

4.5: Covariance and Correlation

Recall that by taking the expected value of various transformations of a random variable, we can measure many interesting characteristics of the distribution of the variable. In this section, we will study an expected value that measures a special type of relationship between two real-valued variables. This relationship is very important both in probability and statistics.

Basic Theory

Definitions

As usual, our starting point is a random experiment modeled by a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Unless otherwise noted, we assume that all expected values mentioned in this section exist. Suppose now that X and Y are real-valued random variables for the experiment (that is, defined on the probability space) with means $\mathbb{E}(X)$, $\mathbb{E}(Y)$ and variances $\text{var}(X)$, $\text{var}(Y)$, respectively.

The *covariance* of (X, Y) is defined by

$$\text{cov}(X, Y) = \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]) \quad (4.5.1)$$

and, assuming the variances are positive, the *correlation* of (X, Y) is defined by

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\text{sd}(X)\text{sd}(Y)} \quad (4.5.2)$$

1. If $\text{cov}(X, Y) > 0$ then X and Y are *positively correlated*.
2. If $\text{cov}(X, Y) < 0$ then X and Y are *negatively correlated*.
3. If $\text{cov}(X, Y) = 0$ then X and Y are *uncorrelated*.

Correlation is a scaled version of covariance; note that the two parameters always have the same sign (positive, negative, or 0). Note also that correlation is dimensionless, since the numerator and denominator have the same physical units, namely the product of the units of X and Y .

As these terms suggest, covariance and correlation measure a certain kind of dependence between the variables. One of our goals is a deeper understanding of this dependence. As a start, note that $(\mathbb{E}(X), \mathbb{E}(Y))$ is the center of the joint distribution of (X, Y) , and the vertical and horizontal lines through this point separate \mathbb{R}^2 into four quadrants. The function $(x, y) \mapsto [x - \mathbb{E}(X)][y - \mathbb{E}(Y)]$ is positive on the first and third quadrants and negative on the second and fourth.

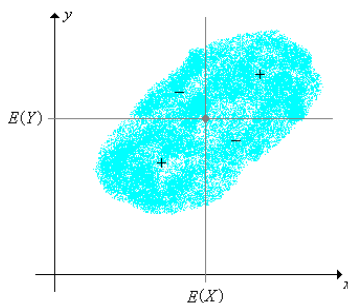


Figure 4.5.1: A joint distribution with $(\mathbb{E}(X), \mathbb{E}(Y))$ as the center of mass

Properties of Covariance

The following theorems give some basic properties of covariance. The main tool that we will need is the fact that expected value is a linear operation. Other important properties will be derived below, in the subsection on the [best linear predictor](#). As usual, be sure to try the proofs yourself before reading the ones in the text. Once again, we assume that the random variables are defined on the common sample space, are real-valued, and that the indicated expected values exist (as real numbers).

Our first result is a formula that is better than the definition for computational purposes, but gives less insight.

$$\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

Proof

Let $\mu = \mathbb{E}(X)$ and $\nu = \mathbb{E}(Y)$. Then

$$\text{cov}(X, Y) = \mathbb{E}[(X - \mu)(Y - \nu)] = \mathbb{E}(XY - \mu Y - \nu X + \mu\nu) = \mathbb{E}(XY) - \mu\mathbb{E}(Y) - \nu\mathbb{E}(X) + \mu\nu = \mathbb{E}(XY) - \mu\nu \quad (4.5.3)$$

From (2), we see that X and Y are uncorrelated if and only if $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, so here is a simple but important corollary:

If X and Y are independent, then they are uncorrelated.

Proof

We showed in Section 1 that if X and Y are independent then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$.

However, the converse fails with a passion: Exercise (31) gives an example of two variables that are *functionally related* (the strongest form of dependence), yet uncorrelated. The computational exercises give other examples of dependent yet uncorrelated variables also. Note also that if one of the variables has mean 0, then the covariance is simply the expected product.

Trivially, covariance is a symmetric operation.

$$\text{cov}(X, Y) = \text{cov}(Y, X).$$

As the name suggests, covariance generalizes variance.

$$\text{cov}(X, X) = \text{var}(X).$$

Proof

Let $\mu = \mathbb{E}(X)$. Then $\text{cov}(X, X) = \mathbb{E}[(X - \mu)^2] = \text{var}(X)$.

Covariance is a linear operation in the first argument, if the second argument is fixed.

If X, Y, Z are random variables, and c is a constant, then

1. $\text{cov}(X + Y, Z) = \text{cov}(X, Z) + \text{cov}(Y, Z)$
2. $\text{cov}(cX, Y) = c \text{cov}(X, Y)$

Proof

We use the computational formula in (2)

$$1. \quad \text{cov}(X + Y, Z) = \mathbb{E}[(X + Y)Z] - \mathbb{E}(X + Y)\mathbb{E}(Z) = \mathbb{E}(XZ + YZ) - [\mathbb{E}(X) + \mathbb{E}(Y)]\mathbb{E}(Z) \quad (4.5.4)$$

$$= [\mathbb{E}(XZ) - \mathbb{E}(X)\mathbb{E}(Z)] + [\mathbb{E}(YZ) - \mathbb{E}(Y)\mathbb{E}(Z)] = \text{cov}(X, Z) + \text{cov}(Y, Z) \quad (4.5.5)$$

$$2. \quad \text{cov}(cX, Y) = \mathbb{E}(cXY) - \mathbb{E}(cX)\mathbb{E}(Y) = c\mathbb{E}(XY) - c\mathbb{E}(X)\mathbb{E}(Y) = c[\mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)] = c \text{cov}(X, Y) \quad (4.5.6)$$

By symmetry, covariance is also a linear operation in the second argument, with the first argument fixed. Thus, the covariance operator is *bi-linear*. The general version of this property is given in the following theorem.

Suppose that (X_1, X_2, \dots, X_n) and (Y_1, Y_2, \dots, Y_m) are sequences of random variables, and that (a_1, a_2, \dots, a_n) and (b_1, b_2, \dots, b_m) are constants. Then

$$\text{cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \text{cov}(X_i, Y_j) \quad (4.5.7)$$

The following result shows how covariance is changed under a linear transformation of one of the variables. This is simply a special case of the basic properties, but is worth stating.

If $a, b \in \mathbb{R}$ then $\text{cov}(a + bX, Y) = b \text{cov}(X, Y)$.

Proof

A constant is independent of any random variable. Hence $\text{cov}(a + bX, Y) = \text{cov}(a, Y) + b \text{cov}(X, Y) = b \text{cov}(X, Y)$.

Of course, by symmetry, the same property holds in the second argument. Putting the two together we have that if $a, b, c, d \in \mathbb{R}$ then $\text{cov}(a + bX, c + dY) = bd \text{cov}(X, Y)$.

Properties of Correlation

Next we will establish some basic properties of correlation. Most of these follow easily from corresponding properties of covariance above. We assume that $\text{var}(X) > 0$ and $\text{var}(Y) > 0$, so that the random variable really are random and hence the correlation is well defined.

The correlation between X and Y is the covariance of the corresponding standard scores:

$$\text{cor}(X, Y) = \text{cov}\left(\frac{X - \mathbb{E}(X)}{\text{sd}(X)}, \frac{Y - \mathbb{E}(Y)}{\text{sd}(Y)}\right) = \mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\text{sd}(X)} \frac{Y - \mathbb{E}(Y)}{\text{sd}(Y)}\right) \quad (4.5.8)$$

Proof

From the definitions and the linearity of expected value,

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\text{sd}(X)\text{sd}(Y)} = \frac{\mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)])}{\text{sd}(X)\text{sd}(Y)} = \mathbb{E}\left(\frac{X - \mathbb{E}(X)}{\text{sd}(X)} \frac{Y - \mathbb{E}(Y)}{\text{sd}(Y)}\right) \quad (4.5.9)$$

Since the standard scores have mean 0, this is also the covariance of the standard scores.

This shows again that correlation is dimensionless, since of course, the standard scores are dimensionless. Also, correlation is symmetric:

$$\text{cor}(X, Y) = \text{cor}(Y, X).$$

Under a linear transformation of one of the variables, the correlation is unchanged if the slope is positive and changes sign if the slope is negative:

If $a, b \in \mathbb{R}$ and $b \neq 0$ then

1. $\text{cor}(a + bX, Y) = \text{cor}(X, Y)$ if $b > 0$
2. $\text{cor}(a + bX, Y) = -\text{cor}(X, Y)$ if $b < 0$

Proof

Let Z denote the standard score of X . If $b > 0$, the standard score of $a + bX$ is also Z . If $b < 0$, the standard score of $a + bX$ is $-Z$. Hence the result follows from the [result above](#) for standard scores.

This result reinforces the fact that correlation is a standardized measure of association, since multiplying the variable by a positive constant is equivalent to a change of scale, and adding a constant to a variable is equivalent to a change of location. For example, in the Challenger data, the underlying variables are temperature at the time of launch (in degrees Fahrenheit) and O-ring erosion (in millimeters). The correlation between these two variables is of fundamental importance. If we decide to measure temperature in degrees Celsius and O-ring erosion in inches, the correlation is unchanged. Of course, the same property holds in the second argument, so if $a, b, c, d \in \mathbb{R}$ with $b \neq 0$ and $d \neq 0$, then $\text{cor}(a + bX, c + dY) = \text{cor}(X, Y)$ if $bd > 0$ and $\text{cor}(a + bX, c + dY) = -\text{cor}(X, Y)$ if $bd < 0$.

The most important properties of covariance and correlation will emerge from our study of the [best linear predictor](#) below.

The Variance of a Sum

We will now show that the variance of a sum of variables is the sum of the pairwise covariances. This result is very useful since many random variables with special distributions can be written as sums of simpler random variables (see in particular the [binomial distribution](#) and [hypergeometric distribution](#) below).

If (X_1, X_2, \dots, X_n) is a sequence of real-valued random variables then

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i, X_j) = \sum_{i=1}^n \text{var}(X_i) + 2 \sum_{\{(i,j): i < j\}} \text{cov}(X_i, X_j) \quad (4.5.10)$$

Proof

From the variance property on (5), and the linear property (7),

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \text{cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right) = \sum_{i=1}^n \sum_{j=1}^n \text{cov}(X_i, X_j) \quad (4.5.11)$$

The second expression follows since $\text{cov}(X_i, X_i) = \text{var}(X_i)$ for each i and $\text{cov}(X_i, X_j) = \text{cov}(X_j, X_i)$ for $i \neq j$ by the symmetry property (4)

Note that the variance of a sum can be larger, smaller, or equal to the sum of the variances, depending on the pure covariance terms. As a special case of (12), when $n = 2$, we have

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2 \text{cov}(X, Y) \quad (4.5.12)$$

The following corollary is very important.

If (X_1, X_2, \dots, X_n) is a sequence of pairwise uncorrelated, real-valued random variables then

$$\text{var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{var}(X_i) \quad (4.5.13)$$

Proof

This follows immediately from (12), since $\text{cov}(X_i, X_j) = 0$ for $i \neq j$.

Note that the last result holds, in particular, if the random variables are independent. We close this discussion with a couple of minor corollaries.

If X and Y are real-valued random variables then $\text{var}(X + Y) + \text{var}(X - Y) = 2[\text{var}(X) + \text{var}(Y)]$.

Proof

From (12),

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y) + 2\text{cov}(X, Y) \quad (4.5.14)$$

Similarly,

$$\text{var}(X - Y) = \text{var}(X) + \text{var}(-Y) + 2\text{cov}(X, -Y) = \text{var}(X) + \text{var}(Y) - 2\text{cov}(X, Y) \quad (4.5.15)$$

Adding gives the result.

If X and Y are real-valued random variables with $\text{var}(X) = \text{var}(Y)$ then $X + Y$ and $X - Y$ are uncorrelated.

Proof

From the linear property (7) and the symmetry property (4),
 $\text{cov}(X + Y, X - Y) = \text{cov}(X, X) - \text{cov}(X, Y) + \text{cov}(Y, X) - \text{cov}(Y, Y) = \text{var}(X) - \text{var}(Y)$

Random Samples

In the following exercises, suppose that (X_1, X_2, \dots) is a sequence of independent, real-valued random variables with a common distribution that has mean μ and standard deviation $\sigma > 0$. In statistical terms, the variables form a random sample from the common distribution.

For $n \in \mathbb{N}_+$, let $Y_n = \sum_{i=1}^n X_i$.

1. $\mathbb{E}(Y_n) = n\mu$
2. $\text{var}(Y_n) = n\sigma^2$

Proof

1. This follows from the additive property of expected value.
2. This follows from the additive property of variance ((13) for independent variables)

For $n \in \mathbb{N}_+$, let $M_n = Y_n/n = \frac{1}{n} \sum_{i=1}^n X_i$, so that M_n is the sample mean of (X_1, X_2, \dots, X_n) .

1. $\mathbb{E}(M_n) = \mu$
2. $\text{var}(M_n) = \sigma^2/n$
3. $\text{var}(M_n) \rightarrow 0$ as $n \rightarrow \infty$
4. $\mathbb{P}(|M_n - \mu| > \epsilon) \rightarrow 0$ as $n \rightarrow \infty$ for every $\epsilon > 0$.

Proof

1. This follows from part (a) of the (16) and the scaling property of expected value.
2. This follows from part (b) of the (16) and the scaling property of variance.
3. This is an immediate consequence of (b).
4. This follows from (c) and Chebyshev's inequality: $\mathbb{P}(|M_n - \mu| > \epsilon) \leq \text{var}(M_n)/\epsilon^2 \rightarrow 0$ as $n \rightarrow \infty$

Part (c) of (17) means that $M_n \rightarrow \mu$ as $n \rightarrow \infty$ in mean square. Part (d) means that $M_n \rightarrow \mu$ as $n \rightarrow \infty$ in probability. These are both versions of the weak law of large numbers, one of the fundamental theorems of probability.

The standard score of the sum Y_n and the standard score of the sample mean M_n are the same:

$$Z_n = \frac{Y_n - n\mu}{\sqrt{n}\sigma} = \frac{M_n - \mu}{\sigma/\sqrt{n}} \quad (4.5.16)$$

1. $\mathbb{E}(Z_n) = 0$
2. $\text{var}(Z_n) = 1$

Proof

The equality of the standard score of Y_n and of Z_n is a result of simple algebra. But recall more generally that the standard score of a variable is unchanged by a linear transformation of the variable with positive slope (a location-scale transformation of the distribution). Of course, parts (a) and (b) are true for any standard score.

The central limit theorem, the other fundamental theorem of probability, states that the distribution of Z_n converges to the standard normal distribution as $n \rightarrow \infty$.

Events

If A and B are events in our random experiment then the covariance and correlation of A and B are defined to be the covariance and correlation, respectively, of their indicator random variables.

If A and B are events, define $\text{cov}(A, B) = \text{cov}(\mathbf{1}_A, \mathbf{1}_B)$ and $\text{cor}(A, B) = \text{cor}(\mathbf{1}_A, \mathbf{1}_B)$. Equivalently,

1. $\text{cov}(A, B) = \mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)$
2. $\text{cor}(A, B) = [\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)] / \sqrt{\mathbb{P}(A)[1 - \mathbb{P}(A)] \mathbb{P}(B)[1 - \mathbb{P}(B)]}$

Proof

Recall that if X is an indicator variable with $\mathbb{P}(X = 1) = p$, then $\mathbb{E}(X) = p$ and $\text{var}(X) = p(1 - p)$. Also, if X and Y are indicator variables then XY is an indicator variable and $\mathbb{P}(XY = 1) = \mathbb{P}(X = 1, Y = 1)$. The results then follow from the definitions.

In particular, note that A and B are positively correlated, negatively correlated, or independent, respectively (as defined in the section on conditional probability) if and only if the indicator variables of A and B are positively correlated, negatively correlated, or uncorrelated, as defined in this section.

If A and B are events then

1. $\text{cov}(A, B^c) = -\text{cov}(A, B)$
2. $\text{cov}(A^c, B^c) = \text{cov}(A, B)$

Proof

These results follow from linear property (7) and the fact that $\mathbf{1}_{A^c} = 1 - \mathbf{1}_A$.

If A and B are events with $A \subseteq B$ then

1. $\text{cov}(A, B) = \mathbb{P}(A)[1 - \mathbb{P}(B)]$
2. $\text{cor}(A, B) = \sqrt{\mathbb{P}(A)[1 - \mathbb{P}(B)] / \mathbb{P}(B)[1 - \mathbb{P}(A)]}$

Proof

These results follow from (19), since $A \cap B = A$.

In the language of the experiment, $A \subseteq B$ means that A implies B . In such a case, the events are positively correlated, not surprising.

The Best Linear Predictor

What linear function of X (that is, a function of the form $a + bX$ where $a, b \in \mathbb{R}$) is closest to Y in the sense of minimizing mean square error? The question is fundamentally important in the case where random variable X (the *predictor variable*) is observable and random variable Y (the *response variable*) is not. The linear function can be used to estimate Y from an observed value of X . Moreover, the solution will have the added benefit of showing that covariance and correlation measure the *linear* relationship between X and Y . To avoid trivial cases, let us assume that $\text{var}(X) > 0$ and $\text{var}(Y) > 0$, so that the random variables really are random. The solution to our problem turns out to be the linear function of X with the same expected value as Y , and whose covariance with X is the same as that of Y .

The random variable $L(Y | X)$ defined as follows is the only linear function of X satisfying properties (a) and (b).

$$L(Y | X) = \mathbb{E}(Y) + \frac{\text{cov}(X, Y)}{\text{var}(X)}[X - \mathbb{E}(X)] \quad (4.5.17)$$

1. $\mathbb{E}[L(Y | X)] = \mathbb{E}(Y)$
2. $\text{cov}[X, L(Y | X)] = \text{cov}(X, Y)$

Proof

By the linearity of expected value,

$$\mathbb{E}[L(Y | X)] = \mathbb{E}(Y) + \frac{\text{cov}(X, Y)}{\text{var}(X)}[\mathbb{E}(X) - \mathbb{E}(X)] = \mathbb{E}(Y) \quad (4.5.18)$$

Next, by the linearity of covariance and the fact that a constant is independent (and hence uncorrelated) with any random variable,

$$\text{cov}[X, L(Y | X)] = \frac{\text{cov}(X, Y)}{\text{var}(X)}\text{cov}(X, X) = \frac{\text{cov}(X, Y)}{\text{var}(X)}\text{var}(X) = \text{cov}(X, Y) \quad (4.5.19)$$

Conversely, suppose that $U = a + bX$ satisfies $\mathbb{E}(U) = \mathbb{E}(Y)$ and $\text{cov}(X, U) = \text{cov}(X, Y)$. Again using linearity of covariance and the uncorrelated property of constants, the second equation gives $b\text{cov}(X, X) = \text{cov}(X, Y)$ so $b = \text{cov}(X, Y)/\text{var}(X)$. Then the first equation gives $a = \mathbb{E}(Y) - b\mathbb{E}(X)$, so $U = L(Y | X)$.

Note that in the presence of part (a), part (b) is equivalent to $\mathbb{E}[XL(Y | X)] = \mathbb{E}(XY)$. Here is another minor variation, but one that will be very useful: $L(Y | X)$ is the only linear function of X with the same mean as Y and with the property that $Y - L(Y | X)$ is uncorrelated with every linear function of X .

$L(Y | X)$ is the only linear function of X that satisfies

1. $\mathbb{E}[L(Y | X)] = \mathbb{E}(Y)$
2. $\text{cov}[Y - L(Y | X), U] = 0$ for every linear function U of X .

Proof

Of course part (a) is the same as part (a) of (22). Suppose that $U = a + bX$ where $a, b \in \mathbb{R}$. From basic properties of covariance and the previous result,

$$\text{cov}[Y - L(Y|X), U] = b \text{cov}[Y - L(Y|X), X] = b(\text{cov}(Y, X) - \text{cov}[L(Y|X), X]) = 0 \quad (4.5.20)$$

Conversely, suppose that V is a linear function of X and that $\mathbb{E}(V) = \mathbb{E}(Y)$ and $\text{cov}(Y - V, U) = 0$ for every linear function U of X . Letting $U = X$ we have $\text{cov}(Y - V, X) = 0$ so $\text{cov}(V, X) = \text{cov}(Y, X)$. Hence $V = L(Y|X)$ by (22).

The variance of $L(Y|X)$ and its covariance with Y turn out to be the same.

Additional properties of $L(Y|X)$:

1. $\text{var}[L(Y|X)] = \text{cov}^2(X, Y) / \text{var}(X)$
2. $\text{cov}[L(Y|X), Y] = \text{cov}^2(X, Y) / \text{var}(X)$

Proof

1. From basic properties of variance,

$$\text{var}[L(Y|X)] = \left[\frac{\text{cov}(X, Y)}{\text{var}(X)} \right]^2 \text{var}(X) = \frac{\text{cov}^2(X, Y)}{\text{var}(X)} \quad (4.5.21)$$

2. From basic properties of covariance,

$$\text{cov}[L(Y|X), Y] = \frac{\text{cov}(X, Y)}{\text{var}(X)} \text{cov}(X, Y) = \frac{\text{cov}^2(X, Y)}{\text{var}(X)} \quad (4.5.22)$$

We can now prove the fundamental result that $L(Y|X)$ is the linear function of X that is closest to Y in the mean square sense. We give two proofs; the first is more straightforward, but the second is more interesting and elegant.

Suppose that U is a linear function of X . Then

1. $\mathbb{E}[(Y - L(Y|X))^2] \leq \mathbb{E}[(Y - U)^2]$
2. Equality occurs in (a) if and only if $U = L(Y|X)$ with probability 1.

Proof from calculus

Let $\text{mse}(a, b)$ denote the mean square error when $U = a + bX$ is used as an estimator of Y , as a function of the parameters $a, b \in \mathbb{R}$:

$$\text{mse}(a, b) = \mathbb{E}[(Y - (a + bX))^2] \quad (4.5.23)$$

Expanding the square and using the linearity of expected value gives

$$\text{mse}(a, b) = a^2 + b^2 \mathbb{E}(X^2) + 2ab \mathbb{E}(X) - 2a \mathbb{E}(Y) - 2b \mathbb{E}(XY) + \mathbb{E}(Y^2) \quad (4.5.24)$$

In terms of the variables a and b , the first three terms are the *second-order terms*, the next two are the *first-order terms*, and the last is the *zero-order term*. The second-order terms define a *quadratic form* whose standard symmetric matrix is

$$\begin{bmatrix} 1 & \mathbb{E}(X) \\ \mathbb{E}(X) & \mathbb{E}(X^2) \end{bmatrix} \quad (4.5.25)$$

The determinant of this matrix is $\mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \text{var}(X)$ and the diagonal terms are positive. All of this means that the graph of mse is a paraboloid opening upward, so the minimum of mse will occur at the unique critical point. Setting the first derivatives of mse to 0 we have

$$-2\mathbb{E}(Y) + 2b\mathbb{E}(X) + 2a = 0 \quad (4.5.26)$$

$$-2\mathbb{E}(XY) + 2b\mathbb{E}(X^2) + 2a\mathbb{E}(X) = 0 \quad (4.5.27)$$

Solving the first equation for a gives $a = \mathbb{E}(Y) - b\mathbb{E}(X)$. Substituting this into the second equation and solving gives $b = \text{cov}(X, Y) / \text{var}(X)$.

Proof using properties

1. We abbreviate $L(Y|X)$ by L for simplicity. Suppose that U is a linear function of X . Then

$$\mathbb{E}[(Y - U)^2] = \mathbb{E}[(Y - L) + (L - U)]^2 = \mathbb{E}[(Y - L)^2] + 2\mathbb{E}[(Y - L)(L - U)] + \mathbb{E}[(L - U)^2] \quad (4.5.28)$$

Since $Y - L$ has mean 0, the middle term is $\text{cov}(Y - L, L - U)$. But L and U are linear functions of X and hence so is $L - U$. Thus $\text{cov}(Y - L, L - U) = 0$ by (23). Hence

$$\mathbb{E}[(Y - U)^2] = \mathbb{E}[(Y - L)^2] + \mathbb{E}[(L - U)^2] \geq \mathbb{E}[(Y - L)^2] \quad (4.5.29)$$

2. Equality occurs in (a) if and only if $\mathbb{E}[(L - U)^2] = 0$, if and only if $\mathbb{P}(L = U) = 1$.

The mean square error when $L(Y | X)$ is used as a predictor of Y is

$$\mathbb{E} \left([Y - L(Y | X)]^2 \right) = \text{var}(Y) [1 - \text{cor}^2(X, Y)] \quad (4.5.30)$$

Proof

Again, let $L = L(Y | X)$ for convenience. Since $Y - L$ has mean 0,

$$\mathbb{E} [(Y - L)^2] = \text{var}(Y - L) = \text{var}(Y) - 2\text{cov}(L, Y) + \text{var}(L) \quad (4.5.31)$$

But $\text{cov}(L, Y) = \text{var}(L) = \text{cov}^2(X, Y) / \text{var}(X)$ by (24). Hence

$$\mathbb{E} [(Y - L)^2] = \text{var}(Y) - \frac{\text{cov}^2(X, Y)}{\text{var}(X)} = \text{var}(Y) \left[1 - \frac{\text{cov}^2(X, Y)}{\text{var}(X)\text{var}(Y)} \right] = \text{var}(Y) [1 - \text{cor}^2(X, Y)] \quad (4.5.32)$$

Our solution to the best linear predictor problems yields important properties of covariance and correlation.

Additional properties of covariance and correlation:

1. $-1 \leq \text{cor}(X, Y) \leq 1$
2. $-\text{sd}(X)\text{sd}(Y) \leq \text{cov}(X, Y) \leq \text{sd}(X)\text{sd}(Y)$
3. $\text{cor}(X, Y) = 1$ if and only if, with probability 1, Y is a linear function of X with positive slope.
4. $\text{cor}(X, Y) = -1$ if and only if, with probability 1, Y is a linear function of X with negative slope.

Proof

Since mean square error is nonnegative, it follows from (26) that $\text{cor}^2(X, Y) \leq 1$. This gives parts (a) and (b). For parts (c) and (d), note that if $\text{cor}^2(X, Y) = 1$ then $Y = L(Y | X)$ with probability 1, and that the slope in $L(Y | X)$ has the same sign as $\text{cor}(X, Y)$.

The last two results clearly show that $\text{cov}(X, Y)$ and $\text{cor}(X, Y)$ measure the *linear* association between X and Y . The equivalent inequalities (a) and (b) above are referred to as the *correlation inequality*. They are also versions of the *Cauchy-Schwarz inequality*, named for Augustin Cauchy and Karl Schwarz

Recall from our previous discussion of variance that the best *constant* predictor of Y , in the sense of minimizing mean square error, is $\mathbb{E}(Y)$ and the minimum value of the mean square error for this predictor is $\text{var}(Y)$. Thus, the difference between the variance of Y and the *mean square error* above for $L(Y | X)$ is the reduction in the variance of Y when the linear term in X is added to the predictor:

$$\text{var}(Y) - \mathbb{E} \left([Y - L(Y | X)]^2 \right) = \text{var}(Y) \text{cor}^2(X, Y) \quad (4.5.33)$$

Thus $\text{cor}^2(X, Y)$ is the *proportion* of reduction in $\text{var}(Y)$ when X is included as a predictor variable. This quantity is called the (distribution) *coefficient of determination*. Now let

$$L(Y | X = x) = \mathbb{E}(Y) + \frac{\text{cov}(X, Y)}{\text{var}(X)} [x - \mathbb{E}(X)], \quad x \in \mathbb{R} \quad (4.5.34)$$

The function $x \mapsto L(Y | X = x)$ is known as the *distribution regression function* for Y given X , and its graph is known as the *distribution regression line*. Note that the regression line passes through $(\mathbb{E}(X), \mathbb{E}(Y))$, the center of the joint distribution.

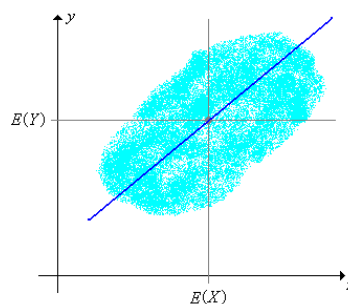


Figure 4.5.2: The distribution regression line

However, the choice of predictor variable and response variable is crucial.

The regression line for Y given X and the regression line for X given Y are not the same line, except in the trivial case where the variables are perfectly correlated. However, the coefficient of determination is the same, regardless of which variable is the predictor and which is the response.

Proof

The two regression lines are

$$y - \mathbb{E}(Y) = \frac{\text{cov}(X, Y)}{\text{var}(X)} [x - \mathbb{E}(X)] \quad (4.5.35)$$

$$x - \mathbb{E}(X) = \frac{\text{cov}(X, Y)}{\text{var}(Y)} [y - \mathbb{E}(Y)] \quad (4.5.36)$$

The two lines are the same if and only if $\text{cov}^2(X, Y) = \text{var}(X)\text{var}(Y)$. But this is equivalent to $\text{cor}^2(X, Y) = 1$.

Suppose that A and B are events with $0 < \mathbb{P}(A) < 1$ and $0 < \mathbb{P}(B) < 1$. Then

1. $\text{cor}(A, B) = 1$ if and only if $\mathbb{P}(A \setminus B) + \mathbb{P}(B \setminus A) = 0$. (That is, A and B are equivalent events.)
2. $\text{cor}(A, B) = -1$ if and only if $\mathbb{P}(A \setminus B^c) + \mathbb{P}(B^c \setminus A) = 0$. (That is, A and B^c are equivalent events.)

Proof

Recall from (19) that $\text{cor}(A, B) = \text{cor}(\mathbf{1}_A, \mathbf{1}_B)$, so if $\text{cor}^2(A, B) = 1$ then from (27), $\mathbf{1}_B = L(\mathbf{1}_B | \mathbf{1}_A)$ with probability 1. But $\mathbf{1}_A$ and $\mathbf{1}_B$ each takes values 0 and 1 only. Hence the only possible regression lines are $y = 0$, $y = 1$, $y = x$ and $y = 1 - x$. The first two correspond to $\mathbb{P}(B) = 0$ and $\mathbb{P}(B) = 1$, respectively, which are excluded by the hypotheses.

1. In this case, the slope is positive, so the regression line is $y = x$. That is, $\mathbf{1}_B = \mathbf{1}_A$ with probability 1.
2. In this case, the slope is negative, so the regression line is $y = 1 - x$. That is, $\mathbf{1}_B = 1 - \mathbf{1}_A = \mathbf{1}_{A^c}$ with probability 1.

The concept of best linear predictor is more powerful than might first appear, because it can be applied to *transformations* of the variables. Specifically, suppose that X and Y are random variables for our experiment, taking values in general spaces S and T , respectively. Suppose also that g and h are real-valued functions defined on S and T , respectively. We can find $L[h(Y) | g(X)]$, the linear function of $g(X)$ that is closest to $h(Y)$ in the mean square sense. The results of this subsection apply, of course, with $g(X)$ replacing X and $h(Y)$ replacing Y . Of course, we must be able to compute the appropriate means, variances, and covariances.

We close this subsection with two additional properties of the best linear predictor, the *linearity properties*.

Suppose that X , Y , and Z are random variables and that c is a constant. Then

1. $L(Y + Z | X) = L(Y | X) + L(Z | X)$
2. $L(cY | X) = cL(Y | X)$

Proof from the definitions

These results follow easily from the linearity of expected value and covariance.

$$1. \quad L(Y + Z | X) = \mathbb{E}(Y + Z) + \frac{\text{cov}(X, Y + Z)}{\text{var}(X)} [X - \mathbb{E}(X)] \quad (4.5.37)$$

$$= \left(\mathbb{E}(Y) + \frac{\text{cov}(X, Y)}{\text{var}(X)} [X - \mathbb{E}(X)] \right) + \left(\mathbb{E}(Z) + \frac{\text{cov}(X, Z)}{\text{var}(X)} [X - \mathbb{E}(X)] \right) \quad (4.5.38)$$

$$= \mathbb{E}(Y | X) + \mathbb{E}(Z | X) \quad (4.5.39)$$

$$2. \quad L(cY | X) = \mathbb{E}(cY) + \frac{\text{cov}(X, cY)}{\text{var}(X)} [X - \mathbb{E}(X)] = c\mathbb{E}(Y) + c \frac{\text{cov}(X, Y)}{\text{var}(X)} [X - \mathbb{E}(X)] = cL(Y | X) \quad (4.5.40)$$

Proof by characterizing properties

1. We show that $L(Y | X) + L(Z | X)$ satisfy the properties that characterize $L(Y + Z | X)$.

$$\mathbb{E}[L(Y | X) + L(Z | X)] = \mathbb{E}[L(Y | X)] + \mathbb{E}[L(Z | X)] = \mathbb{E}(Y) + \mathbb{E}(Z) = \mathbb{E}(Y + Z) \quad (4.5.41)$$

$$\text{cov}[X, L(Y | X) + L(Z | X)] = \text{cov}[X, L(Y | X)] + \text{cov}[X, L(Z | X)] = \text{cov}(X, Y) + \text{cov}(X, Z) = \text{cov}(X, Y + Z) \quad (4.5.42)$$

2. Similarly, we show that $cL(Y | X)$ satisfies the properties that characterize $L(cY | X)$

$$\mathbb{E}[cL(Y | X)] = c\mathbb{E}[L(Y | X)] = c\mathbb{E}(Y) = \mathbb{E}(cY) \quad (4.5.43)$$

$$\text{cov}[X, cL(Y | X)] = c \text{cov}[X, L(Y | X)] = c \text{cov}(X, Y) = \text{cov}(X, cY) \quad (4.5.44)$$

There are several extensions and generalizations of the ideas in the subsection:

- The corresponding statistical problem of estimating a and b , when these distribution parameters are unknown, is considered in the section on Sample Covariance and Correlation.
- The problem finding the function of X that is closest to Y in the mean square error sense (using *all* reasonable functions, not just linear functions) is considered in the section on Conditional Expected Value.
- The best linear prediction problem when the predictor and response variables are random vectors is considered in the section on Expected Value and Covariance Matrices.

The use of characterizing properties will play a crucial role in these extensions.

Examples and Applications

Uniform Distributions

Suppose that X is uniformly distributed on the interval $[-1, 1]$ and $Y = X^2$. Then X and Y are uncorrelated even though Y is a function of X (the strongest form of dependence).

Proof

Note that $\mathbb{E}(X) = 0$ and $\mathbb{E}(Y) = \mathbb{E}(X^2) = 1/3$ and $\mathbb{E}(XY) = \mathbb{E}(X^3) = 0$. Hence $\text{cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) = 0$.

Suppose that (X, Y) is uniformly distributed on the region $S \subseteq \mathbb{R}^2$. Find $\text{cov}(X, Y)$ and $\text{cor}(X, Y)$ and determine whether the variables are independent in each of the following cases:

1. $S = [a, b] \times [c, d]$ where $a < b$ and $c < d$, so S is a rectangle.
2. $S = \{(x, y) \in \mathbb{R}^2 : -a \leq y \leq x \leq a\}$ where $a > 0$, so S is a triangle
3. $S = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq r^2\}$ where $r > 0$, so S is a circle

Answer

1. $\text{cov}(X, Y) = 0$, $\text{cor}(X, Y) = 0$. X and Y are independent.
2. $\text{cov}(X, Y) = \frac{a^2}{9}$, $\text{cor}(X, Y) = \frac{1}{2}$. X and Y are dependent.
3. $\text{cov}(X, Y) = 0$, $\text{cor}(X, Y) = 0$. X and Y are dependent.

In the bivariate uniform experiment, select each of the regions below in turn. For each region, run the simulation 2000 times and note the value of the correlation and the shape of the cloud of points in the scatterplot. Compare with the results in the last exercise.

1. Square
2. Triangle
3. Circle

Suppose that X is uniformly distributed on the interval $(0, 1)$ and that given $X = x \in (0, 1)$, Y is uniformly distributed on the interval $(0, x)$. Find each of the following:

1. $\text{cov}(X, Y)$
2. $\text{cor}(X, Y)$
3. $L(Y | X)$
4. $L(X | Y)$

Answer

1. $\frac{1}{24}$
2. $\sqrt{\frac{3}{7}}$
3. $\frac{1}{2}X$
4. $\frac{2}{7} + \frac{6}{7}Y$

Dice

Recall that a *standard die* is a six-sided die. A *fair die* is one in which the faces are equally likely. An *ace-six flat die* is a standard die in which faces 1 and 6 have probability $\frac{1}{4}$ each, and faces 2, 3, 4, and 5 have probability $\frac{1}{8}$ each.

A pair of standard, fair dice are thrown and the scores (X_1, X_2) recorded. Let $Y = X_1 + X_2$ denote the sum of the scores, $U = \min\{X_1, X_2\}$ the minimum scores, and $V = \max\{X_1, X_2\}$ the maximum score. Find the covariance and correlation of each of the following pairs of variables:

1. (X_1, X_2)
2. (X_1, Y)
3. (X_1, U)
4. (U, V)
5. (U, Y)

Answer

1. 0, 0
2. $\frac{35}{12}$, $\frac{1}{\sqrt{2}} = 0.7071$
3. $\frac{35}{24}$, 0.6082
4. $\frac{1369}{1296}$, $\frac{1369}{2555} = 0.5358$
5. $\frac{35}{12}$, 0.8601

Suppose that n fair dice are thrown. Find the mean and variance of each of the following variables:

1. Y_n , the sum of the scores.
2. M_n , the average of the scores.

Answer

1. $\mathbb{E}(Y_n) = \frac{7}{2}n$, $\text{var}(Y_n) = \frac{35}{12}n$
2. $\mathbb{E}(M_n) = \frac{7}{2}$, $\text{var}(M_n) = \frac{35}{12n}$

In the dice experiment, select fair dice, and select the following random variables. In each case, increase the number of dice and observe the size and location of the probability density function and the mean \pm standard deviation bar. With $n = 20$ dice, run the experiment 1000 times and compare the sample mean and standard deviation to the distribution mean and standard deviation.

1. The sum of the scores.
2. The average of the scores.

Suppose that n ace-six flat dice are thrown. Find the mean and variance of each of the following variables:

1. Y_n , the sum of the scores.
2. M_n , the average of the scores.

Answer

1. $n\frac{7}{2}$, $n\frac{15}{4}$
2. $\frac{7}{2}$, $\frac{15}{4n}$

In the dice experiment, select ace-six flat dice, and select the following random variables. In each case, increase the number of dice and observe the size and location of the probability density function and the mean \pm standard deviation bar. With $n = 20$ dice, run the experiment 1000 times and compare the sample mean and standard deviation to the distribution mean and standard deviation.

1. The sum of the scores.
2. The average of the scores.

A pair of fair dice are thrown and the scores (X_1, X_2) recorded. Let $Y = X_1 + X_2$ denote the sum of the scores, $U = \min\{X_1, X_2\}$ the minimum score, and $V = \max\{X_1, X_2\}$ the maximum score. Find each of the following:

1. $L(Y | X_1)$
2. $L(U | X_1)$
3. $L(V | X_1)$

Answer

1. $\frac{7}{2} + X_1$
2. $\frac{7}{9} + \frac{1}{2}X_1$
3. $\frac{49}{19} + \frac{1}{2}X_1$

Bernoulli Trials

Recall that a *Bernoulli trials process* is a sequence $\mathbf{X} = (X_1, X_2, \dots)$ of independent, identically distributed indicator random variables. In the usual language of reliability, X_i denotes the outcome of trial i , where 1 denotes success and 0 denotes failure. The probability of success $p = \mathbb{P}(X_i = 1)$ is the basic parameter of the process. The process is named for Jacob Bernoulli. A separate chapter on the Bernoulli Trials explores this process in detail.

For $n \in \mathbb{N}_+$, the number of successes in the first n trials is $Y_n = \sum_{i=1}^n X_i$. Recall that this random variable has the binomial distribution with parameters n and p , which has probability density function f given by

$$f_n(y) = \binom{n}{y} p^y (1-p)^{n-y}, \quad y \in \{0, 1, \dots, n\} \quad (4.5.45)$$

The mean and variance of Y_n are

1. $\mathbb{E}(Y_n) = np$
2. $\text{var}(Y_n) = np(1-p)$

Proof

These results could be derived from the PDF of Y_n , of course, but a derivation based on the sum of IID variables is much better. Recall that $\mathbb{E}(X_i) = p$ and $\text{var}(X_i) = p(1-p)$ so the results follow immediately from theorem (16).

In the binomial coin experiment, select the number of heads. Vary n and p and note the shape of the probability density function and the size and location of the mean \pm standard deviation bar. For selected values of the parameters, run the experiment 1000 times and compare the sample mean

and standard deviation to the distribution mean and standard deviation.

For $n \in \mathbb{N}_+$, the proportion of successes in the first n trials is $M_n = Y_n/n$. This random variable is sometimes used as a statistical estimator of the parameter p , when the parameter is unknown.

The mean and variance of M_n are

1. $\mathbb{E}(M_n) = p$
2. $\text{var}(M_n) = p(1-p)/n$

Proof

Recall that $\mathbb{E}(X_i) = p$ and $\text{var}(X_i) = p(1-p)$ so the results follow immediately from theorem (17).

In the binomial coin experiment, select the proportion of heads. Vary n and p and note the shape of the probability density function and the size and location of the mean \pm standard deviation bar. For selected values of the parameters, run the experiment 1000 times and compare the sample mean and standard deviation to the distribution mean and standard deviation.

As a special case of (17) note that $M_n \rightarrow p$ as $n \rightarrow \infty$ in mean square and in probability.

The Hypergeometric Distribution

Suppose that a population consists of m objects; r of the objects are type 1 and $m-r$ are type 0. A sample of n objects is chosen at random, without replacement. The parameters $m, n \in \mathbb{N}_+$ and $r \in \mathbb{N}$ with $n \leq m$ and $r \leq m$. For $i \in \{1, 2, \dots, n\}$, let X_i denote the type of the i th object selected. Recall that (X_1, X_2, \dots, X_n) is a sequence of identically distributed (but *not* independent) indicator random variables.

Let Y denote the number of type 1 objects in the sample, so that $Y = \sum_{i=1}^n X_i$. Recall that this random variable has the hypergeometric distribution, which has probability density function f_n given by

$$f(y) = \frac{\binom{r}{y} \binom{m-r}{n-y}}{\binom{m}{n}}, \quad y \in \{0, 1, \dots, n\} \quad (4.5.46)$$

For distinct $i, j \in \{1, 2, \dots, n\}$,

1. $\mathbb{E}(X_i) = \frac{r}{m}$
2. $\text{var}(X_i) = \frac{r}{m} \left(1 - \frac{r}{m}\right)$
3. $\text{cov}(X_i, X_j) = -\frac{r}{m} \left(1 - \frac{r}{m}\right) \frac{1}{m-1}$
4. $\text{cor}(X_i, X_j) = -\frac{1}{m-1}$

Proof

Recall that $\mathbb{E}(X_i) = \mathbb{P}(X_i = 1) = \frac{r}{m}$ for each i and $\mathbb{E}(X_i X_j) = \mathbb{P}(X_i = 1, X_j = 1) = \frac{r}{m} \frac{r-1}{m-1}$ for each $i \neq j$. Technically, the sequence of indicator variables is exchangeable. The results now follow from the definitions and simple algebra.

Note that the event of a type 1 object on draw i and the event of a type 1 object on draw j are negatively correlated, but the correlation depends only on the population size and not on the number of type 1 objects. Note also that the correlation is perfect if $m = 2$. Think about these result intuitively.

The mean and variance of Y are

1. $\mathbb{E}(Y) = n \frac{r}{m}$
2. $\text{var}(Y) = n \frac{r}{m} \left(1 - \frac{r}{m}\right) \frac{m-n}{m-1}$

Proof

Again, a derivation from the representation of Y as a sum of indicator variables is far preferable to a derivation based on the PDF of Y . These results follow immediately from (45), the additive property of expected value, and Theorem (12).

Note that if the sampling were *with* replacement, Y would have a binomial distribution, and so in particular $E(Y) = n \frac{r}{m}$ and $\text{var}(Y) = n \frac{r}{m} \left(1 - \frac{r}{m}\right)$. The additional factor $\frac{m-n}{m-1}$ that occurs in the variance of the hypergeometric distribution is sometimes called the *finite population correction factor*. Note that for fixed m , $\frac{m-n}{m-1}$ is decreasing in n , and is 0 when $n = m$. Of course, we know that we must have $\text{var}(Y) = 0$ if $n = m$, since we would be sampling the entire population, and so deterministically, $Y = r$. On the other hand, for fixed n , $\frac{m-n}{m-1} \rightarrow 1$ as $m \rightarrow \infty$. More generally, the hypergeometric distribution is well approximated by the binomial when the population size m is large compared to the sample size n . These ideas are discussed more fully in the section on the hypergeometric distribution in the chapter on Finite Sampling Models.

In the ball and urn experiment, select sampling without replacement. Vary m, r , and n and note the shape of the probability density function and the size and location of the mean \pm standard deviation bar. For selected values of the parameters, run the experiment 1000 times and compare the sample mean and standard deviation to the distribution mean and standard deviation.

Exercises on Basic Properties

Suppose that X and Y are real-valued random variables with $\text{cov}(X, Y) = 3$. Find $\text{cov}(2X - 5, 4Y + 2)$.

Answer

24

Suppose X and Y are real-valued random variables with $\text{var}(X) = 5$, $\text{var}(Y) = 9$, and $\text{cov}(X, Y) = -3$. Find

1. $\text{cor}(X, Y)$
2. $\text{var}(2X + 3Y - 7)$
3. $\text{cov}(5X + 2Y - 3, 3X - 4Y + 2)$
4. $\text{cor}(5X + 2Y - 3, 3X - 4Y + 2)$

Answer

1. $-\frac{1}{\sqrt{5}} \approx -0.4472$
2. 65
3. 45
4. $\frac{15}{\sqrt{2929}} \approx 0.2772$

Suppose that X and Y are independent, real-valued random variables with $\text{var}(X) = 6$ and $\text{var}(Y) = 8$. Find $\text{var}(3X - 4Y + 5)$.

Answer

182

Suppose that A and B are events in an experiment with $\mathbb{P}(A) = \frac{1}{2}$, $\mathbb{P}(B) = \frac{1}{3}$, and $\mathbb{P}(A \cap B) = \frac{1}{8}$. Find each of the following:

1. $\text{cov}(A, B)$
2. $\text{cor}(A, B)$

Answer

1. $-\frac{1}{24}$
2. $-\sqrt{2}/8$

Suppose that X , Y , and Z are real-valued random variables for an experiment, and that $L(Y | X) = 2 - 3X$ and $L(Z | X) = 5 + 4X$. Find $L(6Y - 2Z | X)$.

Answer

$2 - 26X$

Suppose that X and Y are real-valued random variables for an experiment, and that $\mathbb{E}(X) = 3$, $\text{var}(X) = 4$, and $L(Y | X) = 5 - 2X$. Find each of the following:

1. $\mathbb{E}(Y)$
2. $\text{cov}(X, Y)$

Answer

1. -1
2. -8

Simple Continuous Distributions

Suppose that (X, Y) has probability density function f given by $f(x, y) = x + y$ for $0 \leq x \leq 1$, $0 \leq y \leq 1$. Find each of the following

1. $\text{cov}(X, Y)$
2. $\text{cor}(X, Y)$
3. $L(Y | X)$
4. $L(X | Y)$

Answer

1. $-\frac{1}{144}$
2. $-\frac{1}{11} \approx -0.0909$
3. $\frac{7}{11} - \frac{1}{11}X$
4. $\frac{7}{11} = \frac{1}{11}Y$

Suppose that (X, Y) has probability density function f given by $f(x, y) = 2(x + y)$ for $0 \leq x \leq y \leq 1$. Find each of the following:

1. $\text{cov}(X, Y)$
2. $\text{cor}(X, Y)$
3. $L(Y | X)$
4. $L(X | Y)$

Answer

1. $\frac{1}{48}$
2. $\frac{5}{\sqrt{129}} \approx 0.4402$
3. $\frac{26}{43} + \frac{15}{43}X$
4. $\frac{5}{9}Y$

Suppose again that (X, Y) has probability density function f given by $f(x, y) = 2(x + y)$ for $0 \leq x \leq y \leq 1$.

1. Find $\text{cov}(X^2, Y)$.
2. Find $\text{cor}(X^2, Y)$.
3. Find $L(Y | X^2)$.
4. Which predictor of Y is better, the one based on X or the one based on X^2 ?

Answer

1. $\frac{7}{360}$
2. 0.448
3. $\frac{1255}{1920} + \frac{245}{634}X$
4. The predictor based on X^2 is slightly better.

Suppose that (X, Y) has probability density function f given by $f(x, y) = 6x^2y$ for $0 \leq x \leq 1, 0 \leq y \leq 1$. Find each of the following:

1. $\text{cov}(X, Y)$
2. $\text{cor}(X, Y)$
3. $L(Y | X)$
4. $L(X | Y)$

Answer

Note that X and Y are independent.

1. 0
2. 0
3. $\frac{2}{3}$
4. $\frac{3}{4}$

Suppose that (X, Y) has probability density function f given by $f(x, y) = 15x^2y$ for $0 \leq x \leq y \leq 1$. Find each of the following:

1. $\text{cov}(X, Y)$
2. $\text{cor}(X, Y)$
3. $L(Y | X)$
4. $L(X | Y)$

Answer

1. $\frac{5}{336}$
2. 0.05423
3. $\frac{30}{51} + \frac{20}{51}X$
4. $\frac{3}{4}Y$

Suppose again that (X, Y) has probability density function f given by $f(x, y) = 15x^2y$ for $0 \leq x \leq y \leq 1$.

1. Find $\text{cov}(\sqrt{X}, Y)$.
2. Find $\text{cor}(\sqrt{X}, Y)$.
3. Find $L(Y | \sqrt{X})$.
4. Which of the predictors of Y is better, the one based on X or the one based on \sqrt{X} ?

Answer

1. $\frac{10}{1001}$
2. $\frac{24}{169}\sqrt{14}$

3. $\frac{5225}{13\ 182} + \frac{1232}{2197}X$

4. The predictor based on X is slightly better.

This page titled [4.5: Covariance and Correlation](#) is shared under a [CC BY 2.0](#) license and was authored, remixed, and/or curated by [Kyle Siegrist \(Random Services\)](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.