SSN COLLEGE OF ENGINEERING

Department of Mathematics

UNIT – 2: TWO- DIMENSIONAL RANDOM VARIABLES

SEC: 1 JOINT PDF, CDF & MARGINAL DISTRIBUTIONS

- **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.***

- **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 $(x_i, y_j), i = 1,2, \dots, n;\ j = 1,2, \dots, m.$

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.***

- **Joint probability function of (X, Y):**

If (X, Y) is a two-dimensional discrete r.v such that $P(X = x_i, Y = y_j) = p(x_i y_j) = p_{ij}$, then p_{ij} is called ***joint probability mass function*** or simply ***joint probability function*** of (X, Y) provided the following conditions are satisfied

(i) $p_{ij} \geq 0$ for all i,j.

(ii) $\sum_j \sum_i p_{ij} = 1$

- ## Joint probability distribution of (X, Y):

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

$x \backslash y$	y_1	y_2	$\cdots$	y_m	$p(x_i)$	
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}	$\boldsymbol{p_{n\bullet}}$	$\boldsymbol{P(X = x_n)}$
$\mathbf{p}(y_j)$	$\boldsymbol{p_{\bullet 1}}$	$\boldsymbol{p_{\bullet 2}}$	$\cdots$	$\boldsymbol{p_{\bullet m}}$	**1**	
	$\boldsymbol{P(Y = y_1)}$	$\boldsymbol{P(Y = y_2)}$	$\cdots$	$\boldsymbol{P(Y = y_m)}$		

- ## Marginal probability function of (X, Y):

If the joint probability distribution of two random variables X and Y is given, then the marginal probability of X is given by the set $\{x_i, p_{i\bullet}\}$ which can be given in the form of table as follows:

X	x_1	x_2	$\cdots$	x_n
$\boldsymbol{P(X = x_i)}$ or $\boldsymbol{p(x_i)}$	$\boldsymbol{p_{1\bullet}}$	$\boldsymbol{p_{2\bullet}}$	$\cdots$	$\boldsymbol{p_{n\bullet}}$

`Similarly, the set $\{y_j, p_{\bullet j}\}$ is called the marginal distribution of Y and is given in the form of table as follows:

Y	y_1	y_2	$\cdots$	y_m
$\boldsymbol{P(Y = y_j)}$ or $\boldsymbol{p(y_j)}$	$\boldsymbol{p_{\bullet 1}}$	$\boldsymbol{p_{\bullet 2}}$	$\cdots$	$\boldsymbol{p_{\bullet m}}$

- ## Conditional probabilities:

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

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

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

It can also be denoted by

$$f(x/y) = \frac{f(x,y)}{f(y)}$$

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

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

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

It can also be denoted by

$$f(y/x) = \frac{f(x,y)}{f(x)}$$

NOTE:

Two random variables X and Y are said to be independent if $p_{ij} = p_{i\bullet}.p_{\bullet j}$ for all i and j.

- ## 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

$$P[a_1 \le X \le b_1, a_2 \le Y \le b_2] = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f(x, y)dydx$$

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

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

- ### 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:

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

- ### Relation between joint pdf and joint cdf:

$$f(x, y) = \frac{\partial^2}{\partial x \partial y} F(x, y)$$

- ### Marginal density function of X

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

- ### Marginal density function of Y

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

NOTE:

When finding the marginal density function of X and Y, if variables limits are given either for X or for Y in the joint p.d.f $f(x, y)$, sketch the region of integration to get the limit in terms of x for y, to find $f(x)$ and vice-versa.

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

SEC 2: FUNCTION OF RANDOM VARIABLES:

Here two random variables X and Y with their joint p.d.f $f(x,y)$ will be given. Let the two new random variables U and V are given by the transformation $U = u(x,y)$ and $V = v(x,y)$. Now our problem is to find the p.d.f of U and / or p.d.f of V.

The joint p.d.f of the transformed variables U and V is given by $g(u,v) = |J|f(x,y)$ where

$J = \frac{\partial(x,y)}{\partial(u,v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$ is the Jacobian of the transformation.

Marginal density function of U and V can be obtained from the joint p.d.f. $g(u,v)$

NOTE:

If the r.v U is of the form XY or $X + Y$, we preferably take the auxiliary r.v V as X. If U is of the form $\frac{X}{Y}$ or $X - Y$ then we take V as Y.

SEC 3: COVARIANCE, CORRELATION & REGRESSION

- **Covariance:**

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

If X & Y are discrete,

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

If X & Y are continuous,

$E(X) = \int_{-\infty}^{\infty} xf(x)dx$ $\quad E(Y) = \int_{-\infty}^{\infty} yf(y)dy$ $\quad E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xyf(x,y)dxdy$

If X & Y are discrete with probability values,

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

NOTE:

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

- **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:

- $-1 \le r(X,Y) \le 1$
- If $r(X,Y) = 0$, then X and Y are said to be uncorrelated.

- **Regression:**

There are 2 lines of regression

1. Line of regression of Y on X

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

2. Line of regression of X on Y

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

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

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

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

NOTE:

1. $r = \sqrt{b_{xy}b_{yx}}$
2. 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.
3. If θ is the angle between 2 regression lines of 2 variables X and Y, then

$$tan\theta = \left(\frac{1-r^2}{r}\right)\frac{\sigma_x\sigma_y}{\sigma_x{}^2+\sigma_y{}^2}$$

- **Regression Curve (NOT IN SYLLABUS):**

The regression curve Y on X is $y = E(Y/x) = \int_{-\infty}^{\infty} yf(y/x)dy$

The regression curve X on Y is $x = E(X/y) = \int_{-\infty}^{\infty} xf(x/y)dy$

SEC.4: CENTRAL LIMIT THEOREM (CLT)

Liapounoff's form:

If $X_1, X_2, \dots, X_n$ is a sequence of independent random variables with $E(X_i) = \mu_i$ and $Var(X_i) = \sigma_i^2$, $i = 1,2, \dots, n$ and if $S_n = X_1 + X_2 + \cdots + X_n$ then under certain general conditions S_n follows a normal distribution with mean $\mu = \sum_{i=1}^{n}\mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n}\sigma_i^2$ as n tends to infinity.

Lindeberg-Levy's form:

If $X_1, X_2, \dots, X_n$ is a sequence of independent identically distributed (i.i.d) random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2$, $i = 1,2, \dots, n$ and if $S_n = X_1 + X_2 + \cdots + X_n$ then under certain general conditions S_n follows a normal distribution with mean $n\mu$ and variance $n\sigma^2$ as n tends to infinity.

REMARK:

If $\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$ then the mean or average $\bar{X}$ follows normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

NOTE:

For problems, n is the sample size, $E(X_i)$ is the mean of the sample and $Var(X_i)$ is the variance of the sample then

- Sum, S_n follows a normal distribution with mean $\mu = nE(X_i)$ and variance $\sigma^2 = nVar(X_i)$ as n tends to infinity.
- Average $\bar{X}$ follows normal distribution with mean $\mu = E(X_i)$ and variance $\sigma^2 = \frac{Var(X_i)}{n}$.
