Joint Probability Distribution

Joint Probability Distribution

- Previously we have seen one-dimensional sample space in which there was only one random variable.
- We can assume an experiment having simultaneous outcomes of more than one random variable.
- If we observe an experiment having two random variables, we will find 2 dimensional sample space (or a plane).
- For example
  - we might measure the amount of precipitate P and volume V of gas released from a controlled chemical experiment.
  - we might be interested in the hardness H and tensile strength T of cold-drawn copper

Joint Probability Distribution

- If X and Y are two discrete random variables, the probability distribution for their simultaneous occurrence can be represented by a function f(x, y)
- It contains pair of values (x, y) within the range of the random variables X and Y
- This function will be called joint probability distribution of X and Y .

Joint Probability Distribution

- **Definition 3.8:**

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. $\sum_x \sum_y f(x, y) = 1$,
3. $P(X = x, Y = y) = f(x, y)$.

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

Compare with single RV PMF

1. $f(x) \geq 0$.
2. $\sum_x f(x) = 1$.
3. $P(X = x) = f(x)$.

Joint Probability Distribution

- **Definition 3.9:**

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.

Compare with single RV pdf

1. $f(x) \geq 0$; $\quad$ for all $x \in R$.
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1$.
3. $P(a < X < b) = \int_a^b f(x)\, dx$.

Joint Probability Distribution

- **Example 3.14**
  - 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) Joint probability function $f(x, y)$

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

Joint Probability Distribution

- **Example 3.14-Solution**

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

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

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

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

Table 3.1: Joint Probability Distribution for Example 3.14

	$f(x,y)$	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

Joint Probability Distribution

- **Example 3.15:**
  - 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:

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

(a) Verify condition 2 of Definition 3.9.

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

Joint Probability Distribution

- **Example 3.15 - Solution:**

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

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

$$= \int_0^1 \left(\frac{2x^2}{5} + \frac{6xy}{5}\right)\Bigg|_{x=0}^{x=1} dy$$

$$= \int_0^1 \left(\frac{2}{5} + \frac{6y}{5}\right) dy = \left(\frac{2y}{5} + \frac{3y^2}{5}\right)\Bigg|_0^1 = \frac{2}{5} + \frac{3}{5} = 1.$$

Joint Probability Distribution

- **Example 3.15 - Solution:**

(b) To calculate the probability, we use

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

Marginal Distributions

- Given the joint probability distribution f(x, y) of the discrete random variables X and Y.
- Probability distribution **g(x)** of **X** alone is obtained by summing **f(x, y)** over the values of **Y**.

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

- Similarly, the probability distribution **h(y)** of **Y** alone is obtained by summing **f(x, y)** over the values of **X**.
- We define g(x) and h(y) to be the marginal distributions of X and Y , respectively

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

- If working in continuous space, we will do integration.
- The fact that the marginal distributions g(x) and h(y) are indeed the probability distributions. Their value will be 1.

Marginal Distributions

- **Definition 3.10:**

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.

Marginal Distributions

- **Example 3.16:**
  - Show that the column and row totals of given table give the marginal distribution of X alone and of Y alone.

	$f(x, y)$	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

Marginal Distributions

- **Example 3.16 - Solution:**

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

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

which are just the column totals of Table 3.1. In a similar manner we could show that the values of $h(y)$ are given by the row totals. In tabular form, these marginal distributions may be written as follows:

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

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

Marginal Distributions

- **Example 3.17:**
  - Find g(x) and h(y) for the joint density function of

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

- **Solution:**

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

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

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

Conditional Distribution

- **Definition 3.11:**

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

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

Compare with Conditional Probability:

$$\boxed{P(B|A) = \frac{P(A \cap B)}{P(A)}, \text{ provided } P(A) > 0,}$$

Conditional Distribution

- **Example 3.18:**
  - Referring to Example 3.14, find the conditional distribution of X, given that Y = 1,
  - and use it to determine P(X = 0 | Y = 1).
- **Ref. Example:**

$f(x,y)$		$x=0$	$x=1$	$x=2$	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

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

Conditional Distribution

- **Example 3.18-Solution:**

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

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

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

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

Conditional Distribution

- **Example 3.18-Solution:**

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

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

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

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,

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

Conditional Distribution

- **Example 3.19:**

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.

Conditional Distribution

- **Example 3.19-Solution:**

(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,

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

Conditional Distribution

- **Example 3.20:**

Given the joint density function

$$f(x, y) = \begin{cases} \frac{x(1+3y^2)}{4}, & 0 < x < 2,\ 0 < y < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

find $g(x)$, $h(y)$, $f(x|y)$, and evaluate $P(\frac{1}{4} < X < \frac{1}{2} \mid Y = \frac{1}{3})$.

Conditional Distribution

- **Example 3.20-Solution:**

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

$$h(y) = \int_{-\infty}^{\infty} f(x,y)\, dx = \int_0^2 \frac{x(1+3y^2)}{4} dx = \left(\frac{x^2}{8} + \frac{3x^2y^2}{8} \right) \Bigg|_{x=0}^{x=2} = \frac{1+3y^2}{2}.$$

Therefore, using the conditional density definition, for $0 < x < 2$,

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

and

$$P\left(\frac{1}{4} < X < \frac{1}{2} \;\middle|\; Y = \frac{1}{3} \right) = \int_{1/4}^{1/2} \frac{x}{2} dx = \frac{3}{64}.$$

Statistically Independent

- **Definition 3.12**
  - 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

  - Compare with

$$P(A \mid B) = P(A) \quad \text{and} \quad P(B \mid A) = P(B).$$

Statistically Independent

- **Example 3.21:**
  - Show that the random variables of Example 3.14 are not statistically independent.

	$f(x, y)$	x = 0	x = 1	x = 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

Statistically Independent

- **Example 3.21-Proof:**

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

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

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

$$\boxed{f(0,1) \neq g(0)h(1),}$$

Therefore X and Y are not statistically independent.