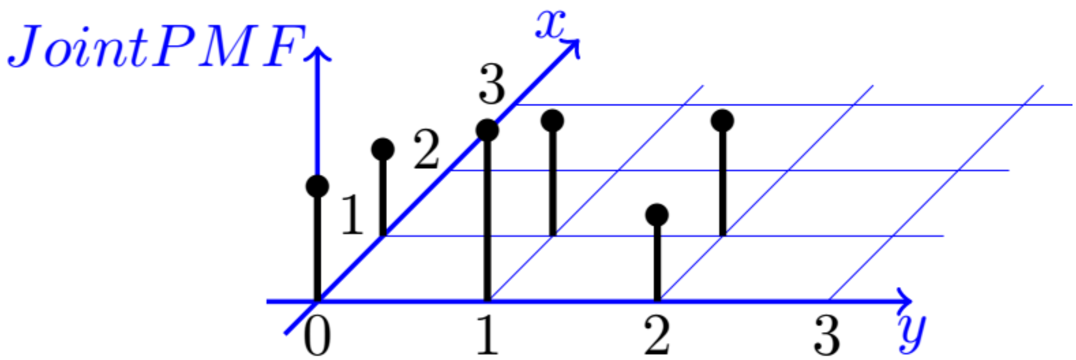
Lecture 12

Joint Probability Distributions **Two discrete R.Vs.**

In the discrete case, $f(x, y) = P(X = x \text{ and } Y = y)$; that is, the values $f(x, y)$ give the probability that outcomes x and y occur at the same time.



Definition: The function $f(x, y)$ is a **joint probability distribution** or **joint 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, $P[(X, Y) \in A] = \sum\sum_A f(x,y)$.

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

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

for the discrete case.

Statistical Independence

Definition: 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

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

for all (x, y) within their range.

Definition: Let $X_1, X_2, \ldots, X_n$ be n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \ldots, x_n)$ and marginal distribution $f_1(x_1), f_2(x_2), \ldots, f_n(x_n)$, respectively. The random variables $X_1, X_2, \ldots, X_n$ are said to be **mutually statistically independent** if and only if

$$\boxed{f(x_1, x_2, \ldots, x_n) = f_1(x_1)f_2(x_2) \cdots f_n(x_n)}$$

for all $(x_1, x_2, \ldots, x_n)$ within their range.

Example: The joint probability density function of two discrete random variables X and Y is given by:

$$f(x, y) = \begin{cases} c(x+y), & x = 0,1,2 \quad y=0,1,2,3, \\ 0, & \text{otherwise.} \end{cases}$$

Find

(a) c.

(b) $P(X \geq 1, Y \leq 2)$.

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

(d) Check independence of the two random variables X and Y.

$f(x, y)$		x			Row Totals
		0	1	2	
y	0	0	c	2c	3c (y = 0)
	1	c	2c	3c	6c (y = 1)
	2	2c	3c	4c	9c (y = 2)
	3	3c	4c	5c	12c (y = 3)
Column Totals		6c (x= 0)	10c (x= 1)	14c (x= 2)	$\sum_x \sum_y f(x,y) = 30c$

Solution: (a) $\sum_x \sum_y f(x,y) = 1 \Rightarrow \sum_{x=0}^{2} \sum_{y=0}^{3} c(x+y) = 1 \Rightarrow \boxed{c = \frac{1}{30}}$

(b) $$P(X \geq 1, Y \leq 2) = \frac{15}{30}$$

(c) $$g(x) = \sum_y f(x,y) = \sum_{y=0}^{3} c(x+y) = \frac{1}{15}(2x+3), \quad x = 0,1,2$$

$$h(y) = \sum_x f(x,y) = \sum_{x=0}^{2} c(x+y) = \frac{1}{10}(y+1), \quad y = 0,1,2,3$$

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

$\boxed{\mathbf{0}}$ $\quad \boxed{\frac{6}{30}}$ $\quad \boxed{\frac{3}{30}}$ $\quad \Rightarrow X$ and Y are not independent.

Also, independence of the two random variables X and Y can be checked as follows:

$f(x,y) \neq g(x)\, h(y)$

$\boxed{\frac{1}{30}(x+y)}$ $\quad \boxed{\frac{1}{15}(2x+3)}$ $\quad \boxed{\frac{1}{10}(y+1)}$ $\quad \Rightarrow X$ and Y are not independent.

Definition: Let X and Y be random variables with joint probability distribution $f(x, y)$. The mean, or expected value, of the random variable $g(X, Y)$ is

$$\mu_{g(X,Y)} = E[g(X, Y)] = \sum_x \sum_y g(x,y) f(x,y)$$ if X and Y are discrete.

Note that if $g(X, Y) = X$, we have

$$E(X) = \sum_x \sum_y x f(x,y) = \sum_x x g(x) \quad \text{(discrete case)},$$ where $g(x)$ is the marginal distribution of X.

Similarly, we define

$$E(Y) = \sum_x \sum_y y f(x,y) = \sum_x y h(y) \quad \text{(discrete case)},$$ where $h(y)$ is the marginal distribution Y.

Definition: Let X and Y be random variables with joint probability distribution $f(x, y)$. The covariance of X and Y is

$$\sigma_{XY} = E[(X-\mu_X)(Y-\mu_Y)] = \sum_x \sum_y (x-\mu_X)(y-\mu_Y) f(x,y)$$

if X and Y are discrete.

Theorem: The covariance of two random variables X and Y with means μ_X and μ_Y, respectively, is given by

$$\text{Cov}(X, Y) = \sigma_{XY} = E(XY) - \mu_X \mu_Y = E(XY) - E(X)E(Y).$$

- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- $\text{Cov}(X, c) = 0$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(aX, bY) = ab\,\text{Cov}(X, Y)$
- $\text{Cov}(X \pm a, Y \pm b) = \text{Cov}(X, Y)$

$$\text{Var}(aX+bY+c) = \sigma^2_{aX+bY+c} = a^2\sigma^2_X + b^2\sigma^2_Y + 2ab\sigma_{XY}.$$

- $\sigma^2_{aX+c} = a^2\sigma^2_X = a^2\sigma^2$
- $\sigma^2_{X+c} = \sigma^2_X = \sigma^2$
- $\sigma^2_{aX} = a^2\sigma^2_X = a^2\sigma^2$

If X and Y are independent random variables, then

$$\sigma^2_{aX\pm bY} = a^2\sigma^2_X + b^2\sigma^2_Y.$$

Theorem: Let X and Y be two independent random variables. Then ***E(XY) = E(X)E(Y).***

Corollary: Let X and Y be two independent random variables. Then $\sigma_{XY} = 0.$

Definition: Let X and Y be random variables with covariance σ_{XY} and standard deviations σ_X and σ_Y, respectively. The correlation coefficient of X and Y is

$$\boxed{\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}. \qquad -1 \le \rho_{XY} \le 1}$$

$$\boxed{\sigma_X = \sqrt{E(X^2) - E^2(X)}} \qquad \boxed{\sigma_{XY} = E(XY) - E(X)E(Y)} \qquad \boxed{\sigma_y = \sqrt{E(Y^2) - E^2(Y)}}$$

- $\rho(X,Y) = \rho(Y,X)$
- $\rho(X,X) = 1$
- $\rho(aX,bY) = \rho(X,Y)$
- $\rho(X \pm a, Y \pm b) = \rho(X,Y)$

Example 12.1: 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 $\{(x, y)|x + y \le 1\}$.
(c) $g(x)$ and $h(y)$ then check independence of the two random variables X and Y.
(d) the correlation coefficient between X and Y.

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

(a) Now, $f(0,1)$, for example, represents the probability that a red and a green pens are selected:

$$f(0, 1) = = \frac{2}{8} \times \frac{3}{7} + \frac{3}{8} \times \frac{2}{7} = \frac{6}{28} = \frac{3}{14}.$$

Similar calculations yield the probabilities for the other cases.

$f(x, y)$		$x=0$	$x=1$	$x=2$	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		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	$\sum_x \sum_y f(x,y) = 1$

(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

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

which are just the column totals of the table. 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
$h(y)$	$\frac{15}{28}$	$\frac{3}{7}$	$\frac{1}{28}$

From the table, 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.

(d) $$E(X) = \mu_X = \sum_{x=0}^{2} x g(x) = (0)(\frac{5}{14}) + (1)(\frac{15}{28}) + (2)(\frac{3}{28}) = \frac{3}{4},$$

and

$$E(Y) = \mu_Y = \sum_{y=0}^{2} y h(y) = (0)(\frac{15}{28}) + (1)(\frac{3}{7}) + (2)(\frac{1}{28}) = \frac{1}{2}.$$

$$E(X^2) = (0^2)\left(\frac{5}{14}\right) + (1^2)\left(\frac{15}{28}\right) + (2^2)\left(\frac{3}{28}\right) = \frac{27}{28}$$

and

$$E(Y^2) = (0^2)\left(\frac{5}{28}\right) + (1^2)\left(\frac{3}{7}\right) + (2^2)\left(\frac{1}{28}\right) = \frac{4}{7},$$

$$\boxed{\Rightarrow \sigma_X^2 = \frac{27}{28} - \left(\frac{3}{4}\right)^2 = \frac{45}{112} \text{ and } \sigma_Y^2 = \frac{4}{7} - \left(\frac{1}{2}\right)^2 = \frac{9}{28}.}$$

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

$$E(XY) = \sum_{x=0}^{2}\sum_{y=0}^{2} xy\,f(x,y)$$
$$= (0)(0)f(0,0) + (0)(1)f(0,1) + (0)(2)f(0,2) + (1)(0)f(1,0) + \underline{(1)(1)f(1,1)} + (1)(2)f(1,2) + (2)(0)f(2,0) + (2)(1)f(2,1) + (2)(2)f(2,2)$$
$$= f(1,1) = \boxed{\frac{3}{14}}.$$

Therefore,

$$\sigma_{XY} = E(XY) - \mu_X\mu_Y = \frac{3}{14} - \left(\frac{3}{4}\right)\left(\frac{1}{2}\right) = -\frac{9}{56}.$$

Therefore, the correlation coefficient between X and Y is

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{-\frac{9}{56}}{\sqrt{\left(\frac{45}{112}\right)\left(\frac{9}{28}\right)}} = -\frac{1}{\sqrt{5}}.$$

Example 12.14: If X and Y are random variables with variances $\sigma_X^2 = 2$ and $\sigma_Y^2 = 4$ and covariance $\sigma_{XY} = -2$, find the variance of the random variable $Z = 3X - 4Y + 8$.

***Solution*:**

$$\sigma_Z^2 = \sigma_{3X-4Y+8}^2 = \sigma_{3X-4Y}^2$$
$$= 9\sigma_X^2 + 16\sigma_Y^2 - 24\sigma_{XY}$$
$$= 9(2) + 16(4) - 24(-2) = 130.$$

Example 12.15: Let X and Y denote the amounts of two different types of impurities in a batch of a certain chemical product. Suppose that X and Y are independent random variables with variances $\sigma_X^2 = 2$ and $\sigma_Y^2 = 3$. Find the variance of the random variable $Z = 3X - 2Y + 5$.

***Solution*:**

$$\sigma_Z^2 = \sigma_{3X-2Y+5}^2 = \sigma_{3X-2Y}^2$$
$$= 9\sigma_X^2 + 4\sigma_Y^2$$
$$= 9(2) + 4(3) = 30.$$