

A Matrix Formulation for Computing Covariances, Variances, and Correlations for Linear Combinations of Random Variables

Gerald E. Rubin

Marshall University, Huntington, West Virginia 25755, U.S.A.

The covariance function is extremely important in both theoretical and applied statistics. Difficulties are often encountered in the computation of covariances, as well as variances and correlations, for linear combinations of random variables. In this paper, a diagrammatic matrix formulation is presented which should simplify such computations, as well as provide some more insight and understanding of some of the concepts involved.

Consider the random variables U_1, U_2, \dots, U_m and V_1, V_2, \dots, V_n . Suppose we want to compute

$$\text{cov} \left(\sum_{i=1}^m U_i, \sum_{j=1}^n V_j \right).$$

We know that

$$\text{cov} \left(\sum_{i=1}^m U_i, \sum_{j=1}^n V_j \right) = \sum_{i=1}^m \sum_{j=1}^n \text{cov}(U_i, V_j).$$

We can construct an $m \times n$ matrix which will contain all of the required terms. Letting $c_{U_i, V_j} = \text{cov}(U_i, V_j)$, we have

$$(1) \quad \begin{array}{cccccc} & V_1 & V_2 & \dots & V_j & \dots V_n \\ U_1 & c_{U_1, V_1} & & \dots & & c_{U_1, V_n} \\ U_2 & & & & & \\ \vdots & \vdots & & & & \vdots \\ & & c_{U_i, V_j} & & & \\ U_m & c_{U_m, V_1} & \dots & & c_{U_m, V_n} \end{array}$$

The desired covariance is the sum of all the terms in the matrix in (1).

Covariance of General Linear Combinations of X_i 's and Y_j 's

Consider the linear combinations:

$$U = \sum_{i=1}^m a_i X_i \quad (\text{where } a_i \text{'s are constants})$$

$$V = \sum_{j=1}^n b_j Y_j \quad (\text{where } b_j \text{'s are constants}).$$

We know that

$$\begin{aligned}\text{cov}(U, V) &= \text{cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n b_j Y_j\right) \\ &= \sum_{i=1}^m \sum_{j=1}^n \text{cov}(a_i X_i, b_j Y_j) \\ &= \sum_{i=1}^m \sum_{j=1}^n a_i b_j \text{cov}(X_i, Y_j)\end{aligned}$$

We call illustrate this by an $m \times n$ matrix, as in (1), as follows:

$$(2) \quad \begin{array}{ccccccc} & b_1 Y_1 & b_2 Y_2 & \dots & b_j Y_j & \dots & b_n Y_n \\ a_1 X_1 & a_1 b_1 c_{X_1, Y_1} & & & \dots & & a_1 b_n c_{X_1, Y_n} \\ \vdots & & & & & & \\ a_i X_i & & & & a_i b_j c_{X_i, Y_j} & & \vdots \\ \vdots & & & & & & \\ a_m X_m & a_m b_1 c_{X_m, Y_1} & & \dots & & & a_m b_n c_{X_m, Y_n} \end{array}$$

The (i, j) th element is $a_i b_j c_{X_i, Y_j}$ and represents the $\text{cov}(a_i X_i, b_j Y_j)$. One then *adds* all of the elements of (2) to find the desired covariance. We illustrate the simplicity of the computations in the following example.

Example 1: Consider $X_1, X_2, X_3, Y_1, Y_2, Y_3, Y_4$ such that

$$\begin{aligned}\text{cov}(X_1, Y_1) &= -3, & \text{cov}(X_1, Y_2) &= -2, & \text{cov}(X_1, Y_3) &= 1, & \text{cov}(X_1, Y_4) &= 0 \\ \text{cov}(X_2, Y_1) &= 0, & \text{cov}(X_2, Y_2) &= -1, & \text{cov}(X_2, Y_3) &= 2, & \text{cov}(X_2, Y_4) &= 1 \\ \text{cov}(X_3, Y_1) &= 1, & \text{cov}(X_3, Y_2) &= 0, & \text{cov}(X_3, Y_3) &= 3, & \text{cov}(X_3, Y_4) &= -2.\end{aligned}$$

Find $\text{cov}(2X_1 + 3X_2 - 2X_3, -3Y_1 + 4Y_2 - 2Y_3 + 3Y_4)$.

Solution: Using (2):

$$\begin{array}{ccccccccc} & -3Y_1 & 4Y_2 & -2Y_3 & 3Y_4 & & & & \\ 2X_1 & -6(-3) & 8(-2) & -4(1) & 6(0) & & -2 & & \\ 3X_2 & -9(0) & 12(-1) & -6(2) & 9(1) & & & & \\ -2X_3 & 6(1) & -8(0) & 4(3) & -6(-2) & & & & \\ & & & & & \xrightarrow{\text{adding row entries}} & -15 & \xrightarrow{\text{adding column figures}} & \underline{\underline{13}} \\ & & & & & & 30 & & \end{array}$$

Variance of Linear Combinations of Random Variables

Consider $U = \sum_{i=1}^m a_i X_i$. We know that

$$\text{var } U = \text{var}\left(\sum_{i=1}^m a_i X_i\right) = \text{cov}\left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^m a_j X_j\right).$$

We will let $c_{ij} = \text{cov}(X_i, X_j)$, $i, j = 1, \dots, m$. Note that

$$c_{ii} = \text{cov}(X_i, X_i) = \text{var } X_i, \quad i = 1, \dots, m.$$

Thus, by (2), we can construct the following $m \times m$ matrix for finding the desired variance.

$$(3) \quad \begin{array}{cccccccc} & a_1 X_1 & a_2 X_2 & \dots & a_i X_i & \dots & a_j X_j & \dots & a_m X_m \\ a_1 X_1 & a_1^2 c_{11} & a_1 a_2 c_{12} & & \dots & & & & a_1 a_m c_{1m} \\ a_2 X_2 & a_2 a_1 c_{21} & a_2^2 c_{22} & & & & & & \\ & & & \ddots & & & & & \\ a_i X_i & & & & a_i^2 c_{ii} & & a_i a_j c_{ij} & & \\ \vdots & \vdots & & & \ddots & & & & \vdots \\ a_j X_j & & & & a_j a_i c_{ji} & & a_j^2 c_{jj} & & \\ \vdots & & & & & & & \ddots & \\ a_m X_m & a_m a_1 c_{m1} & & \dots & & & & & a_m^2 c_{mm} \end{array}$$

The sum of all the entries in (3) yields the desired variance.

Notes:

(1) The main diagonal elements $a_1^2 c_{11}, a_2^2 c_{22}, \dots, a_m^2 c_{mm}$ represent individual variances of $a_1 X_1, a_2 X_2, \dots$, since

$$a_1^2 c_{11} = a_1^2 \text{var}(X_1) = \text{var}(a_1 X_1)$$

and in general

$$a_i^2 c_{ii} = a_i^2 \text{var}(X_i) = \text{var}(a_i X_i).$$

(2) The matrix in (3) is symmetric (the (i, j) th entry = (j, i) th entry for all i, j) since for all i, j ,

$$a_i a_j c_{ij} = a_j a_i c_{ji}$$

($c_{ij} = c_{ji}$ by definition of covariances). Thus, the terms above the main diagonal are exactly the same as those below it. Thus, we could just double the sum of those above the main diagonal and add this to the sum of the diagonal entries to get the desired variance.

Example 2: Consider X_1, X_2, X_3 such that

$$\begin{aligned} \text{var } X_1 &= 4, & \text{var } X_2 &= 1, & \text{var } X_3 &= 9, \\ \text{cov}(X_1, X_2) &= -1, & \text{cov}(X_1, X_3) &= 2, & \text{cov}(X_2, X_3) &= -2. \end{aligned}$$

Find $\text{var}(2X_1 + 3X_2 - 2X_3)$.

Solution: Applying (3):

$$\begin{array}{ccccccc} & 2X_1 & 3X_2 & -2X_3 & & & \\ 2X_1 & 4(4) & 6(-1) & -4(2) & & 2 & \\ 3X_2 & 6(-1) & 9(1) & -6(-2) & \longrightarrow \text{adding} & 15 & \longrightarrow \text{adding} \\ -2X_3 & -4(2) & -6(-2) & 4(9) & \text{row entries} & 40 & \text{column} \\ & & & & & & \text{figures} \\ & & & & & & \underline{\underline{57}} \end{array}$$

Note: As mentioned, we can also add the main diagonal sum (here $16 + 9 + 36 = 61$) to double the sum of the entries above the diagonal (here: $2(-6 - 8 + 12) = -4$) to get the desired variance (here: $-4 + 61 = 57$).

Example 3: Consider Y_1, Y_2, Y_3, Y_4 such that

$$\begin{aligned} \text{var } Y_1 &= -3, & \text{var } Y_2 &= 2, & \text{var } Y_3 &= 9, & \text{var } Y_4 &= 4, \\ \text{cov}(Y_1, Y_2) &= 1, & \text{cov}(Y_1, Y_3) &= -1, & \text{cov}(Y_1, Y_4) &= 0, \\ \text{cov}(Y_2, Y_3) &= -2, & \text{cov}(Y_2, Y_4) &= -1, & \text{cov}(Y_3, Y_4) &= 2. \end{aligned}$$

Find $\text{var}(-3Y_1 + 4Y_2 - 2Y_3 + 3Y_4)$.

Solution: Using (3)

$$\begin{array}{ccccc} & -3Y_1 & 4Y_2 & -2Y_3 & 3Y_4 \\ -3Y_1 & 9(1) & -12(1) & 6(-1) & -9(0) \\ 4Y_2 & -12(1) & 16(2) & -8(-2) & 12(-1) \\ -2Y_3 & 6(-1) & -8(-2) & 4(9) & -6(2) \\ 3Y_4 & -9(0) & 12(-1) & -6(2) & 9(4) \end{array} \rightarrow \begin{array}{c} \text{adding} \\ \text{row entries} \end{array} \begin{array}{c} -9 \\ 24 \\ 34 \\ 12 \end{array} \rightarrow \begin{array}{c} \text{adding} \\ \text{column} \\ \text{figures} \end{array} \underline{\underline{61}}$$

Note: Alternatively, double the sum of the entries above the main diagonal (here, $2(-26) = -52$) plus the diagonal sum (here 113) yields the variance (here, $-52 + 113 = 61$).

Finding Correlations for Linear Combinations of Random Variables

The Pearsonian correlation coefficient of linear combinations, $U = \sum_{i=1}^m a_i X_i$, $V = \sum_{j=1}^n b_j Y_j$, can be expressed as

$$\begin{aligned} \rho(U, V) &= \rho \left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j \right) \\ &= \frac{\text{cov} \left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j \right)}{\sqrt{\text{var} \left(\sum_{i=1}^m a_i X_i \right)} \sqrt{\text{var} \left(\sum_{j=1}^n b_j Y_j \right)}} \\ &= \frac{\text{cov} \left(\sum_{i=1}^m a_i X_i, \sum_{j=1}^n b_j Y_j \right)}{\sqrt{\text{cov} \left(\sum_{i=1}^m a_i X_i, \sum_{k=1}^m a_k X_k \right)} \sqrt{\text{cov} \left(\sum_{j=1}^n b_j Y_j, \sum_{t=1}^n b_t Y_t \right)}} \end{aligned} \quad \begin{array}{l} (4) \\ (5) \end{array}$$

Thus now we can use the matrix formulation in (2) for the numerator covariance expression, and the matrix in (3) for each of the denominator expressions in (4) and (5).

Example 4: For $X_1, X_2, X_3, Y_1, Y_2, Y_3, Y_4$ described in examples 1-3, find the correlation coefficient $\rho_{U,V}$, where $U = 2X_1 + 3X_2 - 2X_3$, $V = -3Y_1 + 4Y_2 - 2Y_3 + 3Y_4$.

Solution: We will use formula (4) or (5). We have actually already computed all the

covariances and variance expressions involved:

$$\text{cov}(U, V) = \text{cov}(2X_1 + 3X_2 - 2X_3, -3Y_1 + 4Y_2 - 2Y_3 + 3Y_4)$$

$$= 13 \text{ (from example 1).}$$

$$\text{var}(U) = \text{var}(2X_1 + 3X_2 - 2X_3)$$

$$= 57 \text{ (from example 2).}$$

$$\text{var}(V) = \text{var}(-3Y_1 + 4Y_2 - 2Y_3 + 3Y_4)$$

$$= 61 \text{ (from example 3).}$$

$$\therefore \rho_{U,V} = \frac{\text{cov}(U, V)}{\sqrt{\text{var } U} \sqrt{\text{var } V}} = \frac{13}{\sqrt{57} \sqrt{61}} = \frac{13}{\sqrt{3477}} \approx \frac{13}{58.966091} \approx 0.220466$$