PROBABILITY AND STATISTICS (UCS401)

Lecture-18

(Conditional mean and variance, Correlation and regression)

Two-dim. r.v.'s and Joint Distributions (Unit -V)

THAPAR INSTITUTE
OF ENGINEERING & TECHNOLOGY
(Deemed to be University)

Dr. Rajanish Rai

Assistant Professor

School of Mathematics

Thapar Institute of Engineering and Technology, Patiala

Contents:

- Independent Random Variables
- Conditional Means
- Conditional Variances
- Examples

Independent Random Variables

Let (X ,Y) be two dimensional random variable with joint probability function $f_{(X,Y)}(x,y)$= P[X = x, Y = y].Then (X,Y) is said to be independent if

$$f_{(X,Y)}(x,y) = f_X(x) . f_Y(y)$$

Note: *If X,Y are independent, then*

a) $f_{X|Y}(x|y) = f(x)$,

b) $f_{Y|X}(y|x) = f(y)$,

Example 1: Let X,Y have following joint probability distribution, $f_{(X,Y)}(x,y)$ .Show that X,Y are independent.

	X=2	X= 4
Y=1	0.10	0.15
Y=3	0.20	0.30
Y=5	0.10	0.15

Solution:

$Y \backslash X$	2	4	$P(Y=y)$
1	0.10	0.15	0.25
3	0.20	0.30	0.50
5	0.10	0.15	0.25
$P(X=x)$	0.40	0.60	1

From the above table we can see that

$f_X(x).f_Y(y)$ = *P(X=x).P(Y=y)=* $f_{(X,Y)}(x,y)$ *for each pair of values (x, y).*

e.g for Pair (2,1), we have $f_{(X,Y)}(2,1)$ *=0.10 and P(X=2)=0.40 and P(Y=1)=0.25*

$f_X(x).f_Y(y)$ = *P(X=x).P(Y=y)=0.25*0.40=0.10=* $f_{(X,Y)}(x,y)$

Example 2: If X,Y have the joint PDF

$$f(x) = \begin{cases} x + y, 0 < x < 1, 0 < y < 1 \\ 0 \quad otherwise \end{cases}$$

Check whether X and Y are independent or not.

Solution: The marginal density function of X is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy = \int_0^1 (x+y)dy = x + \frac{1}{2}, \qquad 0 < x < 1$$

The marginal density function of Y is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx = \int_0^1 (x+y)dx = y + \frac{1}{2}, \qquad 0 < y < 1$$

Now,

$$f(x).f(y) = (x + 1/2)(y + 1/2) \neq f(x,y)$$

Ans: X and Y are not independent.

Conditional Mean and Variance

If (X,Y) is a two-dimensional random variable, then the mean or expectation of (X,Y) is defined as follows

- **Case 1:** when X,Y are discrete random variables, then

$$E(X) = \sum_{x_i} x_i . P(X = x_i)$$

$$E(Y) = \sum_{y_j} y_j . P(Y = y_j)$$

$$E(X/Y) = \sum_{x_i} x_i . P(X = x_i/Y = y_j)$$

$$E(Y/X) = \sum_{y_j} y_j . P(Y = y_j/X = x_i)$$

$$E(XY) = \sum_{x_i} \sum_{y_j} x_i y_j P(X = x_i, Y = y_j)$$

Case 2: When X,Y are continuous random variables, then

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

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

Conditional Expected Values

$$E(X/Y) = \int_{-\infty}^{\infty} x f(x/y) dx$$

$$E(Y/X) = \int_{-\infty}^{\infty} y f(y/x) dy$$

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

Conditional Variance: If (X,Y) is a two-dimensional random variable, then the conditional variance of (X,Y) is

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

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

Notes: If X and Y are independent random variables, then

$$E(X/Y) = E(X)$$

$$E(Y/X) = E(Y)$$

$$\mathrm{E}[E(Y/X)] = E(Y)$$

$$\mathrm{E}[E(X/Y)] = E(X)$$

Example 1: The joint distribution of X and Y is given by

$$f(x,y)=\frac{x+y}{21},\ \ x=1,2,3,\ \ y=1,2$$

Find the marginal distributions of X and Y. Find the mean of X and Y also.

Solution:

$X \backslash Y$	1	2	$P(X=x_i)$	
1	$\frac{2}{21}$	$\frac{3}{21}$	$\frac{5}{21}$	$P(X=1)$
2	$\frac{3}{21}$	$\frac{4}{21}$	$\frac{7}{21}$	$P(X=2)$
3	$\frac{4}{21}$	$\frac{5}{21}$	$\frac{9}{21}$	$P(X=3)$
$P(Y=y_j)$	$\frac{9}{21}$	$\frac{12}{21}$	1	
	$P(Y=1)$	$P(Y=2)$		

The marginal distributions of X are

$$P(X=1)=\frac{5}{21},\ P(X=2)=\frac{7}{21},\ P(X=3)=\frac{9}{21}$$

The marginal distributions of Y are

$$P(Y=1)=\frac{9}{21},\ P(Y=2)=\frac{12}{21}$$

$$E(X)=\sum_{x=1}^{3} xP(X=x) = 1P(X=1)+2P(X=2)+3P(X=3)$$

$$=\frac{5}{21}+\frac{14}{21}+\frac{27}{21}=\frac{46}{21}$$

$$E(Y)=\sum_{y=1}^{2} yP(Y=y) = 1P(Y=1)+2P(Y=2)$$

$$=\frac{9}{21}+\frac{24}{21}=\frac{33}{21}=\frac{11}{7}$$

Example 2: The joint PDF of (X,Y) is given by

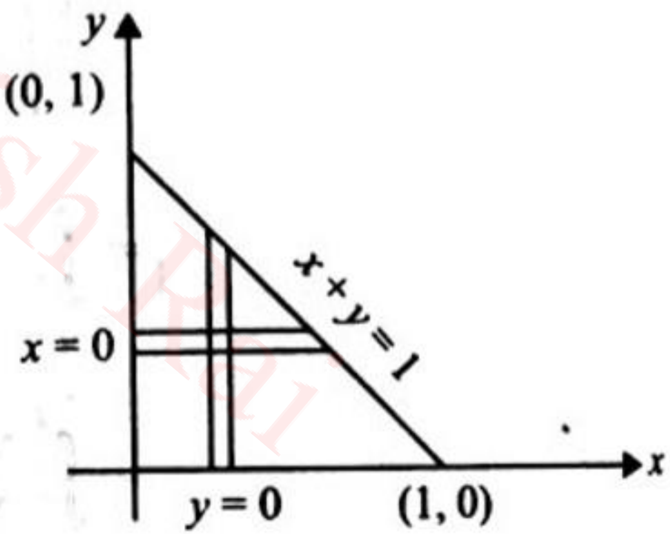
$$f(x,y)=\begin{cases}24xy, 0<x, 0<y, x+y\le 1\\ 0 \quad otherwise\end{cases}$$

Find the conditional mean and variance of Y given X.

Solution:

Given: $f(x, y) = 24xy, \quad x > 0, y > 0, x + y \le 1$

$$\therefore \quad f_X(x) = \int_0^{1-x} 24xy\, dy = 24x \int_0^{1-x} y\, dy,$$



$$= 24x\left[\frac{y^2}{2}\right]_0^{1-x} = 24x\frac{(1-x)^2}{2}$$

$$= 12x\,(1-x)^2, \qquad 0 < x < 1$$

$$f(y/x) = \frac{f(x, y)}{f_X(x)} = \frac{2y}{(1-x)^2}, \qquad 0 < y < 1 - x$$

$$E(Y/X) = \int_0^{1-x} y f(y/x)\,dy$$

$$= \int_0^{1-x} \frac{2y^2}{(1-x)^2}\,dy = \frac{2}{(1-x)^2}\left[\frac{y^3}{3}\right]_0^{1-x} = \frac{2}{3}(1-x), \qquad x > 0$$

$$E(Y^2/X = x) = \int_0^{1-x} y^2 f(y/x)\,dy$$

$$= \int_0^{1-x} y^2 \frac{2y}{(1-x)^2}\,dy = \frac{2}{(1-x)^2}\left[\frac{y^4}{4}\right]_0^{1-x} = \frac{1}{2}(1-x)^2, \qquad x > 0$$

$$\mathrm{Var}(Y/X) = E(Y^2/X) - [E(Y/X)]^2$$

$$= \frac{1}{2}(1-x)^2 - \frac{4}{9}(1-x)^2 = \frac{1}{18}(1-x)^2, \qquad x > 0$$

Correlation and regression

Covariance: Let X and Y be two random variables defined on the same probability space. The covariance of X and Y, denoted as $cov(X,Y)$, is defined as

$$cov(X,Y) = E[\{X - E[X]\}\{Y - E[Y]\}] \quad \textbf{OR}$$

$$cov(X,Y) = E[XY] - E[X].E[Y]$$

Correlation coefficient: The correlation coefficient of two random variables X and Y, denoted by $\rho(X,Y)$, is defined as $\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{var[X].var[Y]}} = \frac{cov(X,Y)}{\sqrt{\sigma_X.\sigma_Y}}$, where $\sigma_X > 0$ and $\sigma_Y > 0$.

It can be verify that $-1 \leq \rho(X,Y) \leq 1$.

Two random variables X and Y are uncorrelated if

$$\rho(X,Y) = 0 \quad or \quad cov(X,Y) = 0$$

$\Rightarrow$ X and Y are independent random variables.

Correlation and regression contd..

In particular, $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy.f_{X,Y}(x,y)\,dxdy$, $E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x.f_{X,Y}(x,y)\,dxdy$ and

$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y.f_{X,Y}(x,y)\,dxdy$.

Conditional Expectation: Let (X,Y) be a two dimensional continuous random variable with joint continuous density function $f_{X,Y}(x,y)$, then the conditional expectation of $g(X,Y)$ given $X = x$, denoted by $E[g(X,Y)|X = x]$, is defined as $E[g(X,Y)|X = x] = \int_{-\infty}^{\infty} g(x,y).f_{Y|X}(y|x)\,dy$.

In particular, $E[Y|X = x] = \int_{-\infty}^{\infty} y.f_{Y|X}(y|x)\,dy$.

Regression curve:

1. The relation $y = E[Y|X = x]$ is called the regression curve of Y on $X = x$.
2. The relation $x = E[X|Y = y]$ is called the regression curve of X on $Y = y$.

Example 2. Let the joint p.d.f. of X and Y be $f(x,y) = \begin{cases} (x+y), & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & otherwise. \end{cases}$

Find **(i)** $P\left[0 < X < \frac{1}{2}, 0 < Y < \frac{1}{4}\right]$, **(ii)** $E[X],\ E[Y], E[XY]\ and\ E[X+Y]$ **(iii)** $\sigma(X,Y)$.

Solution: We have $P\left[0 < X < \frac{1}{2}, 0 < Y < \frac{1}{4}\right] = \int_0^{\frac{1}{2}}\left[\int_0^{\frac{1}{4}}(x+y)\,dy\right]dx$

$$= \int_0^{1/2}\left|xy + \frac{1}{2}y^2\right|_0^{1/4} dx = \int_0^{1/2}\left(\frac{1}{4}x + \frac{1}{32}\right)dx = \frac{3}{64}.$$

Now $E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x.f(x,y)\,dxdy = \int_0^1\int_0^1 x.(x+y)\,dxdy = \int_0^1\left(x^2 + \frac{1}{2}x\right)dx = \frac{7}{12}.$

Similarly, $E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y.f(x,y)\,dxdy = \int_0^1\int_0^1 y.(x+y)\,dxdy = \frac{7}{12}.$

Now $E[XY] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy.f(x,y)\,dxdy = \int_0^1\int_0^1 xy.(x+y)\,dxdy = \frac{1}{3}.$

$$E[X+Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x+y).f(x,y)\,dxdy = \int_0^1\int_0^1(x+y).(x+y)\,dxdy = \frac{7}{6}.$$

We have $cov[X,Y] = E[XY] - E[X].E[Y] = -\frac{1}{144}.$

$$E[X^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^2.f(x,y)\,dxdy = \int_0^1\int_0^1 x^2.(x+y)\,dxdy = \frac{5}{12}.$$

$$E[Y^2] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y^2.f(x,y)\,dxdy = \int_0^1\int_0^1 y^2.(x+y)\,dxdy = \frac{5}{12}.$$

Now $var[X] = E[X^2] - (E[X])^2 = \frac{5}{12} - \left(\frac{7}{12}\right)^2 = \frac{11}{144}$.

$$var[Y] = E[Y^2] - (E[Y])^2 = \frac{11}{144}.$$

Hence $\rho(X,Y) = \frac{cov(X,Y)}{\sqrt{var[X].var[Y]}} = \frac{-\frac{1}{144}}{\sqrt{\frac{11}{144}.\frac{11}{144}}} = -\frac{1}{11}$.

THANK YOU