3. EXPECTATION AND VARIANCE OF A DISCRETE RANDOM VARIABLE

Concept

If a random variable X takes values $x_1, x_2, x_3, \ldots, x_n$ with corresponding probabilities $p_1, p_2, p_3, \ldots, p_n$ respectively, the expectation of the random variable "X" is denoted by E(X) and is given by:

$$E(X) = \sum_{i=1}^{n} x_i p_i \text{ where } \sum_{i=1}^{n} p_i = 1$$

Note that E (X) = arithmetic mean of X

$$\therefore\ E(X) = \overline{X} = \frac{\sum_{i=1}^{n} f_i x_i}{\sum_{i=1}^{n} f_i}$$ where f_i is the frequency corresponding to x_i

Note that when frequencies are given and $N = \sum_{i=1}^{n} f_i$ = total number of observations, and $p_i = \frac{f_i}{N}$

Variance of random variable X is given by :

$$V(X) = E[X - E(X)]^2 = [E(X^2)] - [E(X)]^2 \text{ where } E(X^2) = \sum_{i=1}^{n} x_i^2 p_i$$

Square root of variance is called standard deviation. $\therefore$ S.D = $\sqrt{V(X)}$

Certain Properties of expected value :

1. E $(aX) = a$E (X), where "a" is a real constant.
2. E $(aX + b) = a$E $(X) + b$, where a and b are real constants.
3. E (X + Y) = E (X) + E (Y) where X and Y are independent variables
4. E (X • Y) = E (X) • E (Y) where X and Y are independent variables

Certain Properties of Variance :

1. V $(X + a)$ = V (X), where "a" is a real constant.
2. V $(X - a)$ = V (X).
3. V $(aX) = a^2$V (X)
4. V $(aX \pm b) = a^2$V (X)
5. V $(a) = 0$
6. V (X ± Y) = V (X) + V (Y) where X and Y are independent variables.

The Probability Mass Function :

For a discrete random variable the probability function is called as probability mass function, where the probability function can be listed in a tabular form. The probability mass function P(X) satisfies the condition : $P(X) = \sum_{i=1}^{n} p_i = 1$.

The Probability Density Function :

For a continuous random variable the probability function is called as probability density function. Here probability function cannot be listed in a tabular form. As the continuous random variable can assume any value within a given range, the distribution probabilities are measured over intervals and not single points. The area under the curve between two distinct points defines the probability for that interval.

Let "X" be a continuous random variable taking values over the interval $[c, d]$, then the probability density function P(X) of the variable "X" satisfies the following given properties:

1. $P(X) = \geq 0 \ \forall x \in [c, d]$
2. For any two distinct values $a, b \in [c, d]$, $P(a \leq X \leq b)$ = area under the probability curve between a and b.
3. Total area under the probability curve is equal to unity (one).
4. $\int_{-\infty}^{\infty} P(x)\, dx = 1$

Note : **For continuous random variable :**

1. $P(c \leq X \leq d) = \int_{c}^{d} P(X)\, dx$
2. $P(a) = 0$, where "a" is a real constant.
3. $P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b)$
4. $E(X) = \int_{-\infty}^{\infty} xP(x)\, dx$
5. $E(X^2) = \int_{-\infty}^{\infty} x^2 P(x)\, dx$
6. $V(X) = E[X - E(X)]^2 = [E(X^2)] - [E(X)]^2$

The expected value measures the center of the probability distribution - center of mass, while variance measures the amount of spread of data from the center value.

Note : If the cumulative distribution (of simply distribution function) is given say F(x), then the relation between PDF (probability density function) P(x) and F(x) is :

$$P(x) = \frac{d}{dx} F(x)$$

4. NUMERICALS : DISCRETE RANDOM VARIABLE

Problem : 1

For a random variable X the following data is available, find its probability mass function, expected value and the variance :

x_i	f_i
5	10
10	5
15	20
20	10
25	5

Solution :

The random variable X is discrete. The Probability mass function is given by

x_i	f_i	$P(X = x_i) = \frac{f_i}{N} = p_i$	$x_i p_i$	$x_i^2 p_i$
5	10	0.2	1	5
10	5	0.1	1	10
15	20	0.4	6	90
20	10	0.2	4	80
25	5	0.1	2.5	62.5
Total	$\sum f_i = 50 = N$	$\sum P(X = x_i) = 1$	$\sum_{i=1}^{n} x_i p_i = 14.5$	$\sum x^2{}_i p_i = 247.5$

The expected value is

$$E(X) = E(X) = \sum_{i=1}^{n} x_i p_i = 14.5$$

Variance is

$$V(X) = E[X - E(X)]^2 = [E(X^2)] - [E(X)]^2 = 247.5 - (14.5)^2 = 247.5 - 210.25 = 37.25$$

Problem : 2

In a game of throwing fair dice, A wins ₹ 60 if a 6 is thrown. He gains ₹ 30 if the dice shows 2 or 4 and he loses ₹ 30 if odd number occurs on the uppermost face of the dice. Find the probability mas function, expected gain and the variance.

Solution :

Number on the uppermost face of the die	Gain to A x_i (in ₹)	$P(X = x_i) = p_i$	$x_i p_i$	$x_i^2 p_i$
6	60	$\frac{1}{6}$	10	600
2 or 4	30	$\frac{2}{6}$	10	300
1,3 or 5	-30	$\frac{3}{6}$	-15	450
Total		$\sum P(X = x_i) = 1$	$\sum_{i=1}^{n} x_i p_i = 5$	$\sum x^2{}_i p_i = 1350$

The expected value is

$$E(X) = E(X) = \sum_{i=1}^{n} x_i p_i = 5$$

Variance is $V(X) = E[X - E(X)]^2 = [E(X^2)] - [E(X)]^2 = 1350 - 25 = 1325$

Problem : 3

For the following probability distribution :

Obtain (i) $P(X > 2)$ (ii) $P(X \leq 1)$ (iii) $P(X = 2 \text{ or } 3)$ (iv) $E(X)$ and (v) $V(X)$

X	–2	–1	0	1	2	3
P (X)	0.1	0.2	0.2	0.3	0.15	0.05

Solution :

i) $P(X > 2) = P(X = 3) = 0.05$

ii) $P(X \leq 1) = (P(X = -2) + P(X = -1) + P(X = 0) + P(X = 1) = 0.1 + 0.2 + 0.2 + 0.3 = 0.8$

iii) $P(X = 2 \text{ or } 3) = P(X = 2) + P(X = 3) = 0.15 + 0.05 = 0.2$

iv) E (X) :

X	–2	–1	0	1	2	3	TOTAL
$P(X) = p_i$	0.1	0.2	0.2	0.3	0.15	0.05	
$x_i p_i$	–0.2	–0.2	0	0.3	0.3	0.15	$\sum_{i=1}^{n} x_i p_i = 0.35$
$x_i^2 p_i$	0.4	0.2	0	0.3	0.6	0.45	$\sum x^2_i p_i = 1.95$

$$E(X) = E(X) = \sum_{i=1}^{n} x_i p_i = 0.35$$

Variance is $V(X) = E[X - E(X)]^2 = [E(X^2)] - [E(X)]^2 = 1.95 - 0.1225 = 1.8275$

Problem : 4

An unbiased coin is tossed thrice. X denotes the number of heads in the three tosses. If the probability distribution of X is given as :

$$\begin{aligned}(P(X = x_i) &= \frac{5}{16};\ x_i = 0,1 \\ &= \frac{1}{8};\ x_i = 2 \\ &= \frac{1}{4};\ x_i = 3 \\ &= 0;\ \text{otherwise}\end{aligned}$$

Find E (X) and V (X)

Solution :

$X = x_i$ (number of heads)	$P(X = x_i) = p_i$	$x_i p_i$	$x_i^2 p_i$
0	$\frac{5}{16}$	0	0
1	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{5}{16}$
2	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{3}{4}$	$\frac{9}{4}$
Total	$\sum P(X = x_i) = 1$	$\sum_{i=1}^{n} x_i p_i = \frac{21}{16}$	$\sum x^2_i p_i = \frac{49}{16}$

$$E(X) = E(X) = \sum_{i=1}^{n} x_i p_i = \frac{21}{16} = 1.3125$$

Variance is $V(X) = E[X - E(X)]^2 = [E(X^2)] - [E(X)]^2 = \frac{49}{16} - (1.3125)^2 = 3.0625 - 1.7227 =$ 1.3398

Problem : 5

Let X be a discrete random variable with probability mass function defined as :

$P(X = x_i) = \frac{x+1}{10}$; $x_i = 0, 1, 2, 3$

$= 0$; otherwise

Find expected value of X and Standard Deviation of X.

Solution :

$X = x_i$	$P(X = x_i) = p_i$	$x_i p_i$	$x_i^2 p_i$
0	$\frac{1}{10}$	0	0
1	$\frac{2}{10}$	$\frac{2}{10}$	$\frac{2}{10}$
2	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{12}{10}$
3	$\frac{4}{10}$	$\frac{12}{10}$	$\frac{36}{10}$
Total	$\sum P(X = x_i) = 1$	$\sum_{i=1}^{n} x_i p_i = \frac{20}{10} = 2$	$\sum x^2_i p_i = \frac{50}{10} = 5$

$$E(X) = E(X) = \sum_{i=1}^{n} x_i p_i = 2$$

Variance is $V(X) = E[X - E(X)]^2 = [E(X^2)] - [E(X)]^2 = 5 - 4 = 1$

$\therefore$ S.D. $= \sqrt{V(X)} = 1$

Problem : 6

If X denotes the number of heads appearing when an unbiased coin is tossed thrice, find the expected value and variance of X, if $P(X = 0) = 0.125$, $P(X = 1) = P(X = 2) = 0.375$, $P(X = 3) = p$.

Solution :

$\sum P(X = x_i) = 1$

$0.125 + 0.375 + 0.375 + p = 1$

$p = 0.125$

To find the expected value of X :

$X = x_i$	$P(X = x_i) = p_i$	$x_i p_i$	$x_i^2 p_i$
0	0.125	0	0
1	0.375	0.375	0.375
2	0.375	0.75	1.5
3	0.125	0.375	1.125
Total	$\sum P(X = x_i) = 1$	$\sum_{i=1}^{n} x_i p_i = 1.5$	$\sum x^2_i p_i = 3$

$$E(X) = E(X) = \sum_{i=1}^{n} x_i p_i = 1.5$$

Variance is $V(X) = E[X - E(X)]^2 = [E(X^2)] - [E(X)]^2 = 3 - 2.25 = 0.75$