BITS Pilani
K K Birla Goa Campus

Probability & Statistics,

Dr. Jajati Keshari Sahoo
Department of Mathematics

Joint Distributions

- In many statistical investigations, one is frequently interested in studying the relationship between two or more random variables, such as the relationship between annual income and yearly savings per family or the relationship between occupation and hypertension.

Discrete joint distributions

- For two discrete random variables *X* and *Y*, the probability that *X* will take the value *x* and *Y* will take the value *y* is written as $P(\boldsymbol{X = x, Y = y})$.

- Consequently, $P(X = x, Y = y)$ is the probability of the intersection of the events $X = x$ and $Y = y$.

- If *X* and *Y* are discrete random variables, the function given by $f_{XY}(x, y) = P(X = x, Y = y)$ for each pair of values (x, y) within the range of *X* and *Y* is called the **joint probability distribution or joint density function** of *X* and *Y*.

Discrete joint distributions

Necessary and Sufficient conditions for a function to be act as joint discrete density

$$1.\ f_{XY}(x, y) \geq 0,$$

$$2.\ \sum_{\text{all } x} \sum_{\text{all } y} f(x, y) = 1.$$

Discrete joint distributions

Joint cumulative distribution function:

If *X* and *Y* are discrete random variables, the function given by

$$F(x,y) = P(X \le x, Y \le y) = \sum_{s \le x} \sum_{t \le y} f(s,t)$$

is called the **joint distribution function,** or the **joint cumulative distribution** of *X* and *Y*.

Discrete joint distributions

Marginal distribution: If X and Y are discrete random variables with joint density $f(x, y)$ then

$$P(X = x) = f_X(x) = \sum_{\text{all } y} f(x, y)$$

is called the **marginal distribution** of X.

Similarly, $P(Y = y) = f_Y(y) = \sum_{\text{all } x} f(x, y)$

is called the **marginal distribution** of Y.

Discrete joint distributions

Mean:

For given random variables X, and Y with joint density function $f(x, y)$ the function $H(X, Y)$ is also a random variable.

The random variable $H(X, Y)$ has expected value, or mean, given by

$$E[H(X,Y)] = \sum_{\text{all } x} \sum_{\text{all } y} H(x,y) f(x,y)$$

provided this summation exists.

Discrete joint distributions

Univariate Averages Found Via the Joint Density

$$E[X] = \sum_{\text{all } x} \sum_{\text{all y}} x f(x, y)$$

$$E[Y] = \sum_{\text{all } x} \sum_{\text{all y}} y f(x, y)$$

Discrete joint distributions

Example 1:

There are 8 similar chips in a bowl: 3 marked (0,0), 2 marked with (0,1), and one marked (1,1). A player selects a chip at random and is given the sum of the two coordinates in rupees. If X and Y represents those coordinates, respectively. Find the joint density of X,Y. Also calculate the expected payoff and marginal densities.

Discrete joint distributions

Example 2: Two scanners are needed for an experiment. Of the five, two have electronic defects, another one has a defect in memory, and two are in good working order. Two units are selected at random.

(a) Find the joint probability distribution of X (the number with electronic defects) and Y (the number with a defect in memory.

(b) Find the probability of 0 or 1 total defects among the two selected.

(c) Find the marginal probability distribution of X.

(d) Find mean of Y.

Discrete joint distributions

Solution: (a)

$$f(x,y) = \frac{\binom{2}{x}\binom{1}{y}\binom{2}{2-x-y}}{\binom{5}{2}}$$

where $x = 0, 1, 2$ and $y = 0, 1$

$0 \le x + y \le 2$

The joint distribution table:

$f(x,y)$		y = 0	y = 1	$f_X(x)$
x	0	0.1	0.2	0.3
	1	0.4	0.2	0.6
	2	0.1	0.0	0.1
$f_Y(y)$		0.6	0.4	1

Discrete joint distributions

(b) Let A be the event that $X + Y$ equal to 0 or 1

$$P(A) = f(0, 0) + f(0, 1) + f(1, 0)$$
$$= 0.1 + 0.2 + 0.4 = 0.7$$

(c) The marginal probability distribution of X is given by

$$f_X(x) = \sum_{y=0}^{y=1} f(x, y) = f(x, 0) + f(x, 1)$$

$$f_X(0) = f(0,0) + f(0,1) = 0.1 + 0.2 = 0.3$$

$$f_X(1) = f(1,0) + f(1,1) = 0.4 + 0.2 = 0.6$$

$$f_1(2) = f(2,0) + f(2,1) = 0.1 + 0.0 = 0.1$$

Discrete joint distributions

(d) The mean of *Y* is given by

$$E(Y) = \sum_{x=0}^{2} \sum_{y=0}^{1} y f(x, y)$$

$$= \sum_{x=0}^{2} f(x,1)$$

$$= f(0,1) + f(1,1) + f(2,1)$$

$$= 0.2 + 0.2 + 0.0 = 0.4$$

Discrete joint distributions

Example 3:

A coin is tossed 3 times. Let Y denotes the number of heads and X denotes the absolute difference between the number of heads and tails. Find the joint density of X and Y

Discrete joint distributions

Solution 3:

y

x

$f(x,y)$	0	1	2	3	$f_X(x)$
1	0	3/8	3/8	0	3/4
3	1/8	0	0	1/8	1/4
$f_Y(y)$	1/8	3/8	3/8	1/8	1

Continuous joint distributions

There are many situation in which we describe an outcome by giving the value of several continuous variables.

- For instance, we may measure the weight and the hardness of a rock, the volume, pressure and temperature of a gas, or the thickness, color, compressive strength and potassium content of a piece of glass.

Continuous joint distributions

Definition:

Let *X* and *Y* be continuous random variables. A function f_{XY} such that

1. $f_{XY}(x, y) \geq 0$

2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\,dy\,dx = 1$

3. $P[a \leq X \leq b \text{ and } c \leq Y \leq d] = \int_{a}^{b} \int_{c}^{d} f_{XY}(x, y)\,dy\,dx$

is called the joint density for (X, Y).

Continuous joint distributions

The **joint cumulative distribution function** of *X* and *Y*, is defined by

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

$$= \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,dt\,ds$$

Also partial differentiation yields

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

whenever these partial derivatives exists.

Continuous joint distributions

Marginal densities: If X and Y are continuous random variables with joint density $f(x, y)$ then

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

is called the **marginal distribution** of X.

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

is called the **marginal distribution** of Y.

Continuous joint distributions

Mean:

For given random variables X, and Y with joint density function $f(x, y)$ the function $H(X, Y)$ is also a continuous random variable.

The random variable $H(X, Y)$ has expected value, or mean, given by

$$E[H(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} H(x,y)\, f(x,y)\, dy\, dx$$

provided this integral exists.

Continuous distributions

Univariate Averages Found Via the Joint Density

$$E[X] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xf(x,y)\,dy\,dx$$

$$E[Y] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} yf(x,y)\,dy\,dx$$

Continuous distributions

Independent Random Variables

Definition: Let X and Y be two random variables with joint density $f(x, y)$ and marginal densities $f_X(x)$ and $f_Y(y)$ respectively then X and Y are independent

if and only if

$$f(x, y) = f_X(x) f_Y(y).$$

Continuous distributions

Example-4

A gun is aimed at a certain point (origin of a coordinate system). Because of random failure, the actual hit can be any point *(X,Y)* in a circle of radius *R* about the origin. Assume that joint density is uniform over the circle

(a) Find the joint density

(b) Find the marginal densities

(c) Are *X* and *Y* are independent ?

Continuous distributions

Example-5

The joint density for (X, Y), is given by

$$f(x,y)=\begin{cases} xy, & 0<x<1, 0<y<2, \\ 0, & \text{elsewhere.} \end{cases}$$

Calculate $P[X<1 \cup Y<1]$ and $P[X+Y<1]$.

Continuous distributions

Example-6

Two points are selected randomly on a line of length L so as to be on opposite sides of the midpoint of the line. find the probability that the distance between the two points is greater than $L/3$.

Ans:0.77

Continuous distributions

Example-7

A man and woman decide to meet at a certain location. If each person independently arrives at a time uniformly distributed btween 10.00 and 11.00AM, find the probability that the first to arrive has to wait atleast 10 minutes.

Ans: 25/36

Continuous distributions

Example-8

An accident occurs at a point X that is uniformly distributed on a road of length L. At the time of the accident, an ambulance is at a location Y that is also uniformly distributed on the road. Assuming that X and Y are independent, find the expected distance between the ambulance and the point of the accident. Ans: L/3

Continuous distributions

Conditional density

Definition: Let X and Y be two random variables with joint density $f(x, y)$ and marginal densities $f_X(x)$ and $f_Y(y)$ respectively. Then the conditional density of X given $Y=y$ is defined as

$$f_{X|y}(x) = \frac{f(x, y)}{f_Y(y)}, \text{ provided } f_Y(y) > 0.$$

Continuous distributions

Conditional density

Definition: Let X and Y be two random variables with joint density $f(x, y)$ and marginal densities $f_X(x)$ and $f_Y(y)$ respectively. Then the conditional density of Y given $X=x$ is defined as

$$f_{Y|x}(y) = \frac{f(x, y)}{f_X(x)}, \text{ provided } f_X(x) > 0.$$

Continuous distributions

Conditional means

Definition: The the conditional mean of X given $Y=y$ is defined as

$$\mu_{X|y} = E\left[X \mid_{Y=y} \right]$$

$$= \begin{cases} \int_{-\infty}^{\infty} x f_{X|y}(x)\,dx \ \text{ if continuous,} \\ \sum_{\text{all } x} x f_{X|y}(x) \ \text{ if discrete.} \end{cases}$$

Continuous distributions

Conditional means

Definition: The the conditional mean of Y given $X=x$ is defined as

$$\mu_{Y|x} = E\left[Y|_{X=x}\right]$$

$$= \begin{cases} \int_{-\infty}^{\infty} y f_{Y|x}(y)\,dy \text{ if continuous,} \\ \sum_{\text{all y}} y f_{Y|x}(y) \text{ if discrete.} \end{cases}$$

Continuous distributions

Curves of regression

Definition 1:

The graph of $\mu_{X|y}$ is called curve of regression of X on Y.

Definition 2:

The graph of $\mu_{Y|x}$ is called curve of regression of Y on X.

Continuous distributions

Example-9

If the joint density of X and Y is given by

$$f(x,y)=\frac{1}{x},\ 0<y<x<1.$$

(a) Find condtional densities.

(b) Find conditional means.

(c) Find curves of regression.

(d) Are X and Y independent ?

Continuous distributions

Example-10

If the joint density of X and Y is given by

$$f(x,y)=\frac{1}{10},\ 1\le y\le x\le 4.$$

(a) Find condtional densities.

(b) Find conditional means.

(c) Find curves of regression.

(d) Are X and Y independent ?

Continuous distributions

Example-11

Choose a number X at random from the set of numbers $\{1,2,3,4,5\}$. Now choose another number at random from the subset $\{1,\cdots,X\}$. Call this second number Y. Find the joint density of (X,Y).

Continuous distributions

Covariance

Definition:

Let X and Y be two random variables with means μ_X and μ_Y respectively. The covariance of X and Y is defined as

$$\text{Cov}(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right].$$

Computation formula for Covariance:

$$\text{Cov}(X, Y) = E[XY] - E[X]E[Y].$$

Continuous distributions

Theorems:

Theorem 1:

Let X and Y be any two random variables, then

$$\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).$$

Theorem 2:

If X and Y are independent then

$$E[XY] - E[X]E[Y].$$

Continuous distributions

Observations:

- If X and Y are independent then Cov$(X,Y)=0$.
- Cov$(X,Y)=0$ does not imply that X and Y are independent

Example 12:

y

x

$f(x,y)$	1	-1	$f(x)$
1	0.5	0.5	1
$f(y)$	0.5	0.5	1

Here $X = Y^2, E(X) = 1, E(XY) = 0, E(Y) = 0.$

Continuous distributions

Observations:

Example 13:

Suppose $X \sim U(-1,\ 1)$ and $Y = X^2$ (so X and Y are clearly dependent).

But $E[X] = \int_{-1}^{1} \frac{x}{2} dx = 0$ and

$$E[XY] = E[X^3] = \int_{-1}^{1} \frac{x^3}{2} dx = 0,$$

so $\text{Cov}(X, Y) = E[XY] - E[X]E[Y] = 0.$

Continuous distributions

Correlation:

Definition:

Let X and Y be two random variables.

The correlation between X and Y is

defined as $\rho = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$.

Continuous distributions

Example-14

If the joint density of X and Y is given by

$$f(x, y) = \frac{1}{200},\ 20 < y < x < 40.$$

(a) Find $\mathrm{Cov}(X, Y)$.

(b) Find correlation of X and Y.