

# Exercise Walkthrough: Properties of Variance and Covariance

Justin Lanfermann

June 25, 2025

# Contents

<b>1</b>	<b>Introduction to the Exercise</b>	<b>1</b>
<b>2</b>	<b>Part 1: Variance of a Linear Combination</b>	<b>1</b>
2.1	The Goal . . . . .	1
2.2	The Step-by-Step Derivation . . . . .	1
<b>3</b>	<b>Part 2: Bilinearity of Covariance</b>	<b>3</b>
3.1	The Goal . . . . .	3
3.2	The Step-by-Step Derivation . . . . .	3
<b>4</b>	<b>Check Your Understanding</b>	<b>4</b>
<b>5</b>	<b>Summary and Takeaways</b>	<b>5</b>

# 1 Introduction to the Exercise

Hey there! In this walkthrough, we'll tackle a classic exercise that's fundamental for working with random variables: deriving the formulas for the variance and covariance of linear combinations. These rules are incredibly useful and appear everywhere in probability, statistics, and machine learning. A solid grasp of these manipulations will make many future topics much easier to understand.

Our goals are:

1. To prove that for scalar random variables  $X$  and  $Y$ :

$$\text{var}[aX + bY] = a^2 \text{var}[X] + b^2 \text{var}[Y] + 2ab \text{cov}[X, Y]$$

2. To prove that for random vectors  $Z_1, Z_2, Z_3$ :

$$\text{cov}[aZ_1 + bZ_2, Z_3] = a \text{cov}[Z_1, Z_3] + b \text{cov}[Z_2, Z_3]$$

We will proceed step-by-step, carefully applying the definitions and properties laid out in your "Discrete Probability Theory" script by Niki Kilbertus.

## 2 Part 1: Variance of a Linear Combination

### 2.1 The Goal

We want to prove the formula for the variance of a weighted sum of two random variables,  $aX + bY$ . This is a generalization of the property ' $\text{var}[X+Y]$ ' and is extremely common.

### 2.2 The Step-by-Step Derivation

Let's start by recalling the essential tools from the script.

#### Key Definitions and Properties:

- **Definition 2.5 (Variance):**  $\text{var}[Z] = \mathbb{E}[(Z - \mathbb{E}[Z])^2]$ .
- **Remark 2.6 (Computational Formula for Variance):** A more convenient formula for algebraic manipulation is  $\text{var}[Z] = \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2$ . We will use this one.
- **Proposition 2.4 (Linearity of Expectation):** The expectation operator is linear. For any random variables  $X, Y$  and constants  $a, b \in \mathbb{R}$ , we have  $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$ . <sup>[1]</sup>
- **Remark 2.10 (Covariance Formula):**  $\text{cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ . <sup>[2]</sup>

Now, let's apply these to our variable of interest,  $aX + bY$ .

1. **Apply the Computational Formula for Variance.** We treat  $aX + bY$  as a single random variable. According to Remark 2.6, its variance is:

$$\text{var}[aX + bY] = \mathbb{E}[(aX + bY)^2] - (\mathbb{E}[aX + bY])^2$$

Our strategy is to calculate the two terms on the right-hand side separately.

2. **Calculate the Second Term:**  $(\mathbb{E}[aX + bY])^2$ . This part is straightforward using the linearity of expectation (Proposition 2.4).

$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Now, we square this result. This is just expanding a binomial:  $(u + v)^2 = u^2 + 2uv + v^2$ .

$$(\mathbb{E}[aX + bY])^2 = (a\mathbb{E}[X] + b\mathbb{E}[Y])^2 = a^2(\mathbb{E}[X])^2 + b^2(\mathbb{E}[Y])^2 + 2ab\mathbb{E}[X]\mathbb{E}[Y] \quad (1)$$

3. **Calculate the First Term:**  $\mathbb{E}[(aX + bY)^2]$ . First, we expand the square *inside* the expectation.

$$(aX + bY)^2 = a^2X^2 + b^2Y^2 + 2abXY$$

Next, we apply the expectation to this whole expression and use its linearity property again.

$$\mathbb{E}[(aX + bY)^2] = \mathbb{E}[a^2X^2 + b^2Y^2 + 2abXY] = a^2\mathbb{E}[X^2] + b^2\mathbb{E}[Y^2] + 2ab\mathbb{E}[XY] \quad (2)$$

4. **Combine and Simplify.** Now we subtract the result from equation (1) from the result of equation (2).

$$\begin{aligned} \text{var}[aX + bY] &= (a^2\mathbb{E}[X^2] + b^2\mathbb{E}[Y^2] + 2ab\mathbb{E}[XY]) \\ &\quad - (a^2(\mathbb{E}[X])^2 + b^2(\mathbb{E}[Y])^2 + 2ab\mathbb{E}[X]\mathbb{E}[Y]) \end{aligned}$$

The final step is to group the terms by the constants  $a^2$ ,  $b^2$ , and  $2ab$ . This is where the magic happens!

$$\begin{aligned} \text{var}[aX + bY] &= a^2(\mathbb{E}[X^2] - (\mathbb{E}[X])^2) \\ &\quad + b^2(\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2) \\ &\quad + 2ab(\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]) \end{aligned}$$

We can now see that each of these lines corresponds exactly to the definitions of variance and covariance.

- The first line is  $a^2\text{var}[X]$ .
- The second line is  $b^2\text{var}[Y]$ .
- The third line is  $2ab\text{cov}[X, Y]$ .

Putting it all together, we have proven our desired result:

$$\text{var}[aX + bY] = a^2\text{var}[X] + b^2\text{var}[Y] + 2ab\text{cov}[X, Y] \quad \blacksquare$$

## 3 Part 2: Bilinearity of Covariance

### 3.1 The Goal

Here, we want to prove that covariance is linear in its first argument for random vectors. Specifically, we want to show:

$$\text{cov}[aZ_1 + bZ_2, Z_3] = a \text{cov}[Z_1, Z_3] + b \text{cov}[Z_2, Z_3]$$

This property is known as **bilinearity** (Proposition 2.12 (iii)) because it is linear in both arguments (due to symmetry, linearity in the first implies linearity in the second).

### 3.2 The Step-by-Step Derivation

The hint in the exercise mentions that the covariance between two random vectors  $Z_1, Z_2 : \Omega \rightarrow \mathbb{R}^n$  is a matrix defined as  $\text{cov}[Z_1, Z_2] = \mathbb{E}[(Z_1 - \mathbb{E}[Z_1])(Z_2 - \mathbb{E}[Z_2])^\top]$ . <sup>[3]</sup>

The solution provided in the prompt cleverly suggests a component-wise approach, which is often clearer. A matrix equation holds if and only if it holds for every component. So, we will prove that for any component  $(i, j)$  of the resulting covariance matrix, the equality holds.

#### Key Definitions and Properties:

- **Covariance Matrix (from Hint):** For random vectors  $X, Y \in \mathbb{R}^n$ , the  $(i, j)$ -th entry of the covariance matrix is  $(\text{cov}[X, Y])_{ij} = \text{cov}[X_i, Y_j]$ .
- **Definition 2.9 (Covariance):**  $\text{cov}[X_i, Y_j] = \mathbb{E}[(X_i - \mathbb{E}[X_i])(Y_j - \mathbb{E}[Y_j])]$ .
- **Proposition 2.4 (Linearity of Expectation):** As before, this is our main tool.

Let's proceed with the component-wise proof.

1. **Set up the component-wise calculation.** Let  $(Z_1)_i, (Z_2)_i, (Z_3)_j$  be the  $i$ -th and  $j$ -th components of the respective vectors. We want to compute the  $(i, j)$ -th entry of the matrix  $\text{cov}[aZ_1 + bZ_2, Z_3]$ .

$$(\text{cov}[aZ_1 + bZ_2, Z_3])_{ij} = \text{cov}[(aZ_1 + bZ_2)_i, (Z_3)_j]$$

Note that  $(aZ_1 + bZ_2)_i = a(Z_1)_i + b(Z_2)_i$ . So we are calculating:

$$\text{cov}[a(Z_1)_i + b(Z_2)_i, (Z_3)_j]$$

2. **Apply the definition of scalar covariance.** Using Definition 2.9, we expand the expression. For brevity, let's write  $X_i = (Z_1)_i$ ,  $Y_i = (Z_2)_i$ , and  $W_j = (Z_3)_j$ . We are looking at  $\text{cov}[aX_i + bY_i, W_j]$ .

$$\begin{aligned} & \text{cov}[aX_i + bY_i, W_j] \\ &= \mathbb{E}[(aX_i + bY_i - \mathbb{E}[aX_i + bY_i]) \cdot (W_j - \mathbb{E}[W_j])] \end{aligned}$$

3. **Use Linearity of Expectation on the first term.** Let's simplify the first part inside the expectation,  $(aX_i + bY_i) - \mathbb{E}[aX_i + bY_i]$ .

$$\begin{aligned}(aX_i + bY_i) - \mathbb{E}[aX_i + bY_i] &= (aX_i + bY_i) - (a\mathbb{E}[X_i] + b\mathbb{E}[Y_i]) \\ &= aX_i - a\mathbb{E}[X_i] + bY_i - b\mathbb{E}[Y_i] \\ &= a(X_i - \mathbb{E}[X_i]) + b(Y_i - \mathbb{E}[Y_i])\end{aligned}$$

4. **Substitute back and expand.** Now we substitute this back into the covariance expression.

$$\begin{aligned}\text{cov}[aX_i + bY_i, W_j] \\ &= \mathbb{E}[(a(X_i - \mathbb{E}[X_i]) + b(Y_i - \mathbb{E}[Y_i])) \cdot (W_j - \mathbb{E}[W_j])]\end{aligned}$$

We can now distribute the  $(W_j - \mathbb{E}[W_j])$  term:

$$\begin{aligned}&= \mathbb{E}[a(X_i - \mathbb{E}[X_i])(W_j - \mathbb{E}[W_j]) \\ &\quad + b(Y_i - \mathbb{E}[Y_i])(W_j - \mathbb{E}[W_j])]\end{aligned}$$

5. **Apply Linearity of Expectation one last time.** We can split the expectation over the sum:

$$\begin{aligned}&= a \cdot \mathbb{E}[(X_i - \mathbb{E}[X_i])(W_j - \mathbb{E}[W_j])] \\ &\quad + b \cdot \mathbb{E}[(Y_i - \mathbb{E}[Y_i])(W_j - \mathbb{E}[W_j])]\end{aligned}$$

We recognize these terms! The first is  $a \cdot \text{cov}[X_i, W_j]$  and the second is  $b \cdot \text{cov}[Y_i, W_j]$ .

6. **Conclude the proof.** Reverting to our original notation, we have shown that for any component  $(i, j)$ :

$$(\text{cov}[aZ_1 + bZ_2, Z_3])_{ij} = a(\text{cov}[Z_1, Z_3])_{ij} + b(\text{cov}[Z_2, Z_3])_{ij}$$

Since this equality holds for every component, the matrix equality must hold as well.

$$\text{cov}[aZ_1 + bZ_2, Z_3] = a \text{cov}[Z_1, Z_3] + b \text{cov}[Z_2, Z_3] \quad \blacksquare$$

## 4 Check Your Understanding

Now it's your turn to quickly test your understanding.

*Using the result from Part 1, what does  $\text{var}[X - Y]$  simplify to?  
Be careful with the signs!*

*Answer: We can write  $X - Y$  as  $aX + bY$  with  $a = 1$  and  $b = -1$ . Plugging this into our formula gives:  $\text{var}[X - Y] = 1^2 \text{var}[X] + (-1)^2 \text{var}[Y] + 2(1)(-1) \text{cov}[X, Y] = \text{var}[X] + \text{var}[Y] - 2 \text{cov}[X, Y]$ .*

## 5 Summary and Takeaways

We've successfully proven two very important properties. The key takeaways are:

- **Variance of a Sum:** The variance of a sum is *not* just the sum of variances unless the variables are uncorrelated ( $\text{cov}[X, Y] = 0$ ). You must always account for how the variables move together via the covariance term. This is a very common mistake! If they are independent, then their covariance is 0, and the formula simplifies.
- **Bilinearity of Covariance:** Covariance is a bilinear operator. This means it behaves like multiplication—it's linear in both its first and second arguments. This property is what allows us to "distribute" covariance calculations across sums, which is incredibly useful for simplifying complex expressions.

Mastering these rules is all about being comfortable with applying the linearity of expectation. It's the engine that drives both of these proofs.

## In-Depth Explanations

Here are more detailed explanations of the concepts we used, referenced by the numbers in the text.

### [1] Linearity of Expectation

The property  $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$  is arguably one of the most important properties of the expectation. It states that the expected value of a weighted sum of random variables is the weighted sum of their individual expected values.

**Intuition:** Think of expectation as the "center of mass" or "long-run average". If you take a set of numbers (the outcomes of  $X$ ) and scale them all by  $a$ , it makes sense that their average also gets scaled by  $a$ . If you add two sets of numbers (outcomes of  $X$  and  $Y$ ) element-wise, the average of the result is the sum of the individual averages. Linearity combines these two intuitive ideas.

**Significance:** Crucially, this property holds **regardless of whether  $X$  and  $Y$  are independent**. This is a powerful feature that is not shared by variance.

### [2] Covariance and its Computational Formula

**Definition:** Covariance,  $\text{cov}[X, Y] = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$ , measures the joint variability of two random variables. A positive covariance means that as  $X$  tends to be above its mean,  $Y$  also tends to be above its mean (they move together). A negative covariance means they tend to move in opposite directions. Zero covariance means there is no linear relationship.

**Computational Formula:** The formula  $\text{cov}[X, Y] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$  is derived directly from the definition by expanding the terms and using the linearity of expectation, just like we did for variance. It is almost always easier to use in calculations than the original definition, as calculating  $\mathbb{E}[XY]$  is often simpler than working with the centered variables inside the expectation.

### [3] The Covariance Matrix

When dealing with random vectors, which are just collections of random variables (e.g.,  $Z = (Z_1, Z_2, \dots, Z_n)$ ), we need a way to capture not only the variance of each component but also the covariance between every pair of components. The covariance matrix does exactly this.

**Structure:** For a random vector  $Z \in \mathbb{R}^n$ , the covariance matrix  $\Sigma = \text{cov}[Z, Z]$  is an  $n \times n$  matrix where:

- The diagonal entries  $(\Sigma)_{ii}$  are the variances:  $\text{cov}[Z_i, Z_i] = \text{var}[Z_i]$ .



- The off-diagonal entries  $(\Sigma)_{ij}$  are the covariances:  $\text{cov}[Z_i, Z_j]$  for  $i \neq j$ .

**Formula:** The compact formula  $\mathbb{E}[(Z - \mathbb{E}[Z])(Z - \mathbb{E}[Z])^\top]$  is a neat way to express this using matrix multiplication. If you write it out, you'll see that the  $(i, j)$ -th element of the resulting matrix is exactly  $\mathbb{E}[(Z_i - \mathbb{E}[Z_i])(Z_j - \mathbb{E}[Z_j])]$ , which is  $\text{cov}[Z_i, Z_j]$ .