

## 12.2: Covariance and the Correlation Coefficient

The mean value  $\mu_X = E[X]$  and the variance  $\sigma_X^2 = E[(X - \mu_X)^2]$  give important information about the distribution for real random variable  $X$ . Can the expectation of an appropriate function of  $(X, Y)$  give useful information about the joint distribution? A clue to one possibility is given in the expression

$$\text{Var}[X \pm Y] = \text{Var}[X] + \text{Var}[Y] \pm 2(E[XY] - E[X]E[Y])$$

The expression  $E[XY] - E[X]E[Y]$  vanishes if the pair is independent (and in some other cases). We note also that for  $\mu_X = E[X]$  and  $\mu_Y = E[Y]$

$$E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X\mu_Y$$

To see this, expand the expression  $(X - \mu_X)(Y - \mu_Y)$  and use linearity to get

$$E[(X - \mu_X)(Y - \mu_Y)] = E[XY - \mu_Y X - \mu_X Y + \mu_X \mu_Y] = E[XY] - \mu_Y E[X] - \mu_X E[Y] + \mu_X \mu_Y$$

which reduces directly to the desired expression. Now for given  $\omega$ ,  $X(\omega) - \mu_X$  is the variation of  $X$  from its mean and  $Y(\omega) - \mu_Y$  is the variation of  $Y$  from its mean. For this reason, the following terminology is used.

### Definition: Covariance

The quantity  $\text{Cov}[X, Y] = E[(X - \mu_X)(Y - \mu_Y)]$  is called the *covariance* of  $X$  and  $Y$ .

If we let  $X' = X - \mu_X$  and  $Y' = Y - \mu_Y$  be the ventered random variables, then

$$\text{Cov}[X, Y] = E[X'Y']$$

Note that the variance of  $X$  is the covariance of  $X$  with itself.

If we standardize, with  $X^* = (X - \mu_X)/\sigma_X$  and  $Y^* = (Y - \mu_Y)/\sigma_Y$ , we have

### Definition: Correlation Coefficient

The *correlation coefficient*  $\rho = \rho[X, Y]$  is the quantity

$$\rho[X, Y] = E[X^*Y^*] = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

Thus  $\rho = \text{Cov}[X, Y]/\sigma_X \sigma_Y$ . We examine these concepts for information on the joint distribution. By Schwarz' inequality (E15), we have

$$\rho^2 = E^2[X^*Y^*] \leq E[(X^*)^2]E[(Y^*)^2] = 1 \text{ with equality iff } Y^* = cX^*$$

Now equality holds iff

$$1 = c^2 E^2[(X^*)^2] = c^2 \text{ which implies } c = \pm 1 \text{ and } \rho = \pm 1$$

We conclude  $-1 \leq \rho \leq 1$ , with  $\rho = \pm 1$  iff  $Y^* = \pm X^*$

### Relationship between $\rho$ and the joint distribution

- We consider first the distribution for the standardized pair  $(X^*, Y^*)$
- Since  $P(X^* \leq r, Y^* \leq s) = P\left(\frac{X - \mu_X}{\sigma_X} \leq r, \frac{Y - \mu_Y}{\sigma_Y} \leq s\right)$

$$= P(X \leq t = \sigma_X r + \mu_X, Y \leq u = \sigma_Y s + \mu_Y)$$

we obtain the results for the distribution for  $(X, Y)$  by the mapping

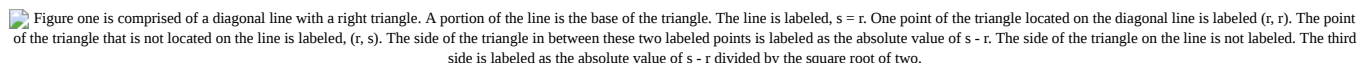
$$\begin{aligned} t &= \sigma_X r + \mu_X \\ u &= \sigma_Y s + \mu_Y \end{aligned}$$

**Joint distribution for the standardized variables**  $(X^*, Y^*)$ ,  $(r, s) = (X^*, Y^*)(\omega)$

$\rho = 1$  iff  $X^* = Y^*$  iff all probability mass is on the line  $s = r$ .

$\rho = -1$  iff  $X^* = -Y^*$  iff all probability mass is on the line  $s = -r$ .

If  $-1 < \rho < 1$ , then at least some of the mass must fail to be on these lines.

 Figure one is comprised of a diagonal line with a right triangle. A portion of the line is the base of the triangle. The line is labeled,  $s = r$ . One point of the triangle located on the diagonal line is labeled  $(r, r)$ . The point of the triangle that is not located on the line is labeled,  $(r, s)$ . The side of the triangle in between these two labeled points is labeled as the absolute value of  $s - r$ . The side of the triangle on the line is not labeled. The third side is labeled as the absolute value of  $s - r$  divided by the square root of two.

**Figure 12.2.1.** Distance from point  $(r, s)$  to the line  $s = r$ .

The  $\rho = \pm 1$  lines for the  $(X, Y)$  distribution are:

$$\frac{u - \mu_Y}{\sigma_Y} = \pm \frac{t - \mu_X}{\sigma_X} \text{ or } u = \pm \frac{\sigma_Y}{\sigma_X} (t - \mu_X) + \mu_Y$$

Consider  $Z = Y^* - X^*$ . Then  $E[\frac{1}{2}Z^2] = \frac{1}{2}E[(Y^* - X^*)^2]$ . Reference to Figure 12.2.1 shows this is the average of the square of the distances of the points  $(r, s) = (X^*, Y^*)(\omega)$  from the line  $s = r$  (i.e. the variance about the line  $s = r$ ). Similarly for  $W = Y^* + X^*$ .  $E[W^2/2]$  is the variance about  $s = -r$ . Now

$$\frac{1}{2}E[(Y^* \pm X^*)^2] = \frac{1}{2}\{E[(Y^*)^2] + E[(X^*)^2] \pm 2E[X^*Y^*]\} = 1 \pm \rho$$

Thus

$1 - \rho$  is the variance about  $s = r$  (the  $\rho = 1$  line)

$1 + \rho$  is the variance about  $s = -r$  (the  $\rho = -1$  line)

Now since

$$E[(Y^* - X^*)^2] = E[(Y^* + X^*)^2] \text{ iff } \rho = E[X^*Y^*] = 0$$

the condition  $\rho = 0$  is the condition for equality of the two variances.

**Transformation to the  $(X, Y)$  plane**

$$t = \sigma_X r + \mu_X \quad u = \sigma_Y s + \mu_Y \quad r = \frac{t - \mu_X}{\sigma_X} \quad s = \frac{u - \mu_Y}{\sigma_Y}$$

The  $\rho = 1$  line is:

$$\frac{u - \mu_Y}{\sigma_Y} = \frac{t - \mu_X}{\sigma_X} \text{ or } u = \frac{\sigma_Y}{\sigma_X} (t - \mu_X) + \mu_Y$$

The  $\rho = -1$  line is:

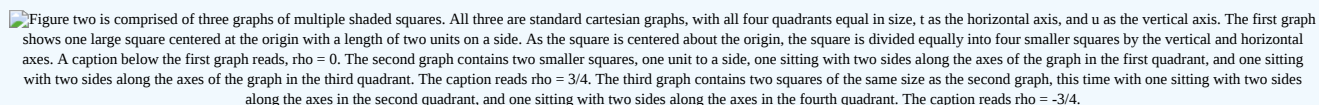
$$\frac{u - \mu_Y}{\sigma_Y} = -\frac{t - \mu_X}{\sigma_X} \text{ or } u = -\frac{\sigma_Y}{\sigma_X} (t - \mu_X) + \mu_Y$$

$1 - \rho$  is proportional to the variance about the  $\rho = 1$  line and  $1 + \rho$  is proportional to the variance about the  $\rho = -1$  line.  $\rho = 0$  iff the variances about both are the same.

### Example 12.2.1 Uncorrelated but not independent

Suppose the joint density for  $\{X, Y\}$  is constant on the unit circle about the origin. By the rectangle test, the pair cannot be independent. By symmetry, the  $\rho = 1$  line is  $u = t$  and the  $\rho = -1$  line is  $u = -t$ . By symmetry, also, the variance about each of these lines is the same. Thus  $\rho = 0$ , which is true iff  $\text{Cov}[X, Y] = 0$ . This fact can be verified by calculation, if desired.

### Example 12.2.2 Uniform marginal distributions

 Figure two is comprised of three graphs of multiple shaded squares. All three are standard cartesian graphs, with all four quadrants equal in size,  $t$  as the horizontal axis, and  $u$  as the vertical axis. The first graph shows one large square centered at the origin with a length of two units on a side. As the square is centered about the origin, the square is divided equally into four smaller squares by the vertical and horizontal axes. A caption below the first graph reads,  $\rho = 0$ . The second graph contains two smaller squares, one unit to a side, one sitting with two sides along the axes of the graph in the first quadrant, and one sitting with two sides along the axes of the graph in the third quadrant. The caption reads  $\rho = 3/4$ . The third graph contains two squares of the same size as the second graph, this time with one sitting with two sides along the axes in the second quadrant, and one sitting with two sides along the axes in the fourth quadrant. The caption reads  $\rho = -3/4$ .

**Figure 12.2.2.** Uniform marginals but different correlation coefficients.

Consider the three distributions in Figure 12.2.2. In case (a), the distribution is uniform over the square centered at the origin with vertices at  $(1,1)$ ,  $(-1,1)$ ,  $(-1,-1)$ ,  $(1,-1)$ . In case (b), the distribution is uniform over two squares, in the first and third quadrants with vertices  $(0,0)$ ,  $(1,0)$ ,  $(1,1)$ ,  $(0,1)$  and  $(0,0)$ ,

$(-1,0)$ ,  $(-1,-1)$ ,  $(0,-1)$ . In case (c) the two squares are in the second and fourth quadrants. The marginals are uniform on  $(-1,1)$  in each case, so that in each case

$$E[X] = E[Y] = 0 \text{ and } \text{Var}[X] = \text{Var}[Y] = 1/3$$

This means the  $\rho = 1$  line is  $u = t$  and the  $\rho = -1$  line is  $u = -t$ .

- By symmetry,  $E[XY] = 0$  (in fact the pair is independent) and  $\rho = 0$ .
- For every pair of possible values, the two signs must be the same, so  $E[XY] > 0$  which implies  $\rho > 0$ . The actual value may be calculated to give  $\rho = 3/4$ . Since  $1 - \rho < 1 + \rho$ , the variance about the  $\rho = 1$  line is less than that about the  $\rho = -1$  line. This is evident from the figure.
- $E[XY] < 0$  and  $\rho < 0$ . Since  $1 + \rho < 1 - \rho$ , the variance about the  $\rho = -1$  line is less than that about the  $\rho = 1$  line. Again, examination of the figure confirms this.

### Example 12.2.3 A pair of simple random variables

With the aid of m-functions and MATLAB we can easily calculate the covariance and the correlation coefficient. We use the joint distribution for Example 9 in "Variance." In that example calculations show

$$E[XY] - E[X]E[Y] = -0.1633 = \text{Cov}[X, Y], \sigma_X = 1.8170 \text{ and } \sigma_Y = 1.9122$$

so that  $\rho = -0.04699$ .

### Example 12.2.4 An absolutely continuous pair

The pair  $\{X, Y\}$  has joint density function  $f_{XY}(t, u) = \frac{6}{5}(t + 2u)$  on the triangular region bounded by  $t = 0$ ,  $u = t$ , and  $u = 1$ . By the usual integration techniques, we have

$$f_X(t) = \frac{6}{5}(1 + t - 2t^2), 0 \leq t \leq 1 \text{ and } f_Y(u) = 3u^2, 0 \leq u \leq 1$$

From this we obtain  $E[X] = 2/5$ ,  $\text{Var}[X] = 3/50$ ,  $E[Y] = 3/4$ , and  $\text{Var}[Y] = 3/80$ . To complete the picture we need

$$E[XY] = \frac{6}{5} \int_0^1 \int_t^1 (t^2 u + 2tu^2) du dt = 8/25$$

Then

$$\text{Cov}[X, Y] = E[XY] - E[X]E[Y] = 2/100 \text{ and } \rho = \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} = \frac{4}{30} \sqrt{10} \approx 0.4216$$

APPROXIMATION

```
tuappr
Enter matrix [a b] of X-range endpoints [0 1]
Enter matrix [c d] of Y-range endpoints [0 1]
Enter number of X approximation points 200
Enter number of Y approximation points 200
Enter expression for joint density (6/5)*(t + 2*u).*(u>=t)
Use array operations on X, Y, PX, PY, t, u, and P
EX = total(t.*P)
EX = 0.4012 % Theoretical = 0.4
EY = total(u.*P)
EY = 0.7496 % Theoretical = 0.75
VX = total(t.^2.*P) - EX^2
VX = 0.0603 % Theoretical = 0.06
VY = total(u.^2.*P) - EY^2
VY = 0.0376 % Theoretical = 0.0375
CV = total(t.*u.*P) - EX*EY
CV = 0.0201 % Theoretical = 0.02
```

```
rho = CV/sqrt(VX*VY)
```

```
rho = 0.4212
```

```
% Theoretical = 0.4216
```

### Coefficient of linear correlation

The parameter  $\rho$  is usually called the correlation coefficient. A more descriptive name would be *coefficient of linear correlation*. The following example shows that all probability mass may be on a curve, so that  $Y = g(X)$  (i.e., the value of  $Y$  is completely determined by the value of  $X$ ), yet  $\rho = 0$ .

#### Example 12.2.5 $Y = g(X)$ but $\rho = 0$

Suppose  $X \sim \text{uniform}(-1, 1)$ , so that  $f_X(t) = 1/2$ ,  $-1 < t < 1$  and  $E[X] = 0$ . Let  $Y = g(X) = \cos X$ . Then

$$\text{Cov}[X, Y] = E[XY] = \frac{1}{2} \int_{-1}^1 t \cos t \, dt = 0$$

Thus  $\rho = 0$ . Note that  $g$  could be any even function defined on  $(-1, 1)$ . In this case the integrand  $tg(t)$  is odd, so that the value of the integral is zero.

### Variance and covariance for linear combinations

We generalize the property (V4) on linear combinations. Consider the linear combinations

$$X = \sum_{i=1}^n a_i X_i \text{ and } Y = \sum_{j=1}^m b_j Y_j$$

We wish to determine  $\text{Cov}[X, Y]$  and  $\text{Var}[X]$ . It is convenient to work with the centered random variables  $X' = X - \mu_X$  and  $Y' = Y - \mu_Y$ . Since by linearity of expectation,

$$\mu_X = \sum_{i=1}^n a_i \mu_{X_i} \text{ and } \mu_Y = \sum_{j=1}^m b_j \mu_{Y_j}$$

we have

$$X' = \sum_{i=1}^n a_i X_i - \sum_{i=1}^n a_i \mu_{X_i} = \sum_{i=1}^n a_i (X_i - \mu_{X_i}) = \sum_{i=1}^n a_i X'_i$$

and similarly for  $Y'$ . By definition

$$\text{Cov}(X, Y) = E[X'Y'] = E[\sum_{i,j} a_i b_j X'_i Y'_j] = \sum_{i,j} a_i b_j E[X'_i Y'_j] = \sum_{i,j} a_i b_j \text{Cov}(X_i, Y_j)$$

In particular

$$\text{Var}(X) = \text{Cov}(X, X) = \sum_{i,j} a_i a_j \text{Cov}(X_i, X_j) = \sum_{i=1}^n a_i^2 \text{Cov}(X_i, X_i) + \sum_{i \neq j} a_i a_j \text{Cov}(X_i, X_j)$$

Using the fact that  $a_i a_j \text{Cov}(X_i, X_j) = a_j a_i \text{Cov}(X_j, X_i)$ , we have

$$\text{Var}[X] = \sum_{i=1}^n a_i^2 \text{Var}[X_i] + 2 \sum_{i < j} a_i a_j \text{Cov}(X_i, X_j)$$

Note that  $a_i^2$  does not depend upon the sign of  $a_i$ . If the  $X_i$  form an independent class, or are otherwise uncorrelated, the expression for variance reduces to

$$\text{Var}[X] = \sum_{i=1}^n a_i^2 \text{Var}[X_i]$$

This page titled [12.2: Covariance and the Correlation Coefficient](#) is shared under a [CC BY 3.0](#) license and was authored, remixed, and/or curated by [Paul Pfeiffer](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.