

The Covariance and Correlation of Random Variables

The Variance:

The variance of a random variable X is

$$\sigma^2 = E(X^2) - \mu^2.$$

The Covariance:

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

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y.$$

Example 4.9:

Let the random variable X represent the number of defective parts for a machine when 3 parts are sampled from a production line and tested. The following is the probability distribution of X .

x	0	1	2	3
$f(x)$	0.51	0.38	0.10	0.01

Using Theorem 4.2, calculate σ^2 .

Solution:

First, we compute

$$\mu = (0)(0.51) + (1)(0.38) + (2)(0.10) + (3)(0.01) = 0.61.$$

Now,

$$E(X^2) = (0)(0.51) + (1)(0.38) + (4)(0.10) + (9)(0.01) = 0.87.$$

Therefore,

$$\sigma^2 = 0.87 - (0.61)^2 = 0.4979.$$

Example 4.10:

The weekly demand for a drinking-water product, in thousands of liters, from a local chain of efficiency stores is a continuous random variable X having the probability density

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

Find the mean and variance of X .

Solution:

Calculating $E(X)$ and $E(X^2)$, we have

$$\mu = E(X) = 2 \int_1^2 x(x-1) dx = \frac{5}{3}$$

and

$$E(X^2) = 2 \int_1^2 x^2(x-1) dx = \frac{17}{6}.$$

Therefore,

$$\sigma^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

Example 4.13:

Example 3.14 on page 95 describes a situation involving the number of blue refills X and the number of red refills Y . Two refills for a ballpoint pen are selected at random from a certain box, and the following is the joint probability distribution:

$f(x, y)$		x			$h(y)$
		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	$\frac{1}{28}$
$g(x)$		$\frac{5}{14}$	$\frac{15}{28}$	$\frac{3}{28}$	1

Find the covariance of X and Y .

Solution:

From Example 4.6, we see that $E(XY) = 3/14$. Now

$$\mu_x = \sum_{x=0}^2 xg(x) = (0) \left(\frac{5}{14}\right) + (1) \left(\frac{15}{28}\right) + (2) \left(\frac{3}{28}\right) = \frac{3}{4},$$

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

Therefore,

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

Example 4.14:

The fraction X of male runners and the fraction Y of female runners who compete in marathon races are described by the joint density function

$$f(x, y) = \begin{cases} 8xy, & 0 \leq y \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the covariance of X and Y .

Find the correlation coefficient of X and Y

Solution:

We first compute the marginal density functions. They are

$$g(x) = \begin{cases} 4x^3, & 0 \leq x \leq 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and

$$h(y) = \begin{cases} 4y(1 - y^2), & 0 \leq y \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

From these marginal density functions, we compute

$$\mu_X = E(X) = \int_0^1 4x^4 dx = \frac{4}{5} \text{ and } \mu_Y = \int_0^1 4y^2(1 - y^2) dy = \frac{8}{15}.$$

From the joint density function given above, we have

$$E(XY) = \int_0^1 \int_y^1 8x^2 y^2 dx dy = \frac{4}{9}.$$

Then

$$\sigma_{XY} = E(XY) - \mu_X \mu_Y = \frac{4}{9} - \left(\frac{4}{5} \right) \left(\frac{8}{15} \right) = \frac{4}{225}.$$

Because

$$E(X^2) = \int_0^1 4x^5 dx = \frac{2}{3} \text{ and } E(Y^2) = \int_0^1 4y^3(1-y^2) dy = 1 - \frac{2}{3} = \frac{1}{3},$$

we conclude that

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75} \text{ and } \sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}.$$

Hence,

$$\rho_{XY} = \frac{4/225}{\sqrt{(2/75)(11/225)}} = \frac{4}{\sqrt{66}}.$$

The Correlation coefficient:

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

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Example 4.15:

Find the correlation coefficient between X and Y in Example 4.13.

Solution:

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

and

$$E(Y^2) = (0^2) \left(\frac{15}{28}\right) + (1^2) \left(\frac{3}{7}\right) + (2^2) \left(\frac{1}{28}\right) = \frac{4}{7},$$

we obtain

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

Therefore, the correlation coefficient between X and Y is

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

Example 4.16:

Find the correlation coefficient of X and Y in Example 4.14.

Solution:

$$E(X^2) = \int_0^1 4x^5 \, dx = \frac{2}{3} \text{ and } E(Y^2) = \int_0^1 4y^3(1-y^2) \, dy = 1 - \frac{2}{3} = \frac{1}{3},$$

we conclude that

$$\sigma_X^2 = \frac{2}{3} - \left(\frac{4}{5}\right)^2 = \frac{2}{75} \text{ and } \sigma_Y^2 = \frac{1}{3} - \left(\frac{8}{15}\right)^2 = \frac{11}{225}.$$

Hence,

$$\rho_{XY} = \frac{4/225}{\sqrt{(2/75)(11/225)}} = \frac{4}{\sqrt{66}}.$$



Practice Problems from Book

4.34 Let X be a random variable with the following probability distribution:

x	-2	3	5
$f(x)$	0.3	0.2	0.5

Find the standard deviation of X .

4.35 The random variable X , representing the number of errors per 100 lines of software code, has the following probability distribution:

x	2	3	4	5	6
$f(x)$	0.01	0.25	0.4	0.3	0.04

Using Theorem 4.2 on page 121, find the variance of X .

4.36 Suppose that the probabilities are 0.4, 0.3, 0.2, and 0.1, respectively, that 0, 1, 2, or 3 power failures will strike a certain subdivision in any given year. Find the mean and variance of the random variable X representing the number of power failures striking this subdivision.

4.49 Consider the situation in Exercise 4.32 on page 119. The distribution of the number of imperfections per 10 meters of synthetic failure is given by

x	0	1	2	3	4
$f(x)$	0.41	0.37	0.16	0.05	0.01

Find the variance and standard deviation of the number of imperfections.

4.50 For a laboratory assignment, if the equipment is working, the density function of the observed outcome X is

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

Find the variance and standard deviation of X .

4.39 The total number of hours, in units of 100 hours, that a family runs a vacuum cleaner over a period of one year is a random variable X having the density function given in Exercise 4.13 on page 117. Find the variance of X .

4.51 For the random variables X and Y in Exercise 3.39 on page 105, determine the correlation coefficient between X and Y .

4.52 Random variables X and Y follow a joint distribution

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

Determine the correlation coefficient between X and Y .