamma distribution with parameters r > 0 and λ > 0 if the p.d.f. of X is

$f(x) = \dfrac{\lambda^r}{\Gamma(r)} x^{r-1}e^{-\lambda x}$, for x > 0, and f(x) = 0, for x ≤ 0.

The mean and variance of X are given by $\mu = E[X] = \dfrac{r}{\lambda}$ and $\sigma^2 = V(X) = \dfrac{r}{\lambda^2}$.

We write X ~ gamma(r, λ) to denote that X has a gamma distribution with parameters r and λ.

It may be easily shown that the integral of the gamma p.d.f. over the interval (0, +∞) is 1, using the definition of the gamma function.

The gamma distribution is very important in statistical inference, both in its own right and because it is the basis for constructing some other distributions useful in inference. For example, the "signal-to-noise" ratio statistic that we will use in analyzing the results of scientific experiments is based on a ratio of random variables which have gamma distributions of a particular form.
The graphs of some gamma p.d.f.'s are shown on p. 72.

Defn: A continuous r.v. X is said to have a chi-squared distribution with k degrees of freedom if X ~ gamma(k, 0.5).

WEIBULL DISTRIBUTION:

Defn: A continuous r.v. X is said to have a Weibull distribution with parameters δ > 0 and β > 0 if the p.d.f. of X is

$f(x) = \dfrac{\beta}{\delta}\left(\dfrac{x}{\delta}\right)^{\beta-1} \exp\left[-\left(\dfrac{x}{\delta}\right)^{\beta}\right]$, for *x* > 0, and $f(x) = 0$, for *x* ≤ 0. The mean and variance of X are

$\mu = E[X] = \delta\Gamma\left(1+\dfrac{1}{\beta}\right)$ and $\sigma^2 = V(X) = \delta^2\Gamma\left(1+\dfrac{2}{\beta}\right) - \mu^2$. We write X ~ Weibull(δ, β).

The c.d.f. for a Weibull(δ, β) distribution is given by $F(x) = 1 - \exp\left[-\left(\dfrac{x}{\delta}\right)^{\beta}\right]$, for *x* > 0, and

F(x) = 0, for *x* ≤ 0.

The Weibull distribution is used to model the reliability of many different types of physical systems. Different combinations of values of the two parameters lead to models with either a) increasing failure rates over time, b) decreasing failure rates over time, or c) constant failure rates over time.

THE UNIFORM DISTRIBUTION

Consider a continuous r.v. X whose distribution has p.d.f. $f(x) = \dfrac{1}{b-a}$, for $a \le x \le b$, and $f(x) = 0$, otherwise.

We say that X has a uniform distribution on the interval (a, b), abbreviated

X ~ Uniform(a, b). If we take a measurement of X, we are equally likely to obtain any value within the interval. Hence, for some subinterval $(c, d) \subseteq (a, b)$, we have $P(c \le x \le d) = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$.

The mean of the uniform distribution is $\mu = \int_{-\infty}^{+\infty} xf(x)dx = \int_a^b \frac{x}{b-a} dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{a+b}{2}$, the midpoint of the interval (a, b).

The second moment of the distribution is

$$E\left[X^2\right] = \int_{-\infty}^{+\infty} x^2 f(x)dx = \frac{1}{b-a}\int_a^b x^2 dx = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)\left(b^2+ab+a^2\right)}{3(b-a)}.$$

Then the variance is

$$\sigma^2 = E\left[X^2\right] - \mu^2 = \frac{b^2+ab+a^2}{3} - \frac{b^2-2ab+a^2}{4} = \frac{(b-a)^2}{12}$$, and the standard deviation is $\sigma = \frac{b-a}{2\sqrt{3}}$.

Note: the longer the interval (a, b), the larger the values of the variance and standard deviation.