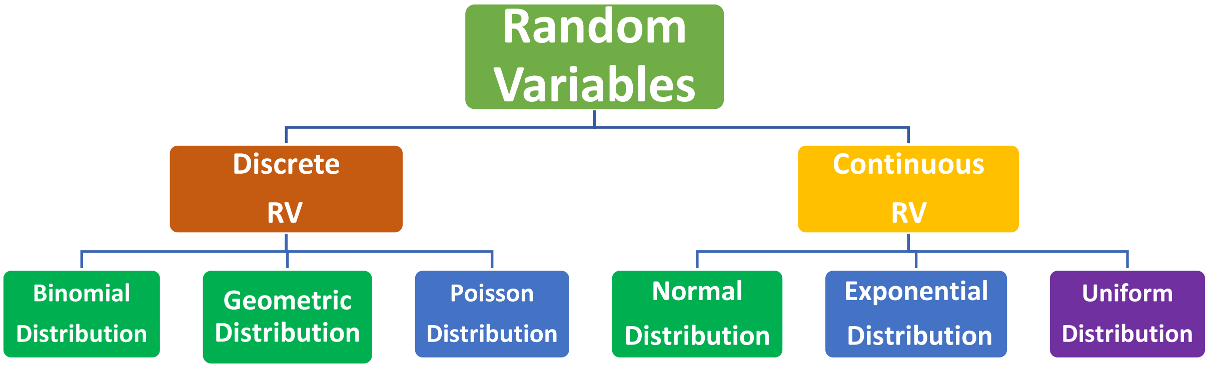
PROBABILITY & STATISTICS

UNIT – 2

Random Variable and Probability Distribution



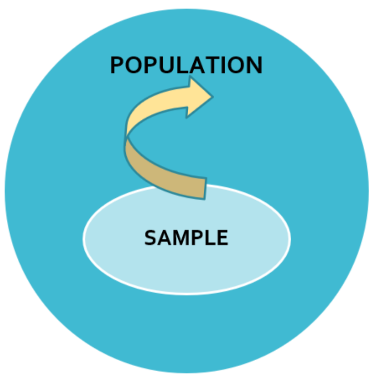
Why we need Probability Distributions???

Most important projects and scientific research studies are conducted with **sample** data rather than with data from an entire **population.**

What is a **Probability Distribution?**

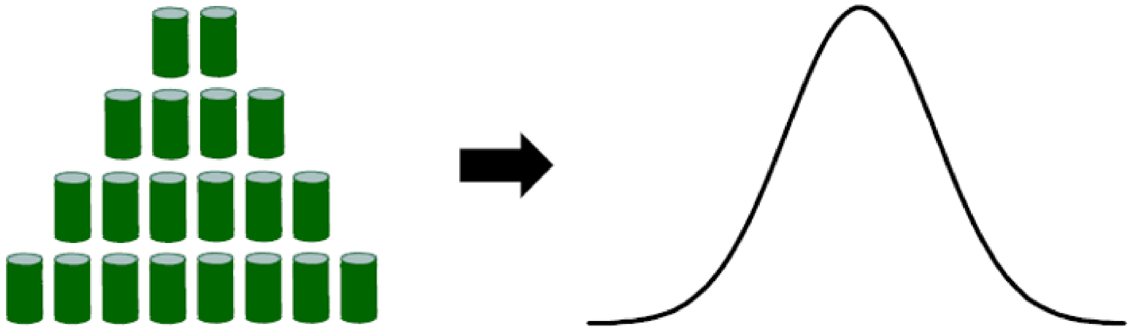
- It is a way to shape the sample data to make predictions and draw conclusions about an entire population.
- It refers to the **frequency** at which some events or experiments occur.
- Probability Distribution helps finding all the possible values a random variable can take between the minimum and maximum possible values.
- Probability Distributions are used to model real-life events for which the outcome is uncertain.

Once we find the appropriate distribution, we can use it to make **inferences** and **predictions**.



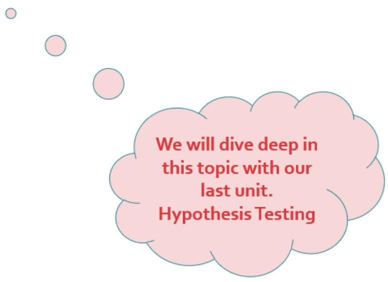
You might be certain if you examine the whole **population**.

- But often times, you only have **samples** to work with.
- To draw conclusions from sample data, you should compare values obtained from the sample with the theoretical values obtained from the probability distribution.



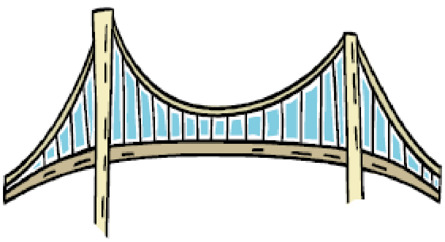
There will always be a **risk** of drawing false conclusions or making false predictions.

- We need to be sufficiently confident before taking any decision by setting **confidence levels**.
- Often set at **90** percent, **95** percent or **99** percent.



1. <u>Probability Distributions:</u>

- Many probability distributions can be defined by **factors** such as the mean and standard deviation of the data.
- Each probability distribution has a formula.
- There are different shapes, models and classifications of probability distributions.



<u>They are often classified into two categories:</u>

- Discrete Probability Distributions.
- Continuous Probability Distributions.

2. Discrete Probability Distributions:

A **Discrete Probability Distribution** relates to discrete data.

- It is often used to model uncertain events where the possible values for the variable are either **attribute** or **countable**.
- The two common discrete probability distributions are **Binomial** and **Poisson** distributions.

3. Binomial Distribution:

3.1. Binomial Experiment:

A binomial experiment is an experiment that satisfies the following conditions.

- The experiment is repeated for a fixed number of trials, where each trial is independent of other trials.
- There are only two possible outcomes of interest for each trial. The outcomes can be classified as a success (S) or as a failure (F).
- The probability of a success P (success) is the same for each trial.
- The random variable x counts the number of successful trials.

3.2. Notation for Binomial Experiments:

Symbol	Description
n	The number of times a trial is repeated
p	The probability of success in a single trial
q	The probability of failure in a single trial. ($q = 1 – p$)
x	The random variable represents a count of the number of successes in n trials: $x = 0, 1, 2, 3, \ldots, n$.

Decide whether the following experiment is a binomial experiment. If it is, specify the values of *n*, *p*, and *q*, and list the possible values of the random variable *x*. If it is not a binomial experiment, explain why?

Examples 1: You randomly select a card from a deck of cards, and note if the card is a king. You then put the card back and repeat this process 8 times.

Yes…This is a binomial experiment. Each of the 8 selections represent an independent trial because the card is replaced before the next one is drawn. There are only two possible outcomes: either the card is a king or not.

$$n = 8 \qquad q = 1 - \frac{1}{13} = \frac{12}{13}$$

$$p = \frac{4}{52} = \frac{1}{13} \qquad x = 0,1,2,3,4,5,6,7,8$$

Examples 2: You roll a die 10 times and note the number the die lands on.

No…. This is not a binomial experiment. While each trial (roll) is independent, there are more than two possible outcomes: 1, 2, 3, 4, 5, and 6.

So, Binomial Distribution is…

A **discrete** probability distribution that is used for data which can only take one of two values, i.e.

- Pass or fail.
- Yes or no.
- Good or defective.

- It allows to compute the probability of **the number of successes for a given number of trials.**
- **Success could mean anything you want to consider as a positive or negative outcome.**

So, now we are ready & loaded to discuss about Binomial Distribution.

3.3. <u>Binomial Distribution Formula:</u>

In a binomial experiment, the probability of getting exactly *x* successes in *n* trials is:

$$P(X = x) = \binom{n}{x} p^x q^{n-x}$$

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

Where n = number of trials

x = number of success

p = probability of success

q = probability of failure

- Binomial distribution is fully defined if we know '*n*' and '*p*', so *n* and *p* are called parameters of Binomial distribution.
- Note that *n* is a discrete parameter whereas *p* is a continuous parameter as $0 < p < 1$.

3.4. Mean of Binomial Distribution:

Mean (Expected value) *E(X)* for binomial distribution is…

$$
\begin{aligned}
E(X) &= \sum x\ p(x) \\
&= \sum_{x=0}^{n} x\ {}^{n}C_x\ p^x\ q^{n-x} \\
&= 0 \cdot {}^{n}C_0\ p^0\ q^n + 1 \cdot {}^{n}C_1\ p q^{n-1} + 2 \cdot {}^{n}C_2\ p^2\ q^{n-2} + \cdots + n\ p^n \\
&= np\ [q^{n-1} + {}^{(n-1)}C_1\ q^{n-2}\ p + {}^{(n-1)}C_2\ q^{n-3}\ p^2 + \cdots + p^{n-1}] \\
&= np\ (q+p)^{\ n-1} \\
&= np \qquad\qquad [\because\ \ p+q=1]
\end{aligned}
$$

So, $E(X) = np$

Similarly, Variance of Binomial distribution Var(X) is...

Var(*X*) = npq

Standard Deviation = $\sqrt{\mathrm{Var}(X)}$

SD = $\sqrt{npq}$

3.5. Recurrence Relation for the Binomial Distribution:

For binomial distribution:

$$
\begin{aligned}
P(X=x) &= {}^{n}C_x\ p^x\ q^{n-x} \\
P(X=x+1) &= {}^{n}C_{x+1}\ p^{x+1}\ q^{n-x-1} \\
\frac{P(X=x+1)}{P(X=x)} &= \frac{{}^{n}C_{x+1}\ p^{x+1}\ q^{n-x-1}}{{}^{n}C_x\ p^x\ q^{n-x}} \\
&= \frac{n!}{(x+1)!\,(n-x-1)!} \times \frac{x!\,(n-x)!}{n!} \cdot \frac{p}{q} \\
&= \frac{(n-x)\,(n-x-1)!\,x!}{(x+1)\,x!\,(n-x-1)!} \cdot \frac{p}{q} \\
&= \frac{n-x}{x+1} \cdot \frac{p}{q}
\end{aligned}
$$

$$
P(X=x+1) = \frac{n-x}{x+1} \cdot \frac{p}{q} \cdot P(X=x)
$$

Examples 1. The mean and standard deviation of a binomial distribution are 5 and 2. Determines the distribution.

Examples 2. The mean and variance of a binomial variate are 8 and 6. Find $P(X \geq 2)$.

Examples 3. With the usual notation, find p for a binomial distribution if $n = 6$ and $9P(X = 4) = P(X = 2)$.

Examples 4. Two dice are thrown five times. Find the probability of getting the sum as 7 (i) at least once, (ii) two times, and (iii) $P(1 < X < 5)$.

Examples 5. The incidence of corona in an industry is such that the workers have a 20% chance of suffering from it. What is the probability that out of 6 workers chosen at random, four or more will suffer from the corona?

Examples 6. If hens of a certain breed lay eggs on 5 days a week on an average, find how many days during a season of 100 days a will poultry keeper with 5 hens of this breed expect to receive at least 4 eggs.

4. Poisson Distribution Formula:

A random variable X is said to follow Poisson distribution if the probability of *x* is given by

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

Where λ is called the parameter of Poisson distribution.

Actually, Poisson is the limiting case of Binomial distribution.

How….??? Let's see….

4.1. Poisson Approximation to the Binomial Distribution:

Poisson distribution is a limiting case of **Binomial distribution** under the following conditions:

(i) The number of trials should be infinitely large, i.e., $n \to \infty$.

(ii) The probability of successes p for each trial should be very small, i.e., $p \to 0$.

(iii) $np = \lambda$ should be finite. Where λ is a constant.

We know that binomial distribution is…

$$P(X=x)=\binom{n}{x}p^x q^{n-x}$$

$$=\binom{n}{x}\left(\frac{p}{q}\right)^x q^n$$

$$=\binom{n}{x}\left(\frac{p}{1-p}\right)^x (1-p)^n$$

Now, let

$$\lambda = np$$

$$\therefore\ p=\frac{\lambda}{n}$$

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

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

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

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

Since,

$$\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n-x}=e^{-\lambda}$$

And

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)=\lim_{n\to\infty}\left(1-\frac{2}{n}\right)=1$$

Now taking limits of both sides as $n\to\infty$

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

4.2. <u>The Poisson distribution holds under the following conditions:</u>

(i) The random variable X should be discrete.
(ii) The numbers of trials *n* are very large.
(iii)The probability of success *p* is very small (very close to zero).
(iv) $\lambda = np$ is finite.
(v) The occurrences are rare.

4.3. <u>Following are some examples of Poisson approximation</u>:

(i) Number of defective bulbs produced by a reputed company.
(ii) Number of telephone calls per minute.
(iii) Number of cars passing a certain point in one minute.
(iv) Number of printing mistakes per page in a large text.
(v) Number of persons born blind per year in a large city.

4.4. <u>Mean of the Poisson Distribution:</u>

Mean (Expected value) ***E(X)*** for **Poisson distribution** is…

$$\begin{aligned}
E(X) &= \sum_{x=0}^{\infty} x\, p(x) \\
&= \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\,\lambda^{x}}{x!} \\
&= \sum_{x=0}^{\infty} \frac{x e^{-\lambda}\,\lambda\,\lambda^{x-1}}{x!} \\
&= e^{-\lambda}\cdot\lambda \sum_{x=1}^{\infty} \frac{x\,\lambda^{x-1}}{x!} \\
&= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \qquad \left[\because \quad \frac{x}{x!} = \frac{1}{(x-1)!}\right] \\
&= \lambda e^{-\lambda}\left(1+\lambda+\frac{\lambda^{2}}{2!}+\cdots\right) \\
&= \lambda e^{-\lambda}\, e^{\lambda} \\
&= \lambda
\end{aligned}$$

4.5. Variance of the Poisson Distribution:

Var*(X)* for Poisson distribution is…

$$\begin{aligned}
\operatorname{Var}(X) &= E(X^2) - \mu^2 \\
&= \sum_{x=0}^{\infty} x^2\, p(x) - \mu^2 \\
&= \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2 \\
&= \sum_{x=0}^{\infty} x\left[(x-1)+x\right]\frac{e^{-\lambda}\lambda^x}{x!} - \lambda^2 \\
&= \sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\lambda^x}{x!} + \sum_{x=0}^{\infty}\frac{x e^{-\lambda}\lambda^x}{x!} - \lambda^2 \\
&= \sum_{x=0}^{\infty} \frac{x(x-1)e^{-\lambda}\,\lambda^{x-2}\,\lambda^2}{x(x-1)(x-2)\cdots 1} + \lambda - \lambda^2 \\
&= e^{-\lambda}\lambda^2 \sum_{x=2}^{\infty}\frac{\lambda^{x-2}}{(x-2)!} + \lambda - \lambda^2 \\
&= e^{-\lambda}\,\lambda^2\left(1+\lambda+\frac{\lambda^2}{2!}+\cdots\right) + \lambda - \lambda^2 \\
&= -e^{\lambda}\, e^{-\lambda}\, \lambda^2 + \lambda - \lambda^2 \\
&= \lambda^2 + \lambda - \lambda^2 \\
&= \lambda
\end{aligned}$$

4.6. Standard deviation of the Poisson Distribution:

As we know that,

$$\begin{aligned}
\text{Standard deviation} &= \sqrt{\text{variance}} \\
&= \sqrt{\lambda}
\end{aligned}$$

4.8. Recurrence Relation for the Poisson Distribution:

As we discussed, for the Poisson distribution:

$$p(x) = \frac{e^{-\lambda}\;\lambda^x}{x!}$$

$$p(x+1) = \frac{e^{-\lambda}\;\lambda^{x+1}}{(x+1)!}$$

$$\frac{p(x+1)}{p(x)} = \frac{e^{-\lambda}\;\lambda^{x+1}}{(x+1)!}\cdot\frac{x!}{e^{-\lambda}\;\lambda^x}$$

$$= \frac{\lambda}{x+1}$$

$$p(x+1) = \frac{\lambda}{x+1}\,p(x)$$

Which is known as a Recurrence relation for Poisson distribution.

Examples 7. If the mean of a Poisson variable is 1.8, find (i) $P(X > 1)$, (ii) $P(X = 5)$ and (iii) $P(0 < X < 5)$

Examples 8. If a random variable has a Poisson distribution such that $P(X = 1) = P(X = 2)$, find (i) the mean of the distribution, (ii) $P(X \geq 1)$, and (iii) $P(1 < X < 4)$.

Examples 9. If X is a Poisson variate such that $P(X = 0) = P(X = 1)$, find $P(X = 0)$ and using recurrence relation formula, find the probabilities at $X = 1, 2, 3, 4$, and 5.

Examples 10. If X is a Poisson variate such that $P(X = 2) = 9P(X = 4) + 90P(X = 6)$. Find (i) the mean of X, (ii) the variance of X.

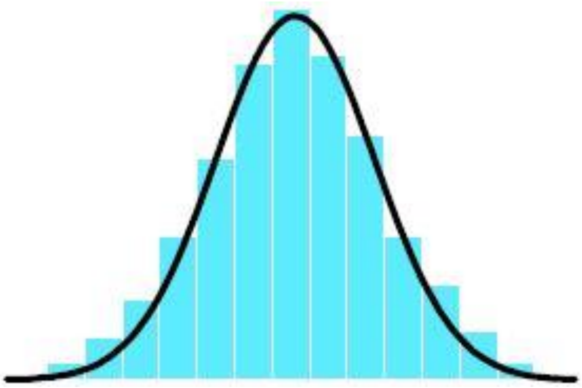
Examples 11. An insurance company insured 4000 people against loss of both eyes in a car accident. Based on previous data, the rates were computed on the assumption that on the average, 10 persons in 100000 will have car accidents each year that result in this type of injury. What is the probability that more than 3 of the insured will collect on their policy in a given year?

Examples 12. Suppose a book of 585 pages contains 43 typographical errors. If these errors are randomly distributed throughout the book, what is the probability that 10 pages, selected at random, will be free from errors?

5. <u>**Normal Distribution:**</u>

- A **symmetrical** probability distribution.
- Most results are located in the **middle** and few are spread on both sides.
- Has the shape of "**a bell**"?
- Can entirely be described by its **mean** and **standard deviation**.

Normality is an important assumption when conducting statistical analysis so that they can be applied in the right manner.



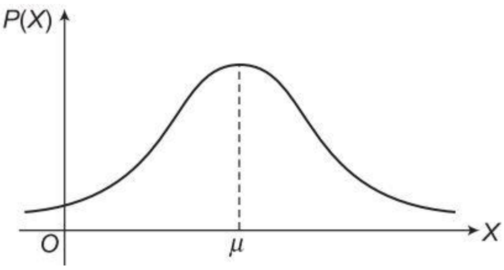
5.1. Normal Distribution Function:

A continuous random variable X is said to follow **normal distribution** with mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma^2}$, if its probability function is given by...

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \text{where } -\infty < X < \infty, -\infty < \mu < \infty, \sigma > 0$$

Here, μ and σ are called parameters of the normal distribution.

The curve representing the normal distribution is called the normal curve.



That means…….

If X is a normal random variable with mean μ and standard deviation σ. Then probability of X lying in the interval (X_1, X_2) is given by...

$$P(x_1 \le X \le x_2) = \int_{x_1}^{x_2} f(x)dx$$

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

5.2 Properties of the Normal Distribution:

A **normal probability curve**, or **normal curve**, has the following properties:

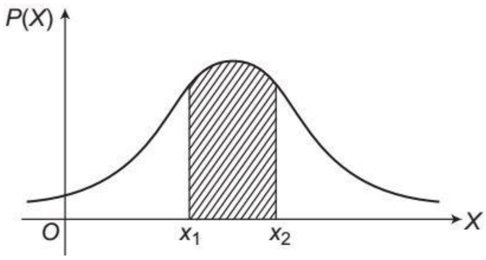
(i) It is a **bell-shaped symmetrical curve** about the ordinate $X = \mu$. The ordinate is **maximum** at $X = \mu$.

(ii) It is a **unimodal curve** and its tails extend infinitely in both the directions, i.e., the curve is asymptotic to $X -$ axis in both the directions.

(iii) All the three measures of central tendency **coincide**, i.e.,
$\text{mean} = \text{median} = \text{mode}$.

(iv) The **total area** under the curve gives the **total probability** of the random variable X taking values between $-\infty$ to ∞. Mathematically,

$$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \mathrm{d}x = 1$$



(v) The ordinate at $X = \mu$ divides the area under the normal curve into two equal parts, i.e.,

$$\int_{-\infty}^{\mu} f(x)\,\mathrm{d}x = \int_{\mu}^{\infty} f(x)\,\mathrm{d}x = \frac{1}{2}$$

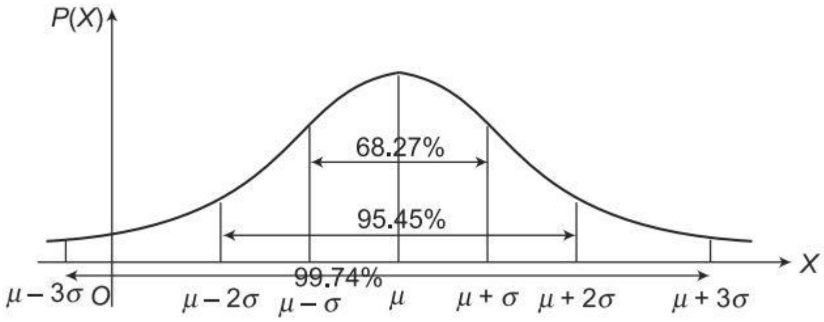
(vi) The value of f(x) is always nonnegative for all values of X, because the whole curve lies above the $X-$ axis.

(vii) The area under the normal curve is distributed as follows:

(a) The area between the ordinates at $\mu - \sigma$ and $\mu + \sigma$ is 68.27%

(b) The area between the ordinates at $\mu - 2\sigma$ and $\mu + 2\sigma$ is 95.45%

(c) The area between the ordinates at $\mu - 3\sigma$ and $\mu + 3\sigma$ is 99.74%



5.3 Parameters of the Normal Distribution:

- **Mathematical expectation/Mean for Normal distribution….**

$$E(\mathrm{X}) = \int_{-\infty}^{\infty} x f(x) dx$$

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

$$= \ldots\ldots$$

$$= \ldots.$$

$$E(\mathrm{X}) = \mu$$

Moreover; $\mathbf{Median} = \mathbf{Mode} = \mu$

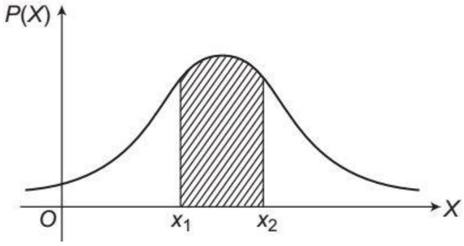
So, Normal distribution is **symmetrical distribution.**

5.4 Probability of a Normal Random Variable in an Interval:

As we discussed earlier…

If $\boldsymbol{X}$ is a normal random variable with mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$. Then probability of $\boldsymbol{X}$ lying in the interval (x_1, x_2) is given by…

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



It looks very difficult to deal with this integration… Isn't it?????

Now, $P(x_1 \le X \le x_2)$ can be evaluated easily by converting a normal random $\boldsymbol{X}$ variable into **another random variable Z**.

Let, $Z = \frac{x-\mu}{\sigma}$ be a new random variable.

$$E(Z) = E\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma}[E(X) - \mu] = 0$$

$$\mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X-\mu}{\sigma}\right) = \frac{1}{\sigma^2}\mathrm{Var}(X-\mu) = \frac{1}{\sigma^2}\mathrm{Var}(X) = 1$$

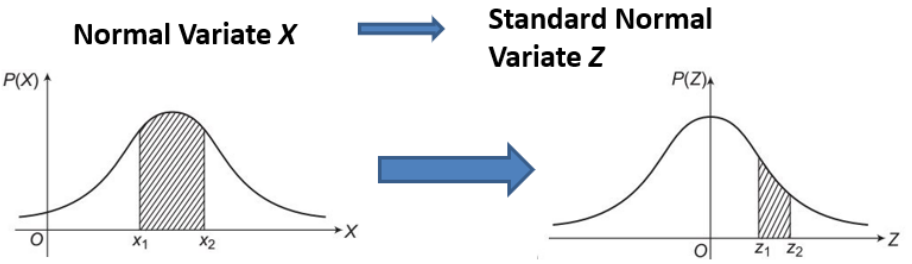
- The distribution of Z is also **normal**.
- Thus, if $\boldsymbol{X}$ is a normal random variable with mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$ then $\boldsymbol{Z}$ is a **normal random variable** with mean **0** and standard deviation **1**.
- Since the parameters of the distribution of $\boldsymbol{Z}$ are **fixed**, it is a known distribution and is termed **standard normal distribution.** Further, $\boldsymbol{Z}$ is termed as a standard normal variate.

Thus, the distribution of any **normal variate *X*** can **always be transformed** into the distribution of the **standard normal variate *Z***.

$$P(x_1 \le X \le x_2) = P\left(\frac{x_1-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{x_2-\mu}{\sigma}\right)$$
$$= P(z_1 \le Z \le z_2)$$

Where $z_1 = \frac{x_1-\mu}{\sigma}$ and $z_2 = \frac{x_2-\mu}{\sigma}$

So, the probability $P(\mathrm{x}_1 \leq X \leq \mathrm{x}_2)$ is equal to the **area under standard normal curve** between the ordinates at $Z = z_1$ and $Z = z_2$.



<u>Case I:</u> - If both z_1 and z_2 are positive (or both negative)

$$P(x_1 \leq X \leq x_2) = P(z_1 \leq Z \leq z_2)$$
$$= P(0 \leq Z \leq z_2) - P(0 \leq Z \leq z_1)$$

$$\boldsymbol{P(x_1 \leq X \leq x_2) = (Area\ under\ normal\ curve\ from\ 0\ to\ z_2) - (Area\ under\ normal\ curve\ from\ 0\ to\ z_1)}$$

<u>Case II</u>: - If $z_1 < 0$ and $z_2 > 0$

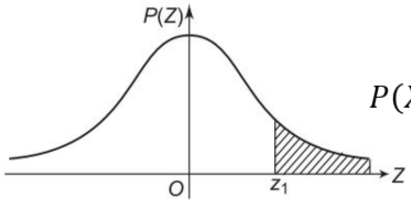
$$P(x_1 \leq X \leq x_2) = P(-z_1 \leq Z \leq z_2)$$
$$= P(-z_1 \leq Z \leq 0) + P(0 \leq Z \leq z_2)$$
$$= P(0 \leq Z \leq z_1) + P(0 \leq Z \leq z_2)$$

$$\boldsymbol{P(x_1 \leq X \leq x_2) = (Area\ under\ normal\ curve\ from\ 0\ to\ z_1) + (Area\ under\ normal\ curve\ from\ 0\ to\ z_2)}$$

Some other cases for $P(X > x_1)$:-

(I) If $z_1 > 0$

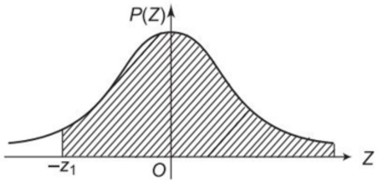
$$P(X > x_1) = P(Z > z_1)$$
$$= 0.5 - P(0 \le Z \le z_1)$$



$P(X > x_1) = 0.5 -$ (area under the curve from 0 to z_1)

(II) If $z_1 < 0$

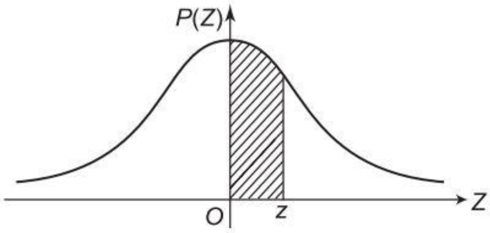
$$P(X > x_1) = P(Z > -z_1)$$
$$= 0.5 + P(-z_1 < Z < 0)$$
$$= 0.5 + P(0 < Z < z_1)$$



$P(X > x_1) = 0.5 +$ (area under the curve from 0 to z_1)

So, we may deal with Z, according to the situation in any other cases also.

Standard Normal *Z* - Table



Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3990	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4115	0.4131	0.4147	0.4162
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

5.5. Uses of Normal Distribution:

5. The normal distribution can be used to approximate binomial and Poisson distributions.
6. It is used extensively in sampling theory. It helps to estimate parameters from statistics and to find confidence limits of the parameter.
7. It is widely used in testing statistical hypothesis and tests of significance in which it is always assumed that the population from which the samples have been drawn should have normal distribution.
   - It serves as a guiding instrument in the analysis and interpretation of statistical data.
   - It can be used for smoothing and graduating a distribution which is not normal simply by contracting a normal curve.

Examples 13. What is the probability that a standard normal variate Z will be (i) greater than 1.09, (ii) less than or equal –1.65, (iii) lying between –1 and1.96, (iv) lying between 1.25 and 2.75?

Examples 14. If X is a normal variate with a mean of 30 and an SD of 5, find the probabilities that (i) $26 \leq X \leq 40,$ and (ii) $X \geq 45$?

Examples 15. A manufacturer knows from his experience that the resistances of resistors he produces is normal with $\text{mean} = 100 \text{ ohms}$ and $\text{SD} = 2 \text{ ohms}$. What percentage of resistors will have resistances between 98 ohms and 102 ohms?

Examples 16. The average seasonal rainfall in a place is 16 inches with an SD of 4 inches. What is the probability that the rainfall in that place will be between 20 and 24 inches in a year?

Important Note:

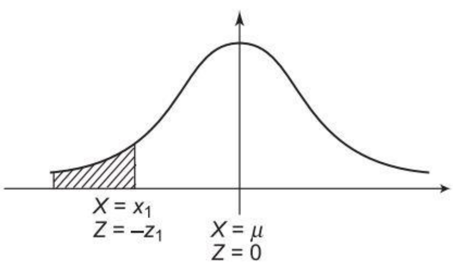
After dealing with previous examples, lets bag some important information.

Here it is…

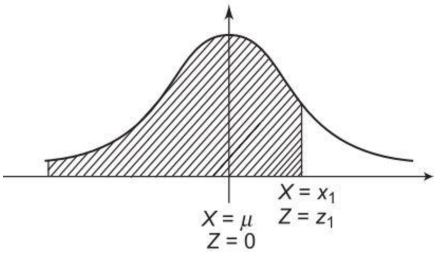
(1) $P(X < x_1) = F(x_1) = \int_{-\infty}^{x_1} f(x)dx$

Hence, $P(X < x_1)$ represent the area under the curve from $X = -\infty$ to $X = x_1$.

(2) If $P(X < x_1) < 0.5$, then the point x_1 lies to the left of $X = \mu$ and the corresponding value of standard normal variate will be negative.



(3) If $P(X < x_1) > 0.5$, then the point x_1 lies to the right of $X = \mu$ and the corresponding value of standard normal variate will be positive.



Examples 17. If X is a normal variate with a mean of 120 and a standard deviation of 10, find c such that (i) $P(X > c) = 0.02$, and (ii) $P(X < c) = 0.05$.

Examples 18. Assume that the mean height of Indian soldiers is 68.22 inches with a variance of 10.8 inches. How many soldiers in a regiment of 1000 would you expect to be over 6 feet tall?

6. <u>Exponential distribution:</u>

A continuous random variable X is said to follow the Exponential distribution if its probability function is given by…

$$f(x)=\begin{cases}\lambda e^{-\lambda x}; & x>0\\ 0 & ;x\le 0\end{cases} \text{ where } \lambda>0 \text{ is rate of distribution.}$$

<u>Note:</u>

- When times between random events follows the **Exponential distribution** with rate λ, then the total number of events in a time period of length t follows the **Poisson distribution** with parameter λt.
- Exponential distribution is **memoryless** distribution (will explain later in this session).

6.1. <u>Relation between Poisson & Exponential:</u>

<u>Poisson</u>

- **Number of hits** to Marwadi University's website in **one minute**.
- **Number of soldiers** killed by horse-kick **per year**.
- **Number of customers** arriving at first floor's Tea Post **in one hour**.

✓ **So, Events per single unit of time.**

<u>Exponential</u>

- Number of minutes **between two hits** to Marwadi University's website.
- Number of years **between horse-kick** deaths of soldier.
- Number of hours **between two customers arrive** at first floor's Tea Post.

✓ **So, Time per single event.**

i.e. **Exponential variate is actually time between the events which are in Poisson distribution. (i.e. you may think like that.. 'inverse' of Poisson)**

6.2. <u>Parameters of Exponential Distribution</u>:

- **Mean of Exponential Distribution:**

$$\begin{aligned}E(X)&=\int_{-\infty}^{\infty} x\, f(x)\mathrm{d}x\\ &=\int_0^{\infty} x\,\lambda e^{-\lambda x}\mathrm{d}x\\ &=\lambda\left| x\cdot\frac{e^{-\lambda x}}{-\lambda}-1\cdot\frac{e^{-\lambda x}}{\lambda^2}\right|_0^{\infty}\\ &=\lambda\cdot\frac{1}{\lambda^2}\\ &=\frac{1}{\lambda}\end{aligned}$$

- **Variance of Exponential Distribution:**

$$\text{Var}(X) = E(X^2) - [E(X)]^2$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x$$

$$= \int_0^{\infty} x^2 \lambda e^{-\lambda x}\,\mathrm{d}x$$

$$= \lambda \left| x^2 \frac{e^{-\lambda x}}{-\lambda} - 2x \frac{e^{-\lambda x}}{\lambda^2} + 2\frac{e^{-\lambda x}}{-\lambda^3} \right|_0^{\infty}$$

$$= \lambda\left(\frac{2}{\lambda^3}\right)$$

$$= \frac{2}{\lambda^2}$$

$$\text{Var}(X) = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2} \qquad \left[\because\ \mu = \frac{1}{\lambda}\right]$$

- **Standard Deviation of Exponential Distribution:**

$$\text{SD} = \sqrt{\text{Var}(X)} = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda}$$

Examples 19. Let X be the Exponential random variate with probability density function $f(x) = \begin{cases} \frac{1}{5}e^{-\frac{x}{5}}; & x > 0 \\ 0; & \text{otherwise} \end{cases}$ Find (i) $P(X > 5)$, (ii)$P(3 \le X \le 6)$, (iii)Mean, (iv)Variance.

Examples 20. Let X be the Exponential random variate with probability density function,

$$f(x) = \begin{cases} ce^{-2x}; x > 0 \\ 0; \text{otherwise} \end{cases}.$$ Find (i) $P(X > 2)$ (ii)$P\left(X < \frac{1}{c}\right)$.

Examples 21. The mileage which car owners get with a certain kind of radial tire is a random variable having an exponential distribution with mean 4000 km. Find the probabilities that one of these tires will last (i) at least 2000 km (ii) at most 3000 km.

Examples 22. The daily consumption of milk in excess of 20000 gallons is approximately exponentially distributed with mean = 3000 gallons. The city has a daily stock of 35000 gallons. What is the probability that of 2 days selected at random, the stock is insufficient for both the days.

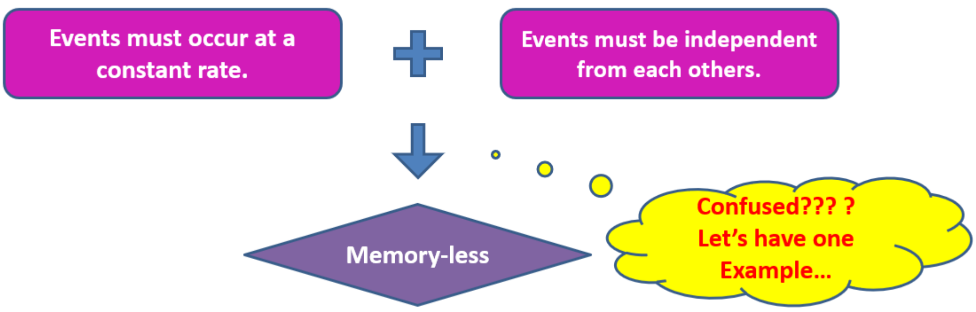
6.3. <u>Memory-less ness of Exponential distribution:</u>

As we know that, An Exponential variate is **time between the Poisson process**.

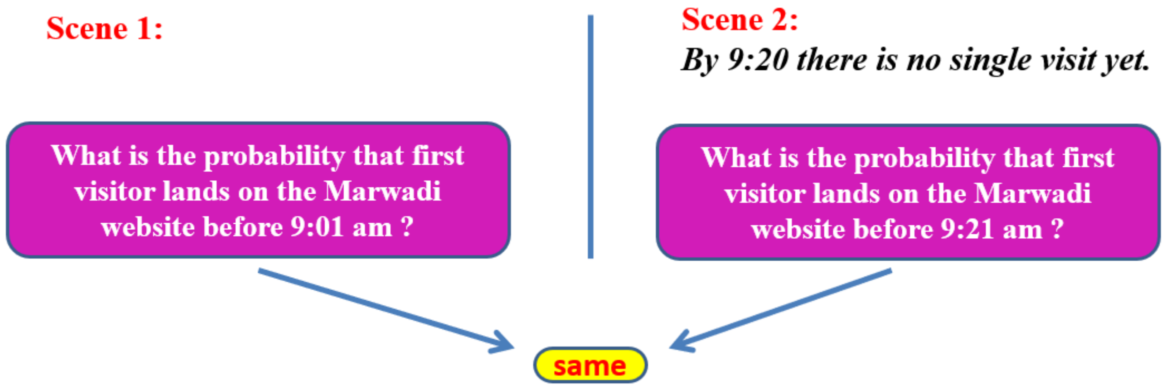
i.e. In other words, we can say Exponential variate is "**inverse**" of Poisson.

Let's **brush up** our knowledge…

Some requirements for Poisson and Exponential variate is…



Example: Suppose Dr. Chetan start monitoring visit to Marwadi University website from 9:00 am.



So, Mathematically…

- The **exponential distribution** has the memory-less(forgetfulness) property.
- This property indicates that the distribution is **independent of its part**, that means future happening of an event has **no relation** to whether or not this even has happened in the past.
- <u>This property is as follows</u>:

If X is exponentially distributed and **s & t** are two positive real numbers then,

$$P[(X > s+t)|(X > s)] = P(X > t)$$

Example 23. The time (in hours) required to repair a machine is exponentially distributed with mean = 2 hours (i) What is the probability that the repair time exceeds 2 hours? (ii) What is the probability that a repair takes at least 11 hours given that its direction exceeds 8 hours?

7. <u>Gamma Distribution:</u>

- Suppose that a system consisting of one original and $(r-1)$ spare components such that in the case of failure of original component, one of the $(r-1)$ spare components can be used.
- The process will continue till we use last component.
- When last component fails, then the whole system fails.
- Let $X_1, X_2, X_3, \ldots. X_r$ be the lifetimes of the *r* components.
- Let each of the random variables $X_1, X_2, X_3, \ldots, X_r$ have the exponential distribution with parameter, and also are probabilistically independent.

Then the lifetime (time until failure) of the entire system is given by…

$$T = \sum_{i=1}^{r} X_i$$

And this whole **system lifetime** has **gamma distribution** with probability density function…

$$f(x) = \begin{cases} \dfrac{\lambda^r}{\Gamma(\mathrm{r})} x^{r-1} e^{-\lambda x} & ; x > 0 \\ 0 & ; x \le 0 \end{cases} \quad \text{for } \lambda, r > 0.$$

- The parameter **r** is called **shape parameter.**
- The parameter λ is called the **rate parameter** (which is same we used in **Exponential distribution**).

Thus, the sum of **r** independent exponential random variables has a **gamma distribution.**

If we take $r = 1$ in above pdf then

$$f(x) = \begin{cases} \lambda e^{-\lambda x}; & x > 0 \\ 0 & ; x \le 0 \end{cases}$$

Which is the probability density function for **Exponential distribution.**

7.1. <u>Parameters of Gamma Distribution:</u>

- **Mean of Gamma Distribution:**

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x \\
&= \int_0^{\infty} x \frac{\lambda^r}{\Gamma r} x^{r-1} e^{-\lambda x}\,\mathrm{d}x \\
&= \frac{\lambda^r}{\Gamma r} \int_0^{\infty} e^{-\lambda x} x^r\,\mathrm{d}x \\
&= \frac{\lambda^r}{\Gamma r} \frac{\Gamma(r+1)}{\lambda^{r+1}} \qquad \left[\because \int_0^{\infty} e^{-kx} x^{n-1}\,\mathrm{d}x = \frac{\Gamma n}{k^n}\right] \\
&= \frac{\lambda^r r \Gamma r}{\Gamma r \cdot \lambda^{r+1}} \\
&= \frac{r}{\lambda}
\end{aligned}
$$

- **Variance of Gamma Distribution:**

$$
\begin{aligned}
\mathrm{Var}(X) &= E(X^2) - [E(X)]^2 \\
E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x \\
&= \int_0^{\infty} x^2 \frac{\lambda^r}{\Gamma r} x^{r-1} e^{-\lambda x}\,\mathrm{d}x \\
&= \frac{\lambda^r}{\Gamma r} \int_0^{\infty} e^{-\lambda x} x^{r+1}\,\mathrm{d}x \\
&= \frac{\lambda^r}{\Gamma r} \frac{\Gamma(r+2)}{\lambda^{r+2}} \qquad \left[\because \int_0^{\infty} e^{-kx} x^{n-1}\,\mathrm{d}x = \frac{\Gamma n}{k^n}\right] \\
&= \frac{(r+1) r \Gamma r}{\Gamma r \cdot \lambda^2} \\
&= \frac{r^2 + r}{\lambda^2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Var}(X) &= \frac{r^2 + r}{\lambda^2} - \frac{r^2}{\lambda^2} \\
&= \frac{r}{\lambda^2}
\end{aligned}
$$

- **SD of Gamma Distribution:**

$$
\mathrm{SD} = \sqrt{\mathrm{Var}(X)} = \sqrt{\frac{r}{\lambda^2}} = \frac{\sqrt{r}}{\lambda}
$$

Example 24. Given a Gamma random variable X with $r = 3$ and $\lambda = 2$. Compute *E(X)*, Var*(X)* and $P(X \leq 1.5 \text{ years})$.

Example 25. In a certain city, the daily consumption of electric power in millions of kilowatt hours can be treated as a random variable having gamma distribution with parameters $\lambda = 1/2$ and $r = 3$. If the power plant of this city has a daily capacity of 12 million kilowatt-hours, what is the probability that this power supply will be inadequate on any given day?