Unit 1 (ii)

Joint Probability Distributions

If X and Y are two discrete random variables the probability distribution for their simultaneous occurrence can be represented by a function with values **f (x, y)** for any pair of values (**x, y**) within the range of the random variables **X** and **Y** .

Definition1

The function $f(x, y)$ is a **joint probability distribution** or **probability mass function** of the discrete random variables X and Y if

1. $f(x, y) \geq 0$ for all (x, y),
2. $\sum\limits_x \sum\limits_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

For any region A in the xy plane, $P[(X, Y) \in A] = \sum \sum\limits_A f(x, y)$.

Definition2

The function $f(x, y)$ is a **joint density function** of the continuous random variables X and Y if

1. $f(x, y) \geq 0$, for all (x, y),
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1$,
3. $P[(X, Y) \in A] = \int \int_A f(x, y)\, dx\, dy$, for any region A in the xy plane.

Definition3

The **marginal distributions** of X alone and of Y alone are

$$g(x) = \sum_y f(x, y) \quad \text{and} \quad h(y) = \sum_x f(x, y)$$

for the discrete case, and

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \quad \text{and} \quad h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

for the continuous case.

Note

$$P(a < X < b) = P(a < X < b, -\infty < Y < \infty)$$
$$= \int_a^b \int_{-\infty}^{\infty} f(x, y)\, dy\, dx = \int_a^b g(x)\, dx.$$

Definition4

Let X and Y be two random variables, discrete or continuous. The **conditional distribution** of the random variable Y given that $X = x$ is

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

Similarly, the conditional distribution of X given that $Y = y$ is

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

Also

$$P(a < X < b \mid Y = y) = \sum_{a<x<b} f(x|y),$$

$$P(a < X < b \mid Y = y) = \int_a^b f(x|y)\, dx.$$

Example1

Two ballpoint pens are selected at random from a box that contains 3 blue pens, 2 red pens, and 3 green pens. If X is the number of blue pens selected and Y is the number of red pens selected, find

(a) the joint probability function $f(x, y)$,

(b) $P[(X, Y) \in A]$, where A is the region $\{(x, y)|x + y \leq 1\}$.

c) give the marginal distribution of X alone and of Y alone.

d) find the conditional distribution of X, given that $Y = 1$, and use it to determine $P(X = 0 \mid Y = 1)$.

Sol:

The possible pairs of values (x, y) are $(0,0)$, $(0,1)$, $(1,0)$, $(1,1)$, $(0,2)$, and $(2,0)$.

$$f(x,y) = \frac{\binom{3}{x}\binom{2}{y}\binom{3}{2-x-y}}{\binom{8}{2}},$$

for $x = 0, 1, 2$; $y = 0, 1, 2$; and $0 \le x + y \le 2$.

	$f(x,y)$	x = 0	1	2	Row Totals
y	0	$\frac{3}{28}$	$\frac{9}{28}$	$\frac{3}{28}$	$\frac{15}{28}$
	1	$\frac{3}{14}$	$\frac{3}{14}$	0	$\frac{3}{7}$
	2	$\frac{1}{28}$	0	0	$\frac{1}{28}$
Column Totals		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

(b) The probability that (X, Y) fall in the region A is

$$P[(X,Y) \in A] = P(X+Y \le 1) = f(0,0) + f(0,1) + f(1,0)$$
$$= \frac{3}{28} + \frac{3}{14} + \frac{9}{28} = \frac{9}{14}.$$

c)

For the random variable X, we see that

$$g(0) = f(0,0) + f(0,1) + f(0,2) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14},$$
$$g(1) = f(1,0) + f(1,1) + f(1,2) = \frac{9}{28} + \frac{3}{14} + 0 = \frac{15}{28},$$

and

$$g(2) = f(2,0) + f(2,1) + f(2,2) = \frac{3}{28} + 0 + 0 = \frac{3}{28},$$

x	0	1	2
$g(x)$	$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$

y	0	1	2
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

d)

We need to find $f(x|y)$, where $y = 1$. First, we find that

$$h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.$$

Now

$$f(x|1) = \frac{f(x,1)}{h(1)} = \left(\frac{7}{3}\right) f(x,1), \quad x = 0,1,2.$$

Therefore,

$$f(0|1) = \left(\frac{7}{3}\right) f(0,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2}, \; f(1|1) = \left(\frac{7}{3}\right) f(1,1) = \left(\frac{7}{3}\right)\left(\frac{3}{14}\right) = \frac{1}{2},$$

$$f(2|1) = \left(\frac{7}{3}\right) f(2,1) = \left(\frac{7}{3}\right)(0) = 0,$$

and the conditional distribution of X, given that $Y = 1$, is

x	0	1	2
$f(x\|1)$	$\frac{1}{2}$	$\frac{1}{2}$	0

Finally,

$$P(X = 0 \mid Y = 1) = f(0|1) = \frac{1}{2}.$$

Therefore, if it is known that 1 of the 2 pen refills selected is red, we have a probability equal to 1/2 that the other refill is not blue. ■

Example2

A privately owned business operates both a drive-in facility and a walk-in facility. On a randomly selected day, let X and Y, respectively, be the proportions of the time that the drive-in and the walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{5}(2x+3y), & 0 \le x \le 1, 0 \le y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Verify condition 2 of Definition 3.9.

(b) Find $P[(X,Y) \in A]$, where $A = \{(x,y) \mid 0 < x < \frac{1}{2}, \frac{1}{4} < y < \frac{1}{2}\}$.

c) Find $g(x)$ and $h(y)$

(a) The integration of $f(x,y)$ over the whole region is

$$\begin{aligned}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy &= \int_0^1\int_0^1 \frac{2}{5}(2x+3y)\,dx\,dy \\ &= \int_0^1 \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Big|_{x=0}^{x=1} dy \\ &= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5}\right) dy = \left(\frac{2y}{5} + \frac{3y^2}{5}\right)\Big|_0^1 = \frac{2}{5} + \frac{3}{5} = 1.\end{aligned}$$

(b) To calculate the probability, we use

$$
\begin{aligned}
P[(X,Y) \in A] &= P\left(0 < X < \frac{1}{2}, \frac{1}{4} < Y < \frac{1}{2}\right) \\
&= \int_{1/4}^{1/2} \int_{0}^{1/2} \frac{2}{5}(2x+3y)\ dx\ dy \\
&= \int_{1/4}^{1/2} \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1/2} dy = \int_{1/4}^{1/2} \left(\frac{1}{10} + \frac{3y}{5}\right) dy \\
&= \left(\frac{y}{10} + \frac{3y^2}{10}\right)\Bigg|_{1/4}^{1/2} \\
&= \frac{1}{10}\left[\left(\frac{1}{2} + \frac{3}{4}\right) - \left(\frac{1}{4} + \frac{3}{16}\right)\right] = \frac{13}{160}.
\end{aligned}
$$

c)

$$
g(x) = \int_{-\infty}^{\infty} f(x,y)\ dy = \int_0^1 \frac{2}{5}(2x+3y)\ dy = \left(\frac{4xy}{5} + \frac{6y^2}{10}\right)\Bigg|_{y=0}^{y=1} = \frac{4x+3}{5},
$$

for $0 \le x \le 1$, and $g(x) = 0$ elsewhere. Similarly,

$$
h(y) = \int_{-\infty}^{\infty} f(x,y)\ dx = \int_0^1 \frac{2}{5}(2x+3y)\ dx = \frac{2(1+3y)}{5},
$$

for $0 \le y \le 1$, and $h(y) = 0$ elsewhere.

Example 3

The joint density for the random variables (X, Y), where X is the unit temperature change and Y is the proportion of spectrum shift that a certain atomic particle produces, is

$$
f(x,y) = \begin{cases} 10xy^2, & 0 < x < y < 1, \\ 0, & \text{elsewhere.} \end{cases}
$$

(a) Find the marginal densities $g(x)$, $h(y)$, and the conditional density $f(y|x)$.

(b) Find the probability that the spectrum shifts more than half of the total observations, given that the temperature is increased by 0.25 unit.

(a) By definition,

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_{x}^{1} 10xy^2\, dy$$
$$= \frac{10}{3}xy^3\Big|_{y=x}^{y=1} = \frac{10}{3}x(1 - x^3),\ 0 < x < 1,$$
$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_{0}^{y} 10xy^2\, dx = 5x^2y^2\Big|_{x=0}^{x=y} = 5y^4,\ 0 < y < 1.$$

Now

$$f(y|x) = \frac{f(x, y)}{g(x)} = \frac{10xy^2}{\frac{10}{3}x(1 - x^3)} = \frac{3y^2}{1 - x^3},\ 0 < x < y < 1.$$

(b) Therefore,

$$P\left(Y > \frac{1}{2}\ \middle|\ X = 0.25\right) = \int_{1/2}^{1} f(y \mid x = 0.25)\, dy = \int_{1/2}^{1} \frac{3y^2}{1 - 0.25^3}\, dy = \frac{8}{9}.$$

Definition 5

Let X and Y be two random variables, discrete or continuous, with joint probability distribution $f(x, y)$ and marginal distributions $g(x)$ and $h(y)$, respectively. The random variables X and Y are said to be **statistically independent** if and only if

$$f(x, y) = g(x)h(y)$$

for all (x, y) within their range.

Example

Using data of example 1 above

Let us consider the point $(0, 1)$. From Table 3.1 we find the three probabilities $f(0, 1)$, $g(0)$, and $h(1)$ to be

$$f(0,1) = \frac{3}{14},$$

$$g(0) = \sum_{y=0}^{2} f(0,y) = \frac{3}{28} + \frac{3}{14} + \frac{1}{28} = \frac{5}{14},$$

$$h(1) = \sum_{x=0}^{2} f(x,1) = \frac{3}{14} + \frac{3}{14} + 0 = \frac{3}{7}.$$

Clearly,

$$f(0,1) \neq g(0)h(1),$$

and therefore X and Y are not statistically independent.

Solve following questions

1.

3.38 If the joint probability distribution of X and Y is given by

$$f(x,y) = \frac{x+y}{30}, \quad \text{for } x = 0,1,2,3;\ y = 0,1,2,$$

find

(a) $P(X \leq 2, Y = 1)$;
(b) $P(X > 2, Y \leq 1)$;
(c) $P(X > Y)$;
(d) $P(X + Y = 4)$.

2. A fast-food restaurant operates both a drive through facility and a walk-in facility. On a randomly selected day, let X and Y , respectively, be the proportions of the time that the drive-through and walk-in facilities are in use, and suppose that the joint density function of these random variables is

$$f(x,y) = \begin{cases} \frac{2}{3}(x+2y), & 0 \leq x \leq 1,\ 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the marginal density of X.
(b) Find the marginal density of Y.
(c) Find the probability that the drive-through facility is busy less than one-half of the time.

3.

A candy company distributes boxes of chocolates with a mixture of creams, toffees, and cordials. Suppose that the weight of each box is 1 kilogram, but the individual weights of the creams, toffees, and cordials vary from box to box. For a randomly selected box, let X and Y represent the weights of the creams and the toffees, respectively, and suppose that the joint density function of these variables is

$$f(x, y) = \begin{cases} 24xy, & 0 \le x \le 1,\ 0 \le y \le 1,\ x + y \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(a) Find the probability that in a given box the cordials account for more than 1/2 of the weight.
(b) Find the marginal density for the weight of the creams.
(c) Find the probability that the weight of the toffees in a box is less than 1/8 of a kilogram if it is known that creams constitute 3/4 of the weight.

4. Let X and Y denote the lengths of life, in years, of two components in an electronic system. If the joint density function of these variables is

$$f(x, y) = \begin{cases} e^{-(x+y)}, & x > 0,\ y > 0, \\ 0, & \text{elsewhere,} \end{cases}$$

find $P(0 < X < 1 \mid Y = 2)$.

Also show that X and Y are statistically independent.

5. Let X denote the number of times a certain numerical control machine will malfunction: 1, 2, or 3 times on any given day. Let Y denote the number of times a technician is called on an emergency call. Their joint probability distribution is given as

		x		
$f(x, y)$		1	2	3
	1	0.05	0.05	0.10
	3	0.05	0.10	0.35
y	5	0.00	0.20	0.10

(a) Evaluate the marginal distribution of X.
(b) Evaluate the marginal distribution of Y.
(c) Find $P(Y = 3 \mid X = 2)$.