UNIT -2
Two Dimensional Random Variable

Two-dimensional random variable:

Let S be the sample space associated with a random experiment E. Let $X = X(s)$ and $Y = Y(s)$ be two functions each assigning a real number to each outcomes $s \in S$. Then, the pair (X, Y) is called a ***two-dimensional random variable (r.v)*** or a ***bivariate r.v.***

$$R_{XY} = \{(x, y) \backslash s \in S \,\& X(s) = x, Y(s) = y\}$$

Discrete and continuous two-dimensional r.v:

- If the possible values of (X, Y) are finite or countably infinite. (X, Y) is called a ***two-dimensional discrete r.v.*** When (X, Y) is a two-dimensional discrete r.v, the possible values of (X, Y) may be represented as $(x_i, y_j), i = 1,2, \ldots, n;\ j = 1,2, \ldots, m$.
- If (X, Y) can assume all values in a specified region R in the xy-plane, (X, Y) is called a ***two-dimensional continuous r.v.***

CLASSIFICATION

- **Discrete bivariate random variable**

If both the random variables X and Y are discrete then (X,Y) is called a *discrete bivariate random variable*.

- **Continuous bivariate random variable**

If both the random variables X and Y are continuous then (X,Y) is called a *continuous bivariate random variable*.

- **Mixed bivariate random variable**

If one of X and Y is discrete while the other one is continuous, then (X,Y) is a *mixed bivariate random variable.*

EXAMPLE

- *EXAMPLE: I*

Consider the experiment of tossing a coin twice. The sample space is

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

Let X denotes the number of heads obtained in the first toss and Y denotes the number of heads in the second toss. Then

s	HH	HT	TH	TT
X(s)	1	1	0	0
Y(s)	1	0	1	0

EXAMPLE (contd…)

The range space of *(X, Y)* is {(1,1), (1,0), (0,1), (0,0)} which is finite.

(X, Y) is a two-dimensional discrete random variable.

EXAMPLE: II

Consider the random experiment in which two fair dice are thrown simultaneously. Define a random variable *X* as the number on the first die and *Y* as the number on the second die. Then *(X,Y)* is a *bivariate random variable*.

Joint probability mass function of $(\boldsymbol{X},\boldsymbol{Y})$

- *If* (X, Y) *is a two-dimensional discrete r.v such that* $P(X = x_i, Y = y_j) = P(X = x_i \cap Y = y_j) = p(x_i, y_j) = p_{ij}$, *then* p_{ij} *is called* ***joint probability mass function*** *or simply* ***joint probability function*** *of* (X, Y) *provided the following conditions are satisfied*
- $p_{ij} \geq 0$ *for all i,j.*
- $\sum_j \sum_i p_{ij} = 1$

The set of triples $\{x_i y_j, p_{ij}\}, i = 1,2, \dots, n;\ j = 1,2, \dots, m$ *is called the joint probability distribution of* (X, Y) *and it can be given in the form of table as given below*

y / x	y_1	y_2	$\cdots$	y_m	$\boldsymbol{p(x_i)}$	
x_1	p_{11}	p_{12}	$\cdots$	p_{1m}	$\boldsymbol{p_{1\bullet}}$	$\boldsymbol{P(X = x_1)}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2m}	$\boldsymbol{p_{2\bullet}}$	$\boldsymbol{P(X = x_2)}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$	$\vdots$
x_n	p_{n1}	p_{n2}	$\cdots$	p_{nm}	$\boldsymbol{p_{n\bullet}}$	$\boldsymbol{P(X = x_n)}$
p(y_j)	$\boldsymbol{p_{\bullet 1}}$	$\boldsymbol{p_{\bullet 2}}$	$\cdots$	$\boldsymbol{p_{\bullet m}}$	1	
	$\boldsymbol{P(Y = y_1)}$	$\boldsymbol{P(Y = y_2)}$	$\cdots$	$\boldsymbol{P(Y = y_m)}$		

MARGINAL PROBABILITY MASS FUNCTION

Let $P_{XY}(x_i, y_j)$ is the joint probability mass function of (X, Y). Suppose for a fixed value $X = x_i$ the random variable *Y* can take the possible values $y_j, j = 1,2,3, \dots m$. Then, the probability distribution of *X* is

$$P_X(x_i) = P(X = x_i) = \sum_{y_j} P_{XY}(x_i, y_j)$$

and is called the marginal probability mass function of *X*. Similarly

$$P_Y(y_j) = P(Y = y_j) = \sum_{x_i} P_{XY}(x_i, y_j)$$

is called the marginal probability mass function of *Y*

Conditional probabilities

- The conditional probability function of X given $Y = y_j$ is given by

$$P(\boldsymbol{X} = \boldsymbol{x_i} | \boldsymbol{Y} = y_j) = P(\boldsymbol{X} = \boldsymbol{x_i} \cap \boldsymbol{Y} = y_j) / \boldsymbol{P}(\boldsymbol{Y} = y_j)$$

$$= \frac{p_{ij}}{\boldsymbol{p_{\bullet j}}}$$

- Similarly the conditional probability function of Y given $X = x_i$ is given by

$$P(\boldsymbol{Y} = y_j | \boldsymbol{X} = \boldsymbol{x_i}) = P(\boldsymbol{Y} = y_j \cap \boldsymbol{X} = \boldsymbol{x_i}) / \boldsymbol{P}(\boldsymbol{X} = \boldsymbol{x_i})$$

$$= \frac{p_{ij}}{\boldsymbol{p_{i\bullet}}}$$

Joint probability density function:

- If X and Y are continuous random variables then $f(x, y)$ is said to be joint probability function or joint pdf of two random variables X and Y, if
- $P[a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y)dydx$

provided (i) $f(x, y) \geq 0$

(ii) $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)dydx = 1$

Marginal density function of X

$$f(x) = \int_{-\infty}^{\infty} f(x, y)dy$$

Marginal density function of Y

$$f(y) = \int_{-\infty}^{\infty} f(x, y)dx$$

Joint probability distribution for continuous r.vs X and Y:

The joint probability distribution function of two-dimensional r.vs (X, Y) is defined by

$$F(x, y) = P(X \leq x, Y \leq y)$$

$$= \int_{-\infty}^{y} \int_{-\infty}^{x} f(x, y) dxdy$$

Properties:

- $F(-\infty, y) = 0$
- (ii) $F(x, \infty) = 0$
- (iii) $F(\infty, \infty) = 1$

Conditional probability distribution

- The conditional probability function of Y given X, where X and Y are continuously distributed is given by $f\left(\frac{y}{x}\right) = \frac{f(x,y)}{f(x)}$, provided $f(x) > 0$.
- The conditional probability function of X given Y, where X and Y are continuously distributed is given by $f\left(\frac{x}{y}\right) = \frac{f(x,y)}{f(y)}$, provided $f(y) > 0$.

NOTE:

Two r.vs X and Y are independent if $f(x, y) = f(x).f(y)$.

- $P(a < X < b/Y = y) = \int_a^b [f(x/y)]_{Y=y} dx$
- $P[(a < X < b) \cap (c < Y < d)] = \int_c^d \int_a^b f(x, y)dxdy$
- $P[(a < X < b)/(c < Y < d)] = \frac{P[(a<X<b)\cap(c<Y<d)]}{P(c<Y<d)}$

Covariance:

$COV(X,Y) = E(XY) - E(X).E(Y)$

NOTE: If *X* and *Y* are independent then $COV(X,Y) = 0$, but not vice-versa.

X & *Y* are raw data

$$E(X) = \frac{\sum x_i}{n}$$

$$E(Y) = \frac{\sum y_j}{n} \qquad E(XY) = \frac{\sum x_i y_j}{n}$$

X & *Y* are continuous

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

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

$$E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy$$

X & *Y* are discrete

Then $E(X) = \sum xp(x)$ $\qquad E(Y) = \sum yp(y)$ $\qquad E(XY) = \sum\sum x_i\, y_j p(x_i y_j)$

Coefficient of correlation:

$$r(X,Y) = \frac{COV(X,Y)}{\sigma_x \sigma_y}$$

Where $\sigma_x^2 = var(X) = E(X^2) - (E(X))^2$

$$\sigma_y^2 = var(Y) = E(Y^2) - (E(Y))^2$$

Also, If X is discrete, $E(X^2) = \sum x^2 p(x)$ if probability is given, otherwise

$$E(X) = \frac{\sum {x_i}^2}{n}$$

If X is continuous, $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) dx$. Similarly, we can define $E(Y^2)$.

NOTE:

$-1 \leq r(X,Y) \leq 1$

If $r(X,Y) = 0$, then *X* and *Y* are said to be uncorrelated.

Regression:

There are 2 lines of regression

Line of regression of *Y* on *X*

$$y - \bar{y} = b_{yx}(x - \bar{x})$$

Line of regression of *X* on *Y*

$$x - \bar{x} = b_{xy}(y - \bar{y})$$

b_{yx} and b_{xy} are said to be ***regression coefficient*** of *Y* on *X* and *X* on *Y* respectively and are given by

$b_{yx} = r\frac{\sigma_y}{\sigma_x}$ and $b_{xy} = r\frac{\sigma_x}{\sigma_y}$.

$\bar{x} = E(X)$ and $\bar{y} = E(Y)$

NOTE:

$$r = \sqrt{b_{xy}b_{yx}}$$

The point of intersection of 2 regression lines is $(\bar{x}, \bar{y})$, that is solving 2 regression lines, we get mean of *X* and mean of *Y*.

If θ is the angle between 2 regression lines of 2 variables *X* and *Y*, then $tan\theta = \left(\frac{1-r^2}{r}\right)\frac{\sigma_x\sigma_y}{\sigma_x{}^2+\sigma_y{}^2}$


Thank
You