Distributions

Discrete Uniform Random Variable:

A discrete random variable X, which assumes the values x_1, x_2, …, x_n each with probability $\frac{1}{n}$, is called a discrete uniform random variable.

Continuous Uniform Random Variable :

A continuous random variable X is said to be a continuous uniform random variable on the interval (a, b) if its probability density function is given by

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a < x < b \\ 0 & \text{otherwise} \end{cases}$$

Binomial distribution :

Let n be the number of trials.

Let p be the probability of success in each trial.

Let $q = 1 - p$ be the probability of failure in each trial.

Let us define $\quad p(x) = nC_x\, p^x\, q^{n-x}, \quad x = 0,1,2,\ldots,n$

Here $p(x)$ is the probability of x successes out of n trials.

$$\begin{aligned} & p(0) + p(1) + p(2) + \ldots + p(n) \\ & = nC_0 p^0 q^n + nC_1 p^1 q^{n-1} + nC_2 p^2 q^{n-2} + \ldots + nC_n p^n q^0 \\ & = (q+p)^n = 1 \qquad \text{since} \quad q + p = 1 \end{aligned}$$

Therefore, $p(x)$ represents the p.d.f. of discrete r.v. X.

Since $\quad p(x)$, $x = 0,1,2,\ldots,n \quad$ represent the terms of the binomial expansion of $(q+p)^n$, the r.v. X is called binomial variable.

Note : To represent the p.d.f. of a binomial variable, we need two parameters n and p.

Moment generating function of the binomial distribution about origin :

MGF of BD about origin is given by

$$M_0(t) = E(e^{tx}) = \sum_{x=0}^{n} {}^nC_x\, p^x\, q^{n-x}\, e^{tx} = \sum_{x=0}^{n} {}^nC_x \left(pe^t\right)^x q^{n-x} = \left(q + pe^t\right)^n$$

Mean and variance of the binomial distribution :

$$\text{Mean} = E(X) = \frac{d}{dt} M_0(t)\Big|_{t=0} = n\left(q + pe^t\right)^{n-1} pe^t\Big|_{t=0} = n(q+p)^{n-1} p = np$$

$$\begin{aligned}E(X^2) &= \frac{d^2}{dt^2} M_0(t)\Big|_{t=0} = \left[\frac{d}{dt}\left[n\left(q + pe^t\right)^{n-1} pe^t\right]\right]_{t=0} \\ &= \left[n(n-1)\left(q + pe^t\right)^{n-2}(pe^t)^2 + n\left(q + pe^t\right)^{n-1} pe^t\right]_{t=0} \\ &= n(n-1)p^2 + np\end{aligned}$$

$$\begin{aligned}\text{Variance} &= E(X^2) - [E(X)]^2 = n(n-1)p^2 + np - n^2p^2 \\ &= np[(n-1)p + 1 - np] = np(-p+1) = npq\end{aligned}$$

Mode of the binomial distribution :

Mode is that value of x for which $p(x)$ is maximum. Consider

$$\begin{aligned}\frac{p(x)}{p(x-1)} &= \frac{{}^nC_x p^x q^{n-x}}{{}^nC_{x-1} p^{x-1} q^{n-x+1}} = \frac{\left(\dfrac{n!}{x!(n-x)!}\right)}{\left(\dfrac{n!}{(x-1)!(n-x+1)!}\right)}\frac{p}{q} \\ &= \frac{(n-x+1)p}{xq} = \frac{(n+1)p - xp}{xq} = \frac{(n+1)p - x(1-q)}{xq} \qquad (1) \\ &= \frac{xq + (n+1)p - x}{xq} = 1 + \frac{(n+1)p - x}{xq}\end{aligned}$$

Here $x = 1, 2, \ldots, n$

Case (i) : When $(n+1)p$ is not an integer.

Let m = integral part of $(n+1)p$

$$(1) \Rightarrow \frac{p(1)}{p(0)} > 1, \frac{p(2)}{p(1)} > 1, \ldots, \frac{p(m)}{p(m-1)} > 1, \frac{p(m+1)}{p(m)} < 1, \ldots, \frac{p(n)}{p(n-1)} < 1$$

$$\therefore \; p(0) < p(1) < p(2) < \ldots < p(m-1) < p(m) > p(m+1) > \ldots > p(n)$$

From the above relation, it is clear that $p(x)$ is maximum when $x = m$

Case (ii) : When $(n+1)p$ is an integer.

Let $m = (n+1)p$

$$(1) \Rightarrow \frac{p(1)}{p(0)} > 1, \frac{p(2)}{p(1)} > 1, \ldots, \frac{p(m-1)}{p(m-2)} > 1, \frac{p(m)}{p(m-1)} = 1, \frac{p(m+1)}{p(m)} < 1, \ldots, \frac{p(n)}{p(n-1)} < 1$$

$$\therefore \; p(0) < p(1) < p(2) < \ldots < p(m-1) = p(m) > p(m+1) > \ldots > p(n)$$

From the above relation, it is clear that

$p(x)$ is maximum when $x = m$ and m-1

Therefore,

$$\text{Mode of the B.D.} = \begin{cases} \text{integral part of } (n+1)p, & \text{if } (n+1)p \text{ is not an integer} \\ (n+1)p \text{ and } (n+1)p-1, & \text{if } (n+1)p \text{ is an integer} \end{cases}$$

Note : For other measures of central tendency and measures of dispersion, we have to use the p.d.f. as we have done in the case of a general r.v.

Problems on binomial distribution :

1. The probability that a pen manufactured by a company will be defective is $\frac{1}{10}$.

   If 12 such pens are manufactured, then find the probability that

   (i) exactly two pens will be defective

   (ii) atleast three will be defective

   (iii) no pen will be defective

2. 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 1,000 such samples, how many samples would be expected to contain atleast 3 defectives.

3. An irregular six faced die is thrown and the expectation that in 10 throws it will give five even numbers is twice the expectation that it will give four even numbers. How many times in 10,000 sets of 10 such throws would you expect it to give no even number.

4. Comment on the following data :

   The mean and variance of a binomial distribution are 3 and 4 respectively.

5. If a binomial distribution has mean and variance as 4 and 3 respectively, then find the mode of the distribution.

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

7. The latest nationwide political poll indicates that the Americans who are randomly selected, the probability that they are conservative is 0.55, the probability that they are liberal is 0.3 and the probability that they are middle-of-the-road is 0.15. Assuming that these probabilities are accurate, answer the following questions pertaining to a randomly chosen group of 10 Americans.

(i) What is the probability that four are liberal?

(ii) What is the probability that none are conservative?

(iii) What is the probability that two are middle-of-the-road?

What is the probability that at least eight are liberal?

8. Fit the binomial distribution to the following data:

x :	0	1	2	3	4	5	6	7
Frequency :	7	6	19	35	30	23	7	1

Poisson distribution :

Poisson distribution arises in three occasions. They are

1. Special case of a binomial distribution.
2. Arrival pattern of a customer in queuing system
3. Service pattern to a customer in queuing system

Poisson distribution as a special case of binomial distribution :

We know that n and p are the parameters of a binomial variable X.

If $n \to \infty$ and $p \to 0$ such that $\lambda = np$ is finite, then binomial variable X is referred as Poisson variable. Here λ is the parameter for the Poisson distribution.

Probability density function for Poisson distribution :

P.d.f. for binomial distribution

$$= {}^{n}C_{x} p^{x} q^{n-x} = \frac{n(n-1)(n-2)\dots(n-x+1)}{x!}\left(\frac{\lambda}{n}\right)^{x}\left(1-\frac{\lambda}{n}\right)^{n-x}$$

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

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

As n $\to \infty$, we get the p.d.f. for BD as

$$\frac{\lambda^{x}}{x!}\underset{n\to\infty}{Lt}\left[\frac{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{x-1}{n}\right)}{\left(1-\frac{\lambda}{n}\right)^{x}}\left(1-\frac{\lambda}{n}\right)^{n}\right] = \frac{\lambda^{x}}{x!}e^{-\lambda} \qquad (1)$$

(1) is the p.d.f. for Poisson distribution with parameter λ.

Arrival pattern of a customer in queuing system represents Poisson distribution :

If λ represents mean arrival rate or uniform arrival rate of a customer in queuing system, then

$P_n(t)$ = probability of n customers arrived in time t

$$= \frac{e^{-\lambda t}(\lambda t)^n}{n!}, n = 0,1,2,...$$

Here $P_n(t)$ represents a Poisson distribution with parameter λt.

Service pattern to a customer in queuing system represents Poisson distribution :

If μ represents mean service rate or uniform service rate to a customer in queuing system, then

$P_n(t)$ = probability of n customers served in time t

$$= \frac{e^{-\mu t}(\mu t)^n}{n!}, n = 0,1,2,...$$

Here $P_n(t)$ represents a Poisson distribution with parameter μt.

Moment generating function of the Poisson distribution about origin :

Let λ be the parameter of the Poisson distribution.

MGF for the PD about origin

$$= E\left(e^{tx}\right) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = \sum_{x=0}^{\infty} \frac{e^{-\lambda}\left(\lambda e^t\right)}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\left(\lambda e^t\right)}{x!} = e^{-\lambda} e^{\lambda e^t}$$

Mean and variance of the Poisson distribution :

$$\text{Mean} = E(X) = \frac{d}{dt} M_0(t)\Big|_{t=0} = \left[e^{-\lambda} e^{\lambda e^t} \lambda e^t\right]_{t=0} = e^{-\lambda} e^{\lambda} \lambda = \lambda$$

$$E(X^2) = \frac{d^2}{dt^2} M_0(t)\Big|_{t=0} = \left[\frac{d}{dt}\left(e^{-\lambda} e^{\lambda e^t} \lambda e^t\right)\right]_{t=0}$$

$$= \left[e^{-\lambda} e^{\lambda e^t} \lambda e^t + e^{-\lambda} e^{\lambda e^t} \left(\lambda e^t\right)^2\right]_{t=0} = \lambda + \lambda^2$$

$$\text{Variance} = E(X^2) - [E(X)]^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$$

Mode of the binomial distribution :

Mode is that value of x for which $p(x)$ is maximum where $p(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$.

$$\text{Consider} \quad \frac{p(x)}{p(x-1)} = \frac{\left(\dfrac{e^{-\lambda}\lambda^x}{x!}\right)}{\left(\dfrac{e^{-\lambda}\lambda^{x-1}}{(x-1)!}\right)} = \frac{\lambda}{x} \qquad (1)$$

Here $x = 1,2,3,\ldots$

Case (i) : If λ is not an integer, then let m = integral part of λ

$$(1) \Rightarrow \frac{p(1)}{p(0)} > 1, \frac{p(2)}{p(1)} > 1, \ldots, \frac{p(m)}{p(m-1)} > 1, \frac{p(m+1)}{p(m)} < 1, \frac{p(m+2)}{p(m+1)} < 1, \ldots$$

$$\therefore \; p(0) < p(1) < p(2) < \ldots < p(m-1) < p(m) > p(m+1) > p(m+2) > \ldots$$

From the above relation, it is clear that $p(x)$ is maximum when $x = m$

Case (ii) : If λ is an integer, then let $m = \lambda$

$$(1) \Rightarrow \frac{p(1)}{p(0)} > 1, \frac{p(2)}{p(1)} > 1, \ldots, \frac{p(m-1)}{p(m-2)} > 1, \frac{p(m)}{p(m-1)} = 1, \frac{p(m+1)}{p(m)} < 1, \frac{p(m+2)}{p(m+1)} < 1, \ldots$$

$$\therefore \; p(0) < p(1) < p(2) < \ldots < p(m-1) = p(m) > p(m+1) > p(m+2) > \ldots$$

From the above relation, it is clear that

$p(x)$ is maximum when $x = m$ and m-1

$$\therefore \; \text{Mode of the P.D.} = \begin{cases} \text{integral part of } \lambda, \text{ if } \lambda \text{ is not an integer} \\ \lambda \text{ and } \lambda - 1, \text{ if } \lambda \text{ is an integer} \end{cases}$$

Note :

1. For other measures of central tendency and measures of dispersion, we have to use the p.d.f. as we have done in the case of a general r.v.

2. In n and p are the parameters of the binomial distribution, then the binomial distribution is referred as the Poisson distribution

   if $\dfrac{n}{p} \geq 500$ or $n \geq 20, p \leq 0.05$

Problems on Poisson distribution :

1. If the probability of a bad reaction from a injection is 0.001, determine the probability that out of 2,000 individuals more than two will get a bad reaction.

2. In a certain factory turning out razor blades, there is a chance of 0.002 of any blade to be defective. The blades are supplied in packets of 10. Calculate approximate number of packets containing no defective, one defective and two defective blades respectively in a consignment of 10,000 packets.

3. If the prices of new cars increase an average of four times every 3 years, find the probability of
   - (i) no price hikes in a randomly selected period of 3 years.
   - (ii) two price hikes in a randomly selected period of 3 years.
   - (iii) four price hikes in a randomly selected period of 3 years.
   - (iv) 5 or more price hikes in a randomly selected period of 3 years.

4. Guy Ford, production supervisor for the Wistead Company's Charlottesville plant, is worried about an elderly employer's ability to keep up the minimum work pace in the mechanical section. In addition to the normal daily breaks, this employee stops for rest periods at an average of 4.1 times per hour. The rest period is a fairly consistent 3 minutes each time. Ford has decided that if the probability of the employee resting for 12 minutes (not including normal breaks) or more per hour is greater than 0.5, he will move the employee to a different section. Should he do so?

5. A car hire firm has 4 cars which it hires out every day. The mean number of demands for a car on each day is 1.5. Calculate the probability that
(i) any car is not used (ii) some demand is refused
in a day.

6. A manufacturer of cotter pins knows that 5% of his product is defective. If he sells cotter pins in boxes of 100 and guarantees that no more than 10 pins will be defective, what is the probability that a box will fail to meet the guaranteed quality?

7. A Poisson distribution has a double mode at $x = 1$ and $x = 2$. Find the probability that the variable X will have these two values.

8. Find the mean deviation about unit mean of a Poisson variable.

9. Fit the Poisson distribution to the following data :

x :	0	1	2	3	4	5	6	7	8
Frequency :	56	156	132	92	37	22	4	0	1

Normal distribution :

Normal r.v. is a continuous r.v.

Let us suppose that the normal r.v. X has mean μ and S.D. σ .

(Here mean μ and S.D. σ are the parameters of the normal distribution)

The p.d.f. $f(x)$ for the normal variable X is given by

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

Mode of the normal distribution :

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

$$f'(x) = 0 \quad \Rightarrow \quad x = \mu$$

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

$$f''(\mu) = \frac{1}{\sqrt{2\pi}\sigma} \left(-\frac{1}{\sigma^2} \right) < 0$$

$\Rightarrow \quad x = \mu$ is the mode of the normal distribution.

Median of the normal distribution :

If M is the median of the normal distribution, then we have

$$\int_{-\infty}^{M} f(x)\,dx = \int_{M}^{\infty} f(x)\,dx = \frac{1}{2}$$

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

Substitute $z=\dfrac{X-\mu}{\sigma}$, then (1) can be written as

$$\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{z^2}{2}}\sigma\,dz=2\int_{0}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz=1 \;\Rightarrow\; \int_{0}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz=\frac{1}{2} \qquad (2)$$

We have $\displaystyle\int_{M}^{\infty}f(x)\,dx=\frac{1}{2} \;\Rightarrow\; \int_{M}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu\}^2}{2\sigma^2}}dx=\frac{1}{2}$ (3)

Substitute $z=\dfrac{X-\mu}{\sigma}$, then (3) can be written as

$$\int_{\frac{M-\mu}{\sigma}}^{\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz=\frac{1}{2} \qquad (4)$$

Comparing (2) and (4), we get $\dfrac{M-\mu}{\sigma}=0 \;\Rightarrow\; M=\mu$

Mean deviation of the normal distribution :

$$\text{Mean deviation} = E(|X-\mu|)=\int_{-\infty}^{\infty}|x-\mu|\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu\}^2}{2\sigma^2}}dx \qquad (5)$$

Substitute $z=\dfrac{X-\mu}{\sigma}$, then (5) can be written as

$$\text{Mean deviation} = \int_{-\infty}^{\infty}|\sigma z|\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}dz=\frac{2\sigma}{\sqrt{2\pi}}\int_{0}^{\infty}z\,e^{-\frac{z^2}{2}}dz$$

$$=\sqrt{\frac{2}{\pi}}\sigma\left[-e^{-\frac{z^2}{2}}\right]_{z=0}^{\infty}=\sqrt{\frac{2}{\pi}}\sigma=0.7979\sigma\approx 0.8\sigma$$

In normal distribution,

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

Standard normal variable :

A normal variable X is said to be a standard normal variable if its mean is zero and s.d. is equal to 1.

Let $z = \frac{X-\mu}{\sigma}$

Here mean of z is zero and s.d. of z is 1.

Therefore $z = \frac{X-\mu}{\sigma}$ is a standard normal variable.

To find $P(a \le X \le b)$

Let $z_1 = \frac{a-\mu}{\sigma}$ and $z_2 = \frac{b-\mu}{\sigma}$

$$P(a \le X \le b) = \int_a^b \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{z_1}^{z_2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = P(z_1 \le z \le z_2)$$

Here the values of $\int_0^{z_0} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz$, where $z_0 > 0$ (for various values of z_0) are available in the form of a table.

$$P(a \le X \le b) = P(z_1 \le z \le z_2)$$

$$= \begin{cases} P(0 \le z \le z_2) - P(0 \le z \le z_1), \text{ if } z_1, z_2 > 0 \\ P(0 \le z \le -z_1) + P(0 \le z \le z_2), \text{ if } z_1 < 0, z_2 > 0 \\ P(0 \le z \le -z1) - P(0 \le z \le -z2), \text{ if } z_1, z_2 < 0 \end{cases}$$

Note : A binomial variable X with parameters n and p is referred as a normal variable if n is very large and neither p nor q is small. i.e., if $n \ge 30$ and $p,q > 0.05$, then corresponding binomial variable X is referred as a normal variable.

Problems on normal distribution :

1. Given that a random variable, X, has a binomial distribution with $n = 50$ and $p = 0.25$. Find

   (i) $P(X > 10)$ (ii) $P(X < 18)$ (iii) $P(X > 21)$ (iv) $P(9 < X < 14)$

2. If X is a normal variable with mean 12 and s.d 4, then

   (a) find the probability of (i) $X \geq 20$ (ii) $0 \leq X \leq 12$

   (b) find x_1, if $P(X > x_1) = 0.24$

   (c) find x_0 and x_1 when $P(x_0 < X < x_1) = 0.5$ and $P(X > x_1) = 0.25$

3. If X is a normal variable with mean 30 and s.d 5, find the probability that

   (i) $26 \leq X \leq 40$ (ii) $X \geq 45$ (iii) $|X - 30| \geq 5$

4. In a distribution exactly normal, 7% of the items are under 35 and 89% of the items are under 63. What are mean and s.d. of the distribution?

5. Of a large group of people, 5% are under 60 inches in height and the height of the 40% of the people lie between 60 and 65 inches. Find the mean and s.d. of the distribution.

6. The Quickie Sales Corporation has just been given two conflicting estimates of sales for the upcoming quarter. Estimate I says that sales (in millions of dollars) will be normally distributed with $\mu = 325$ and $\sigma = 60$. Estimate II says that sales will be normally distributed with $\mu = 300$ and $\sigma = 50$. The board of directors finds that each estimate appears to be equally believable a priori. In order to determine which estimate should be used for future predictions, the board of directors has decided to meet again at the end of the quarter to use updated sales information to make a statement about the credibility of each estimate.

   (a) Assuming that Estimate I is correct, what is the probability that Quickie will have quarterly sales in excess of $350 million?

   (b) Rework part (a) assuming that Estimate II is correct.

   (c) At the end of the quarter, the board of directors finds that Quickie Sales Corporation has had sales in excess of $350 million. Given this updated information, what is the probability that Estimate I was originally the accurate one?

   (d) Rework part (c) for Estimate II.

7. Show that for a Normal distribution the quartile deviation, mean deviation and standard deviation are approximately in the ratio 10 : 12 : 15.

8. Fit the normal distribution to the following data :

Class Interval	60-65	65-70	70-75	75-80	80-85	85-90	90-95	95-100
Frequency	3	21	150	335	326	135	26	4

Exponential distribution :

A r.v. X is said to follow exponential distribution with parameter $\theta > 0$ if its p.d.f. is given by

$$f(x) = \begin{cases} \theta e^{-\theta x}, x \geq 0 \\ 0, otherwise \end{cases}$$

M.G.F. for exponential distribution about origin:

$$M_0(t) = E(e^{tx}) = \int_0^\infty e^{tx} f(x)\,dx = \int_0^\infty e^{tx}\,\theta e^{-\theta x}\,dx = \int_0^\infty \theta e^{(t-\theta)x}\,dx$$

$$= \left[\theta \frac{e^{(t-\theta)x}}{t-\theta}\right]_{x=0}^{\infty} = \frac{\theta}{\theta - t} = \frac{1}{1-\frac{t}{\theta}} = 1 + \frac{t}{\theta} + \frac{t^2}{\theta^2} + \frac{t^3}{\theta^3} + \ldots$$

Mean and variance of exponential distribution :

$$\text{Mean} = E(X) = \frac{d}{dt} M_0(t)\Big|_{t=0} = \frac{1}{\theta}$$

$$E(X^2) = \frac{d^2}{dt^2} M_0(t)\Big|_{t=0} = \frac{2}{\theta^2}$$

$$\text{Variance} = E(X^2) - [E(X)]^2 = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

Important result on exponential distribution :

Show that the exponential distribution "lacks memory", i.e., if X follows an exponential distribution, then for every constant $a \geq 0$, it has

$P(Y \leq x / X \geq a) = P(X \leq x)$ for all x, where $Y = X - a$.

Proof : The p.d.f. for the exponential distribution with parameter θ is given by

$f(x) = \theta e^{-\theta x}, \quad \theta > 0, \quad x \geq 0$

We have $\quad P(Y \leq x / X \geq a) = \dfrac{P(Y \leq x \cap X \geq a)}{P(X \geq a)}$

$$\begin{aligned} P(Y \leq x \cap X \geq a) &= P(X - a \leq x \cap X \geq a) \qquad (\text{since } Y = X - a) \\ &= P(X \leq a + x \cap X \geq a) = P(a \leq X \leq a + x) \\ &= \int_a^{a+x} \theta e^{-\theta x}\, dx = \left[\theta \frac{e^{-\theta x}}{-\theta}\right]_{x=a}^{a+x} = -e^{-\theta(a+x)} + e^{-\theta a} = e^{-\theta a}(1 - e^{-\theta x}) \end{aligned}$$

$$P(X \geq a) = \int_a^{\infty} f(x)\, dx = \int_a^{\infty} \theta e^{-\theta x}\, dx = \left[\theta \frac{e^{-\theta x}}{-\theta}\right]_{x=a}^{\infty} = e^{-a\theta}$$

$$\therefore \quad P(Y \leq x / X \geq a) = \frac{P(Y \leq x \cap X \geq a)}{P(X \geq a)} = 1 - e^{-\theta x} \qquad (1)$$

$$P(X \leq x) = \int_0^{x} f(x)\, dx = \int_0^{x} \theta e^{-\theta x}\, dx = \left[\theta \frac{e^{-\theta x}}{-\theta}\right]_{x=0}^{x} = 1 - e^{-\theta x} \qquad (2)$$

From (1) and (2), we have $\quad P(Y \leq x / X \geq a) = P(X \leq x)$

Hypergeometric distribution :

When the population is finite and sampling is done without replacement, so that the events are stochastically dependent (although random), we obtain hypergeometric distribution.

Consider an urn with N balls, M of which are white and N-M are non-white. Suppose that we draw a sample of n balls from the urn at random without replacement, then the probability of getting k white balls out of n ($k \le n$) drawn balls is given by

$$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}} \quad \text{where} \binom{M}{k} = {}^{M}\mathrm{C}_{\mathrm{k}}$$

Def : A r.v. X is said to follow the hypergeometric distribution if it assumes only non-negative values and its p.d.f. is given by

$$p(x) = P(X = x) = h(x; N, M, n) = \begin{cases} \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}, x = 0,1,2,\ldots,n \\ 0, \quad \text{otherwise} \end{cases}$$

Hypergeometric distribution approximated as a binomial distribution :

Hypergeometric distribution tends to a binomial distribution if

$$N \to \infty \text{ and } \frac{M}{N} \to p$$

Problems on hypergeometric distribution :

1. A taxi cab has 12 Maruthi Swift cars and 8 Tata Vista cars. If 5 of these cars in the shop are in repair, then find the probability that
    (i) 3 of them are Maruthi Swift cars and 2 are Tata Vista cars
    (ii) atleast 3 of them are Maruthi Swift cars
    (iii) all 5 of them are of the same make.

2. 200 students of first year MCA class in a certain college are divided at random into 20 batches of 10 each for the annual practical examination in statistics. Suppose that the class consists of 40 resident students and 160 non-resident students. Let R denote the number of resident students in a batch. Find the probability that $R \geq 3$.

3. Find the probability that the income tax official will catch 3 income tax returns with illegitimate deductions, if the official randomly selects 5 returns among from 12 returns of which 6 returns contain illegitimate deductions.