EEF 271E
Probability and Statistics

Week #5

(PREVIOUS LECTURE)

DEFINITION If X and Y are discrete random variables, the function given by $f(x, y) = P(X = x, Y = y)$ for each pair of values (x, y) within the range of X and Y is called the **joint probability distribution** of X and Y.

THEOREM A bivariate function can serve as the joint probability distribution of a pair of discrete random variables X and Y if and only if its values, $f(x, y)$, satisfy the conditions

1. $f(x, y) \geq 0$ for each pair of values (x, y) within its domain;

2. $\sum_x \sum_y f(x, y) = 1$, where the double summation extends over all possible pairs (x, y) within its domain.

		$x=0$	$x=1$	$x=2$
y	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$
	1	$\frac{2}{9}$	$\frac{1}{6}$	
	2	$\frac{1}{36}$		

EXAMPLE

Determine the value of k for which the function given by

$$f(x, y) = kxy \qquad \text{for } x = 1, 2, 3; \quad y = 1, 2, 3$$

can serve as a joint probability distribution.

Solution Substituting the various values of x and y, we get

$$f(1,1) = k, \quad f(2,2) = 4k,$$
$$f(1,2) = 2k, \quad f(2,3) = 6k,$$
$$f(1,3) = 3k, \quad f(3,1) = 3k,$$
$$f(2,1) = 2k, \quad f(3,2) = 6k,$$
$$f(3,3) = 9k.$$

constant k must be nonnegative, satisfy

$$k + 2k + 3k + 2k + 4k + 6k$$
$$+ 3k + 6k + 9k = 1$$

and $k = \frac{1}{36}$.

As in the univariate case, there are many problems in which it is of interest to know the probability that the values of two random variables are less than or equal to some real numbers x and y.

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

$$F(x, y) = P(X \leqq x, Y \leqq y) = \sum_{s \leqq x} \sum_{t \leqq y} f(s, t) \qquad \text{for} \quad -\infty < x < \infty, \; -\infty < y < \infty$$

where $f(s, t)$ is the value of the joint probability distribution of X and Y at (s, t), is called the **joint distribution function**, or the **joint cumulative distribution**, of X and Y.

EXAMPLE

With reference to Example about Red, Yellow, Blue balls find $F(1, 1)$.

		x		
		0	1	2
y	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$
	1	$\frac{2}{9}$	$\frac{1}{6}$	
	2	$\frac{1}{36}$		

Solution

$$F(1, 1) = P(X \leqq 1, Y \leqq 1)$$

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

$$= \frac{1}{6} + \frac{2}{9} + \frac{1}{3} + \frac{1}{6} \quad = \frac{8}{9}$$

Let us now extend the various concepts introduced in this section to the continuous case.

DEFINITION A bivariate function with values $f(x, y)$, defined over the xy-plane, is called a **joint probability density function** of the continuous random variables X and Y if and only if

$$P[(X, Y) \in A] = \int_A \int f(x, y)\, dx\, dy$$

for any region A in the xy-plane.

THEOREM A bivariate function can serve as a joint probability density function of a pair of continuous random variables X and Y if its values, $f(x, y)$, satisfy the conditions

1. $f(x, y) \geqq 0 \quad \text{for } -\infty < x < \infty, \quad -\infty < y < \infty;$

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

EXAMPLE

Given the joint probability density function

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

of two random variables X and Y, find $P[(X, Y) \in A]$, where A is the region $\{(x, y) | 0 < x < \frac{1}{2}, 1 < y < 2\}$.

Solution

$$P[(X,Y) \in A] = P\left(0 < X < \frac{1}{2}, 1 < Y < 2\right)$$

$$= \int_1^2 \int_0^{\frac{1}{2}} \frac{3}{5} x(y+x)\, dx\, dy$$

$$= \int_1^2 \frac{3x^2y}{10} + \frac{3x^3}{15}\bigg|^{x=\frac{1}{2}} dy$$

$$= \int_1^2 \left(\frac{3y}{40} + \frac{1}{40}\right) dy = \frac{3y^2}{80} + \frac{y}{40}\bigg|_1^2$$

$$= \frac{11}{80}$$

Definition If X and Y are continuous random variables, the function given by

$$F(x, y) = P(X \leqq x, Y \leqq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s, t)\, ds\, dt \qquad \text{for } -\infty < x < \infty, \; -\infty < y < \infty$$

where $f(s, t)$ is the value of the joint probability density of X and Y at (s, t), is called the **joint distribution function** of X and Y.

EXAMPLE

If the joint probability density of X and Y is given by

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

Solution

If either $x < 0$ or $y < 0$, it follows immediately that $F(x, y) = 0$.

For $0 < x < 1$ and $0 < y < 1$

$$F(x, y) = \int_0^y \int_0^x (s + t)\, ds\, dt = \frac{1}{2}xy(x + y)$$

for $x > 1$ and $0 < y < 1$

$$F(x, y) = \int_0^y \int_0^1 (s + t)\, ds\, dt = \frac{1}{2}y(y + 1)$$

for $x > 1$ and $0 < y < 1$

$$F(x, y) = \int_0^y \int_0^1 (s + t)\, ds\, dt = \frac{1}{2}y(y + 1)$$

for $0 < x < 1$ and $y > 1$

$$F(x, y) = \int_0^1 \int_0^x (s + t)\, ds\, dt = \frac{1}{2}x(x + 1)$$

and for $x > 1$ and $y > 1$

$$F(x, y) = \int_0^1 \int_0^1 (s + t)\, ds\, dt = 1$$

$$F(x, y) = \begin{cases} 0 & \text{for } x \leqq 0 \text{ or } y \leqq 0 \\ \frac{1}{2}xy(x + y) & \text{for } 0 < x < 1, 0 < y < 1 \\ \frac{1}{2}y(y + 1) & \text{for } x \geqq 1, 0 < y < 1 \\ \frac{1}{2}x(x + 1) & \text{for } 0 < x < 1, y \geqq 1 \\ 1 & \text{for } x \geqq 1, y \geqq 1 \end{cases}$$

MARGINAL DISTRIBUTIONS

Definition If X and Y are discrete random variables and $f(x, y)$ is the value of their joint probability distribution at (x, y), the function given by

$$g(x) = \sum_{y} f(x, y)$$

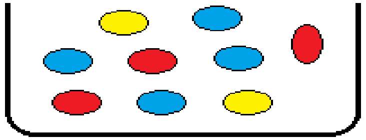
for each x within the range of X is called the **marginal distribution** of X.

Correspondingly, the function given by

$$h(y) = \sum_{x} f(x, y)$$

for each y within the range of Y is called the **marginal distribution** of Y.

EXAMPLE



Two balls are selected at random from a box containing 3 Red 2 Yellow and 4 Blue balls. If X and Y are respectively, the numbers Red and Yellow balls, find the probabilities associated with all possible pairs of values of X and Y.

Solution

The possible pairs are

(0, 0), (0, 1), (1, 0),
(1, 1), (0, 2), and (2, 0).

we obtain the values shown in the following table:

			x	
		0	1	2
	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$
y	1	$\frac{2}{9}$	$\frac{1}{6}$	
	2	$\frac{1}{36}$		

The column totals are the probabilities that X will take on the values 0, 1, and 2. In other words, they are the values

$$g(x) = \sum_{y=0}^{2} f(x, y) \qquad \text{for } x = 0, 1, 2$$

of the probability distribution of X.

together with the marginal totals,

$$\begin{array}{cc|ccc|}
 & & & x & \\
 & & 0 & 1 & 2 \\
\hline
 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{12} \\
y & 1 & \frac{2}{9} & \frac{1}{6} & \\
 & 2 & \frac{1}{36} & & \\
\hline
 & & \frac{5}{12} & \frac{1}{2} & \frac{1}{12}
\end{array}$$

Similarly, the row totals are the values

$$h(y) = \sum_{x=0}^{2} f(x, y) \qquad \text{for } y = 0, 1, 2$$

of the probability distribution of Y.

together with the marginal totals,

		$x=0$	$x=1$	$x=2$	
y	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{7}{12}$
	1	$\frac{2}{9}$	$\frac{1}{6}$		$\frac{7}{18}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$	

When X and Y are continuous random variables, the probability distributions are replaced by probability densities, the summations are replaced by integrals, and we get

DEFINITION If X and Y are continuous random variables and $f(x, y)$ is the value of their joint probability density at (x, y), the function given by

$$g(x) = \int_{-\infty}^{\infty} f(x, y)\,dy \qquad \text{for } -\infty < x < \infty$$

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

Correspondingly, the function given by

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\,dx \qquad \text{for } -\infty < y < \infty$$

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

EXAMPLE

Given the joint probability density

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

find the marginal densities of X and Y.

Solution Performing the necessary integrations, we get

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

for $0 < x < 1$ and $g(x) = 0$ elsewhere.

Likewise,

$$h(y) = \int_{-\infty}^{\infty} f(x, y)\, dx = \int_0^1 \frac{2}{3}(x + 2y)\, dx = \frac{1}{3}(1 + 4y)$$

for $0 < y < 1$ and $h(y) = 0$ elsewhere.

When we are dealing with more than two random variables, we can speak not only of the marginal distributions of the individual random variables, but also of the **joint marginal distributions** of several of the random variables.

If the joint probability distribution of the discrete random variables $X_1, X_2, \ldots$, and X_n has the values $f(x_1, x_2, \ldots, x_n)$, the marginal distribution of X_1 alone is given by

$$g(x_1) = \sum_{x_2} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n)$$

for all values within the range of X_1,

the joint marginal distribution of X_1, X_2, and X_3 is given by

$$m(x_1, x_2, x_3) = \sum_{x_4} \cdots \sum_{x_n} f(x_1, x_2, \ldots, x_n)$$

for all values within the range of X_1, X_2, and X_3, and other marginal distributions can be defined in the same way.

For the continuous case, probability distributions
are replaced by probability densities, summations are replaced by integrals, and if the joint probability density of the continuous random variables $X_1, X_2, \ldots,$ and X_n has the values $f(x_1, x_2, \ldots, x_n)$, the marginal density of X_2 alone is given by

$$h(x_2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_1\, dx_3 \cdots dx_n$$

for $-\infty < x_2 < \infty$,

the joint marginal density of X_1 and X_n is given by

$$\varphi(x_1, x_n) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n)\, dx_2\, dx_3 \cdots dx_{n-1}$$

for $-\infty < x_1 < \infty$ and $-\infty < x_n < \infty$, and so forth.

EXAMPLE

If the **trivariate** probability density of X_1, X_2, and X_3 is given by

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & \text{for } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find $P[(X_1, X_2, X_3) \in A]$, where A is the region

$$\left\{(x_1, x_2, x_3) | 0 < x_1 < \frac{1}{2}, \frac{1}{2} < x_2 < 1, x_3 < 1\right\}$$

Solution

$$P[(X_1, X_2, X_3) \in A] = P\left(0 < X_1 < \frac{1}{2}, \frac{1}{2} < X_2 < 1, X_3 < 1\right)$$

$$= \int_0^1 \int_{\frac{1}{2}}^1 \int_0^{\frac{1}{2}} (x_1 + x_2)e^{-x_3}\, dx_1\, dx_2\, dx_3$$

$$= \int_0^1 \int_{\frac{1}{2}}^1 \left(\frac{1}{8} + \frac{x_2}{2}\right) e^{-x_3}\, dx_2\, dx_3$$

$$= \int_0^1 \frac{1}{4} e^{-x_3}\, dx_3$$

$$= \frac{1}{4}(1 - e^{-1}) = 0.158$$

EXAMPLE

Considering again the trivariate probability density ,

$$f(x_1, x_2, x_3) = \begin{cases} (x_1 + x_2)e^{-x_3} & \text{for } 0 < x_1 < 1, 0 < x_2 < 1, x_3 > 0 \\ 0 & \text{elsewhere} \end{cases}$$

find the joint marginal density of X_1 and X_3 and the marginal density of X_1 alone.

Solution Performing the necessary integration, we find that the joint marginal density of X_1 and X_3 is given by

$$m(x_1, x_3) = \int_0^1 (x_1 + x_2)e^{-x_3}\, dx_2 = \left(x_1 + \frac{1}{2}\right) e^{-x_3}$$

for $0 < x_1 < 1$ and $x_3 > 0$ and $m(x_1, x_3) = 0$ elsewhere.

Using this result, we find
that the marginal density of X_1 alone is given by

$$g(x_1) = \int_0^\infty \int_0^1 f(x_1, x_2, x_3)\, dx_2\, dx_3 = \int_0^\infty m(x_1, x_3)\, dx_3$$

$$= \int_0^\infty \left(x_1 + \frac{1}{2}\right) e^{-x_3}\, dx_3 = x_1 + \frac{1}{2}$$

CONDITIONAL DISTRIBUTIONS

we defined the conditional probability of event A, given event B, as

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

provided $P(B) \neq 0$.

Suppose now that A and B are the events $X = x$ and $Y = y$ so that we can write

$$\begin{aligned} P(X = x|Y = y) &= \frac{P(X = x, Y = y)}{P(Y = y)} \\ &= \frac{f(x, y)}{h(y)} \end{aligned}$$

provided $P(Y = y) = h(y) \neq 0$,

where $f(x, y)$ is the value of the joint probability distribution of X and Y at (x, y), and $h(y)$ is the value of the marginal distribution of Y at y. Denoting the conditional probability by $f(x|y)$ to indicate that x is a variable and y is fixed, let us now make the following definition.

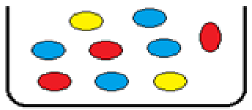
DEFINITION If $f(x, y)$ is the value of the joint probability distribution of the discrete random variables X and Y at (x, y), and $h(y)$ is the value of the marginal distribution of Y at y, the function given by

$$f(x|y) = \frac{f(x, y)}{h(y)} \qquad h(y) \neq 0$$

for each x within the range of X, is called the **conditional distribution** of X given $Y = y$.

EXAMPLE

With reference to Example



find the conditional distribution of X given $Y = 1$.

		x = 0	x = 1	x = 2	
y	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{12}$	$\frac{7}{12}$
	1	$\frac{2}{9}$	$\frac{1}{6}$		$\frac{7}{18}$
	2	$\frac{1}{36}$			$\frac{1}{36}$
		$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{12}$	

$$f(x|y) = \frac{f(x, y)}{h(y)} \qquad h(y) \neq 0$$

Solution Substituting the appropriate values from the table , we get

$$f(0|1) = \frac{f(0,1)}{h(1)}$$

$$f(0,1) = \frac{2}{9},$$

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

$$f(0|1) = \frac{\frac{2}{9}}{\frac{7}{18}} = \frac{4}{7}$$

$$f(1|1) = \frac{\frac{1}{6}}{\frac{7}{18}} = \frac{3}{7} \qquad f(2|1) = \frac{0}{\frac{7}{18}} = 0$$

EXAMPLE

With reference to the joint probability density

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

find the conditional
density of X given $Y = y$, and use it to evaluate $P(X \leq \frac{1}{2} | Y = \frac{1}{2})$.

Solution Using the results obtained , we have

$$f(x|y) = \frac{f(x, y)}{h(y)} = \frac{\frac{2}{3}(x+2y)}{\frac{1}{3}(1+4y)} = \frac{2x+4y}{1+4y}$$

for $0 < x < 1$ and $f(x|y) = 0$ elsewhere.
Now,

$$f\left(x \middle| \frac{1}{2}\right) = \frac{2x + 4 \cdot \frac{1}{2}}{1 + 4 \cdot \frac{1}{2}} = \frac{2x+2}{3}$$

and we can write

$$P\left(X \leq \frac{1}{2} \middle| Y = \frac{1}{2}\right) = \int_0^{\frac{1}{2}} \frac{2x+2}{3}\, dx = \frac{5}{12}$$