Probability Distribution

Random Variable

- A variable used to denote the numerical value of the outcome of an experiment is called the random variable, abbreviated as r.v.

- A random variable is denoted by capital letters X, Y, Z, ……. and its values are denoted by $x_1, x_2, \dots y_1, y_2, \dots z_1, z_2, \dots$ etc.

Discrete Random Variable

- A random variable is called a **discrete random variable** if it takes discrete (distinct) values

  $x_1, x_2, \ldots x_n \ldots in\ (a, b)$.

For example, number of floors in a building, number of kids in a family etc.

Continuous Random Variable

- A random variable is called a **continuous random variable** if it takes all values between an interval (a, b).
- For example, age, height, weight are continuous random variables.

Probability Distribution Of A discrete Random Variable

- Let X be a **discrete random variable** .
- Let $x_1, x_2, \ldots x_n \ldots$ be the possible values of X. with each possible outcome x_i we associate a number $p(x_i) = p(X = x_i) = p_i$ called the Probability of x_i.

- The numbers $p(x_i), i = 1, 2, \ldots n \ldots$ must satisfy the following conditions :
  1. $p(x_i) \geq 0 \; for \; all \; i$
  2. $\sum_{i=1} p(x_i) = 1.$
- The function p is called the **probability function** or **probability mass function (p.m.f.)** **or** **probability density function (p.d.f.)** of the random variable X and the set of pairs (x_i, p_i) is called the probability distribution of X.

- The probability distribution of a discrete random variable X taking values $x_1, x_2, x_3 \ldots\ x_n \ldots$ with probabilities $p_1, p_2, p_3 \ldots p_n \ldots where\ p_1 \geq 0$ and $\sum p_i = 1$ can be given in tabular form as.

X	x_1	x_2	x_3………	x_n……….
$P(X = x_i)$	p_1	p_2	p_3………	p_n……….

Distribution Function of A Discrete Random Variable X

- Suppose, X is a random variable taking values $x_1, x_2 \ldots x_n$ with probabilities $p(x_i), i = 1,2, \ldots n \ldots$ such that
  (i) $p(x_i) \geq 0$ for all i , **(ii)** $\sum p(x_i) = 1$
  Consider F defined by
  $$F(x_i) = P(X \leq x_i), i = 1,2,3, \ldots \ldots$$
  i.e.$F(x_i) = P(x_1) + P(x_2) + \cdots + P(x_i)$
  then the function F is called the
  cumulative distribution function
  or simply distribution function

- The set of pairs $\{x_i, F(x_i)\}$ is called the **cumulative probability distribution**.
- Consider the following table

X	x_1	x_2	x_3………	x_n………..
$F(x_i) = F(X = x_i)$	$P(x_1)$	$\sum_1^2 P(x_i)$	$\sum_1^3 P(x_i)$….	$\sum_1^n P(x_i)$….

The function F is called the distribution function

Probability Density Function Of A Continuous Random Variable

- A continuous function $y = f(x)$ such that

(i) $f(x)$ is integrable.

(ii) $f(x) \geq 0$

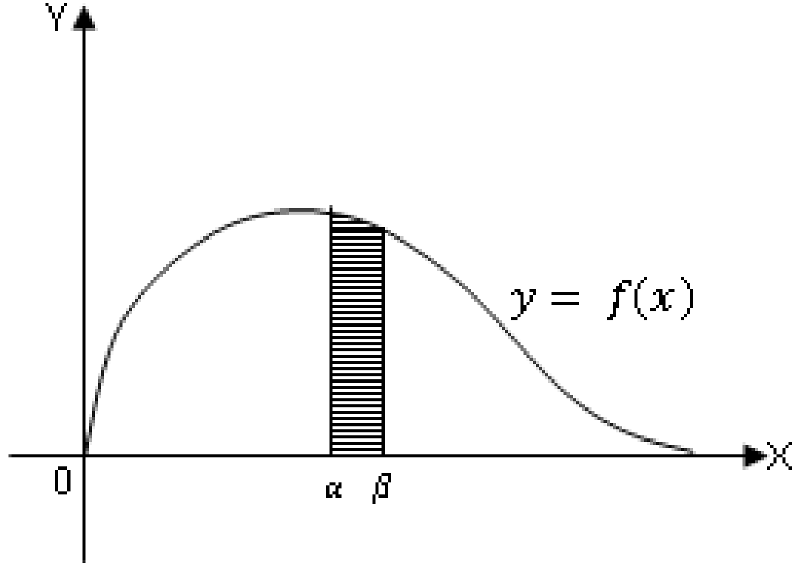
(iii) $\int_a^b f(x)dx = 1\ if\ X\ lies\ in\ [a, b]$ and

(iv) $\int_\alpha^\beta f(x)dx = P(\alpha \leq X \leq \beta)$

$$where\ a < \alpha < \beta < b$$

is called **probability desnity function** of a continuous random variable X.

- The curve given by $y = f(x)$ is called the **probability density curve** or simply **probability curve**.
- The expression $f(x)dx$ is usually denoted by $df(x)$ and is known as **probability differential**



- You know that for discrete random variable the probability at $X = C$ may not be zero.
- But, in a continuous random variable $P(X = C)$ is always zero because

$$P(X = C) = \int_C^C f(x)dx$$

and this definite integral is zero.
- For a continuous random variable X

$$P(\alpha \leq X \leq \beta) = P(\alpha < X < \beta)$$
$$= P(\alpha < X \leq \beta) = P(\alpha \leq X < \beta)$$

In other words we may include or may not include the end points in the interval.

Continuous Distribution Function

- If X is a continuous random variable X, having the probability density function f(x) then the function

$$F(x) = P(X \leq x)$$
$$= \int_{-\infty}^{x} f(t)dt, -\infty < x < \infty$$

is called **distribution function** or **cumulative distribution function** of the random variable X.

Expectation of a Random Variable

- If a discrete random variable X assumes values $x_1, x_2 \ldots\ x_n \ldots$ with probability $P_1, P_2 \ldots\ P_n \ldots$ respectively then the mathematical expectation of X denoted by $E(x)$ is defined by

$$\boldsymbol{E(X) = P_1X_1 + P_2X_2 + \cdots + P_nX_n \ldots}$$

$$\boldsymbol{= \sum P_iX_i} \qquad \text{where } \sum P_i = 1.$$

This value is also referred to as **mean** value of X.

- Let X be a continuous random variable with probability density function f (x).

  Then the mathematical expectation of X, denoted by E(X) is defined by

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

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

Laws of Expectation

- If X is a discrete random variable such that $x_i \geq 0$ for all i, then $\boldsymbol{E(X) \geq 0}$
- If X is a discrete (or continuous) random variable, a and b are constants then $\boldsymbol{E(aX + b) = aE(X) + b}$
- Putting $a = 0$. $E(b) = b$ i.e. expectation of a constant is the constant itself.
- Putting $b = 0$. $E(ax) = aE(X)$ i.e. for calculations the constant can be taken out.

- Putting $a = 1, b = -\bar{X}$, $E(X - \bar{X}) = 0$.
- **Theorem of Addition:**

  The expectation of the sum (or difference) of two (discrete or continuous) variates is equal to the sum (or difference) of their expectations.

  In symbols, $\boldsymbol{E(X \pm Y) = E(X) \pm E(Y)}$
- **Theorem of Multiplication**:

  The expectation of the product of **two independent** variates (discrete or continuous) is equal to the product of their expectations if the expectation exist.
- In symbols, $\boldsymbol{E(XY) = E(X).E(Y)}$

Note: It should be noted that the converse of the above theorem is not. If $E(XY) = E(X).E(Y)$ then it does not mean that X, Y are independent.

VARIANCE

- **Var (X)** $= E(X - \bar{X})^2$

$$= E[X - E(X)]^2$$

$$= E[X^2 - 2XE(X) + \{E(X)\}^2]$$

$$= E(X^2) - 2E(X).E(X) + [E(X)]^2$$

Var (X) $= \boldsymbol{E\left(X^2\right) - [E(X)]^2}$

Properties of Variance

- Variance of a constant is zero, $V(C) = 0$.
- If X is a random variate and a, b are constants

  then $\boldsymbol{V(aX+b) = a^2V(X)}$
- $\boldsymbol{V(aX) = a^2V(X)}$
- $\boldsymbol{V(X+b) = V(X)}$

- Note that $E(aX+b)=aE(X)+b$
  we do not have V $(aX+b)=aV(X)+b$.
  Instead, we have V$(aX+b)=a^2V(X)$.
- $\boldsymbol{V(a_1X_1+a_2X_2)=a_1{}^2V(X_1)+a_2{}^2\,V(X_2)}$
  Where X_1 and X_2 are independent random variates.
- If $a_1=1, a_2=1$, we get
  $\boldsymbol{V(X_1+X_2)=V(X_1)+V(X_2)}$ &
- If $a_1=1, a_2=-1$, we get
  $\boldsymbol{V(X_1-X_2)=V(X_1)+V(X_2)}$

EX.1 If X_1 has mean 4 and variance 9 and X_2 has mean -2 variance 4, and the two are independent, find $E(2X_1 + X_2 - 3)$ and $V(2X_1 + X_2 - 3)$

Solution: We have $E(X_1) = 4, V(X_1) = 9,$

$E(X_2) = -2$ and $V(X_2) = 4$

$$\begin{aligned}\therefore E(2X_1 + X_2 - 3) &= E(2X_1 + X_2) - 3 \\ &= 2E(X_1) + E(X_2) - 3 \\ &= 2(4) + (-2) - 3 = 3\end{aligned}$$

$$\therefore V(2X_1 + X_2 - 3) = V(2X_1 + X_2)$$
$$= 2^2 V(X_1) + V(X_2)$$
$$= 4(9) + 4 = 40$$

MEDIAN

- The median is the size of the item which lies at the middle
- For a continuous distribution the median M divides the area under the curve from x = a to x = b in to two equal parts.
- If M is the median then $\int_a^M f(x)\,dx = \int_M^b f(x)dx = \frac{1}{2}$
- By solving any one of the equations

$$\int_a^M f(x)\,dx = \frac{1}{2}, \quad or \quad \int_M^b f(x)\,dx = \frac{1}{2}$$

We can get the median M.

MODE

- Mode is the size of the item having maximum frequency
- Using the theory of maxima,

mode is obtained by solving the equation

$\frac{dy}{dx} = 0$ i.e. $f'(x) = 0$ with the condition that $\frac{d^2y}{dx^2} < 0\ i.e. f''(x) < 0$ and that x lies in the interval $[a, b]$ of X.

EX.2 Given the following probability function of a discrete random variable X

X	0	1	2	3	4	5	6	7
$P(X = x)$	0	c	2c	2c	3c	c^2	$2c^2$	$7c^2 + c$

Find **(i)** c **(ii)** $P(X \geq 6)$,

(iii) $P(X < 6)$,

(iv) k if, $P(X \leq k) > 1/2$, where $k \in N$,

(v) $P(1.5 < X < 4.5/X > 2)$

(vi) $E(X)$ **(vii)** $\mathrm{V(X)}$

Solution: Since $\sum p_i = 1$

$$\therefore 0 + c + 2c + 2c + 3c + c^2 + 2c^2 + 7c^2 + c = 1$$

$$\therefore 10c^2 + 9c = 1$$

$$c = -1 \; or \; 0.1$$

$c = -1$ is not possible as it represents probability

$$\therefore c = 0.1$$

Hence the probability distribution of X is

X	0	1	2	3	4	5	6	7
P $(X = x)$	0	0.1	0.2	0.2	0.3	0.01	0.02	0.17

(ii) $P(X \geq 6) = P(X = 6) + P(X = 7)$

$$= 0.02 + 0.17 = 0.19$$

(iii) $P(X < 6) = 1 - P(X \geq 6)$

$$= 1 - 0.19 = 0.81$$

(iv) To find k if, $P(X \leq k) > 1/2,$ where $k \in N$

$$P(X \leq 0) = 0$$

$$P(X \leq 1) = 0.1$$

$$P(X \leq 2) = 0.3$$

$$P(X \leq 3) = 0.5$$

$$P(X \leq 4) = 0.8$$

$$\therefore k = 4$$

(v) $P(1.5 < X < 4.5/X > 2)$

$$= \frac{P[(1.5 < X < 4.5) \cap (X > 2)]}{P(X > 2)}$$

$$= \frac{P(2 < X < 4.5)}{P(X > 2)}$$

$$= \frac{P(X = 3) + P(X = 4)}{P(X > 2)}$$

$$= \frac{0.5}{0.7} = \frac{5}{7}$$

(vi) $E(X) = \sum p_i x_i$

$$= 0 \times 0 + 1 \times 0.1 + 2 \times 0.2 + 3 \times 0.2 + 4 \times 0.3 + 5 \times 0.01 + 6 \times 0.02 + 7 \times 0.17$$

$$= 3.66$$

(vii) To find $V(X)$

$$E(X^2) = \sum p_i {x_i}^2$$

$$= 1 \times 0.1 + 4 \times 0.2 + 9 \times 0.2 + 16 \times 0.3 + 25 \times 0.01 + 36 \times 0.02 + 49 \times 0.17$$

$$= 16.8$$

$$V(x) = E(X^2) - [E(X)]^2$$

$$= 16.8 - (3.66)^2$$

$$= 3.4044$$

EX.3 The probability distribution of a discrete random variable is given by

X	0	1	2	3
$P(X = x)$	k	0.3	0.5	k

If $Y = X^2 + 2X$, find the probability distribution of Y, mean and variance of Y. What is the probability that $1 < y < 10$?

Solution: Since $\sum p_i = 1$

$$k + 0.3 + 0.5 + k = 1$$

$$k = 0.1$$

Hence, the p.d.f. of X is

X	0	1	2	3	Total
$P(X = x)$	0.1	0.3	0.5	0.1	1

Now, when X takes values $0, 1, 2, 3$.

$Y = X^2 + 2X$ takes values $0, 3, 8, 15$ resply

with respective probabilities

Y	0	3	8	15	Total
$P(Y=y)$	0.1	0.3	0.5	0.1	1

$$\therefore P(1 < Y < 10) = P(y = 3) + P(y = 8)$$
$$= 0.3 + 0.5 = 0.8$$

Now, mean $\bar{Y} = \sum p_i y_i$

$$= 0(0.1) + 3(0.3) + 8(0.5) + 15(0.1)$$
$$= 0 + 0.9 + 4.0 + 1.5 = 6.4$$

$$E(Y^2) = \sum p_i {y_i}^2$$
$$= 0^2(0.1) + 3^2(0.3) + 8^2(0.5) + 15^2(0.1)$$
$$= 0 + 2.7 + 32 + 22.5 = 57.2$$

$\therefore$ Variance $= E(Y^2) - \left(E(Y)\right)^2$

$$= 57.2 - 40.96 = 16.24$$

EX.4 A r.v. X has the following probability distribution

X	-2	-1	0	1	2
$P(X = x)$	$1/5$	$1/5$	$2/5$	$2/15$	$1/15$

Find the probability distribution of

(i) $V = X^2 + 1,$

(ii) $W = X^2 + 2X + 3$

V	1	2	5
$P(V)$	2/5	1/3	4/15

W	2	3	6	11
$P(W)$	1/5	3/5	2/15	1/15

EX. If the mean of the following distribution is 16 find m, n and variance

X	:	8	12	16	20	24
$P(X = x)$	:	$\frac{1}{8}$	m	n	$\frac{1}{4}$	$\frac{1}{12}$

EX. A function is defined as

$$f(x) = \begin{cases} 0\,, & \text{for } x < 2 \\ \frac{2x+3}{18}\,, & \text{for } 2 \le x \le 4 \\ 0, & \text{for } x > 4 \end{cases}$$

Show that $f(x)$ is a probability density function and find the probability that $2 < x < 3$.

EX. A continuous random variable has probability density function $f(x) = 6(x - x^2),\ 0 \leq x \leq 1$ Find

(i) mean **(ii)** Variance

(iii) median **(iv)** mode

(v) harmonic mean

(vi) $P(|x - m| < \sigma)$

(vii) $P(\mu - 2\sigma < X < \mu + 2\sigma)$

where $m =$ median, $\mu =$ mean, $\sigma =$ S.D.

Solution:

(i) $E(X) = \int_0^1 x \cdot 6(x - x^2)\, dx$

$$= 6\int_0^1 (x^2 - x^3)\, dx$$

$$= 6\left[\frac{x^3}{3} - \frac{x^4}{4}\right]_0^1$$

$$= 6\left(\frac{1}{3} - \frac{1}{4}\right)$$

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

(ii) $E(X^2) = \int_0^1 x^2 \cdot 6(x - x^2)\,dx$

$$= 6\int_0^1 (x^3 - 3x^4)\,dx$$

$$= 6\left[\frac{x^4}{4} - \frac{x^5}{5}\right]_0^1$$

$$= 6\left(\frac{1}{4} - \frac{1}{5}\right)$$

$$= \frac{6}{20} = \frac{3}{10}$$

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

$$= \frac{3}{10} - \frac{1}{4} = \frac{1}{20}$$

(iii) Median M is given by $\int_0^M 6(x - x^2)\, dx = \frac{1}{2}$

$$\therefore 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_0^M = \frac{1}{2}$$

$$\therefore 6\left[\frac{M^2}{2} - \frac{M^3}{3}\right] = \frac{1}{2}$$

$$\therefore 3M^2 - 2M^3 = \frac{1}{2}$$

$$\therefore 4M^3 - 6M^2 + 1 = 0$$

$$\therefore (2M - 1)(2M^2 - 2M - 1) = 0$$

$\therefore M = \frac{1}{2}$ or $\frac{2 \pm \sqrt{12}}{4} = \frac{1 \pm \sqrt{3}}{2}$

Of the three values only $M = 1/2$ lies in $(0, 1)$

Hence, $M = 1/2$

(iv) For mode we must have $f'(x) = 0$

and $f''(x) < 0$

$\therefore f'(x) = 6(1 - 2x) = 0$

$\therefore x = 1/2$

$f''(x) = -12$

$\therefore \text{Mode} = \frac{1}{2}$

(v) Harmonic mean is given by

$$\frac{1}{H} = \int_0^1 \frac{1}{x} \cdot 6(x - x^2)\, dx$$

$$= 6\int_0^1 (1 - x)\, dx$$

$$= 6\left[x - \frac{x^2}{2}\right]_0^1$$

$$= 6\left[1 - \frac{1}{2}\right] = 3$$

$$\therefore H = \frac{1}{3}$$

(vi) $P(|x-M|<\sigma) = P(-\sigma < x - M < \sigma)$

$$= P(M-\sigma < x < M+\sigma)$$

$$= P(0.5 - 0.2236 < x < 0.5 + 0.2230)$$

$$= P(0.2764 < x < 0.7236)$$

$$= 6\int_{0.2764}^{0.7236} (x - x^2)\,dx$$

$$= 6\left[\frac{x^2}{2} - \frac{x^3}{3}\right]_{0.2764}^{0.7236}$$

$$= 6\left[\left(\frac{(0.7236)^2}{2} - \frac{(0.7236)^3}{3}\right) - \left(\frac{(0.2764)^2}{2} - \frac{(0.2764)^3}{3}\right)\right]$$

$$= 6[0.1355 - 0.0311] = 0.6264$$

EX 5. If X is a continuous random variable with probability density function given by

$$f(x) = \begin{cases} k(x - x^3); & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find **(i)** k, **(ii)** mean ,

(iii) variance, **(iv)** median.

(v) Mode

EX. A continuous random variable X has p.d.f.

$$f(x) = kx^2e^{-x}, x \geq 0.$$

Find k, mean and variance

EX. The daily consumption of electric power (in millon kwh) is a random variable X with probability distribution function

$$f(x) = \begin{cases} kxe^{-x/3}, \text{for } x > 0 \\ \quad 0, \quad\quad \text{for } x \leq 0 \end{cases}$$

Find the value of k, the expectation of X and the probability that on a given day the electric consumption is more than expected value.

Ex. Let X be a continuous random variable with p.d.f. $f(x) = kx(1-x), 0 \leq x \leq 1$. Find k and determine a number b such that

$$P(X \leq b) = P(X \geq b)$$