Chapter 9 Mean and Variance

Day – 1

Introduction

Corresponding to every outcome of a random experiment, we can associate a real number. This correspondence between the elements of the sample space associated to a random experiment and the set of real numbers id defined as a random variable. If a random variable assumes countable number of values, it is called a discrete random variable. Otherwise, it is known as continuous random variable. We shall study these two types of random variables in the following sections.

Discrete Random Variable

Let S be the sample space associated with a given random experiment. Then, a real valued function X which assigns to each event $w \in S$ to a unique real number X (w) is called a random variable.

Thus, a random variable associated with a given random experiment associates every event to a unique real number as discussed below.

Consider a random experiment of tossing three coins. The sample space of eight possible outcomes of this experiment is given by

S = {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

Let X be a real valued function on S, defined by

X(w) = number of heads in $w \in S$.

Then, X is a random variable such that:

X(HHH) = 3, X(HHT) = 2, X(HTH) = 2, X(THH) = 2

X (HTT) = 1, X(THT) = 1, X(TTH) = 1, and X(TTT) = 0

Also, if w denotes the event "getting two heads", then

w = {HTH, THH, HHT}

And X(w) = S

Similarly, X associates every other compound event to a unique real number.

For the random variable X, we have range (X) = {0, 1, 2, 3} and we say that X is a random variable such that it assumes values 0, 1, 2, 3. This random variable can also be described as the number of heads in a single throw of three coins.

Now, consider the random experiment of throwing an unbiased die. Let Y be a real valued function defined on the sample space S = {1, 2, 3, 4, 5, 6} associated with the random experiment, defined by

$$Y(w) = \begin{cases} 1, \text{if the outcome is an even number} \\ -1, \text{if the outcome is an odd number} \end{cases}$$

Clearly, Y is a random variable such that:

Y(1) = – 1, Y(2) = 1, Y(3) = – 1, Y(4) = 1, Y(5) = – 1 and Y(6) = 1.

Here, range (Y) = {– 1, 1}. Therefore, we say that Y is a random variable such that it assumes values – 1and 1.

Illustration

Consider a random experiment of tossing three coins. Let X be a real valued function defined on the sample space

S = {HHH, HHT, HTH, THH, HTT, TTH, THT, TTT}such that
X(*w*) = Number of tails in X(*w*) = S.

Then X, is a random variable such that

X(HHH) = 0, X(HHT) = 1, X(HTH) = 1, X(THH) = 1, X(HTT) = 2,
X(THT) = 2, X(TTH) = 2, and X(TTT) = 3

Clearly, range of X is {0, 1, 2, 3}

Illustration

Consider a random experiment of throwing a six faced die. Let X denote the number on the upper face of the die. Then,

X(1) = 1, X(2) = 2, X(3) = 3, X(4) = 4, X(5) = 5 and X(6) = 6

Clearly, X is a random variable which assumes values 1, 2, , 4, 5, 6 i.e., range of X = {1, 2, 3, 4, 5, 6}.

Illustration

Let there be a bag containing 5 white, 4 red and 3 green balls. Three balls are drawn. If X denotes the number of green balls in the draw. Then, X can assume values 0, 1, 2, 3. Clearly, X is a random variable with its range = {0, 1, 2, 3}.

Illustration

A pair of dice is thrown. If X denotes the sum of the numbers on two dice, then X assumes values 2, 3, 4,…,12. Clearly, X is a random variable with its range {2, 3, 4,…,12}.

Probability Distribution

In the previous section, we have defined random variable. Now, consider a random experiment in which three coins are tossed simultaneously (or a coin is tossed three times). Le X be a random variable defined on the sample space.

S = {HHH, THH, HHT, THT, TTH, HTT, HTT, TTT} such that

H(w) = number of heads in $w \in$ S.

Clearly, X assumes value 0, 1, 2, 3.

Now, P(X = 0) = Probability of getting no head $= \mathrm{P(TTT)} = \frac{1}{8}$

P(X = 1) = Probability of getting one head

$= \mathrm{P(HTT\ \ or\ \ THT\ \ or\ \ TTH)} = \frac{3}{8}$

P(X = 2) = Probability of getting two head

$= \mathrm{P(HHT\ \ or\ \ THH\ \ or\ \ HTH)} = \frac{3}{8}$

P(X = 3) = Probability of getting 3 heads

$= \mathrm{P(HHH)} = \frac{1}{8}$

These values of X and the corresponding probabilities can be exhibited as under:

X:	0	1	2	3
P(X):	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

This tabular representation of the values of a random variable X and the corresponding probabilities is known of the probability distribution.

The formal definition of the probability distribution of a random variable is as given below.

Probability Distribution

If a random variable X takes values $x_1, x_2, \ldots\ldots x_n$ with respective probabilities $p_1, p_2, \ldots p_n$, then

X:	x_1	x_2	x_3	….	xn
P(X):	p_1	p_2	p_3	….	p_n

Is known as the probability distribution of X.

Thus, a tabular description giving the values of the random variable along with the corresponding probabilities is called its probability distribution.

Remark – 1

The probability distribution of a random variable X is defined only when we have the various values of the random variable e.g. $x_1, x_2, \ldots x_n$ together with respective probabilities $p_1, p_2, \ldots p_n$ satisfying $\sum_{i=1}^{n} p_i = 1$.

Remark – 2

If X is a random variable with the probability distribution

X:	x_1	x_2	x_3	….	xn
P(X):	p_1	p_2	p_3	….	p_n

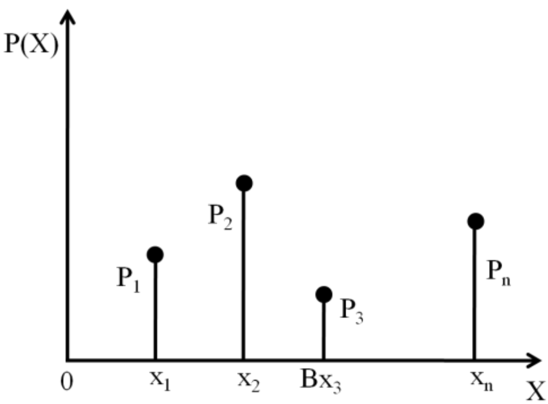
Then,

$$P(X \le x_i) = P(X = x_1) + P(X = x_2) + \ldots + P(X = x_i)$$
$$= p_1 + p_2 + \ldots + p_i$$
$$P(X \le x_i) = P(X = x_1) + P(X = x_2) + \ldots + P(X = x_{i-1})$$
$$= p_1 + p_2 + \ldots + p_{i-1}$$
$$P(X \ge x_i) = P(X = x_i) + P(X = x_{i+1}) + \ldots + P(X = x_n)$$
$$= p_{i+1} + p_{i+2} + \ldots + p_n$$

Also,

$$P(X \le x_i) = 1 - P(X < x_i), P(X > x_i) = 1 - P(X \le x_i)$$
$$P(X \le x_i) = 1 - P(X > x_i) \text{ and } P(X < x_i) = 1 - P(X \ge x_i)$$
$$P\left(x_i \le X \le x_j\right) = P(X = x_i) + P(X = x_{i+1}) + \ldots + P(X = x_j)$$
$$P\left(x_i < X < x_j\right) = P(X = x_{i+1}) + P(X = x_{i+2}) + \ldots + P(X = x_{j-1})$$

The graphical representation of a probability distribution is as follows:



Illustration

Determine which of the following can be probability distribution of a random X:

(i):- X:	0	1	2		
P(X):	0.4	0.4	0.2		
(ii):- X:	0	1	2		
P(X):	0.6	0.1	0.2		
(iii):- X:	0	1	2	3	4

P(X): 0.1 0.5 0.2 – 0.1 0.3

Solution

We have,

(i):- P(X = 0) + P(X = 1) + P(X = 2) = 0.4 + 0.4 + 0.2 = 1.

Hence, the given distribution of probabilities is a probability distribution of random variable X.

(ii):- P(X = 0) + P(X = 1) + P(X = 2) =0.6 + 0.1 + 0.2 = 0.9 ≠ 1.

Hence, the given distribution of probabilities is not a probability distribution.

(iii):- We have,

P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)

= 0.1 + 0.5 + 0.2 – 0.1 + 0.3 = 1

But, P(X = 3) = 0.1 < 0.

So, the given distribution of probabilities is not a probability distribution.

Illustration

An unbiased die is rolled. If the random variable X is defined as

$$\mathrm{X(w)} = \begin{cases} 1, \text{if the outcome w is an even number} \\ 0, \text{if the outcome w is and odd number.} \end{cases}$$

Find the probability distribution of X.

Solution

In a single throw of die either we get an even number or we get an odd number. Thus, the possible values of the random variable X are 0 and 1.

Now,

$$\mathrm{P(X = 0)} = \text{Probability of getting an odd number} = \frac{3}{6} = \frac{1}{2}$$

$$\mathrm{P(X = 1)} = \text{Probability of getting an even number} = \frac{3}{6} = \frac{1}{2}$$

Thus, the probability distribution of the random variable X is given by

X:	0	1
P(X):	$\frac{1}{2}$	$\frac{1}{2}$

Mean of a Discrete Random Variable

If X is a discrete random variable which assumes values x_1, x_2, x_3,…,x_n with respective probabilities p_1, p_2, p_3,…,p_n, then the mean $\bar{X}$ of X is defined as

$$\bar{X} = p_1x_1 + p_2x_2 + \dots + p_nx_n \quad \text{or} \quad \bar{X} = \sum_{i=1}^{n} p_ix_i$$

Remark – 1

The mean of a random variable X is also known as its mathematical expectation or expected value and is denoted by E(X).

Remark – 2

In case of a frequency distribution x_i/f_i; i= 1, 2,…,n the mean $\bar{X}$ is given by

$$\bar{X} = \frac{1}{N}(f_1x_1 + f_2x_2 + \ldots + f_nx_n)$$

$$\Rightarrow \quad \bar{X} = \frac{f_1}{N}x_1 + \frac{f_2}{N}x_2 + \ldots + \frac{f_n}{N}x_n$$

$$\Rightarrow \quad \bar{X} = p_1x_1 + p_2x_2 + \ldots + p_nx_n, \text{ where } p_i = \frac{f_i}{N}$$

Thus, if we replace $\frac{f_i}{N}$ by p_i in the definition of mean, we obtain the mean of a discrete random variable. Consequently, the term 'mean' is appropriate for the sum $\sum_{i=1}^{n} p_i\, x_i$.

Note:-

The mean of a random variable means the mean of its probability distribution.

Illustration

In a single throw of a die, if X denotes the number on its upper face. Find the mean of X.

Solution

Clearly, X can take the values 1, 2, 3, 4, 5, 6each with probabilities 1/6.
So, the probability distribution of X is as given below:

X:	1	2	3	4	5	6
P(X):	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$\therefore \quad \bar{X} = \frac{1}{6}\times 1 + \frac{1}{6}\times 2 + \frac{1}{6}\times 3 + \frac{1}{6}\times 4 + \frac{1}{6}\times 5 + \frac{1}{6}\times 6$$

$$\Rightarrow \quad \bar{X} = \frac{1}{6}\times(1+2+3+4+5+6) = \frac{1}{6}\times\frac{6(6+1)}{2} = \frac{7}{2}$$

Illustration

If a pair of dice is thrown and X denotes the sum of the numbers on them. Find the probability distribution of X. Also, find the expectation of X.

Solution

In a single throw of a pair of dice the sum of number on them can be 2, 3, 4,….,12. So, X can assume values 2, 3, 4,….,12. The probability distribution of X is as given below:

X:	2	3	4	5	6	7	8	9	10	11	12
P(X):	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$\therefore \quad E(X) = \frac{1}{36}\times 2 + \frac{2}{36}\times 3 + \frac{3}{36}\times 4 + \frac{4}{36}\times 5 + \frac{5}{36}\times 6 + \frac{6}{36}\times 7$
$+ \frac{5}{36}\times 8 + \frac{4}{36}\times 9 + \frac{3}{36}\times 10 + \frac{2}{36}\times 11 + \frac{1}{36}\times 12$

$\Rightarrow \quad E(X) = \frac{1}{36}[2 + 6 + 12 + 20 + 30 + 42 + 50 + 36 + 30 + 22 + 12]$

$\Rightarrow \quad E(X) = \frac{252}{36} = 7.$

Variance of a Discrete Random Variable

If X is a discrete random variable which assumes values $x_1, x_2, x_3, \ldots, x_n$ with the respective probabilities $p_1, p_2, \ldots, p_n$, the variance of X defined as

$$\text{Var}(X) = p_1(x_1 - \overline{X})^2 + p_2(x_2 - \overline{X})^2 + \ldots. + p - n\,(x_n - \overline{X})^2$$

$\Rightarrow \quad \text{Var}(X) = \sum_{i=1}^{n} p_1(x_1 - \overline{X})^2$, where $\overline{X} = \sum_{i=1}^{n} p_i x_i$ is the mean of X.

Now, $\quad \text{Var}(X) = \sum_{i=1}^{n} p_1(x_1 - \overline{X})^2$

$\Rightarrow \quad \text{Var}(X) = \sum_{i=1}^{n} p_i(x_i^2 - 2x_i\bar{X} + \overline{X}^2)$

$\Rightarrow \quad \text{Var}(X) = \sum_{i=1}^{n} p_i x_i^2 - 2\bar{X}\left(\sum_{i=1}^{n} p_i x_i\right) + \overline{X^2}\left(\sum_{i=1}^{n} p_i\right)$

$\Rightarrow \quad \text{Var}(X) = \sum_{i=1}^{n} p_i {x_i}^2 - 2\bar{X}.\bar{X} + \overline{X^2} \qquad [\because \sum_{i=1}^{n} p_i]$

$\Rightarrow \quad \text{Var}(X) = \sum_{i=1}^{n} p_i {x_i}^2 - 2\bar{X}^2 + \overline{X^2}$

$\Rightarrow \quad \text{Var}(X) = \sum_{i=1}^{n} p_i {x_i}^2 - \overline{X^2}$

$\Rightarrow \quad \text{Var}(X) = \sum_{i=1}^{n} p_i {x_i}^2 - \left(\sum_{i=1}^{n} p_i x_1\right)^2$

Thus,

$$\text{Var}(X) = \sum_{i=1}^{n} p_i {x_i}^2 - \left(\sum_{i=1}^{n} p_i x_1\right)^2$$

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

Remark

The variance of a discrete random variable X is also known as its second central moment μ_2.

Illustration

Find the mean and variance of the number of heads in the two tosses of a coin.

Solution

Let X denote the number of heads in the two tosses of a coin. Then X can take values 0 , 1 or 2 such that

P(X = 0) = Probability of getting no head = P(TT) = 1/4

P(X = 1) = Probability of getting one head

$\Rightarrow \quad P(X = 1) = P(HT \text{ or } TH) = \frac{2}{4} = \frac{1}{2}$

P(X = 2) = Probability of getting both heads $= P(HH) = \frac{1}{4}$

Thus, the probability distribution of X is as given below:

X:	0	1	2
P(X):	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Computation of mean and variance

x_i	$p_i = P(X = x_i)$	$p_i x_i$	$p_i x_1^2$
0	1/4	0	0
1	1/2	1/2	1/2
2	¼	1/2	1
		$\sum p_i x_i = 1$	$\sum p_i x_i^2 = 3/2$

We have,

$$\sum p_i x_i = 1 \text{ and } \sum p_i x_i^2 = \frac{3}{2}$$

∴ $\bar{X} = \text{Mean} = \sum p_i x_i = 1$

$$\text{Var(X)} = \sum p_i x_i^2 - (\sum p_i x_i)^2 = \frac{3}{2} - 1 = \frac{1}{2}$$

Hence, Mean = 1 and Variance = $\frac{1}{2}$.