MODULE-I

PROBABILITY DISTRIBUTIONS

Topic Learning Objectives:

Upon Completion of this Module, students will be able to:

- To apply the knowledge of the statistical analysis and theory of probability in the study of uncertainties.
- To use probability theory to solve random physical phenomena and implement appropriate distribution models.
- Apply discrete and continuous distributions in analyzing the probability models arising in engineering field.

Some basic definitions:

Trial and Event: Consider an experiment which, though repeated under essentially identical conditions, does not give unique results but may result in any one of the several possible outcomes. The experiment is known as a trial and the outcomes are known as events or casts. For example:

(i) Throwing of a die is a trial and getting 1 (or 2 or 3, ... or 6) is an event.

(ii) Tossing of a coin is a trial and getting head (H) or tail (T) is an event.

(iii) Drawing two cards from a pack of well-shuffled cards is a trial and getting a king and a queen are events.

Exhaustive Events: The total number of possible outcomes in any trial is known as exhaustive events or exhaustive cases. For example:

(i) In tossing of a coin there are two exhaustive cases, *viz.*, head and tai1.

(ii) In throwing of a die, there are six, exhaustive cases since anyone of the 6 faces 1, 2, ... , 6 may come uppermost.

(iii)In drawing two cards from a pack of cards the exhaustive number of cases is 52_{C_2}, since 2 cards can be drawn out of 52 cards in 52_{C_2} ways.

Favourable Events or Cases: The number of cases favourable to an event in a trial is the number of outcomes which entail the happening of the event. For example:

(i) In drawing a card from a pack of cards the number of cases favourable to drawing of an ace is 4, for drawing a spade 13 and for drawing a red card is 26.

(ii) In throwing of two dice, the number of cases favourable to getting the sum 5 is : (1,4) (4,1) (2,3) (3,2), *i.e.*, 4.

Mutually exclusive events: Events are said to be mutually exclusive or incompatible if the happening of any one of them precludes the happening of all the others, i.e., if no two or more of them can happen simultaneously in the same trial. For example:

(i) In throwing a die all the 6 faces numbered 1 to 6 are mutually exclusive since if any one of these faces comes, the possibility of others, in the same trial, is ruled out.

(ii) Similarly in tossing a coin the events head and tail are mutually exclusive.

Equally likely events: Outcomes of a trial are set to be equally likely if taking into consideration all the relevant evidences, there is no reason to expect one in preference to the others. For example:

(i) In tossing an unbiased or uniform coin, head or tail are likely events.

(ii) In throwing an unbiased die, all the six faces are equally likely to come.

Independent events: Several events are said to be independent if the happening (or non-happening) of an event is not affected by the supplementary knowledge concerning the occurrence of any number of the remaining events. For example:

(i) In tossing an unbiased coin the event of getting a head in the first toss is independent of getting a head in the second, third and subsequent throws.

(ii) If one draws a card from a pack of well-shuffled cards and replace it before drawing the second card, the result of the second draw is independent of the first draw. But, however, if the first card drawn is not replaced then the second draw is dependent on the first draw.

There are three systematic approaches to the study of probability as mentioned below.

Mathematical or Classical or 'a priori' Probability: If a trial results in n exhaustive, mutually exclusive and equally likely cases and m of them are favourable to the happening of an event E then the probability 'p' of happening of E is given by

$$p = P(E) = \frac{\text{Favourable number of cases}}{\text{Exhaustive number of cases}} = \frac{m}{n}$$

Since the number of cases favourable to the 'non-happening' of the event E are (n - m) ,the probability 'q' that E will not happen is given by

$$q = \frac{n-m}{n} = 1 - \frac{m}{n} = 1 - p \text{ gives } p + q = 1$$

Obviously p as well as q are non-negative and cannot exceed unity, *i.e*,

$0 \le p \le 1 , 0 \le q \le 1$.

Random Variable: A random variable is a rule which assigns a numerical value to each and every outcomes of the random experiment. It is nothing but a function from the sample space 'S' to the set of all real numbers, denoted as $f : S \to R$. Random Variables are usually denoted by X, Y, Z, ………. The set of all real numbers of a random variable X is called the range of X.

Example: 1. While tossing a coin, suppose the value 1 is assigned for the outcome Head (H) & 0 is assigned for Tail (T), then]

S = {H, T), X(H) =1 and X(T) = 0, then the range of X = {0, 1}

2) Tossing 3 fair coins up, Then S = {HHH, HHT, HTH, THH, TTH, THT, HTT, TTT}

Now X(HHH}= 3,

X(HHT) = X(HTH} = X(THH} = 2,

X(TTH} = X(THT} = X(HTT} = 1 &

X(TTT} = 0

& Range of X = {0, 1, 2, 3}

3) Let the random experiment be throwing a pair of 'die' and the sample space S associated with it is the set of all pair of numbers chosen from 1 to 6.

i.e. $S = \{(x, y) / x, y \in \{1, 2, 3, 4, 5, 6\}\}$

Then to each outcome (x, y) of S let us associate a random variable X = x + y

Now S = {(1, 1), (1, 2), ………, (2, 1), (2, 2), ………, (6, 5), (6, 6)}

Therefore X(1, 1) = 2, X(1, 2) = X(2, 1) = 3, ……….., X(6, 6) = 12

$\Rightarrow$ Range = {2, 3, 4, …….,12}

<u>There are two types of Random Variable:</u>

1. Discrete Random Variable 2. Continuous Random Variables

A variable X which takes finite or countably infinite number of values is called discrete random variable

Example: 1. throwing a die , random variable X is the number obtained

i.e. X(S) = {1, 2 , 3 , 4, 5, 6}, X is discrete random variable.(or whole number)

2. X is the total marks scored by a student in an examination (whole number)

A random variable whose range is uncountably infinite is called random variable. Here the range of variable is an interval of real numbers.

Example: 1. weight of articles

2. Observing the pointer on a speedometer / voltmeter

3. Conducting a survey on the life of electric bulbs.

According to the type of random variable we have two types of probability distributions

1. Discrete probability distribution   2. Continuous probability distribution.

1. <u>Discrete Probability Distribution</u>:

Probability mass function (p.m.f):

Let X be a discrete random variable and $p(x_i) = P[X = x_i]$, then $p(x_i)$ is the probability mass function (p.m.f.) of X if

(i) $p(x_i) \geq 0$ for all x_i

(ii) $\sum_i p(x_i) = 1$. $i.e.$ $p(x_1) + p(x_2) + p(x_3) + ... + p(x_n) = 1$

Discrete Probability Distribution: It is a systematic presentation of value taken by the random variable with the corresponding probabilities**.**

The set of values $[x_i,\ p(x_i)]$ is called a discrete probability distribution of the discrete random variable X.

The distribution function F(x) of the discrete random variable X is defined as

$F(x) = P(X \leq x) = \sum_{i=1}^{x} p(x_i),$ where x is any integer. Also it is known as **cumulative distribution function (c. d. f.)**

Example: The discrete probability distribution for X, the sum of the numbers which turn on tossing a pair of dice is

$X = x_i$	2	3	4	5	6	7	8	9	10	11	12
$p(x_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Note That $p(x_i) \geq 0$ & $\sum_i p(x_i) = 1$

2. <u>Continuous Probability Distribution</u>:

Probability density function (p.d.f.): Let X be a continuous random variable. Then a function f(x) is a p.d.f. of X if

(i) $f(x) \geq 0$, (ii) $\int_{-\infty}^{\infty} f(x)dx = 1$

The distribution function F(x) of the continuous random variable X is defined as

$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)dx$. Also it is known as **cumulative distribution function (c. d. f.)**

Properties of cumulative distribution function:

i) $F^1(x) = f(x) \geq 0, \Rightarrow$ F(x) is a non – decreasing function.

$ii)$ $F(-\infty) = 0;$ $iii)$ $F(\infty) = 1$

iv) If r is any real number, then $P(X \geq r) = \int_{r}^{\infty} f(x)dx$ &

$$P(X < r) = 1 - P(X \geq r) \text{ i.e. } P(X < r) = 1 - \int_{r}^{\infty} f(x)dx$$

$$v)\ P(a \leq X \leq b) = \int_{a}^{b} f(x)dx = \int_{a}^{-\infty} f(x)dx + \int_{-\infty}^{b} f(x)dx = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = F(b) - F(a)$$

Expectation: The mean value (μ) of the probability distribution of a random variable (variate) X is commonly known as its expectation and is denoted by E(X).

If $p(x_i)$ is the probability mass function of the variable X, then

$$E(X) = \sum_i x_i p(x_i) \text{ (for discrete distribution)}$$

If $f(x)$ is the probability density function of the variable X, then

$$E(X) = \int_{-\infty}^{\infty} x f(x)dx \text{ (for continuous distribution)}$$

Variance: Variance of a distribution is given by

$$\sigma^2 = \sum_i (x_i - \mu)^2 p(x_i) \text{ (for discrete distribution) Or}$$

$$= E(X^2) - (E(X))^2 \ \&$$

$$\sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)dx \text{ (for continuous distribution.)}$$

Note that σ is standard deviation of the distribution.

Problems:

1. For the following function

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

Find (i) k, (ii) E(X) & (iii) Var(X)

Solution: (i) $p(x_i) \geq 0$ if $k \geq 0$

$$\sum_i p(x_i) = 1 \Rightarrow 0.1 + k + 0.2 + 2k + 0.3 + k = 1 \Rightarrow 4k + 0.6 = 1 \Rightarrow k = \frac{0.4}{4} = 0.1$$

Then the new table is

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

(ii) $E(X) = \sum_i x_i p(x_i)$

$$= (-2)\times 0.1 + (-1)\times 0.1 + 0\times 0.2 + 1\times 0.2 + 2\times 0.3 + 3\times 0.1 = 0.8$$

(iii) $Var(X) = \sigma^2 = E(X^2) - (E(X))^2$

$$E(X^2) = \sum_i x_i^2 p(x_i) = 4\times 0.1 + 1\times 0.1 + 0 + 1\times 0.2 + 4\times 0.3 + 9\times 0.1 = 2.8$$

$$\therefore \sigma^2 = 2.8 - 0.8^2 = 2.16$$

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

$X = x_i$	0	1	2	3	4	5	6	7
$p(x_i)$	0	k	2k	2k	3k	k^2	$2k^2$	$7k^2 + k$

Find (i) k (ii) E(X) and (iii) P(X < 6)

Solution: (i) We must have $P(x_i) \geq 0 \Rightarrow k \geq 0$

Also, we have $\sum_i p(x_i) = 1$

$$\Rightarrow 0 + k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1 \Rightarrow 10k^2 + 9k - 1 = 0$$

$$\Rightarrow 10k^2 + 10k - k - 1 = 0 \Rightarrow 10k(k+1) - 1(k+1) = 0 \Rightarrow (k+1)(10k-1) = 0$$

$$\Rightarrow k = -1 \quad or \quad k = \frac{1}{10}. \ But\ k = -1\ is\ not\ possible \quad (\because k \geq 0)$$

$\therefore k = \frac{1}{10}$. Then the new table is

$X = x_i$	0	1	2	3	4	5	6	7
$p(x_i)$	0	1/10	1/5	1/5	3/10	1/100	1/50	17/100

(ii) $E(X) = \sum_i x_i p(x_i) = 0 + 1 \times \frac{1}{10} + 2 \times \frac{1}{5} + 3 \times \frac{1}{5} + 4 \times \frac{3}{10} + 5 \times \frac{1}{100} + 6 \times \frac{1}{50} + 7 \times \frac{17}{100}$

$= 3.66$

(iii) $P(X < 6) = 1 - P(X \geq 6) = 1 - [P(6) + P(7)]$

$$=10-\left[\frac{1}{50}+\frac{17}{100}\right]=1-\frac{19}{100}=\frac{81}{100}=0.81$$

3. Find which of the following function is a probability density function.

$(i)\ f_1(x)=\begin{cases}2x & 0<x<1\\ 0 & otherwise\end{cases}$ $\quad (ii)\ f_2(x)=\begin{cases}2x & -1<x<1\\ 0 & otherwise\end{cases}$

$(iii)\ f_3(x)=\begin{cases}|x| & |x|\le 1\\ 0 & otherwise\end{cases}$ $\quad (iv)\ f_4(x)=\begin{cases}2x & 0<x\le 1\\ 4-4x & 1<x<2\\ 0 & otherwise\end{cases}$

Solution: Condition for a p. d. f. are $f(x)\ge 0$ & $\int_{-\infty}^{\infty} f(x)dx=1$

(i) $Here\ f(x)\ge 0$ & $\int_{-\infty}^{\infty} f_1(x)dx=\int_0^1 2xdx=2\left[\frac{x^2}{2}\right]_0^1=1$

Hence $f_1(x)$ is p. d. f.

(ii) $f_2(x)$ can be written in the form $f_2(x)=\begin{cases}2x & -1<x<0\\ 2x & 0<x<1\\ 0 & otherwise\end{cases}$

In $-1<x<0,\ f_2(x)=2x<0$ & $\int_{-\infty}^{\infty} f_2(x)dx=\int_{-1}^{1} 2xdx=\left[x^2\right]_{-1}^{1}=0$

Hence both the conditions are not satisfied $\Rightarrow f_2(x)$ is not a p. d. f.

(iii) $f_3(x)=|x|\ge 0$ & $\int_{-\infty}^{\infty} f_3(x)dx=\int_{-1}^{1}|x|dx=\int_{-1}^{0}|x|dx+\int_0^1|x|dx$

$$=\int_{-1}^{0}-xdx+\int_0^1 xdx=-\left[\frac{x^2}{2}\right]_{-1}^{0}+\left[\frac{x^2}{2}\right]_0^1$$

$$=\frac{1}{2}+\frac{1}{2}=1$$

$\therefore f_3(x)$ is a p. d. f.

iv) $f_4(x)=2x>0\ \ in\ \ 0<x\le 1$. But $f_4(x)=4-4x<0\ \ in\ \ 1<x<2$

The first condition is not satisfied $\Rightarrow f_4(x)$ is not a p. d. f.

4. Find the constant k such that $f(x)=\begin{cases} kx^2, & 0<x<3 \\ 0, & otherwise \end{cases}$ is a p. d. f.

Also compute $(i)\ P(1<X<2),\ (ii)\ P(X\le 1)\ \ (iii)\ P(X>1)$ (iv) Mean (v) Variance.

Solution: The condition for p. d. f. is $f(x)\ge 0\ \&\ \int_{-\infty}^{\infty} f(x)dx=1$

Now $f(x)\ge 0\ \ if\ \ k\ge 0\ \&\ \int_{-\infty}^{\infty} f(x)dx=1 \Rightarrow \int_0^3 kx^2 dx=1 \Rightarrow k\left[\frac{x^3}{3}\right]_0^3=1\Rightarrow 9k=1\Rightarrow k=\frac{1}{9}$

$$\therefore\ f(x)=\begin{cases} \frac{x^2}{9}, & 0<x<3 \\ 0, & otherwise \end{cases}$$

$$(i)\ P(1<X<2)=\int_1^2 f(x)dx=\int_1^2 \frac{x^2}{9}dx=\frac{1}{9}\left[\frac{x^3}{3}\right]_1^2=\frac{1}{27}[8-1]=\frac{7}{27}$$

$$(ii)\ P(X\le 1)=\int_0^1 f(x)dx=\int_0^1 \frac{x^2}{9}dx=\frac{1}{9}\left[\frac{x^3}{3}\right]_0^1=\frac{1}{27}[1-0]=\frac{1}{27}$$

$$(iii)\ P(X>1)=1-P(X\le 1)=1-\frac{1}{27}=\frac{26}{27}$$ Or

$$P(X>1)=\int_1^{\infty} f(x)dx=\int_1^3 \frac{x^2}{9}dx=\frac{1}{9}\left[\frac{x^3}{3}\right]_1^3=\frac{1}{27}(27-1)=\frac{26}{27}.$$

$$(iv)\ Mean=\mu=\int_{-\infty}^{\infty} xf(x)dx=\int_0^3 x.\frac{x^2}{9}dx=\left[\frac{x^4}{36}\right]_0^3=\frac{1}{36}(81-0)=\frac{81}{36}=2.25=E(X)$$

$$(v)\ Variance=\sigma^2=E(X^2)-[E(X)]^2$$

Now $E(X^2)=\int_{-\infty}^{\infty} x^2 f(x)dx=\int_0^3 \frac{x^4}{9}dx=\frac{1}{9}\left[\frac{x^5}{5}\right]_0^3=\frac{1}{45}(243-0)=\frac{243}{45}=5.4$

Hence $Variance=\sigma^2=5.4-(2.25)^2=0.3375$

5. A continuous random variable has the distribution function

$F(x)=\begin{cases} 0 & x\le 1 \\ c(x-1)^4, & 1\le x\le 3 \\ 1, & x>3 \end{cases}$ Find c and also the p. d. f.

Solution: $f(x)=\dfrac{d}{dx}[F(x)]$

$$\therefore f(x)=\begin{cases}0, & x\le 1\\ 4c(x-1)^3, & 1\le x\le 3\\ 0, & x>3\end{cases}$$

Now $f(x)\ge 0$ $for\ c\ge 0$ & $\int_{-\infty}^{\infty} f(x)dx=1$

$$i.e.\quad \int_1^3 4c(x-1)^3dx=1\Rightarrow 4c\left[\frac{(x-1)^4}{4}\right]_1^3=1\Rightarrow c[16-0]=1\Rightarrow 16c=1\Rightarrow c=\frac{1}{16}$$

Thus the p. d. f., $f(x)=\begin{cases}0, & x\le 1\\ \frac{1}{4}(x-1)^3, & 1\le x\le 3\\ 0, & x>3\end{cases}$

6. Suppose that the error in the reaction temperature, in ^{0}C, for a controlled laboratory experiment is a Random variable X having the p. d. f.

$f(x)=\begin{cases}\frac{x^2}{3}, & -1<x<2\\ 0, & otherwise\end{cases}$ Find (i) F(x) and (ii) Use it to evaluate $P(0<X\le 1)$

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

$$F(x)=P[X\le x]=\int_{-\infty}^{x} f(t)dt$$

Case 1: $x\le -1,\ then\ f(x)=0$

Hence $F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{x}0dt=0$

Case 2: $-1<x<2$

Then $Fx)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{-1} f(t)dt+\int_{-1}^{x} f(t)dt=0+\int_{-1}^{x}\frac{t^2}{3}dt=\left[\frac{t^3}{9}\right]_{-1}^{x}=\frac{x^3+1}{9}$

Case 3: $X\ge 2$

Then $F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{-1} f(t)dt+\int_{-1}^{2} f(t)dt+\int_{2}^{x} f(t)dt$

$$=0+\int_{-1}^{2}\frac{t^2}{3}dt=\left[\frac{t^3}{9}\right]_{-1}^{2}=\frac{8+1}{9}=1$$

Hence $F(x)=\begin{cases} 0 & x\le -1 \\ \dfrac{x^3+1}{9}, & -1<x<2 \\ 1, & x\ge 2 \end{cases}$

$$P[0<x\le 1]=F(1)-F(0)=\frac{1+1}{9}-\frac{0+1}{9}=\frac{1}{9}$$

7. If the p. d. f. of a Random variable X is given by $f(x)=\begin{cases} 2kxe^{-x^2}, & for\ x>0 \\ 0, & for\ x\le 0 \end{cases}$

Find (a) the value of k and (b) distribution function for X.

Solution: We know that $\int_{-\infty}^{\infty} f(x)dx=1$

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

$$i.e. \int_{0}^{\infty} 2kxe^{-x^2}dx=1 \qquad (Put\ x^2=t\Rightarrow 2xdx=dt$$

$$\Rightarrow \int_{0}^{\infty} ke^{-t}dt=1\Rightarrow \left[-ke^{-t}\right]_0^{\infty}=1\Rightarrow -k(0-1)=1\Rightarrow k=1$$

Now $F(x)=P[X\le x]=\int_{-\infty}^{x} f(t)dt=0, \quad if\ x\le 0$

If $x>0$ $F(x)=\int_{-\infty}^{0} f(t)dt+\int_{0}^{x} f(t)dt=0+\int_{0}^{x} 2te^{-t^2}dt=\int_{0}^{x^2} e^{-z}dz=-e^{-z}\Big|_0^{x^2}=1-e^{-x^2}$

Hence $F(x)=\begin{cases} 1-e^{-x^2}, & for\ x\ge 0 \\ 0, & otherwise \end{cases}$

8. Find the c. d. f. of the random variable whose p. d. f. is given by

$$f(x)=\begin{cases} \frac{x}{2}, & for\ 0<x<1 \\ \frac{1}{2}, & for 1<x\le 2 \\ \dfrac{3-x}{2}, & for\ 2<x\le 3 \\ 0, & otherwise \end{cases}$$

Solution: Case 1: $X\le 0$, then $F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{0} 0dt=0$

Case 2: $0<x\le 1$, then $F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{0} odt+\int_{0}^{x}\frac{t}{2}dt=0+\left.\frac{t^2}{4}\right]_0^x=\frac{x^2}{4}$

Case 3: $1<x<2$, then

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{0} 0dt+\int_{0}^{1}\frac{t}{2}dt+\int_{1}^{x}\frac{1}{2}dt=\left.\frac{t^2}{4}\right]_0^1+\left.\frac{1}{2}t\right]_1^x=\frac{1}{4}+\frac{x}{2}-\frac{1}{2}=\frac{x}{2}-\frac{1}{4}=\frac{2x-1}{4}$$

Case 4: $2<x<3$, then

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{0} 0dt+\int_{0}^{1}\frac{t}{2}dt+\int_{1}^{2}\frac{1}{2}dt+\int_{2}^{x}\frac{3-t}{2}dt$$

$$=\left.\frac{t^2}{4}\right]_0^1+\left.\frac{1}{2}t\right]_1^2+\left.\frac{6t-t^2}{4}\right]_2^x=\frac{1}{4}+1-\frac{1}{2}+\frac{6x-x^2}{4}-\frac{8}{4}$$

$$=\frac{1+4-2+6x-x^2-8}{4}=\frac{6x-x^2-5}{4}$$

Case 5: $x>3$, then

$$F(x)=\int_{-\infty}^{x} f(t)dt=\int_{-\infty}^{0} 0dt+\int_{0}^{1}\frac{t}{2}dt+\int_{1}^{2}\frac{1}{2}dt+\int_{2}^{3}\frac{3-t}{2}dt+\int_{3}^{x}0dt$$

$$=\left.\frac{t^2}{4}\right]_0^1+\left.\frac{1}{2}t\right]_1^2+\left.\frac{6t-t^2}{4}\right]_2^3+0$$

$$=\frac{1}{4}+1-\frac{1}{2}+\frac{18-9}{4}-\frac{8}{4}=\frac{1+4-2+9-8}{4}=1$$

Hence $F(x)=\begin{cases} 0, & x \le 0 \\ \dfrac{x^2}{4}, & 0 < x \le 1 \\ \dfrac{2x-1}{4}, & 1 < x \le 2 \\ \dfrac{6x-x^2-5}{4}, & 2 < x \le 3 \\ 1, & x > 3 \end{cases}$

Exercise:

1. For the following function

x_i	0	1	2	3	4
$p(x_i)$	0.2	0.35	0.25	0.15	0.05

Find E(X) & V(X).

2. The p. d. f. of a continuous random variable is given by $f(x)=\begin{cases} kx(1-x)e^x; 0 \le x \le 1 \\ 0 \quad\quad\quad otherwise \end{cases}$

Find k and hence find the mean and the S.D.

Answers:

1. $E(X)=\sum_i x_i\, p(x_i)=1.5$, $Var(X)=\sigma^2=E(X^2)-(E(X))^2=1.25$

2. $k=\frac{1}{3}-e, mean=0.55\ \&\ S.D.=0.23$

__**Discrete Probability Distributions**__:

We have two Probability Distributions. They are

(1) Binomial distribution (2) Poisson distribution

Repeated trials:

A random experiment with only two possible outcomes categorized as **success** and **failure** is called a Bernoulli trial where the probability of success 'p' is same for each trial.

If a trial is repeated 'n' times and if 'p' is the probability of a success and 'q' that of a failure, then the probability of 'x' successes and (n – x) failures is given by $p^x q^{n-x}$. But these 'n' successes and (n – x) failures can occur in any of the ^nC_x ways in each of which the probability is same. Thus, the probability of 'x' successes is $^nC_x p^x q^{n-x}$.

__**Binomial Distribution**__: (James Bernoulli)

It is concerned with trials of a repetition nature in which only the occurrence or non – occurrence, success or failure, acceptance or rejection, yes or no of a particular event is of interest.

If we perform a series of independent trials such that for each trial p is the probability of a success and q that of a failure. Then the probability of x successes in a series of n trials is given by

$$P(X=x)=P(x)={}^{n}C_{x}p^{x}q^{n-x};\ \ x=0,1,2,3,\ldots\ldots.,\ n$$

P(X) is a Binomial Probability distribution.

We form the following probability distribution of [x, P(x)], where x = 0, 1, 2, ……, n

x	0	1	2			r			n
p(x)	q^n	${}^nC_1pq^{n-1}$	${}^nC_2p^2q^{n-2}$			${}^nC_rp^rq^{n-r}$			p^n

P(x) for different values of x = 0, 1, 2, ……., n are the successive terms in the binomial expansion of $(q + p)^n$. Therefore, this distribution is called the Binomial Distribution.

Now $\sum P(x)=q^n+{}^nC_1pq^{n-1}+{}^nC_2p^2q^{n-2}+\ldots\ldots\ldots+p^n=(q+p)^n=1^n=1$

Hence p(x) is a probability function. n & p are the parameters of distribution.

<u>Mean (Expectation) & Variance of the Binomial distribution</u>:

Mean:

$$E(X)=\mu=\sum_{x=0}^{n}xp(x)=\sum_{x=0}^{n}x.{}^nC_xp^xq^{n-x}$$

$$=0+1.{}^nC_1pq^{n-1}+2.{}^nC_2p^2q^{n-2}+\ldots\ldots..+{}^nC_np^nq^0$$

$$=npq^{n-1}+n(n-1)p^2q^{n-2}+\ldots\ldots..+np^n$$

$$=np\left(q^{n-1}+(n-1)pq^{n-2}+\ldots\ldots+p^{n-1}\right)$$

$$=np(q+p)^{n-1}=np.1^{n-1}=np$$

Hence

$$E(X)=\mu=np$$

Variance:

$$\sigma^2=E(X^2)-\{E(X)\}^2$$

$$=\sum_{x=0}^{n}x^2p(x)-(np)^2 \qquad \left[Now\ x^2=x(x-1)+x\ \&\ p(x)={}^nC_xp^xq^{n-x}\right]$$

$$=\sum_{x=0}^{n}[x(x-1)+x]\left({}^nC_xp^xq^{n-x}\right)-n^2p^2$$

$$= \sum_{x=0}^{n} x(x-1).\ {}^nC_x p^x q^{n-x} + \sum_{x=0}^{n} x.\ {}^nC_x p^x q^{n-x} - n^2 p^2$$

$$= 0 + 0 + 2.1.\ {}^nC_2 p^2 q^{n-2} + 3.2.\ {}^nC_3 p^3 q^{n-3} + ... + n(n-1).\ {}^nC_n p^n q^0 + np - n^2 p^2$$

$$= n(n-1) p^2 q^{n-2} + 3.2.\frac{n(n-1)(n-2)}{3.2.1} p^3 q^{n-3} + ... + n(n-1) p^n + np - n^2 p^2$$

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

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

$$= n^2 p^2 - np^2 + np - n^2 p^2 = np(1-p) = n\,p\,q$$

$$\therefore \sigma^2 = n\,p\,q$$

$$\sigma = \sqrt{n\,p\,q}$$

Binomial Frequency Distribution:

Fitting Binomial distribution: If ‘n’ independent trials constitute one experiment and this experiment be repeated ‘N’ times, then the frequency of ‘x’ success is $N\,p(x) = N \times \left({}^nC_x p^x q^{n-x}\right)$. The possible number of successes together with these expected frequencies constitutes binomial frequency distribution.

Application of binomial distribution:

This distribution is applied to problems concerning;

(i) Number of defectives in a sample from production line

(ii) Estimation of reliability of system

(iii) Number of rounds fired from a gun hitting a target.

(iv) Radar detection.

Problems:

1. The probability that a pen manufacture by a company will be defective is $\frac{1}{10}$. If 12 such pens are manufactured, find the probability that

(a) exactly two will be defective, (b) at least two will be defective

(c) none will be defective.

Solution: X: Number of defective pens, n = 12 p = probability of a defective pen = $\frac{1}{10} = 0.1$

q = 1 - p = 1- 0.1 = 0.9

$$\therefore P(X = x) = p(x) = {}^nC_x p^x q^{n-x} = {}^{12}C_x (0.1)^x (0.9)^{12-x}$$

(a) $P\,[exactly\,2\,will\,be\,defective] = P(X = 2) = {}^{12}C_2 (0.1)^2 (0.9)^{12-2} = 0.2301$

(b) $P[at\,least\,two\,will\,be\,defective] = P(X \geq 2) = 1 - P(X < 2)$

$$= 1 - [p(0) + p(1)] \quad (x = 0\,or\,1)$$

$$= 1 - [{}^{12}C_0 (0.1)^0 (0.9)^{12-0} + {}^{12}C_1 (0.1)^1 (0.9)^{12-1}]$$

$$= 1 - [(0.9)^{12} + 12 \times 0.1 \times (0.9)^{11}] = 1 - 0.6590 = 0.3410$$

(c) $P[that\,none\,will\,be\,defective] = P(X = 0) = {}^{12}C_0 (0.1)^0 (0.9)^{12} = 0.2824$

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 1000 such samples, how many would be expected to contain at least 3 defective parts?

Solution: Mean number of defective = 2 = n p = 20 p

$\Rightarrow$ Probability of a defective part $= p = \frac{2}{20} = 0.1$

$\Rightarrow$ probability of non-defective parts = $q = 1 - p = 0.9$

n = 20, p = 0.1 & q = 0.9

Hence $p(x) = {}^nC_x p^x q^{n-x} = {}^{20}C_x (0.1)^x (0.9)^{20-x}$

$\therefore$ Probability of at least 3 defectives in a sample of 20 is

$$P(X \geq 3) = 1 - P(X < 3)$$

$$= 1 - [p(0) + p(1) + p(2)]$$

$$= 1 - \left[{}^{20}C_0 (0.1)^0 (0.9)^{20-0} + {}^{20}C_1 (0.1)^1 (0.9)^{20-1} + {}^{20}C_2 (0.1)^2 (0.9)^{20-2}\right]$$

$$= 1 - [0.9^{20} + 20 \times 0.1 \times (0.9)^{19} + 10 \times 19 \times (0.1)^2 \times (0.9)^{18}]$$

$$= 1 - (0.9)^{18} [(0.9)^2 + 20 \times 0.1 \times 0.9 + 190 \times (0.1)^2]$$

$$= 1 - (0.9)^{18} \times 4.51 = 0.3231$$

There in 100 samples, the expected number of samples having at least 3 defectives is = 1000×0.3231=323.1=323

3. Fit a binomial distribution for the following data

X = x	0	1	2	3	4	Total
frequency	30	62	46	10	2	150 = N

Solution:

X = x	f	f.x
0	30	0
1	62	62
2	46	92
3	10	30
4	2	8
Total	$\sum f = N = 150$	$\sum f\,x = 192$

$$p(x) = {}^{n}C_x p^x q^{n-x} \quad x = 0, 1, 2, 3, 4 \ \& \ \ p+q=1$$

$$\bar{x} = \frac{\sum f\,x}{N} = \frac{192}{150} = 1.28 = \mu$$

$$\therefore np = 1.28 \Rightarrow p = \frac{1.28}{n} = \frac{1.28}{4} = 0.32 \ \& \ \ q = 1-p = 1-0.32 = 0.68$$

Hence $p(x) = {}^{4}C_x (0.32)^x (0.68)^{4-x}$, x = 0, 1, 2, 3, 4

Expected frequency = N × p(x)

$$p(0) = {}^{4}C_0 (0.32)^0 (0.68)^{4-0} = (0.68)^4 = 0.2138$$

$$E(X=0) = N \times p(0) = 150 \times 0.2138 = 32.07 \approx 32$$

$$p(1) = {}^{4}C_1 (0.32)^1 (0.68)^{4-1} = 4 \times (0.32)(0.68)^3 = 0.4025$$

$$E(X=1) = N \times p(1) = 150 \times 0.4025 = 60$$

$$p(2) = {}^{4}C_2 (0.32)^2 (0.68)^{4-2} = 6 \times (0.32)^2 (0.68)^2 = 0.2841$$

$$E(X=2) = N \times p(2) = 150 \times 0.2841 = 43$$

$$p(3) = {}^{4}C_3 (0.32)^3 (0.68)^{4-3} = 4 \times (0.32)^3 (0.68) = 0.0891$$

$$E(X=3) = N \times p(3) = 150 \times 0.0891 = 13$$

$$p(4) = {}^{4}C_4 (0.32)^4 (0.68)^{4-4} = 1 \times (0.32)^4 = 0.0105$$

$$E(X=4) = N \times p(4) = 150 \times 0.0105 = 2$$

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

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

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

$p(x \geq 3) = 1 - p(x < 3)$

$= 1 - p(x = 0,1,2) = 0.323$

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

5. The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contracted this disease, what is the probability that (i) at least 10 survive, (ii) from 3 to 8 survive.

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

Given; n = 15, p = 0.4 and q = 1-p = 0.6

(i) $p(X \geq 10) = p(10) + p(11) + p(12) + p(13) + p(14) + p(15) = 0.0338$

(ii) $p(3 \leq X \leq 8) = p(3) + p(4) + p(5) + p(6) + p(7) + p(8) = 0.8779$

Exercises:

1) Fit a binomial distribution to the following frequency distribution.

x	0	1	2	3	4	5
f	2	14	20	34	22	8

2. Out of 800 families with 5 children each, how many would you expect to have (a) 3 boys (b) 5 girls (c) either 2 or 3 boys? Assume equal probabilities for boys & girls.

3. If 10 percent of the rivets produces by a machine are defective, find the probability that out of 12 rivets chosen at random

(i) exactly 2 will be defective

(ii) at least 2 will be defective

(iii) none will be defective.

<u>Poisson Distribution:</u>

[Named after French Mathematician Simon Denis Poison (1781 – 1840)]

It is a distribution related to the probabilities of events which are extremely rare, but which have a large number of independent opportunities for occurrence.

Eg: The arrival of "Customers" is commonly modeled as a Poisson process in the study of simple queuing systems. The number of deaths by horse kick in an army corp.

<u>The Poisson distribution</u>: Poison distribution is limiting case of Binomial distribution. It can be derived as a limiting case of B.D. by making n very large and 'p' very small, keeping np fixed (= m say).

The probability of 'x' success out of n trials in a binomial distribution is

$$P(x) = {}^{n}C_{x} p^{x} q^{n-x}, \text{ where } q = 1 - p$$

$$= \frac{n(n-1)(n-2).......(n-x+1)}{x!} P^{x} q^{n-x}$$

As n tends to infinity (∞) p tends to 0 and taking np as a fixed quantity say np = m, $\Rightarrow p = \frac{m}{n}$.

we get,

$$P(x) = \frac{m\left(m-\frac{m}{n}\right)\left(m-2\frac{m}{n}\right)\left(m-3\frac{m}{n}\right)........\left(m-x\frac{m}{n}\right)}{x!}\left(1-\frac{m}{n}\right)^{n-x}$$

$$= \frac{m^{x}}{x!} \underset{n\to\infty}{Lim} \frac{\left(1-\frac{m}{n}\right)^{n}}{\left(1-\frac{m}{n}\right)^{x}} = \frac{m^{x}}{x!} e^{-m}, \text{ where x = 0, 1, 2, 3, ...}$$

$$\left(\because \underset{n\to\infty}{Lim}\left(1-\frac{m}{n}\right)^{n} = e^{-m}, as \underset{k\to 0}{Lim}(1+k)^{1/k} = e \ \& \ \underset{n\to\infty}{Lim}\left(1-\frac{m}{n}\right)^{x} = 1 \right)$$

'm' is the parameter, the mean number of occurrences.

Note: 1.P(x) is the probability of exactly 'x' occurrences.

2. P(x) is called Poisson probability function and 'x' is a Poisson variate.

The distribution of probability for x = 0, 1, 2, ……. is as follows:

x	0	1	2	3	……	……
$P(x)$	e^{-m}	$\frac{m}{1!}e^{-m}$	$\frac{m^2}{2!}e^{-m}$	$\frac{m^3}{3!}e^{-m}$	……	……

$\therefore$ we have $P(x) \geq 0$ &

$$\sum_{x=0}^{\infty} P(x) = e^{-m} + \frac{me^{-m}}{1!} + \frac{m^2}{2!}e^{-m} + \frac{m^2 e^{-m}}{2!} + ...$$

$$= e^{-m}\left[1+\frac{m}{1!}+\frac{m^2}{2!}+\frac{m^3}{3!}+.......\right] = e^{-m}.e^{m} = 1$$

Hence P(x) is a probability mass function.

Mean and Variance of Poison Distribution:

Mean $\mu = \sum\limits_{x=0}^{\infty} xP(x) = \sum\limits_{x=0}^{\infty} x.\frac{m^x e^{-m}}{x!} = \sum\limits_{x=1}^{\infty} \frac{m^x e^{-m}}{(x-1)!} = me^{-m}\sum\limits_{x=1}^{\infty}\frac{m^{x-1}}{(x-1)!}$

$$= me^{-m}\left[1+\frac{m}{1!}+\frac{m^2}{3!}+........\right] = m.e^{-m}.e^{m} = m$$

$\therefore mean\,(\mu) = m$

Variance (V) = E(X^2) – (E(X))^2

$$= \sum_{x=0}^{\infty} x^2 p(x) - m^2$$

$$= \sum_{x=0}^{\infty}[x(x-1)+x]\frac{m^x e^{-m}}{x!} - m^2$$

$$= \sum_{x=0}^{\infty} x(x-1)\frac{m^x e^{-m}}{x!} + \sum_{x=0}^{\infty} x\frac{m^x e^{-m}}{x!} - m^2$$

$$= \sum_{x=2}^{\infty}\frac{m^x e^{-m}}{(x-2)!} + \sum_{x=0}^{\infty} xp(x) - m^2$$

$$= m^2 e^{-m}\sum_{x=2}^{\infty}\frac{m^{x-2}}{(x-2)!} + m - m^2$$

$$= m^2 e^{-m}\left[1+\frac{m}{1!}+\frac{m^2}{2!}+........\right] + m - m^2$$

$$= m^2 e^{-m} e^{m} + m - m^2 = m^2 + m - m^2 = m$$

Hence $Variance\,(\sigma^2) = m$ & $SD\,(\sigma) = \sqrt{m}$

Problems:

1. A company receives three complaints per day on average. What is the probability of receiving more than one complaint on a particular day?

Solution: Here m = 3, $\therefore P(x) = \frac{m^x e^{-m}}{x!} = \frac{3^x e^{-3}}{x!}, \quad x = 0, 1, 2, 3.........$

Therefore, probability of receiving more than one complaint

$$= P(X>1) = P(2) + P(3) + ...$$

$$=1-P(X\le 1)$$
$$=1-[P(0)+P(1)]$$
$$=1-\left[\frac{3^0e^{-3}}{0!}+\frac{3^1e^{-3}}{1!}\right]=1-[0.0498+0.1494]=0.8008$$

2. If the probability of a bad reaction from a certain injection is 0.001, determine the chance that out of 2000 individuals more than two will get a bad reaction.

Solution: It follows a Poisson distribution as the probability of occurrence is very small.

Mean = m = np = 2000×0.001=2,

Hence $P(x)=\dfrac{2^x e^{-2}}{x!}$

Therefore, probability more than two will get a bad reaction is

$$=P(X>2)=1-P(X\le 2)$$
$$=1-[P(0)+P(1)+P(2)]$$
$$=1-\left[e^{-2}+\frac{2e^{-2}}{1}+\frac{2^2e^{-2}}{2!}\right]$$
$$=1-\frac{1}{e^2}(1+2+2)=1-\frac{5}{e^2}=0.3233$$

3. In a certain factory turning out razor blades, there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10, use Poisson distribution to calculate the approximate number of packets containing no, one defective & 2 defective blades respectively in a consignment of 10, 000 packets.

Solution: Here n = 10, p = 0.002, N = 10, 000

The m = np = 10 (0.002) = 0.02

Therefore $P(x)=\dfrac{m^x e^{-m}}{x!}=\dfrac{(0.02)^x e^{-0.02}}{x!}$

(i) Probability of no defective blade (i.e. X = 0) = $P(0)=\dfrac{(0.02)^0 e^{-0.02}}{0!}=0.9802$

$\therefore$ Number of packets containing no defective blade is $N\times p(0)=10000\times 0.9802=9802$

(ii) Number of packets containing one defective blade is

$$N\times P(1)=10000\times\frac{(0.02)^1e^{-0.02}}{1!}=196.$$

(iii) Number of packets containing two defective blades is

$$N \times P(2) = 10000 \times \frac{(0.02)^2 e^{-0.02}}{2!} = 1.96 \approx 2$$

4. Fit a Poisson distribution to the following data:

x:	0	1	2	3	4
f:	122	60	15	2	1

Find the corresponding theoretical estimation for f.

Solution:

x	f	f.x
0	122	0
1	60	60
2	15	30
3	2	6
4	1	4
Total	N =200	$\sum fx = 100$

$$\bar{x} = \frac{\sum fx}{N} = \frac{100}{200} = 0.5 = m,$$ the mean of Poisson distribution.

Hence $P(x) = \dfrac{m^x e^{-m}}{x!} = \dfrac{(0.5)^x e^{0.5}}{x!}, \; x = 0, 1, 2, 3, 4$

Hence the expected frequency for 'x' successes is

$$E_x = N \times P(x) = \frac{200 \times (0.5)^x e^{-0.5}}{x!},$$ where x = 0, 1, 2, 3, 4.

Putting x = 0, 1, 2, 3 , 4 we get

$$E_0 = N \times P(0) = \frac{200 \times (0.5)^0 e^{-0.5}}{0!} = 121,$$

$$E_1 = N \times P(1) = \frac{200 \times (0.5)^1 e^{-0.5}}{1!} = 61,$$

$$E_2 = N \times P(2) = \frac{200 \times (0.5)^2 e^{-0.5}}{2!} = 15,$$

$$E_3 = N \times P(3) = \frac{200 \times (0.5)^3 e^{-0.5}}{3!} = 3,$$

$$E_4 = N \times P(4) = \frac{200 \times (0.5)^4 e^{-0.5}}{4!} = 0,$$

Hence the theoretical frequencies are

x:	0	1	2	3	4
f:	121	61	15	3	0

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

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

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

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

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

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

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

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

Exercise:

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

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

4. In a certain factory turning out razor blades there is a small chance of 0.002 for any blade to be defective. The blades are supplied in packets of 10. Use Poisson distribution to calculate the approximate number of packets containing no defective, one defective two blades defective respectively in a consignment of 10,000 packets.
5. A manufacturer of cotter pins knows that 5% of his product is defective. If he sells cotter pins in boxes of 100 and guarantees that not more than 10 pins will be defective, what is the approximate probability that a box will fail to meet the guaranteed quality?
6. The number of accidents in a year to taxi drivers in a city follows a Poisson distribution with mean 3. Out of 1000 taxi drivers find approximately the number of the drivers with (i) No accidents in a year?

(ii) Exactly two accidents in a year? (iii) More than three accidents in a year?

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

4. 9802, 196, 2   5. $1 - e^{-5}\sum_{x=0}^{10}\frac{(5)^x}{x!}$ 6.

Normal distribution:

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

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

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

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

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

Remarks:

1. A random variable X with mean μ and variance σ^2 and following the normal law is expressed by $X \sim N(\mu, \sigma^2)$.
2. If $X \sim N(\mu, \sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$, is a standard normal variate with E(Z) = 0 and Var (Z) = 1 and we write $Z \sim N(0,1)$.
3. The p.d.f of standard normal variate Z is given by $\varphi(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}, -\infty < z < \infty$

   and the corresponding distribution function, denoted by $F(z)$ is given by

   $F(z) = P(Z \le z) = \int_{-\infty}^{z}\varphi(u)du = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}e^{-u^2/2}\,du.$

   Also, $F(-z) = \int_{-\infty}^{-z}\varphi(u)du = 1 - F(z)$.
4. The graph of f(x) is a famous 'bell-shaped' curve. The top of the bell is directly above the mean μ. For large values of σ, the curve tends to flatten out and for small values of σ, it has a sharp peak.

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

Properties of Normal Distribution:

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

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

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

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

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

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

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

Normally Distributed Variable:

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

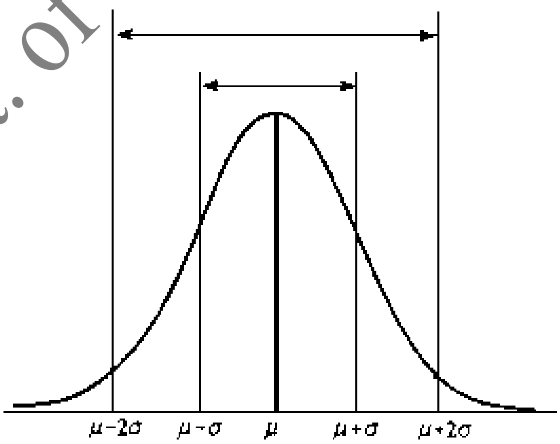

Fig. 4.1 Normal distribution curve

Problems:

1. A sample of 100 battery cells is tested to find the length of life, gave the following results. Mean = 12 hrs. Standard Deviation = 3 hrs. Assuming the data to be normally distributed what % of

battery cells are expected to have life (i) more than 15 hrs. (ii) less than 6hrs. (iii) between 10 & 14 hrs .

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

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

$$= P(z > 1)$$

= 0.5 - 0.3413

= 0.1587 = 16%

(ii) When x = 6

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

$$= P(z < -2)$$

= 0.5 - 0.4772

= 0.0228 = 2.28%

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

$$= P(-0.6667 < z < 0.6667)$$

$$= 2 * P(0 < z < 0.6667)$$

= 2*0.2485 = 0.497 = 50%

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

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

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

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

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

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

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

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

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

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

$$= 1 - 0.86214 = 0.13786$$

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

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

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

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

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

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

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

$$= 0.76539$$

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

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

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

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

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

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

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

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

= 1- (0.84134 - 0.15866) = 0.31732

Exercise:

1. In a test of 2000 electric bulbs, it was found that the life of a particular make was normally distributed with an average life of 2040 hours and standard deviation of 60 hours. Estimate the number of bulbs likely to burn for (i) more than 2150 hours (ii) less than 1950 hours (iii) more than 1920 hours but less than 2060 hours.
2. If X is a random variable which is distributed normally with mean 60 and standard deviation 5, find the probabilities of the following events $(i) 60 \leqslant X \leqslant 70, (ii) X > 40 (iii) X \leqslant 50$

3. Assume that mean height of soldiers to be 68.22 inches with a variance of 10.8 $inches^2$.How many soldiers in a regiment of 1000 would you expect to be (i) over six feet tall ? and (ii) below 5.5 feet ? Assume heights are normally distributed.
4. The daily wages of 1000 workermen are normally distributed around a mean of Rs.70 and with standard deviation of Rs.5. Estimate the number of workers whose daily wages will be (i) between Rs.70 and 72 ? (ii) more than Rs. 75 ? iii) less than Rs.63

Answers: 1. 0.0336 (67 bulbs), 0.0668 (134 bulbs), 0.6065 (1213). 2. 0.4772,0.9987,0.0228 3. 125,250 4. 155, 159, 81.

Exponential distribution:

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

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

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

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

Mean and variance of Exponential distribution

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

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

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

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

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

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

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

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

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

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

Problems:

1. If x is an exponential variate with mean α =5, evaluate (i) $P(0 < X < 1)$ (ii) $P(-\infty < X < 10)$ (iii) $P(X \le 0 \; or \; X \ge 1)$

Solution: $f(x) = \alpha e^{-\alpha x} \quad x > 0$; Given mean = $5 = \frac{1}{\alpha} \Rightarrow \alpha = \frac{1}{5}$

$$\therefore \; f(x) = \frac{1}{5} e^{-\frac{x}{5}}, x > 0$$

Hence (i) $P(0 < X < 1) = \int_0^1 f(x)dx = \int_0^1 \frac{1}{5} e^{-\frac{x}{5}} dx \quad = \frac{1}{5}\left[\frac{e^{-x/5}}{-1/5}\right]_0^1 = -(e^{-1/5} - 1)$

$$= 1 - e^{-0.2} = 0.1813$$

$$\therefore P(0 < X < 1) = 0.1813$$

(ii) $P(-\infty < X < 10) = \int_{-\infty}^{10} f(x)dx \; = \int_{-\infty}^{0} f(x)dx + \int_0^{10} f(x)dx$

$$= \int_0^{10} \frac{1}{5} e^{-x/5} dx \; = \frac{1}{5}\left[\frac{e^{-x/5}}{-1/5}\right]_0^{10}$$

$$= -\left(e^{-2} - 1\right) = 1 - e^{-2} = 0.8647$$

$$\therefore P(-\infty < X < 10) = 0.8647$$

(iii) $P(X \le 0 \; or \; X \ge 1) = P(X \le 0) + P(X \ge 1)$

$$= \int_{-\infty}^{0} f(x)dx + \int_1^{\infty} f(x)dx$$

$$= 0 + \int_1^{\infty} \frac{1}{5} e^{-x/5} dx \; = \frac{1}{5}\left[\frac{e^{-x/5}}{-1/5}\right]_1^{\infty}$$

$$-\left(0 - e^{-1/5}\right) = e^{-1/5} = 0.8187$$

$$\therefore \; P(X \le 0 \; or \; X \ge 1) = 0.8187$$

2. Let the mileage (in thousands of miles) of a particular tyre be a random variable X having the probability density $(x) = \begin{cases} \frac{1}{20} e^{\frac{-x}{20}} & ,x > 0 \\ 0, x \leq 0 \end{cases}$. Find the probability that one of these tyres will last (i) at most 10,000 miles (ii) anywhere between 16,000 to 24,000 miles (iii) at least 30,000 miles. Also, find the mean and variance of the given probability density function.

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

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

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

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

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

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

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

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

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

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

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

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

$= 0.3968$

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

Solution: Given, Mean = 800

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

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

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

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

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

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

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

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

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

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

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

Let X represents the sales per day

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

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

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

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

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

Exercise:

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

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

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

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

Video Links:

https://www.youtube.com/watch?v=82Ad1orN-NA

https://www.youtube.com/watch?v=c06FZ2Yq9rk
https://www.youtube.com/watch?v=N-IVFB8Rlfo
https://www.youtube.com/watch?v=d5iAWPnrH6w
https://www.youtube.com/watch?v=vjXLH7FXrj8