(73)

~~Module~~ I

CHAPTER 2

Random Variables: Probability Distribution and Expectation

2.1 RANDOM VARIABLE

Definition 1: Let S be the sample space (or event space) associated with a given random experiment. Then a real valued function X defined on S is called a **one dimensional random variable** or just a **random variable (r.v.)** or sometimes a **variate**.

In other words, a **random variable** is defined as a variable which takes numerical values determined by the outcomes of a random experiment. So, a random variable can be thought of as a function that maps the points of the sample space into the set of real numbers.

The sample space S is termed as the **domain** of the corresponding random variable and the collection of all the numbers (values of random variable) is termed as the **range** or **spectrum** of the random variable.

***Notes*:** *(i)* One dimensional random variables will be denoted by capital letters X, Y, Z, etc.

(ii) Let w be an outcome of the underlying random experiment, then $X(w)$ represents the real number which the random variable X associates with the outcome w. The values taken by a random variable X are usually denoted by lower case letters x, y, etc.

(iii) Two or more different outcomes might give the same value of X but two different numbers in the range (or spectrum) cannot be assigned to the same outcome.

Example 1: Consider the random experiment of tossing a coin twice. Here the sample space is

$$S = \{HH, HT, TH, TT\}.$$

Let X represents the random variable associated with the outcome 'number of heads'. Then we can assign a number for X as shown in the table below:

Outcome	*HH*	*HT*	*TH*	*TT*
X	2	1	1	0

Example 2: Consider the random experiment of throwing a die. Here the sample space is $S = \{1, 2, 3, 4, 5, 6\}$.

The most natural choice of random variable X is $X(w) = w$, where $w = 1, 2, 3, 4, 5, 6$. Another random variable Y on S can be defined as:

$$Y(w) = \begin{cases} 1, \text{ if } w \text{ is even} \\ 0, \text{ if } w \text{ is odd} \end{cases}$$

So, when the face turned up is 2 or 4 or 6, Y takes the value 1 and when 1 or 3 or 5 come up, Y takes the value 0.

Note: Associated with same sample space different random variables can be defined.

Definition 2 (*Events described by a random variable*): If X is a random variable and x is a fixed real number, then the event $(X = x)$ is defined as:

$$(X = x) = \{w \in S : X(w) = x\}$$

Similarly, we can define the following events:

$$(X \le a) = \{w \in S : X(w) \le a\}$$
$$(X > b) = \{w \in S : X(w) > b\}$$
$$(a < X \le b) = \{w \in S : a < X(w) \le b\}$$

The corresponding probabilities are:

$$P(X = x) = P\{w \in S : X(w) = x\}$$
$$P(X \le a) = P\{w \in S : X(w) \le a\}$$
$$P(X > b) = P\{w \in S : X(w) > b\}$$
$$P(a < X \le b) = P\{w \in S : a < X(w) \le b\}.$$

Notes: (*i*) The symbol '$\in$' stands for the word 'belongs to'. (*ii*) The symbol ':' stands for the word 'such that'.

Example 3: Consider the random experiment given in Example 1.

Let A and B denote the events $(X = 2)$ and $(X \le 1)$ respectively. Here the sample space is $S = \{HH, HT, TH, TT\}$ and X represents the r.v. associated with the outcome 'number of heads'.

$\therefore$ $A = (X = 2) = \{w \in S : X(w) = 2\} = \{HH\}$

and $B = (X \le 1) = \{w \in S : X(w) \le 1\} = \{HT, TH, TT\}$.

Assume that the event points (or sample points or outcomes) are equally likely, we have

$$P(A) = P(X = 2) = \frac{1}{4} \text{ and } P(B) = P(X \le 1) = \frac{3}{4}.$$

Theorem 1: *If X_1 and X_2 are two random variables defined on the same sample space (or event space) S, then $X_1 + X_2$ is also a random variable defined on S.*

Proof: Let X_1 and X_2 be two random variables defined on the sample space S associated with a given random experiment E. Let $w \in S$ be an outcome of E, so $X_1(w)$ and $X_2(w)$ are real numbers.

Hence $X_1(w) + X_2(w)$ is also a real number, *i.e.*, $(X_1 + X_2)(w)$ is also a real number. Thus $X_1 + X_2$ is a function from the sample space S to the set of real numbers. Therefore $X_1 + X_2$ is also a random variable on S.

Theorem 2: *If X_1 and X_2 are two random variables defined on the same sample space S and C, C_1, C_2 are constants, then CX_1, $X_1 X_2$ and $C_1 X_1 + C_2 X_2$ are also random variables defined on S.*

Proof: Proceed as in Theorem 1.

Note: It follows that $X_1 - X_2$ is also a r.v. on S.

Theorem 3: *If X_1 and X_2 are two random variables defined on the same sample space S, then max (X_1, X_2) is also a random variable on S.*

Proof: Proceed as in Theorem 1.

Example 4: A random experiment consists of three independent tosses of a fair (unbiased) coin. The sample space S contains $2^3 = 8$ event points:

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

Let us define two random variables X and Y on S, where X is associated with the outcome 'number of heads' and Y is associated with the outcome 'number of tails'.

Let X takes the values 3, 2, 1, 0 corresponding to three heads, two heads, one head, no head and Y takes the values 3, 2, 1, 0 corresponding to three tails, two tails, one tail, no tail. Then the random variable $X + Y$ takes the value 3 which is shown below:

Outcome	*Value of X*	*Value of Y*	*Value of X + Y*
HHH	3	0	3
THH	2	1	3
HTH	2	1	3
HHT	2	1	3
HTT	1	2	3
THT	1	2	3
TTH	1	2	3
TTT	0	3	3

2.2 DISTRIBUTION FUNCTION OF A RANDOM VARIABLE

Most of the information about a random experiment described by a random variable X is obtained by studying the behaviour of a function known as the **distribution function** which is defined as follows:

Definition 3 (*Distribution function*): Let X be a random variable defined on the sample space S associated with a given random experiment. The **cumulative distribution function (c.d.f.)**, or simply, **distribution function (d.f.)** of X is denoted and defined as

$$F_X(x) = P(X \le x), -\infty < x < \infty$$

where $P(X \le x) = P\{w \in S : X(w) \le x\}$.

Sometimes $F_X(x)$ is simply written as $F(x)$.

Clearly, $0 \le F(x) \le 1$.

Properties of distribution function: Let $F(x)$ is the distribution function of a random variable X and $a < b$ where a, b are any two real numbers.

$\therefore$ $F(a) = P(X \le a)$ and $F(b) = P(X \le b)$

(76)

Property 1: (*i*) $P(a < X \leq b) = F(b) - F(a)$.

Proof: The events $(X \leq a)$ and $(a < X \leq b)$ are mutually exclusive and

$$(X \leq b) = (X \leq a) \cup (a < X \leq b)$$

$$\Rightarrow \quad P(X \leq b) = P\{(X \leq a) \cup (a < X \leq b)\}$$

$$\Rightarrow \quad P(X \leq b) = P(X \leq a) + P(a < X \leq b) \qquad \text{[By addition law]}$$

$$\Rightarrow \quad P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$

$$= F(b) - F(a)$$

(*ii*) $$P(a \leq X \leq b) = P(X = a) + F(b) - F(a).$$

Proof: The events $(X = a)$ and $(a < X \leq b)$ are mutually exclusive and

$$(a \leq X \leq b) = (X = a) \cup (a < X \leq b)$$

$$\Rightarrow \quad P(a \leq X \leq b) = P\{(X = a) \cup (a < X \leq b)\}$$

$$\Rightarrow \quad P(a \leq X \leq b) = P(X = a) + P(a < X \leq b) \qquad \text{[By addition law]}$$

$$\Rightarrow \quad P(a \leq X \leq b) = P(X = a) + F(b) - F(a) \qquad \text{[By (i)]}$$

(*iii*) $$P(a < X < b) = F(b) - F(a) - P(X = b).$$

Proof: The events $(a < X < b)$ and $(X = b)$ are mutually exclusive and

$$(a < X \leq b) = (a < X < b) \cup (X = b)$$

$$\Rightarrow \quad P(a < X \leq b) = P\{(a < X < b) \cup (X = b)\}$$

$$\Rightarrow \quad F(b) - F(a) = P(a < X < b) + P(X = b) \qquad \text{[By (i) and addition law]}$$

$$\Rightarrow \quad P(a < X < b) = F(b) - F(a) - P(X = b).$$

(*iv*) $$P(a \leq X < b) = P(X = a) + F(b) - F(a) - P(X = b).$$

Proof: The events $(X = a)$ and $(a < X < b)$ are mutually exclusive and

$$(a \leq X < b) = (X = a) \cup (a < X < b)$$

$$\Rightarrow \quad P(a \leq X < b) = P\{(X = a) \cup (a < X < b)\}$$

$$\Rightarrow \quad P(a \leq X < b) = P(X = a) + P(a < X < b) \qquad \text{[By addition law]}$$

$$\Rightarrow \quad P(a \leq X < b) = P(X = a) + F(b) - F(a) - P(X = b). \qquad \text{[By (iii)]}$$

Note: When $P(X = a) = P(X = b) = 0$, the probability of all the four events $(a < X < b)$, $(a \leq X < b)$, $(a < X \leq b)$ and $(a \leq X \leq b)$ is same and is equal to $F(b) - F(a)$.

Property 2: (*i*) $0 \leq F(x) \leq 1$.

(*ii*) $x < y \Rightarrow F(x) \leq F(y)$. [*W.B.U.T. 2010*]

Proof: (*i*) $F(x) = P(X \leq x)$.

Now, $0 \leq P(X \leq x) \leq 1$.

$\therefore \quad 0 \leq F(x) \leq 1$.

(*ii*) $x < y \Rightarrow F(y) - F(x) = P(x < X \leq y)$ [By property 1, (*i*)]

$\Rightarrow \quad F(y) - F(x) \geq 0$ [$\because \; P(x < X \leq y) \geq 0$]

$\Rightarrow \quad F(x) \leq F(y)$.

$\Rightarrow$ Distribution function is monotonic non-decreasing everywhere.

Property 3: $F(-\infty) = \lim_{x\to-\infty} F(x) = 0$ and $F(\infty) = \lim_{x\to\infty} F(x) = 1.$ (*W.B.U.T. 2010*)

Proof: By definition of distribution function,

$$F(\infty) = \lim_{x\to\infty} F(x) = \lim_{x\to\infty} P(X \le x) = P(S) = 1,$$

$$F(-\infty) = \lim_{x\to-\infty} F(x) = \lim_{x\to-\infty} P(X \le x) = P(\varphi) = 0.$$

Property 4: The distribution function $F(x)$ is continuous on the right, *i.e.*,

$$\lim_{h\to 0+} F(a+h) = F(a). \qquad \text{(W.B.U.T. 2010)}$$

Proof: By property 1 (*ii*):

$$P(a \le X \le a+h) = P(X=a) + F(a+h) - \text{F}(a)$$

$$\therefore \quad \lim_{h\to 0+} P(a \le X \le a+h) = P(X=a) + \lim_{h\to 0+} F(a+h) - F(a).$$

$$\Rightarrow \quad P(X=a) = P(X=a) + \lim_{h\to 0+} F(a+h) - F(a).$$

$$\Rightarrow \quad \lim_{h\to 0+} F(a+h) = F(a).$$

$\Rightarrow$ $F(x)$ is continuous on the right.

Property 5: $P(X=a) = F(a) - \lim_{h\to 0+} F(a-h) = F(a) - \lim_{h\to a-} F(x)$

$$= F(a) - F(a-0).$$

Proof: For $h > 0$, using property 1 (*i*):

$$P(a-h < X \le a) = F(a) - F(a-h)$$

$$\therefore \quad P(X=a) = \lim_{h\to 0+} P(a-h < X \le a) = \lim_{h\to 0+} \{F(a) - F(a-h)\}$$

$$= F(a) - \lim_{h\to 0+} F(a-h)$$

$$= F(a) - \lim_{x\to a-} F(x) = F(a) - F(a-0).$$

2.3 DISCRETE RANDOM VARIABLES

There are two types of random variables known as:

(*i*) discrete random variables,

(*ii*) continuous random variables.

Definition 4 (*Discrete random variable*): If a random variable takes finite or an infinite sequence (countably infinite) of distinct values it is called a **discrete random variable**.

In other words, a real valued function defined on a discrete sample space is known as a **discrete random variable**.

The **range** or **spectrum** of a discrete random variable is **finite** or **countably infinite**.

Example: The random variable X defined in Example 1 and the random variables X, Y defined in Example 2 are discrete random variables.

Definition 5 [*Probability mass function (p.m.f.)*]: Let X be a discrete random variable which assumes the values $x_0, x_1, x_2, \ldots, x_n, \ldots$ with probabilities $P(X = x_i) = p(x_i) = p_i$. The value of p_i depends on x_i, *i.e.*, on i. This function p_i is called **probability mass function (p.m.f.)** of the random variable X provided the following conditions are satisfied:

(*i*) $p_i \geq 0, \forall i$ (*ii*) $\sum_{i=0}^{\infty} p_i = 1.$

A particular value of p_i is called a **probability mass** and the set of ordered pairs (x_i, p_i) is known as the **discrete probability distribution** of the random variable X.

Definition 6 (*Discrete distribution function*): Let X be a discrete random variable assuming the values $x_0, x_1, x_2, \ldots, x_n, \ldots$, then the **distribution function (d.f.)** of X is given by

$$F(x) = P(X \leq x) = P(X = x_0) + P(X = x_1) + \ldots + P(X = x_i)$$

$$= \sum_{k=0}^{i} p_k \text{, where } x_i \leqslant x < x_{i+1}$$

Example 1: *A random variable X has the following probability mass function:*

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

Determine the value of k. (*W.B.U.T. 2011*)

Solution: We know that $P(X = x)$ is a possible probability mass function if $P(X = x) \geq 0, \forall x$ and $\sum_x P(X = x) = 1$.

Here, $\sum_x P(X = x) = 1$

$\Rightarrow$ $0 + 5k + 3k + k + k = 1$

$\Rightarrow$ $10k = 1$

$\Rightarrow$ $k = \frac{1}{10}$

Also, for $k = \frac{1}{10}$, $P(X = x) \geq 0, \forall x$.

Hence the required value of k is $\frac{1}{10}$.

Example 2: *A random experiment consists of three independent tosses of an unbiased coin. Let X and Y denote the following events:*

$X \equiv$ The number of heads

$Y \equiv$ Consecutive occurrence of at least two heads.

Find the probability mass function (or probability distribution) of (i) X (ii) Y (iii) X + Y (iv) XY.

Solution: The sample space containts $2^3 = 8$ event points. Let us assign a number to each of X and Y as shown in the following table:

Events	X	Y	$X+Y$	XY
HHH	3	1	4	3
THH	2	1	3	2
HTH	2	0	2	0
HHT	2	1	3	2
HTT	1	0	1	0
THT	1	0	1	0
TTH	1	0	1	0
TTT	0	0	0	0

(*i*) X is a random variable which assumes the values 0, 1, 2, 3.

Values of X (x)	0	1	2	3	$p(x)>0,$
$p(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\sum p(x)=1.$

(*ii*) Y is a random variable which assumes the values 0, 1.

Values of Y (y)	0	1	$p(y)>0,$
$p(y)$	$\frac{5}{8}$	$\frac{3}{8}$	$\sum p(y)=1.$

(*iii*) $U=X+Y$ is a random variable which can take the values 0, 1, 2, 3, 4.

Values of U (u)	0	1	2	3	4	$p(u)>0,$
$p(u)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\sum p(u)=1.$

(*iv*) $V=XY$ is a random variable which can take the values 0, 2, 3.

Values of V (v)	0	2	3	$p(v)>0,$
$p(v)$	$\frac{5}{8}$	$\frac{2}{8}$	$\frac{1}{8}$	$\sum p(v)=1.$

Example 3: *A discrete random variable X has the following probability mass function:*

Values of X, x	0	1	2	3	4	5	6	7
$p(x)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

(*i*) *Find k.* (*W.B.U.T. 2011*)

(*ii*) *Evaluate* $P(X<6)$, $P(X\geq 6)$ *and* $P(0<X<5)$. (*W.B.U.T. 2004, 2006, 2007*)

(*iii*) *Determine the distribution function of X.* (*W.B.U.T. 2004*)

Solution: (*i*) Since $\sum_{x=0}^{7} p(x) = 1$, we have

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

$$\Rightarrow \quad 10k^2 + 9k - 1 = 0$$

$$\Rightarrow \quad 10k^2 + 10k - k - 1 = 0$$

$$\Rightarrow \quad (10k - 1)(k + 1) = 0$$

$$\Rightarrow \quad k = \frac{1}{10} \qquad (\because\ k \not< 0, \text{ as } p(x) \geq 0, \forall x)$$

(*ii*) $P(X < 6) = P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5)$

$$= 0 + \frac{1}{10} + \frac{2}{10} + \frac{2}{10} + \frac{3}{10} + \frac{1}{100} \qquad \left(\because\ k = \frac{1}{10}\right)$$

$$= \frac{81}{100}.$$

$$P(X \geqslant 6) = 1 - P(X < 6) = 1 - \frac{81}{100} = \frac{19}{100}.$$

$$P(0 < X < 5) = P(X = 1) + P(X = 2) + P(X = 3) + P(X = 4)$$

$$= 8k = \frac{4}{5}.$$

(*iii*) The distribution function $F(x)$ of X is given below:

$$F(x) = \begin{cases} 0 & , \quad -\infty \le x < 1 \\ \dfrac{1}{10} & , \quad 1 \le x < 2 \\ \dfrac{1}{10} + \dfrac{2}{10} = \dfrac{3}{10} & , \quad 2 \le x < 3 \\ \dfrac{3}{10} + \dfrac{2}{10} = \dfrac{1}{2} & , \quad 3 \le x < 4 \\ \dfrac{1}{2} + \dfrac{3}{10} = \dfrac{8}{10} & , \quad 4 \le x < 5 \\ \dfrac{8}{10} + \left(\dfrac{1}{10}\right)^2 = \dfrac{81}{100} & , \quad 5 \le x < 6 \\ \dfrac{81}{100} + 2\left(\dfrac{1}{10}\right)^2 = \dfrac{83}{100} & , \quad 6 \le x < 7 \\ \dfrac{83}{100} + 7\left(\dfrac{1}{10}\right)^2 + \dfrac{1}{10} = 1 & , \quad x \ge 7. \end{cases}$$

Note: If $P(X \leqslant a) > \frac{1}{2}$, then the minimum value of a is 4. (*W.B.U.T. 2004*)