Random variable

Definition

Let $\mathcal{S}$ be a sample space associated with a random experiment E, then a random variable X is a real valued function defined on $\mathcal{S}$.
i.e., $X : \mathcal{S} \to \mathcal{R}$ such that for each element $s \in \mathcal{S}$, there is a unique real number $X = X(s)$ associated.

Example 1

Consider the random experiment of tossing a coin two times in succession. Then the sample space is $\mathcal{S} = \{HH, HT, TH, TT\}$.
Let X denotes the number of heads obtained, then X is a random variable and for each outcome, its value is given as,

$$X(HH) = 2, X(HT) = 1, X(TH) = 1, X(TT) = 0.$$

In this case we can write, the range of X is $\{0, 1, 2\}$.

Example 2

A person plays a game of tossing a coin thrice. For each head, he is given Rs 2 by the organizer of the game and for each tail, he has to give Rs 1.50 to the organizer. Let X denote the amount gained or lost by the person. Then X is a random variable.

Here, $\mathcal{S} = \{HHH, HTT, THH, TTH, THT, TTT, HHT, HTH\}$. Then,

$X(HHH) = 3 \times 2 = 6$

$X(HTT) = X(THT) = X(TTH) = (1 \times 2) - (2 \times 1.50) = -1$

$X(HHT) = X(HTH) = X(THH) = (2 \times 2) - (1 \times 1.50) = 2.50$

$X(TTT) = 3 \times 1.50 = 4.50.$

Hence the range set of the random variable X is $\{-1, -2.50, 4.50, 6\}$.

There are two types of one dimensional random variables;

1. Discrete random variable
2. Continuous random variable

Discrete random variable

A random variable X is said to be **discrete** if X assumes only finite number of values or countably infinite values.

Example

The number of heads in four tosses of a coin is a discrete random variable. Because, here X takes only the values $0, 1, 2, 3, 4$.

Continuous random variable

A random variable X is said to be **continuous** if X assumes any value within an interval.

Example

The weights of a group of individuals in a class.

Discrete probability distribution

Let X be a discrete random variable which assumes values $x_1, x_2, x_3, ..., x_n, ...$ With each possible x_i we associate a real number $p_i = P(X = x_i)$ for $i = 1, 2, 3, ..., n, ...,$ satisfies the conditions

1. $p_i \geq 0$ for all $i = 1, 2, 3, ..., n, ...$
2. $\sum_{i=1}^{\infty} p_i = 1.$

Then the function p_i is called the **discrete probability distribution** or **probability mass function (p.m.f.)** or **probability density function** of the random variable X.

We can represent the probability mass function of X in a tabular form as below,

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

Note: For a specified $t = x_i$,

$$P(X \leq t) = P(X = x_1) + P(X = x_2) + ... + P(X = x_{i-1}) + P(X = x_i)$$

$$P(X < t) = P(X = x_1) + P(X = x_2) + P(X = x_3) + ... + P(X = x_{i-1})$$

Distribution function or Cumulative distribution function (c.d.f.)

Let X be a any random variable. Then the **distribution function or cumulative distribution function (c.d.f.)** of X is a function F defined by

$$F(t) = P(X \leq t).$$

For a discrete random variable X, the c.d.f. of X is,

$$F(t) = P(X \leq t) = \sum_{x \leq t} P(X = x).$$

Expectation of a discrete random variable

Let X be a discrete random variable which assumes values $x_1, x_2, x_3, ..., x_n, ...$ with p.m.f. $p_i = P(X = x_i)$. Then the **expectation or mean** of X is defined as

$$E(X) = \sum_{i=1}^{\infty} x_i P(X = x_i).$$

Similarly,

$$E(X^2) = \sum_{i=1}^{\infty} x_i^2 P(X = x_i).$$

Continuous probability distribution

Let X be a continuous random variable which assumes values from an interval $I \subseteq \mathbb{R}$. If there exists a function f satisfies the conditions;

1. $f(x) \geq 0$ for all $x \in I$
2. $\int_{-\infty}^{\infty} f(x)\ dx = 1.$

Then the function f is called the **continuous probability distribution** or **probability distribution function (p.d.f.)** of the random variable X.

Note: Let X be a continuous random variable. Then for a specific a, b such that $-\infty < a < b < \infty$ we have

$$\begin{aligned} P(a \leq x \leq b) = P(a < x \leq b) = P(a \leq x < b) &= P(a < x < b) \\ &= \int_a^b f(x)\ dx. \end{aligned}$$

For a continuous random variable X, the c.d.f. of X is,

$$F(t) = P(X \leq t) = \int_{-\infty}^{t} f(x)\ dx.$$

Expectation of a continuous random variable

Let X be a continuous random variable with p.d.f. $f(x)$. Then the **expectation or mean** of X is defined as

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

Similarly,

$$E(X^2) = \int_{-\infty}^{\infty} x^2\ f(x)\ dx.$$

Note: If X and Y are two random variables and α, β are two constants then $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$ and
$E(\alpha X + \beta) = \alpha E(X) + \beta$.

Variance

Let X be a random variable with the probability distribution. Then the **variance** of X is defined as

$$V(X) = E\left[X - E(X)\right]^2$$

i.e.,

$$\begin{aligned} V(X) &= E\left[X - E(X)\right]^2 \\ &= E\left[X^2 - 2XE(X) + (E(X))^2\right] \\ &= E(X^2) - 2E(X)E(X) + -\left[E(X)\right]^2 \\ &= E(X^2) - 2\left[E(X)\right]^2 + \left[E(X)\right]^2 \\ &= E(X^2) - \left[E(X)\right]^2 \end{aligned}$$

The **standard deviation** of X is defined as $S.D.(X) = \sqrt{V(X)}$.

Note: If X and Y are two random variables and α, β are two constants then

$$V(\alpha X + \beta Y) = \alpha^2 V(X) + \beta^2 V(Y)$$

and

$$V(\alpha X + \beta) = \alpha^2 V(X).$$

Problems

Question 1.

Five defective bulbs are accidently mixed with 20 good ones. It is not possible to just look at a bulb and tell whether or not it is defective. Find the probability distribution of the number of defective bulbs, if 4 bulbs are drawn at random from this lot?

Solution: Let X be the random variable denotes the number of defective bulbs out of 4. Then, X takes the values $0, 1, 2, 3, 4$.
Number of defective bulbs =5.
Number of good bulbs=20.
Total number of bulbs = 25.
$P(X = 0) = P(\text{no defective bulb}) = P(\text{all good ones}) = \dfrac{20C_4}{225C_4} = \dfrac{969}{2530}$.

$$P(X=1) = P(1 \text{ defective bulb and 3 good ones}) = \frac{5C_1 \times 20C_3}{225C_4} = \frac{1140}{2530}.$$
$$P(X=2) = P(2 \text{ defective bulb and 2 good ones}) = \frac{5C_2 \times 20C_2}{225C_4} = \frac{380}{2530}.$$
$$P(X=3) = P(3 \text{ defective bulb and 1 good ones}) = \frac{5C_3 \times 20C_1}{225C_4} = \frac{40}{2530}.$$
$$P(X=4) = P(4 \text{ defective bulb }) = \frac{5C_4}{225C_4} = \frac{1}{2530}.$$

Therefore the required p.m.f. is,

X	0	1	2	3	4
$P(X=x)$	$\frac{969}{2530}$	$\frac{1140}{2530}$	$\frac{380}{2530}$	$\frac{40}{2530}$	$\frac{1}{2530}$

Question 2.

The probability mass function (p.m.f.) of a random variable X is given by the following table:

X	-2	-1	0	1	2	3
$P(X = x)$	0.3	k	0.2	$2k$	0.3	k

1. Determine the value of k.
2. Find the mean and variance of X.
3. Find $P(X < 1)$, $\;P(-1 \leq X < 2)$.

Solution: (1) For a valid p.m.f. we have, $\sum_x P(X = x) = 1$ implies

$0.6 + 4k = 1 \implies k = 0.1.$

(2)

$$\begin{aligned}\text{Mean of } X = E(X) &= \sum_{i=1}^{6} x_i P(X = x_i) \\ &= -2(0.1) - 1(k) + 0 + 1(k) + (0.3) + 3(k) \\ &= 0.4 + 4k = 0.8.\end{aligned}$$

$$
\begin{aligned}
E(X^2) &= \sum_{i=1}^{6} x_i^2 P(X = x_i) \\
&= 4(0.1) + 1(k) + 0 + 1(k) + (0.3) + 9(k) \\
&= 1.6 + 12k = 1.8.
\end{aligned}
$$

Therefore,variance of X is,

$$
\begin{aligned}
V(X) &= E(X^2) - (E(X))^2 \\
&= 1.8 - 0.64 = 2.16
\end{aligned}
$$

(3) $P(X < 1) = P(X = -2) + P(X = -1) + P(X = 0) = 0.5 + k = 0.6$
$P(-1 \leq X < 2) = P(X = -1) + P(X = 0) + P(X = 1) = 0.2 + 3k = 0.5$

Question 3.

A box contains 12 balls of which 3 are white and 9 are red. A sample of 3 balls is selected from the box. Let X denote the number of white balls in the sample. Find the p.m.f. of X. Determine the mean and standard deviation of X.

Solution: Here $X = 0, 1, 2, 3$. Then,

$$P(X=0) = \frac{9C_3}{12C_3} = \frac{84}{220}; \quad P(X=1) = \frac{3C_1 \times 9C_2}{12C_3} = \frac{108}{220}$$

$$P(X=2) = \frac{3C_2 \times 9C_1}{12C_3} = \frac{27}{220}; \quad P(X=3) = \frac{3C_3}{12C_3} = \frac{1}{220}$$

Therefore the required p.m.f. is,

X	0	1	2	3
$P(X=x)$	$\frac{84}{220}$	$\frac{108}{220}$	$\frac{27}{220}$	$\frac{1}{220}$

Question 4.

Let X be a continuous random variable with p.d.f.

$$f(x) = \begin{cases} 3x^2, & \text{if } 0 < x < 1, \\ 0, & \text{else where} \end{cases}$$

Then find $E(X)$ and $V(X)$?

Solution: We have,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \ f(x) \ dx \\ &= \int_0^1 x \ (3x^2) \ dx \\ &= 3 \int_0^1 x^3 \ dx \\ &= \frac{3}{4}. \end{aligned}$$

Also,

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2\ f(x)\ dx \\ &= \int_0^1 x^2\ (3x^2)\ dx \\ &= 3\int_0^1 x^4\ dx \\ &= \frac{3}{5}. \end{aligned}$$

Therefore, $V(X) = E(X^2) - (E(X))^2 = \frac{3}{5} - \frac{9}{16} = \frac{3}{80}$.

Question 5.

Let X be a continuous random variable with p.d.f.

$$f(x) = \quad ke^{-|x|}, \quad \text{if } -\infty < x < \infty \ .$$

Then find the value of k, $E(X)$ and $S.D.(X)$?

Solution: We have, $|x| = \begin{cases} -x, & \text{if } x < 0, \\ x, & \text{if} x \geq 0 \end{cases}$ Therefore, the p.d.f. becomes, $f(x) = \begin{cases} ke^{x}, & \text{if } x < 0, \\ ke^{-x}, & \text{if } x \geq 0 \end{cases}$

Since $f(x)$ is a valid p.d.f, we've

$$\int_{-\infty}^{\infty} f(x)\ dx = 1 \implies \int_{-\infty}^{\infty} \underbrace{ke^{-|x|}}_{\text{even function}}\ dx = 1$$

$$\implies 2\int_{0}^{\infty} ke^{|x|}\ dx = 1 \implies 2k\int_{0}^{\infty} e^{-x}\ dx = 1$$

$$\implies 2k\left(-e^{-x}\right)_{0}^{\infty} = 1 \implies k = \tfrac{1}{2}.$$

We have,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x \; f(x) \; dx \\ &= \int_{-\infty}^{\infty} \underbrace{xe^{-|x|}}_{\text{odd function}} \; dx = 0 \end{aligned}$$

Also,

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 \; f(x) \; dx \\ &= \int_{-\infty}^{\infty} \underbrace{x^2 e^{-|x|}}_{\text{even function}} \; dx \\ &= 2\int_{0}^{\infty} x^2 e^{-|x|} \; dx = 2\int_{0}^{\infty} x^2 e^{-x} \; dx = 2 \end{aligned}$$

So, $V(X) = E(X^2) - (E(X))^2 = 2$. Hence, $S.D.(X) = \sqrt{2}$.

Question 6.

Let X be a continuous random variable with p.d.f.
$f(x) = \begin{cases} x, & \text{if } 0 \le x \le 1, \\ 2-x, & 1 < x \le 2 \\ 0, & \text{else where} \end{cases}$ Then find the c.d.f of X and find (i) $P\left(X \ge \dfrac{3}{2}\right)$ (ii) $P\left(\frac{3}{2} < X \le 2\right)$?

Solution: We know the c.d.f of X is $F(t) = P(X \le t) = \displaystyle\int_{-\infty}^{t} f(x)\, dx$.

Case(1): When $-\infty < t < 0$

We have, $F(t) = \displaystyle\int_{-\infty}^{t} f(x)\, dx = \int_{-\infty}^{t} 0\, dx = 0.$

Case(2): When $0 \le t \le 1$

We have, $F(t) = \displaystyle\int_{-\infty}^{t} f(x)\, dx = \int_{-\infty}^{0} f(x)\, dx + \int_{0}^{t} f(x)\, dx =$
$\displaystyle\int_{-\infty}^{0} 0\, dx + \int_{0}^{t} x\, dx = \frac{t^2}{2}.$

Case(3): When $1 < t \leq 2$

We have,

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

$$\int_{-\infty}^{0} 0\ dx + \int_{0}^{1} x\ dx + \int_{1}^{t} (2-x)\ dx = \frac{1}{2} + \left(-\frac{(t-2)^2}{2} + \frac{1}{2}\right) = 1 - \frac{(t-2)^2}{2}.$$

Case(4): When $t \geq 2$

We have, $F(t) = \int_{-\infty}^{t} f(x)\ dx =$

$$\int_{-\infty}^{0} f(x)\ dx + \int_{0}^{1} f(x)\ dx + \int_{1}^{2} f(x)\ dx + \int_{1}^{t} f(x)\ dx =$$

$$\int_{-\infty}^{0} 0\ dx + \int_{0}^{1} x\ dx + \int_{1}^{t} (2-x)\ dx \int_{1}^{t} 0\ dx = 0 + \frac{1}{2} + \frac{1}{2} + 0 = 1.$$

Hence the required c.d.f of X is,

$$F(t) = P(X \leq t) = \begin{cases} 0, & \text{if } -\infty < t < 0, \\ \frac{t^2}{2}, & \text{if } 0 \leq t \leq 1 \\ 1 - \frac{(t-2)^2}{2}, & \text{if } 1 < t \leq 2 \\ 0, & \text{if } t > 2 \end{cases}$$

(i) We know that,

$$\begin{aligned} P(X \geq \frac{3}{2}) &= 1 - P(X < \frac{3}{2}) \\ &= 1 - F\left(\frac{3}{2}\right) \\ &= 1 - \left(1 - \frac{\left(\frac{3}{2} - 2\right)^2}{2}\right) = \frac{1}{8}. \end{aligned}$$

Or,

$$\begin{aligned}
P(X \geq \frac{3}{2}) &= 1 - P(X < \frac{3}{2}) \\
&= 1 - \int_{-\infty}^{\frac{3}{2}} f(x)\ dx \\
&= 1 - \int_{-\infty}^{0} 0\ dx + \int_{0}^{1} x\ dx + \int_{1}^{\frac{3}{2}} (2-x)\ dx = \frac{1}{8}.
\end{aligned}$$

(ii) We have $P\left(\frac{3}{2} < X \leq 2\right) = F(2) - F\left(\frac{3}{2}\right) = 1 - \frac{7}{8} = \frac{1}{8}$.
Or,

$$\begin{aligned}
P\left(\frac{3}{2} < X \leq 2\right) &= \int_{\frac{3}{2}}^{2} f(x)\ dx \\
&= \int_{\frac{3}{2}}^{2} (2-x)\ dx \\
&= \left(2x - \frac{x^2}{2}\right)_{\frac{3}{2}}^{2} \frac{1}{8}.
\end{aligned}$$