MODULE 5: PROBABILITY DISTRIBUTION

CONTENTS:

- Discrete Random variable
- Continuous Random variable
- Mathematical Expectation, Mean, Variance.
- Binomial distributions-problems,
- Poisson distributions-problems
- Normal distributions-problems.

Lab component:

1. Compute Discrete probability distributions and continuous probability distributions using MATLAB.
2. Hands-on random experiments related to Probability distributions.

RBT Level: L1, L2 and L3

Learning Objectives:

- The course will Build a strong foundation in probability theory essential to solve real world random phenomena.

- Illustrate the knowledge of fundamental concepts of Multivariable calculus, Vector differentiation, Multiple Integral, Partial differential equations and Probability

Module Outcomes: - After Completion of this module, student will be able to

- Illustrate the knowledge of fundamental concepts of Probability and distributions

- Apply suitable techniques to solve given engineering and scientific problems related to Probability based on the acquired knowledge.

- Analyse mathematical solutions of engineering and scientific problems related to Probability and predict their behaviour in real-world scenario.

Basic Concepts:

In probability theory,

- an outcome is a possible result of an experiment or trial
- the sample space of an experiment or random trial is the set of all possible outcomes or results of that experiment
- an event is a set of outcomes of an experiment (a subset of the sample space) to which a probability is assigned

- two events (or propositions) are mutually exclusive or disjoint if they cannot both occur at the same time.

Random Variable:

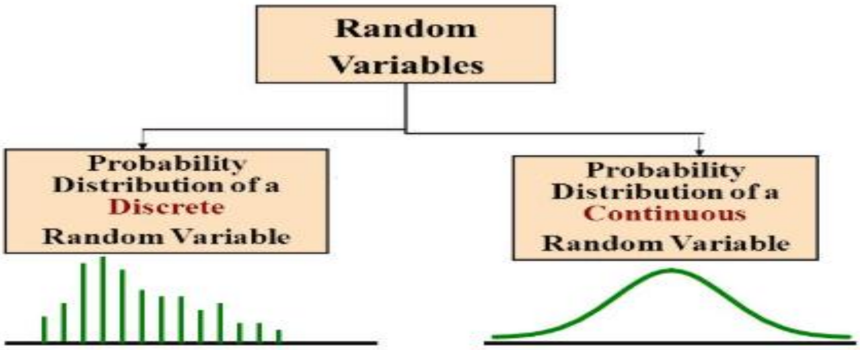
A random variable is a variable whose value is unknown or a function that assigns values to each of an experiment's outcomes. i.e., used to associate rewards every outcome of a random experiment.

Ex 1: Spinning the wheel observing position of the pointer when the wheel halts.

Ex 2: If the random variable Y is the number of heads we get from tossing two coins, then Y could be 0, 1, or 2. This means that we could have no heads, one head, or both heads on a two-coin toss. Tossing two coins land in four different ways: TT, HT, TH, and HH. Then, the P(Y=0) = 1/4 since we have one chance of getting no heads (i.e., two tails [TT] when the coins are tossed). Similarly, the probability of getting two heads (HH) is also 1/4. Notice that getting one head has a likelihood of occurring twice: in HT and TH. In this case, P (Y=1) = 2/4 = ½

Ex 3: For example, in a soccer game we may be interested in the number of goals, shots, shots on goal, corners kicks, fouls, etc. If we consider an entire soccer match as a random experiment, then each of these numerical results gives some information about the outcome of the random experiment.

Random variables can be classified into two main types:



Discrete Random Variables

Discrete random variables take on a countable number of distinct values.

Ex 4: Consider an experiment where a coin is tossed three times. If X represents the number of times that the coin comes up heads, then X is a discrete random variable that can only have the values 0, 1, 2, or 3 (from no heads in three successive coin tosses to all heads). No other value is possible for X.

Probability Mass Function (PMF):

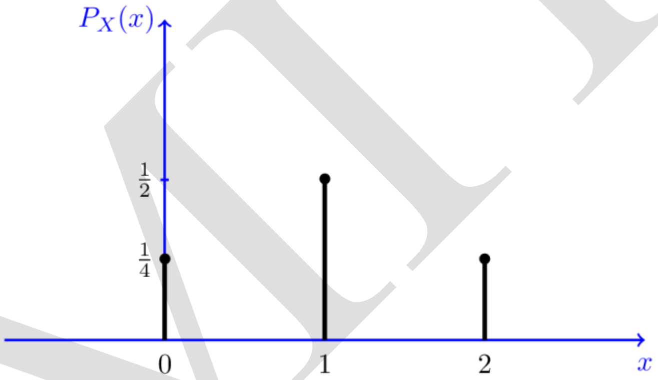
For a discrete random variable, the probability mass function $P(X = x)$ gives the probability that X takes on a particular value x. It's denoted as: $P(X = x) = P(x)$

The PMF must satisfy

- $0 \leq p(X = x) \leq 1$ for all x.
- $\sum P(X = x) = 1$ over all possible values of x.

Ex 5 : Toss a fair coin twice and let X be defined as the number of heads observe.

- sample space is given by, S={HH, HT,TH,TT}.
- The number of heads will be x = 0, 1 or 2
- $P(x = 0) = \frac{1}{4}$; $P(x = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$; $P(x = 2) = \frac{1}{2}$



PMF for random variable

Continuous random variables:

Continuous random variables can represent any value within a specified range or interval and can take on an infinite number of possible values.

Ex 6: An example of a continuous random variable would be an experiment that involves measuring the amount of rainfall in a city over a year or the average height of a random group of 25 people.

Probability Density Function (PDF):

For a continuous random variable X, we use a probability density function $f(x)$, to describe the likelihood of observing a value in a certain interval. Unlike the PMF, the PDF does not directly give probabilities but the probability density at a point x, gives a relative likelihood of X taking a value near x.

The PDF must satisfy

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

Why Are Random Variables Important?

- Random variables produce probability distributions based on experimentation, observation, or some other data-generating process. Random variables, in this way, allow us to understand the world around us based on a sample of data, by knowing the likelihood that a specific value will occur in the real world or at some point in the future.
- Random variables, whether discrete or continuous, are a key concept in statistics and experimentation. Because they are random with unknown exact values, these allow us to understand the probability distribution of those values or the relative likelihood of certain events. As a result, analysts can test hypotheses and make inferences about the natural and social world around us.

Difference Between PMF and PDF:

The main difference between a PMF and a PDF lies in how they are used:

PMF gives the probability that a discrete random variable is equal to a certain value.

PDF gives the density of the probability distribution of a continuous random variable at a particular value

Both PMFs and PDFs are crucial for understanding the behaviour of random variables, analysing data, and making statistical inferences. They form the basis of many statistical models and calculations in fields such as engineering, finance, science, and more.

Mathematical Expectation, Mean and Variance for Discrete random variable:

If a random variable X takes on the values $x_1, x_2, x_3 \ldots x_n$ with the probability

$f(x_1)\,, f(x_2), f(x_3) \ldots f(x_n)$, its mathematical expectation or expected value or mean is E(x) or μ

Expectation / Mean: E(x) or $\mu = \sum x \,.\, f(x)$

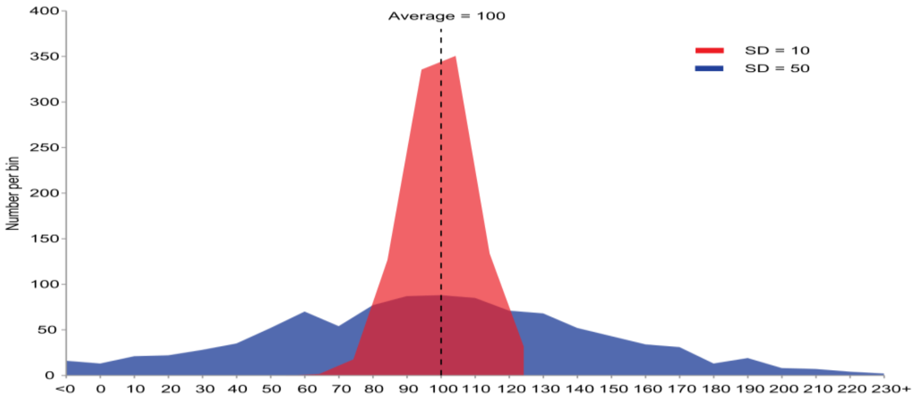
Variance measures the dispersion or spread of the distribution and It's calculated as the average of the squared differences between each data point and the mean.

Variance(x) $\sigma^2 = \sum_i (x_i - \mu)^2 \; f(x) = (\sum_i x_i^{\,2} \; f(x)) - \mu^2$

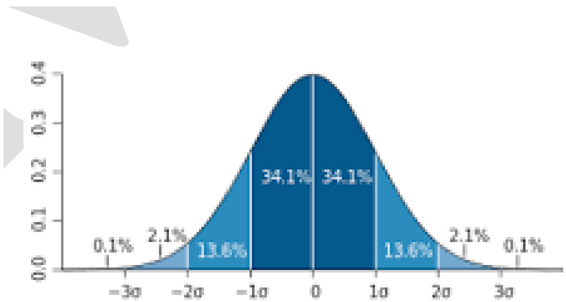
Standard Deviation (S.D) = $\sigma = \sqrt{variance}$

Geometrical interpretation of Variance, S D and Expectation:

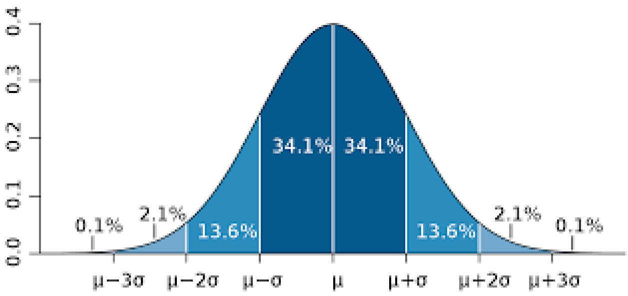
Geometrically, the variance can be seen as a measure of the "spread" or "dispersion" of the probability distribution around its mean. It gives us insight into how much the outcomes of the random variable deviate from their average value.



The geometric interpretation of standard deviation illustrates the extent of dispersion among a set of numbers whose central tendency is represented by the geometric mean.



Geometrically, the expected value represents the balance point or centre of mass of the probability distribution. It's a measure of central tendency that captures where, on average, the values of the random variable tend to concentrate.



Mathematical Expectation, Mean and Variance for Continuous random variable:

Expectation / Mean: $\mu = \int_{-\infty}^{\infty} x \, . f(x) \, dx$

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

Problems:

1. Find the value of k such that the following represents a finite probability distribution. Hence find its mean and standard deviation. Also find $p(x \le 1)$, $p(x > 1)$ and $p(-1 < x \le 2)$.

x	-3	-2	-1	0	1	2	3
$p(x)$	k	$2k$	$3k$	$4k$	$3k$	$2k$	k

Solution:

We must have $p(x) \ge 0$ for all x and $\sum p(x) = 1$. The first condition is satisfied if $k \ge 0$.

K+2k+3k+4k+3k+2k+k = 1 ; $k = \frac{5}{2}$

Mean $\mu = \sum x \, . p(x) = 0$

Variance(x) $\sigma^2 = \sum_i (x_i - \mu)^2 \; p(x) = \frac{5}{2}$

Standard Deviation (S.D) = $\sigma = \sqrt{variance} = \sqrt{\frac{5}{2}}$

Also,

$p(x \le 1) = p(-3) + p(-2) + p(-1) + p(0) + p(1) = \frac{13}{16}$

$p(x > 1) = p(3) + p(2) = \frac{3}{16}$

$p(-1 < x \le 2) = p(0) + p(1) + p(2) = \frac{9}{16}$

2. A random variable x has the density function $p(x) = \begin{cases} kx^2, & 0 \le x \le 3 \\ 0, & elsewhere \end{cases}$. Find (i) value k (ii) $P(x \le 1)$ (iii) $P(1 < x < 2)$ (iv) $P(x > 1)$ (v) mean (vi) variance.

Solution: $P(x) \geq 0$. We must have $\int_{-\infty}^{\infty} f(x)dx = 1$;

$$\int_0^3 k\,x^2 dx = 1; \quad k = \frac{1}{9}.$$

(i) $p(x \leq 1) = \int_0^1 k\,x^2 dx = \frac{1}{27}$

(ii) $P(1 < x < 2) = \int_1^2 k\,x^2 dx = \frac{7}{27}$

(iii) $P(x > 1) = \int_1^3 k\,x^2 dx = \frac{26}{27}$

Mean = $\mu = \int_{-\infty}^{\infty} x\,.f(x)\,dx = \int_0^3 k\,x^3\,dx = \frac{9}{4}$

Variance(x) $\sigma^2 = \int_{-\infty}^{\infty} x^2.p(x)\,dx - \mu^2 = \int_0^3 x^2.\frac{x^2}{9}\,dx - \left(\frac{9}{4}\right)^2 = \frac{27}{80}$

CO1:

1. Identify the set of possible values for the random variable. The number of heads in two tosses of a coin (sol: {0,1,2})

2. A service organization in a large town organizes a raffle each month. One thousand raffle tickets are sold for Rs1 each. Each has an equal chance of winning. First prize is Rs.300, second prize is Rs.200, and third prize is Rs.100. Let X denote the net gain from the purchase of one ticket.
   a. Construct the probability distribution of X
   b. Find the probability of winning any money in the purchase of one ticket.
   c. Find the expected value of x and interpret its meaning.

3. A random experiment of tossing a 'die' twice is performed. Random variable X on this sample space is defined to be the sum of the two numbers turning up on the toss. Find the discrete probability distribution for the random variable X

CO2 :

1. A discrete random variable x has the probability function

x:	0	1	2	3	4	5	6	7	8
f(x) :	K	2k	3k	5k	5k	4k	3k	2k	k

Find the value of E(x).

2. The probability distribution of find the random variable X is given by the following table.

x	-2	-1	0	1	2	3
$p(x)$	0.1	k	0.2	$2k$	0.3	k

Find the value of k, mean and variance.

3. A random variable X has the following probability function for various values of x.

x	0	1	2	3	4	5	6	7
$p(x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

(i) Find k (ii) Evaluate $p(x<6)$, $p(x \geq 6)$ and $p(3<x\leq 6)$.

4. A random variable x has the density function $p(x)=\begin{cases} kx^2, & -3\leq x\leq 3 \\ 0, & elsewhere \end{cases}$. Find (i) value k (ii) $P(x\leq 2)$ (iii) $P(1\leq x\leq 2)$ (iv) $P(x>1)$ (v) mean (vi) variance.

5. Find k such that $f(x)=\begin{cases} kx(1-x), & 0<x<1 \\ 0, & otherwise \end{cases}$ is a *p.d.f.* Find the value of k, mean and variance.

6. Find the value of k such that the following represents a finite probability distribution. Hence find its mean and standard deviation. Also find $p(x\leq 1), p(x>1)$ and $p(-1<x\leq 2)$.

x	-3	-2	-1	0	1	2	3
$p(x)$	k	$2k$	$3k$	$4k$	$3k$	$2k$	k

7. The *p.d.f* of a variate X is given by the following table.

x	0	1	2	3	4	5	6
$p(x)$	k	$3k$	$5k$	$7k$	$9k$	$11k$	$13k$

For what value of k, this represents a valid probability distribution?

CO3:

1. From a sealed box containing a dozen apples it was found that 3 apples are perished. Obtain the probability distribution of the number perished apples when 2 apples are drawn at random. Also find the mean and variance of this distribution.

2. 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 and hence find the expectation.

3. A random experiment of tossing a 'die' twice is performed. Random variable X on this sample space is defined to be the sum of the two numbers turning up on the toss. Find the discrete probability distribution for the random variable X and compute the corresponding mean and standard deviation.

Binomial Distribution

The Binomial distribution is a fundamental concept in probability theory and statistics, often used to model the number of successes in a fixed number of independent Bernoulli trials. It is named after Swiss mathematician Jacob Bernoulli, who first introduced it in the 17th century.

Imagine conducting a series of independent experiments, each with only two possible outcomes, such as success or failure, heads or tails, or yes or no. a child can be male or female; a person can die or not die; a person can be employed or unemployed. These outcomes are often labelled as "success" or "failure." Note that there is no connotation of "goodness" here - for example, when looking at births, the statistician might label the birth of a boy as a "success" and the birth of a girl as a "failure," but the parents wouldn't necessarily see things that way. We may choose to define either outcome as a success. The process is referred to as a **Bernoulli process**. The Binomial distribution allows us to calculate the probability of obtaining a specific number of successes in a given number of trials, given the probability of success for each trial.

This distribution is particularly useful in analysing scenarios such as the number of defective items in a production batch, the likelihood of a certain number of goals scored in a series of penalty kicks, or the probability of passing a test with a certain number of correct answers.

With its versatility and applicability to various real-world situations, the Binomial distribution serves as a cornerstone in statistical theory and plays a vital role in fields such as quality control, genetics, and market research. Its elegant mathematical formulation and practical relevance make it an indispensable tool for understanding and predicting random outcomes in a wide range of contexts.

The notations are $p = probability\ of\ success\ in\ each\ trial$, $q = probability\ of\ failure\ in\ each\ trial = 1 - p$. Note that $p + q = 1$. In statistical terms, **A Bernoulli trial is each repetition of an experiment involving only 2 outcomes.**

independent trial- the result of one trial does not affect the result of another trial.

repeated trial- conditions are the same for each trial, i.e. p and q remain constant across trials.

A binomial distribution gives us the probabilities associated with independent, repeated Bernoulli trials. In a binomial distribution the probabilities of interest are those of receiving a certain number of successes, **r**, in $\boldsymbol{n}$ independent trials each having only two possible outcomes and the same probability, **p**, of success.

for example, we want to find the probability of getting 4 heads in 10 tosses. In this case, we'll call getting a head a "success." Also, in this case, n = 10, the number of successes is r = 4, and the number of failures (tails) is n – r = 10 – 4 = 6.

More generally, if p = probability of success and $q = 1 - p$ = probability of failure, the probability of a specific sequence of outcomes where there are r successes and n-r failures is

$$p^r q^{n-r}$$

Note:

1. The number of different ways that $\boldsymbol{n}$ distinct things may be arranged in order is $\boldsymbol{n}$**!.**

2. The total number of ways of selecting $\boldsymbol{r}$ distinct combinations of $\boldsymbol{n}$ objects, irrespective of order, is

$$\binom{n}{r} = \binom{n}{n-r} = \frac{n!}{r!\,(n-r)!}$$

Definition:

Any random variable X with probability function given by

$$\boldsymbol{P(X = r;\ n, p) = \binom{n}{r} p^r q^{n-r} = \frac{n!}{r!\,(n-r)!} p^r q^{n-r}}.$$

is said to have a binomial distribution with parameters $\boldsymbol{n}$ and $\boldsymbol{p}$. It is also represented by $\boldsymbol{B(n, p)}$.

Note: $\boldsymbol{B(n,p)}$ is a Probability Mass Function.

[In fact,, $\boldsymbol{B(n,p) \geq 0}$, $\forall \boldsymbol{x}$ and $\sum \boldsymbol{B(n,p) = (p+q)^n = 1}$].

Mean and Variance of Binomial Distribution

Mean of Binomial Distribution is given by $\boldsymbol{\mu = n * p}$, where n is number of trials and p is probability of success.

Variance of Binomial Distribution is given by $\boldsymbol{\sigma^2 = n * p * q}$, where n is number of trials, p is probability of success and q is probability of failure. Hence, we have Standard Deviation is given by $\boldsymbol{\sigma = \sqrt{n * p * q}}$.

Problems

CO2 Problems

1. Find the binomial distribution which has mean 2 and variance 4/3. Hence, find the value at $P(X = 6)$.
   Ans: $P(X = r) = \binom{6}{r}\left(\frac{1}{3}\right)^r \left(\frac{2}{3}\right)^{6-r}$. $P(X = 6) = \left(\frac{1}{3}\right)^6$.
2. If mean and s.d. of some experiment with 4096 objects are 2.5 and $\sqrt{1.875}$. Find the value of n and P(X=r). Hence, find $P(0) + P(1) + P(2)$.
   Answer: $n = 10$, $P(x) = \frac{1}{4^{10}}[10C_r 3^{10-r}]$, 538.73.

CO3 Problems

1. **When an honest coin is tossed 4 times, find the probability of getting**
   - **i. Exactly one head**
   - **ii. Atmost 3 heads**
   - **iii. At least two heads**

Solution: $n = 4, success = r = getting\ head, p = \frac{1}{2}\ and\ q = \frac{1}{2}$.

We know that $B(n, p, r) = \binom{n}{r} p^r\, q^{n-r}$

Hence, we have $B\left(4, \frac{1}{2}, r\right) = \binom{4}{r}\frac{1}{2}^r \frac{1}{2}^{n-r} = \binom{4}{r} * \frac{1}{16}$.

i. We have r=1. Therefore, from the above equation we have $B(1) = \binom{4}{1} * \frac{1}{16} = \frac{1}{4}$.

ii. For Atmost 3 heads we have $B(r \leq 3) = B(0) + B(1) + B(2) + B(3) = \frac{15}{16}$.

iii. For At least two heads we have $B(r \geq 2) = B(2) + B(3) + B(4) = \frac{11}{16}$.

2. 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 (a) at least 10 survive, (b) from 3 to 8 survive, and (c) exactly 5 survive?

Solution: $n = 15,\ success = r = recovery,\ p = 0.4\ and\ q = 1 - 0.4 = 0.6.$

We know that $B(n, p, r) = \binom{n}{r} p^r q^{n-r}$

Hence, we have $B(15,\ 0.4,\ r) = \binom{15}{r} 0.4^r\ 0.6^{n-r}$

i. For at least 10 survives, from the above equation we have $B(r \geq 10) = 0.0338$

ii. For 3 to 8 survives we have $B(3 \leq r \leq 8) = B(3) + B(4) + B(5) + B(6) + B(7) + B(8) = 08779.$

iii. For exactly 5 survives we have $B(r = 5) = B(5) = 01859.$

3. The number of telephone lines busy at an instant of time is a binomial variate with probability 0.1 that a line is busy. If 10 lines are chosen at random, what is the probability that (i) no line is busy (ii)all lines are busy (iii) at least one line is busy (iv) at most 2lines are busy.
Solution: (i) $p(x = 0) = 0.3487$, (ii) $p(x = 10) = (0.1)^{10}$, (iii) $p(x \geq 1) = 0.6513$, (iv) $p(x \leq 2) = 0.9298$

4. The number of telephone lines busy at an instant of time is a binomial variate with probability 0.2. If at an instant 10 lines are chosen at random, what is the probability that (i) 5lines are busy? (ii) at most 2 lines are busy? (iii) all lines are busy?
Solution: (i) $p(x = 5) = 0.0264$, (ii) $p(x \leq 2) = 0.6778$,

(iii) $p(x = 10) = 1.024 \times 10^{-7}$

5. A pair of dice is thrown twice. Find the probability of scoring 7 points (i) once (ii) twice, and (iii) at least once.
Solution: (i) $p(x = 1) = \frac{5}{8}$ (ii) $p(x = 2) = \frac{1}{36}$ (iii) $p(x \geq 1) = \frac{11}{36}$

6. The probability that a person aged 60 years live up to 70 is 0.65. What is the probability that out of 10 persons aged 60 at least 7 of them will live up to 70?
Solution: $p(x \geq 7) = 0.5138$

7. The probability that a pen manufactured by a factory be defective is $1/10$. If 12 such pens are manufactured, what is the probability that (i) exactly 2 are defective (ii) at least 2 are defective (iii) none of them are defective.

Solution: (i) $p(x=2)=0.2301$ (ii) $p(x\geq 2)=0.341$ (iii) $p(x=0)=0.2824$

8. In 256 sets of 12 tosses of an honest coin, in how many sets one can expect 8 heads and 4 tails?
   Solution: 31 sets

9. In a quiz contest of answering "yes" or "no" what is the probability of guessing at least 6 answers correctly out of 10 questions asked? Also find the probability of the same if there are 4 options for a correct answer.

   Solution: (i) $p(r\geq 6)=0.377$ (ii) $p(r\geq 6)=0.019$

Some Areas of Application

The binomial distribution finds applications in many scientific fields. An industrial engineer is keenly interested in the "proportion defective" in an industrial process. Often, quality control measures and sampling schemes for processes are based on the binomial distribution. This distribution applies to any industrial situation where an outcome of a process is dichotomous and the results of the process are independent, with the probability of success being constant from trial to trial. The binomial distribution is also used extensively for medical and military applications. In both fields, a success or failure result is important. For example, "cure" or "no cure" is important in pharmaceutical work, and "hit" or "miss" is often the interpretation of the result of firing a guided missile.

Poisson Distributions

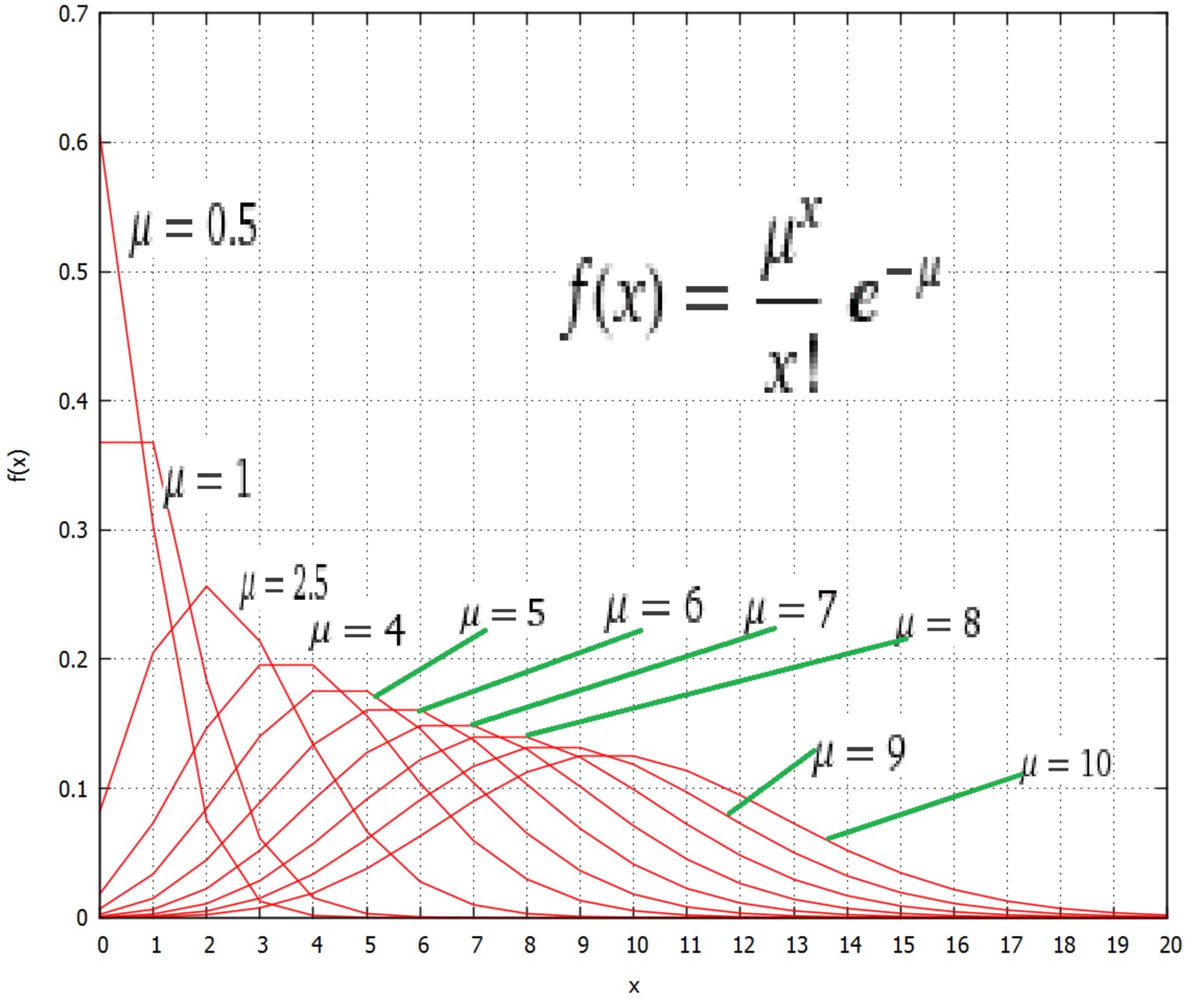
The Poisson distribution, named after French mathematician Simeon Denis Poisson, is a fundamental concept in probability theory and statistics. It is widely used to model the number of events occurring within a fixed interval of time or space, given a known average rate of occurrence. From modelling the number of phone calls received at a call centre in a given hour to estimating the number of customers arriving at a store during a specified time frame, the Poisson distribution provides a powerful tool for analysing rare events or occurrences in various real-world scenarios. With its elegant mathematical formulation and practical applicability, the Poisson distribution continues to play a pivotal role in fields such as insurance, telecommunications, and queueing theory, offering insights into the behaviour of random processes and aiding in decision-making processes.

It is a distribution related to probabilities of events which are extremely rare, but which have a large number of independent opportunities for an occurrence. For example, number of printing mistakes per page, number of blind persons born blind in a large city, bacterial or injection reaction, defective materials produced in a factory etc

Poisson distribution is denoted by $P(\mu,r)$ and is defined as

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

It can be derived as a limiting case of Binomial distribution by letting $n \to \infty$ and $p \to 0$ keeping $\mu = n * p$ fixed.



Mean and Variance of Poisson Distribution

Mean and Variance of Poisson Distribution is given by $\boldsymbol{\mu = \sigma^2 = n * p}$, where n is number of trials and p is probability of success. Hence, we have Standard Deviation is given by

$$\boldsymbol{\sigma = \sqrt{\mu} = \sqrt{n * p}.}$$

Note:

Poisson Distribution gives the probability of r successes with mean μ out of n large trials. (i.e., in P.D. n is very large compared to p).

Problems

1. **In a certain factory turning out razor blades, there is a small chance of 0.002, for a blade to be defective. The blades are supplied in packets of 10. Calculate the approximate number of packets containing (i) no defective, (ii) one defective, and (iii) two defective blades, in a consignment of 10,000 packets.**

Solution: Probability that a blade is defective in a packet is $p = 0.002$.

For 10 blades, i.e., $n = 10$, mean $\mu = n * p = 0.02$.

Hence,

$$\boldsymbol{P(\mu, r)} = \frac{\boldsymbol{\mu^r\, e^{-\mu}}}{\boldsymbol{r!}} = \frac{(0.02)^r\, e^{-0.02}}{r!} = P(r)$$

Let r be the number of defective blades out of 10 blades (1 packet). Then

$$P(0) = 0.9802,\ \ P(1) = 0.0196\ \ and\ \ P(2) = 0.002.$$

Therefore, in a consignment of 10,000 packets, the number of packets containing

(i) no defective blade = 10,000*0.9802 = 9802

(ii) one defective blade = 10,000*0.0196 = 196

(iii) two defective blades = 10,000*0.0002 = 2.

CO1 Problems

2. What is the sum of probability of success and probability of failure in binomial distribution?
   Ans: 1
3. Give real life scenario where binomial/ Poisson distribution find application?
4. Give the expression to find the mean and standard variation of Binomial/Poisson distribution.
5. If the probability of failure of an experiment is 0.35 which follows binomial distribution, what will be probability of success?
6. What is the probability mass function (PMF) of the Poisson distribution/Binomial distribution?
7. What is the condition on n and p which decides the problem can be classified into Poisson distribution?
8. Whether $P(x) = \binom{n}{r} p^r\, q^{n-r}$ is a mass function or not? Justify your answer.
9. In binomial distribution, what is the range of r? Can it be negative?
10. For a quiz contest of answering yes or no, what is the probability of answering a question correctly?

CO2 Problems

11. If X follows a Poisson law such that $P(X = 2) = (2/3)P(X = 1)$, find $P(X) = 0\ and\ P(X = 3)$.
    Solution: (i) $p(x = 0) = 0.2636$ (ii) $p(x = 3) = 0.10414$.

12. If X follows a Poisson law such that $P = 9P(X = 4) + 90P(X = 6)$, find the standard deviation.
    Solution: 1

CO3 Problems

13. The number of accidents in a year to taxi drivers in a city follows a Poisson distribution with mean 3. Out of 100 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: (i) 50 drivers (ii) 350 drivers

14. In sampling a large number of parts manufactured by a company, the mean number of defective in samples of 20 is 2. Out of 1000 such samples how many would be expected to contain at least 3 defective parts.
Solution: 323

15. There is a chance that 5% of the pages of a book contain typographical errors. If 100 pages of the book are chosen at random, find the probability that 2 of these pages contain typographical errors.
Solution: $p(x = 2) = 0.084$

16. 2% of the fuses manufactured by a firm are found to be defective. Find the probability that a box containing 200 fuses contains (i) no defective fuses (ii) 3 or more defective fuses.
Solution: (i) $p(x = 0) = 0.0183$ (ii) $p(x \geq 3) = 0.762$

17. A certain screw making machine has a chance of producing 2 defectives out of 1000.The screws are packed in boxes of 100. Find the approximate number of boxes containing (i) no defective screw (ii) one defective screw, and (iii) two defective screws, in a consignment of 5000 boxes.
Solution: (i) 4094 boxes (ii) 819 boxes (iii) 82 boxes.

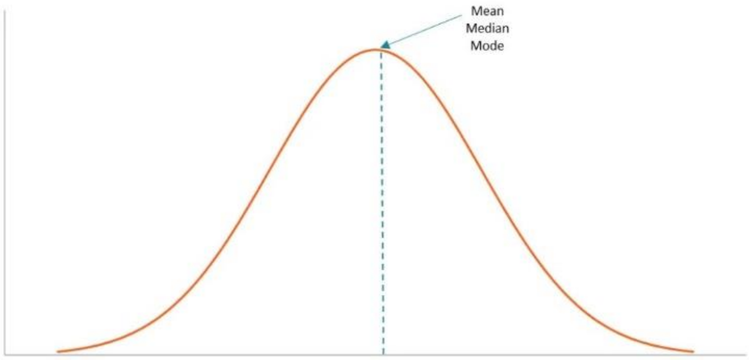
NORMAL DISTRIBUTION

The normal distribution is a continuous distribution. It can be derived from the binomial distribution in the limiting case when $\boldsymbol{n}$, the number of trials is very large and $\boldsymbol{p}$, the probability of a success, is close to $\frac{1}{2}$. The general equation of the normal distribution is given by

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

where the variable $\boldsymbol{x}$ can assume all values from $-\infty$ to $+\infty$. μ and σ, called the parameters of the distribution, are respectively the mean and the standard deviation of the distribution and $-\infty < \mu < +\infty, \sigma > 0$. x is called the normal variate and $f(x)$ is called probability density function of the normal distribution.

If a variable x has the normal distribution with mean μ and standard deviation σ, we briefly write $x: N(\mu, \sigma^2)$.



The graph of the normal distribution is called the normal curve. It is bell-shaped and symmetrical about the mean μ. The two tails of the curve extend to $-\infty$ to $+\infty$ towards the positive and negative directions of the x-axis respectively and gradually approach the x-axis without ever meeting it. The curve is unimodal and the mode of the normal distribution coincides with its mean μ. The line $x = \mu$ divides the area under the normal curve above x-axis into two equal parts. Thus, the median of the distribution also coincides with its mean and mode. The area under the normal curve between any two given ordinates and represents the probability of values falling into the given interval. The area under the normal curve above the x-axis is 1.

BASIC PROPERTIES OF THE NORMAL DISTRIBUTION

The probability density function of the normal distribution is given by

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

(i) $f(x) \geq 0$ and $\int_{-\infty}^{\infty} f(x)\,dx = 1$. i.e., the total area under the normal curve above the x-axis is 1.

(ii) The normal distribution is symmetrical about its mean.

(iv) It is a unimodal distribution. The mean, mode and median of this distribution coincide.

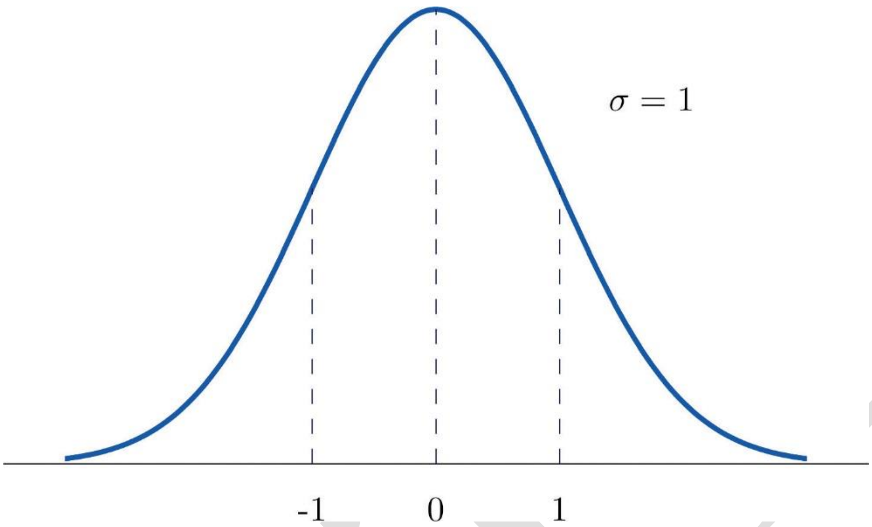
STANDARD FORM OF THE NORMAL DISTRIBUTION

If X is a normal random variable with mean μ and standard deviation σ, then the random variable $\boldsymbol{Z} = \frac{\boldsymbol{X-\mu}}{\boldsymbol{\sigma}}$ has the normal distribution with mean 0 and standard deviation 1. The random variable Z is called the standardized (or standard) normal random variable.

The probability density function for the normal distribution in standard form is given by

$$\boldsymbol{f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}}$$

It is free from any parameter. This helps us to compute areas under the normal probability curve by making use of standard tables.



Note:

1. If $f(z)$ is the probability density function for the normal distribution, then

$P(z_1 \le Z \le z_2) = \int_{z_1}^{z_2} f(z)\, dz = F(z_2) - F(z_1)$, where $F(z) = \int_{-\infty}^{z} f(z)\, dz = P(Z \le z)$.

The function $F(z)$ defined above is called the distribution function for the normal distribution.

2. The probabilities $P(z_1 \le Z \le z_2), P(z_1 < Z \le z_2), P(z_1 \le Z < z_2)$ and $P(z_1 < Z < z_2)$ are all to be the same.

3. $F(-z) = 1 - F(z)$.

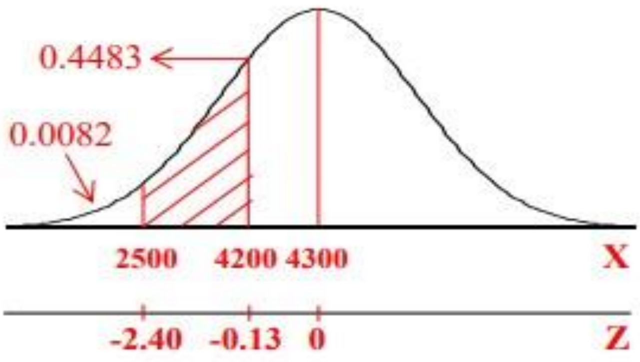
Problems:

The average number of acres burned by forest and range fires in a large New Mexico County is 4,300 acres per year, with a standard deviation of 750 acres. The distribution of the number of acres burned is normal. What is the probability that between 2,500 and 4,200 acres will be burned in any given year?

Solution: Given $\bar{x} = 4300$ and $\sigma = 750$.
We have standard normal variate given by $z = \frac{x-\mu}{\sigma}$
For $x = 2500$ we have $z = -2.40$ and for $x = 4200$ we have $z = -0.1333$.
Hence,

$$\begin{aligned} P(2500 < X < 4200) &= P(-2.40 < z < -0.1333) \\ &= P(z < -0.1333) - P(z < -2.40) \\ &= 0.4483 - 0.0082 = 0.4401. \end{aligned}$$

Note: The percentage of probability is 44.01%



Problems on Normal Distribution:

CO1 QUESTIONS

1. What is the nature of curve for probability function associated with normal distribution? Is it symmetric about any line? If yes, give the equation of the line.
2. In normal distribution, $z = \frac{x-\mu}{\sigma}$ is known as? Is it symmetric about any line? If yes, give the equation of the line.
3. What is the area under the curve of normal distribution?
4. Is normal distribution can be derived as limiting case of binomial distribution? If yes, what are the constraints?

   Answer: $n \to \infty$ and $p = q = \frac{1}{2}$.
5. What are the real-world examples where the normal distribution is commonly observed or applied?
6. What are the properties of the probability density function (PDF) and cumulative distribution function (CDF) of the normal distribution?

   The normal distribution, also known as the Gaussian distribution, is a probability distribution that is symmetric and bell-shaped. Its key characteristics include a mean (μ) that represents the centre of the distribution and a standard deviation (σ) that determines the spread or variability of the data around the mean.
7. Can the normal distribution be skewed? If so, under what circumstances?

   Unlike the binomial or Poisson distributions, which model discrete random variables, the normal distribution models continuous random variables. Additionally, the normal distribution is symmetric, whereas the binomial and Poisson distributions may be skewed.

CO2 Problems

1. Given a random variable X having a normal distribution with $\mu = 50$ and $\sigma = 10$, find the probability that X assumes a value between 45 and 62.

2. Given that X has a normal distribution with $\mu = 300$ and $\sigma = 50$, find the probability that X assumes a value greater than 362.

CO3 Problems

3. The weekly wages of workers in a company are normally distributed with mean of Rs. 700 and standard deviation of Rs.50. Find the probability that the weekly wage of a randomly chosen workers is (i) between Rs.650 and Rs. 750 (ii) more than Rs. 750.

4. The marks of 1000 students in an examination follows a normal distribution with mean 70 and standard deviation 5. Find the number of students whose marks will be (i) less than 65 (ii) more than 75 (iii) between 65 and 75.

5. An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns between 778 and 834 hours.

6. In a test on electric bulb, it was found that the life of a particular brand was distributed normally with an average life of 2000 hours and S.D. of 60 hours. If a firm purchases 2500 bulbs find the number of bulbs that are likely to last for (i) more than 2100 hours (ii) less than 1950 hours (iii) between 1900 to 2100 hours.

7. A certain type of storage battery lasts, on average, 3.0 years with a standard deviation of 0.5 year. Assuming that battery life is normally distributed, find the probability that a given battery will last less than 2.3 years.

8. In an examination taken by 500 candidates, the average and S.D. of marks obtained are 40% and 10% respectively. Assuming normal distribution, find (i) how many have scored above 60%, (ii) how many will pass if 50% is fixed as the minimum marks for passing, (iii) how many will pass if 40% is fixed as the minimum marks for passing .

9. A research scientist reports that mice will live an average of 40 months when their diets are sharply restricted and then enriched with vitamins and proteins. Assuming that the lifetimes of such mice are normally distributed with a standard deviation of 6.3 months, find the probability that a given mouse will live (a) more than 32 months; (b) less than 28 months; (c) between 37 and 49 months.

10. The mean weight of 500 students during a medical examination was found to be 50kgs and S.D. weight 6kgs.Assuming that the weights are normally distributed, find the number of students having weight (i) between 40to 50 kgs (ii) more than 60kgs.
11. A certain number of articles manufactured in a batch were classified into three categories according to some particular characteristics, being less than 50, between 50 and 60 and greater than 60. If this characteristic is known to be normally distributed, determine the mean and standard deviation for this batch if 60%, 35% and 5% were found in these categories.
12. The mean height of 500 students is 151 cm and the S.D. is 15 cm. Assuming that heights are normally distribute, find how many students are between 120 and 155 cm in height.