Random variables

1. Random Variables

A real number associated with outcome of an experiment is known as a random variable.

Example(1)

Suppose two fair coins are tossed. Then S={HH, HT, TH, TT}

$$X=\begin{cases} 0\text{ , } \textit{if atleast one head occurs} \\ 1\text{, } \textit{otherwise} \end{cases}$$

Outcome	HH	HT	TH	TT
X	0	0	0	1

Let X be the random variable corresponding to number of heads. Then X takes values 0,1,2.

Outcome	HH	HT	TH	TT
X	2	1	1	0

Range of random variable

The set of all real numbers of a random variate X is called range of X.

In example(1), Range of $X=\{0,1,2\}$

Example(3)

Consider the random experiment of throwing a dice twice.

The corresponding sample space is

$S=\{(a,b) \mid a=1,2,3,4,5,6 \text{ and } b=1,2,3,4,5,6\}$

Let $X=a+b$ be a random variable, then

Range of $X=\{2,3,4,5,6,7,8,9,10,11,12\}$

Over the same sample space, we can define

$$\mathbf{Y} = \begin{cases} \mathbf{1} & \boldsymbol{if}\ \mathbf{a}+\mathbf{b}\ \boldsymbol{is\ even} \\ -\mathbf{1} & \boldsymbol{if}\ \mathbf{a}+\mathbf{b}\ \boldsymbol{is\ odd} \end{cases}$$

Range of Y={-1,1}

Thus infinitely many random variables can be defined on a given sample space.

2. Discrete and Continuous random variable

If a random variable X takes at most a countable number of values x_1, x_2, x_3 $x_4, \ldots x_n$, then it is called a discrete random variable(DRV).

Examples: (1) and (2).

A random variable X is said to be a continuous random variable(CRV) if it can take any value in an interval which may be finite or infinite.

Example: weight of articles

3. Discrete Probability distribution

Let X be a discrete random variable assuming the values x_1, x_2, x_3 $x_4, \ldots\ldots x_n$. With each possible outcome x_i we associate a number $p_i = P(X=x_i) = P(x_i)$ called the probability of x_i.

Then $P(x_i)$ is called the probability mass function (PMF) of the random variable X if the following conditions are satisfied.

(i) $\mathbf{P(x_i) \geq 0 \quad \forall i}$

(ii) $\mathbf{\sum P(x_i) = 1}$

The set $\{\mathbf{P(xi)}\}$ is called the probability distribution of the random variable.

Example:

Suppose two fair coins are tossed. Let X be the random variable corresponding to number of heads.

Outcome	HH	HT	TH	TT
X	2	1	1	0

The probability distribution for number of heads is given by

X	0	1	2
P(X)	¼	2/4	¼

Clearly $\mathbf{P(xi) \geq 0} \quad \forall \mathbf{i}$ and $\sum \mathbf{P(xi) = 1}$

E(X)=0+(1/2)+(1/2)=1=Mean????

Example:

Suppose THREE fair coins are tossed. Let X be the random variable corresponding to number of heads.

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

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

The probability distribution for number of heads is given by

X	0 TTT	1 TTH,THT, HTT	2 HHT,HTH ,THH	3 HHH
P(X)	1/8	3/8	3/8	1/8

Mean=E(X)=(3/8)+(3/4)+(3/8)

=(3+6+3)/8=12/8= 1.5

4. Continuous Probability distribution

Let X be a continuous random variable assuming values x over an interval.

We assign a real number $f(x)$ satisfying the conditions

(i) $f(x) \geq 0 \quad \forall i$

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

Then $f(x)$ is called the probability density function (PDF) of the continuous variable X.

If (a,b) is a subinterval of the range space of X then the probability of x which lies in (a,b) is denoted by P(a≤ X ≤b) and is defined as

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

5. Cumulative distribution function

Let X be a random variable (discrete or continuous). We define F(x) to be the cumulative distribution function (CDF) or simply distribution function if

$$\mathbf{F(x) = P(X \leq x)}$$

If X is a discrete random variable, then

$$\begin{aligned}\mathbf{F(x_i)} &= \mathbf{P(X \leq x_i)} \\ &= \mathbf{P(x_1) + P(x_2) + \cdots P(x_i)}\end{aligned}$$

If X is a continuous variable with PDF $f(x)$ then

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

Thus, $f(x) = \frac{d}{dx}\left(\mathbf{F(x)}\right)$

6. Mathematical Expectation E(X):

For a DRV **X** having the possible values $x_1, x_2, x_3, x_4, \ldots\ldots x_n$, the expectation of **X** is defined as

$$\boxed{E(\mathbf{X}) = \sum_{i=1}^{n} x_i \mathrm{P}(x_i)}$$

Similarly, expectation of $\mathbf{X}^2$ is defined as

$$\boxed{E(\mathbf{X}^2) = \sum_{i=1}^{n} x_i^2 \mathrm{P}(x_i)}$$

In general,

$$\boxed{E(\phi(\mathbf{X})) = \sum_{i=1}^{n} \phi(x_i)\mathrm{P}(x_i)}$$

For a CRV **X**, having density function $f(x)$ the expectation of **X** is defined as

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

In general,

$$\boxed{E\big(\phi(x)\big) = \int_{-\infty}^{\infty} \phi(x) f(x)\, dx}$$

7. Mean & Variance:

The expectation of X is known as mean value of the probability distribution and it it is denoted as μ.

$$\therefore \boxed{\mu = E(\mathbf{X})}$$

Variance of a distribution is defined as

$$\boxed{Var(X) = \sigma^2 = E\big[(X-\mu)^2\big]}$$

For a DRV

$$\boxed{\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \mathrm{P}(x_i)}$$

For a CRV

$$\boxed{\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, dx}$$

NOTE:

The positive square root of the variance is called the standard deviation and is given by $\sigma = \sqrt{Var(X)} = \sqrt{E[(X - \mu)^2]}$

8. Laws of Expectation:

a) $\boxed{E(cX) = cE(X)}$ where c is constant

b) If X and Y are random variables, then $\boxed{E(X + Y) = E(X) + E(Y)}$

c) If X and Y are independent random variables, then $\boxed{E(XY) = E(X)E(Y)}$

Problems

a. Prove that $\sigma^2 = E(X^2) - [E(X)]^2$

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

$$\Rightarrow \sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 P(x_i)$$

$$\Rightarrow \sigma^2 = \sum_{i=1}^{n} \left(x_i{}^2 + \mu^2 - 2x_i\mu\right) P(x_i)$$

$$\sigma^2 = \sum_{i=1}^{n} \left(x_i{}^2 P(x_i) + \mu^2 P(x_i) - 2x_i\mu P(x_i)\right)$$

$$\sigma^2 = \sum_{i=1}^{n} x_i{}^2 P(x_i) + \mu^2 \sum_{i=1}^{n} P(x_i) - 2\mu \sum_{i=1}^{n} x_i\, P(x_i)$$

$$\sigma^2 = E\left(X^2\right) + \mu^2 - 2\mu^2$$

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

b. The probability distribution of a random variable X is given by the following table. Find k and evaluate the mean.

X	0	1	2	3	4	5
P(X=x)	k	5k	10k	10k	5k	k

Soln: The set $\{P(xi)\}$ is called the probability distribution of the random variable if it satisfies the following conditions.

(i) $P(x_i) \geq 0 \quad \forall i$

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

$$\therefore k + 5k + 10k + 10k + 5k + k = 1$$

$$\therefore 32k = 1 \qquad \Rightarrow \boxed{k = \frac{1}{32}}$$

Mean, $\mu = E(X) = \sum_{i=1}^{n} x_i P(x_i)$

$$\Rightarrow \mu = 0(k) + 1(5k) + 2(10k) + 3(10k) + 4(5k) + 5(k)$$

$$\Rightarrow \mu = 80k = \frac{80}{32} = \frac{5}{2}$$

c. The probability density function of a variate X is

X	0	1	2	3	4	5	6
P(X)	k	3k	5k	7k	9k	11k	13k

Find k

(i) $P(X < 4)$, $P(X \geq 5)$, $P(3 < X < 6)$

(ii) Mean and variance

d. A random variable X has the following probability function

X	1	2	3	4	5	6	7
P(X)	k	2k	2k	3k	k^2	$2k^2$	$7k^2+k$

Find (i) k (ii) $P(X \geq 6)$ (iii) $P(X < 6)$ (iv) $P(1 \leq X < 5)$ (v) E(X)

Soln:

$$k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$$

$$10k^2 + 9k - 1 = 0$$

$$k = \frac{-9 \pm \sqrt{81 + 40}}{20} = \frac{-9 \pm 11}{20}$$

$$k = -1 \quad or \quad 0.1$$

But, $P(x_i) \geq 0$ and hence $\boxed{k = 0.1}$

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

$$ii)\ P(X \geq 6) = P(6) + P(7) = 0.19$$

$$iii)\ P(X < 6) = 1 - P(X \geq 6) = 0.81$$

$$\begin{aligned} iv)\ P(1 \leq X < 5) \\ &= P(1) + P(2) + P(3) + P(4) \\ &= 0.8 \end{aligned}$$

$$iv)\ E(X) = \sum_{i=1}^{n} x_i \mathrm{P}(x_i) = 3.66$$

e. A company has five applicants for two positions. Two women and three men have applied. Suppose that no preference is given for choosing either sex. Let the random variable X be the number of women chosen to fill the two positions. Write the probability mass function of X.

Soln:

$$S = \{M_1M_2, M_1M_3, M_2M_3, W_1W_2, M_1W_1, M_1W_2, M_2W_1, M_2W_2, M_3W_1, M_3W_2\}$$

<table>
<tr><td rowspan="2">X</td><td colspan="2">0</td><td colspan="2">1</td><td colspan="2">2</td></tr>
<tr><td>0W</td><td>2M</td><td>1W</td><td>1M</td><td>2W</td><td>0M</td></tr>
<tr><td>P(X)</td><td colspan="2">3/10</td><td colspan="2">6/10</td><td colspan="2">1/10</td></tr>
</table>

Method-02 {without writing S}

<table>
<tr><td rowspan="2">X</td><td colspan="2">0</td><td colspan="2">1</td><td colspan="2">2</td></tr>
<tr><td>0W</td><td>2M</td><td>1W</td><td>1M</td><td>2W</td><td>0M</td></tr>
<tr><td>P(X)</td><td colspan="2">$\dfrac{{}^2C_0 . {}^3C_2}{{}^5C_2}$
3/10</td><td colspan="2">$\dfrac{{}^2C_1 . {}^3C_1}{{}^5C_2}$
6/10</td><td colspan="2">$\dfrac{{}^2C_2 . {}^3C_0}{{}^5C_2}$
1/10</td></tr>
</table>

f. Find E[X], E[X²] and σ^2 for the probability function P(X) defined by the following table

x_i	1	2	3	....	n
$P(x_i)$	k	2k	3k	....	nk

Where k is an appropriate constant.

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

a) $f_1(x) = \begin{cases} 2x, 0 < x < 1 \\ 0. \ otherwise \end{cases}$

b) $f_2(x) = \begin{cases} 2x, -1 < x < 1 \\ 0, otherwise \end{cases}$

c) $f_3(x) = \begin{cases} |x|, \ |x| \leq 1 \\ 0, otherwise \end{cases}$

d) $f_4(x) = \begin{cases} 2x, & 0 < x \leq 1 \\ 4 - 4x, & 1 < x < 2 \\ 0, & otherwise \end{cases}$

j. A random variable X has the density function $f(x) = \frac{k}{1+x^2}, \quad -\infty < x < \infty$ determine k and evaluate $P(X \geq 0)$.

k. A continuous random variable X has pdf given by $f(x) = \begin{cases} 2e^{-2x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$

Evaluate (i) $E(x)$ and (ii) $E(x^2)$. Hence find the standard deviation.

l. If a continuous random variable X has pdf

$$f(x) = \begin{cases} \frac{1}{4}, & -2 < x < 2 \\ 0, & elsewhere \end{cases}$$

Obtain (i) $P[X < 1]$ (ii) $P[|X| > 1]$ (iii) $P[2X + 3 > 5]$

m. Let the continuous random variable X have the pdf $f(x) = \begin{cases} \frac{k}{1+x^2}, & 1 < x < \infty \\ 0, & elsewhere \end{cases}$. Find the distribution function F(x).

n. The pdf of a random variable X is given by

$$f(x) = \begin{cases} x & , 0 \leq x \leq 1 \\ 2 - x, & 1 < x \leq 2 \\ 0 & , elsewhere \end{cases}$$

Find (i) Probability distribution function F(X) and (ii) P(X ≥ 1.5).

O. Is the function $f(x) = \begin{cases} e^{-x}, & 0 \leq x < \infty \\ 0, & elsewhere \end{cases}$ a density function of the continuous random variable X?

i). If so, determine P(1 ≤ X ≤ 2) .

ii). Also, find the probability distribution function $f(x)$at $x = 2$.

p. The diameter of an electric cable is assumed to be a continuous random variable with pdf $f(x) = 6x\ (1 - x),\ 0 \le x \le 1$, 0 elsewhere. Verify that the above is a pdf. Also find its mean and variance.

q. A random variable X has the probability function $\boldsymbol{p(x) = \frac{1}{2^x}\ ;\ x = 1, 2, 3, 4 \ldots}$
Find (i) Mean & (ii) variance.