

# RANDOM VARIABLE AND ITS PROBABILITY DISTRIBUTION

**Presented  
By  
Pradip Panda  
Asstt. Professor, Deptt. Of Statistics  
Serampore College**

# Random Variable

A random variable is a real valued function defined over the sample space of a random experiment.

Two coins are tossed.



Random variable  $X$  is the number of heads appear

Sample space  $S = \{ HH, HT, TH, TT \}$

$X$  takes the values



## Random Variable



**Discrete random variable**

**Continuous random variable**

### Discrete random variable:

A random variable which can assume a finite or a countably infinite number of values is called a discrete random variable.

**Example:** The sum of points obtained in two throws of a fair die.

## **Continuous random variable:**

A random variable that can take an uncountably infinite number of values is called a continuous random variable.

Example: The decay time for a radioactive particle or the weight of a new born baby is a continuous random variable.

## PROBABILITY DISTRIBUTION

A statement of all possible values or sets of values of a random variable together with the corresponding probabilities gives the probability distribution of the random variable.

**Example: If a fair coin is tossed twice, the number of heads obtained ( $X$ ) will have the following probability distribution:**

Value of $X$	0	1	2	Total
$P( X = x )$	1/4	1/2	1/4	1

The sample space  $S = \{ TT, TH, HT, HH \}$

## BIVARIATE PROBABILITY DISTRIBUTION

A statement of the possible pairs of values assumed by X and Y , together with the corresponding probabilities is called the bivariate probability distribution or joint probability distribution of X and Y.

$$P(X = x_i, Y = y_j) = p_{ij} \quad \forall i, j; i = 1(1)k, j = 1(1)l$$

$$\text{Marginal probability of } X = P(X = x_i)$$

$$= P(X = x_i, Y = y_1) + P(X = x_i, Y = y_2) + \dots +$$

$$P(X = x_i, Y = y_j) + \dots + P(X = x_i, Y = y_l)$$

$$= p_{i1} + p_{i2} + \dots + p_{ij} + \dots + p_{il}$$

$$= \sum_{j=1}^l p_{ij} \quad \text{Mr. Pradip Panda, Asstt. Prof., Deptt. of Statistics, Serampore College}$$

# BIVARIATE PROBABILITY DISTRIBUTION OF X AND Y

X \ Y	$y_1$	$y_2$	.....	$y_j$	.....	$y_l$	Marginal total
$x_1$	$P_{11}$	$P_{12}$	.....	$P_{1j}$	.....	$P_{1l}$	$p_{10}$
$x_2$	$P_{21}$	$P_{22}$	.....	$P_{2j}$	.....	$P_{2l}$	$p_{20}$
.	.....	.....	.....	.....	.....	.....	.....
$x_i$	$P_{i1}$	$P_{i2}$	.....	$P_{ij}$	.....	$P_{il}$	$p_{i0}$
.	.....	.....	.....	.....	.....	.....	.....
$x_k$	$P_{k1}$	$P_{k2}$	.....	$P_{kj}$	.....	$P_{kl}$	$p_{k0}$
Marginal total	$p_{01}$	$p_{02}$	.....	$p_{0j}$	.....	$p_{0l}$	1

Marginal probability of  $Y = P(Y = y_j)$

$$= P(X = x_1, Y = y_j) + P(X = x_2, Y = y_j) + \dots +$$

$$P(X = x_i, Y = y_j) + \dots + P(X = x_k, Y = y_j)$$

$$= p_{1j} + p_{2j} + \dots + p_{ij} + \dots + p_{kj}$$

$$= \sum_{i=1}^k p_{ij}$$

$$= p_{0j}, j = 1(1)l$$

# EXPECTATION

x	$x_1$	$x_2$	$x_3$	.....	$x_i$	.....	$x_k$	Total
Probability	$p_1$	$p_2$	$p_3$	.....	$p_i$	.....	$p_k$	$\sum_{i=1}^k p_i = 1$

$$E(X) = \sum_{i=1}^k x_i p_i \quad \text{provided} \quad \sum_{i=1}^k |x_i p_i| < \infty$$

## Properties :

i) If  $X = C$ , a constant, then  $E(X) = C$

**Proof:**  $E(X) = \sum_{i=1}^k x_i p_i = \sum_{i=1}^k C p_i = C \sum_{i=1}^k p_i = C$

Since,  $\sum_{i=1}^k p_i = 1$

ii) If  $Y = CX$ ,  $C$  is a constant, then  $E(Y) = C.E(X)$

**Proof:**  $E(Y) = \sum_{i=1}^k C.x_i p_i = C \sum_{i=1}^k x_i p_i = CE(X)$

Provided,  $E(X)$  exists.

iii) If  $Z = X+Y$ , where  $X$  and  $Y$  are two random variables, defined on the same sample space, then  $E(Z) = E(X) + E(Y)$

**Proof:**  $E(Z) = \sum_{i=1}^k (x_i + y_i) p_i = \sum_{i=1}^k x_i p_i + \sum_{i=1}^k y_i p_i = E(X) + E(Y)$

Provided, both the expectations,  $E(X)$  &  $E(Y)$  exist.

iv) If  $Y = a + bX$ , then  $E(Y) = a + bE(X)$

**Proof:**  $E(Y) = \sum_{i=1}^k (a + bx_i)p_i = a\sum_{i=1}^k p_i + b\sum_{i=1}^k x_i p_i = a + b.E(X)$

Since,  $\sum_{i=1}^k p_i = 1$

v) If  $Z = XY$ , where X and Y are independent random variables, defined on the same sample space , then  $E(Z) = E(X).E(Y)$

## VARIANCE

$$\begin{aligned}\text{Var}(X) &= E[X - E(X)]^2 \\&= E [ X^2 - 2 X \cdot E(X) + E^2(X) ] \\&= E(X^2) - 2 \cdot E(X) \cdot E(X) + E^2(X) \\&= E(X^2) - 2 \cdot E^2(X) + E^2(X) \\&= E(X^2) - E^2(X)\end{aligned}$$

The positive square root of the variance is called the standard deviation.

### Properties :

i) If  $X = C$ , a constant, then  $\text{Var}(X) = 0$

$$\begin{aligned}\text{Proof: } \text{Var}(X) &= E[X - E(X)]^2 \\&= E[c - c]^2 \\&= 0\end{aligned}$$

ii) If  $Y = C.X$ ,  $C$  is a constant, then  $\text{Var}(Y) = C^2\text{Var}(X)$

**Proof:**  $\text{Var}(Y) = E[Y - E(Y)]^2 = E[C.X - E(C.X)]^2$

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

iii) If  $Y = a + b X$ , then  $\text{Var}(Y) = b^2\text{Var}(X)$

**Proof:**  $\text{Var}(Y) = E[Y - E(Y)]^2 = E[a + b X - E(a + b X)]^2$

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

## COVARIANCE

$$\begin{aligned}\text{Cov}(X,Y) &= E[\{X - E(X)\}\{Y - E(Y)\}] \\&= E[X.Y - X.E(Y) - Y.E(X) + E(X).E(Y)] \\&= E[X.Y] - E(X).E(Y) - E(Y).E(X) + E(X).E(Y) \\&= E[X.Y] - E(X).E(Y)\end{aligned}$$

When , X and Y are independent, then  $\text{Cov}(X,Y) = 0$ . Because,  
 $E(X.Y) = E(X).E(Y)$

**Independence:**

The random variables X and Y are said to be independent if  
 $P(X = x_i, Y = y_j) = P(X = x_i).P(Y = y_j),$   
*i.e.*  $p_{ij} = p_{i0}.p_{0j}, \text{ for all } i, j.$

## SUM LAW OF EXPECTATION

If  $X$  and  $Y$  be two jointly distributed random variables, then  
 $E(X + Y) = E(X) + E(Y)$

**Proof:**

$$\begin{aligned}E(X + Y) &= \sum_{i=1}^k \sum_{j=1}^l (x_i + y_j) P(X = x_i, Y = y_j) \\&= \sum_{i=1}^k \sum_{j=1}^l (x_i + y_j) p_{ij} \\&= \sum_{i=1}^k x_i \sum_{j=1}^l p_{ij} + \sum_{j=1}^l y_j \sum_{i=1}^k p_{ij} \\&= \sum_{i=1}^k x_i p_{i0} + \sum_{j=1}^l y_j p_{0j} = E(X) + E(Y)\end{aligned}$$

## PRODUCT LAW OF EXPECTATION

If X and Y are independent random variables, then  
 $E(XY) = E(X) \cdot E(Y)$

**Proof:**

$$\begin{aligned} E(XY) &= \sum_{i=1}^k \sum_{j=1}^l (x_i \cdot y_j) P(X = x_i, Y = y_j) \\ &= \sum_{i=1}^k \sum_{j=1}^l (x_i \cdot y_j) p_{ij} \\ &= \sum_{i=1}^k x_i p_{i0} \sum_{j=1}^l y_j p_{0j} = E(X) \cdot E(Y) \end{aligned}$$

Since, X and Y are independent random variables, so,

$$p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i) \cdot P(Y = y_j) = p_{i0} \cdot p_{0j}$$

If variables X and Y are independent, then  $\text{Cov}(X,Y) = 0$ .  
So,  $\rho_{XY} = 0$ . But, the converse is not necessarily true.  
If , however, each of X and Y assumes only two distinct values, then the converse is true.

Correlation Coefficient  $\rho_{XY}$ , as a measure of linear association is given by -

$$\begin{aligned}\rho_{XY} &= \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}} \\ &= \frac{E[\{X - E(X)\}\{Y - E(Y)\}]}{\sqrt{E\{X - E(X)\}^2 E\{Y - E(Y)\}^2}} \\ &= \frac{E(XY) - E(X).E(Y)}{\sqrt{\{E(X^2) - E^2(X)\}\{E(Y^2) - E^2(Y)\}}}\end{aligned}$$

For  $k$  variables  $X_1, X_2, \dots, X_k$ , defined on the same sample space ,

$$\text{Var}(X_1 + X_2 + \dots + X_k) = \sum_{i=1}^k \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq k} \sum \text{Cov}(X_i, X_j)$$

**Proof:**  $\text{Var}(X_1 + X_2 + \dots + X_k) = E\left[\sum_{i=1}^k \{X_i - E(X_i)\}\right]^2$

$$= E\left[\sum_i \{X_i - E(X_i)\}^2 + 2 \sum_{1 \leq i < j \leq k} \sum \{X_i - E(X_i)\}\{X_j - E(X_j)\}\right]$$

$$= \sum_i E\{X_i - E(X_i)\}^2 + 2 \sum_{1 \leq i < j \leq k} \sum E[\{X_i - E(X_i)\}\{X_j - E(X_j)\}]$$

$$= \sum_i \text{Var}(X_i) + 2 \sum_{1 \leq i < j \leq k} \sum \text{Cov}(X_i, X_j)$$

# THANK YOU