

# Probability Basics

Lorenzo Naranjo

August 2026

This note introduces the concepts used to model uncertainty: outcomes and events, a probability measure that assigns each outcome a number representing its probability, and random variables that assign payoffs to outcomes. From these we define the expectation, variance, and covariance of a random variable, the tools used throughout the course to measure the return and risk of a portfolio.

In the following, there is only a finite number of future possibilities to simplify the mathematical exposition. For example, the future price of a stock can go up, stay constant, or go down. However, all the results presented in this chapter hold if we relax this assumption and allow for an infinite number of future outcomes.

We start by reviewing basic set theory, since sets are the natural language for describing a collection of possible future outcomes. From there, we build up the notions of events, random variables, and probability, and use them to define the expectation, variance, and covariance of asset payoffs, which are the building blocks for reasoning about risk and diversification in a portfolio.

## Sets

A **set** is a collection of objects. The objects of a set can be anything you want. For example, a set may contain numbers, letters, cars, or pictures. In our case, we will be concerned with sets that contain future possibilities or outcomes that can occur.

One way to define a set is to enumerate its elements. For example, the set of all integers from 1 to 10 is

$$A = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}.$$

Once we have defined a set, we can determine whether an object is an element of the set or not. For example, the number 3 is an element of  $A$  whereas the number 20 is not. We use the symbol  $\in$  to denote membership of a set and  $\notin$  to denote the contrary. Therefore, we have that  $3 \in A$  and  $20 \notin A$ .

Some sets can have an infinite number of elements. For example, the natural numbers are defined as

$$\mathbb{N} = \{0, 1, 2, 3, \dots\},$$

where the triple dots mean that if  $n$  is in  $\mathbb{N}$ , then  $n + 1$  is also in  $\mathbb{N}$ .

Since all elements of  $A$  are also members of  $\mathbb{N}$ , we say that  $A$  is a subset of  $\mathbb{N}$  and write it as  $A \subset \mathbb{N}$ . Using this terminology, we can redefine the set  $A$  defined above in a more *Pythonic* way:

$$A = \{n \in \mathbb{N} : n < 11\}.$$

If we are studying sets of natural numbers, it makes sense to define the **universe** to be  $\mathbb{N}$  and sets under study will be subsets of the universe.

Now, define the set  $B$  as

$$B = \{6, 7, 8, 9, 10, 11, 12, 13, 14, 15\}.$$

The **intersection** between  $A$  and  $B$  is the set denoted  $A \cap B$  whose members are both in  $A$  and  $B$ . Using the sets defined above, we have that

$$A \cap B = \{6, 7, 8, 9, 10\}.$$

The **union** of the sets  $A$  and  $B$  is the set denoted  $A \cup B$  whose members are either in  $A$ ,  $B$ , or both. Thus, using our previously defined sets we have that

$$A \cup B = \{1, 2, 3, \dots, 14, 15\}.$$

The **set difference** of  $A$  and  $B$  is the set denoted  $A \setminus B$  whose members are in  $A$  but are not members of  $B$ . Thus,

$$A \setminus B = \{1, 2, 3, 4, 5\}$$

and

$$B \setminus A = \{11, 12, 13, 14, 15\}.$$

The **complement** of  $A$  is the set denoted by  $A^C$  whose members are not in  $A$ . Of course this statement only makes sense if we define a universe where the elements not in  $A$  can live. If the universe is  $\mathbb{N}$ , then

$$A^C = \mathbb{N} \setminus A = \{11, 12, 13, \dots\}.$$

Similarly,

$$B^C = \{0, 1, 2, 3, 4, 5\} \cup \{16, 17, 18, \dots\}.$$

Note that if you take all the elements of  $A$  out of  $A$  you end up with an empty set, that is  $A \setminus A = \{\}$ . We typically denote the empty set by  $\emptyset$ , but is good to keep in mind that  $\emptyset = \{\}$ . In our universe of natural numbers, no natural number is a member of the empty set. We can write this formally as  $n \notin \emptyset, \forall n \in \mathbb{N}$ . Thus, the empty set is a subset of any subset of  $\mathbb{N}$ .

The **cardinality** of the set  $A$ , denoted by  $|A|$ , counts the number of elements in  $A$ . We then have that  $|A| = |B| = 10$ . The empty set has cardinality 0 whereas the cardinality of  $\mathbb{N}$  is denoted  $\aleph_0$ .

The **power set** of a set  $C$ , denoted by  $\mathcal{P}(C)$ , is the set containing all possible subsets of  $C$ . For example, if  $C = \{1, 2, 3\}$ , then  $|C| = 3$  and

$$\mathcal{P}(C) = \{\{\}, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{2, 3\}, \{1, 3\}, \{1, 2, 3\}\}.$$

Clearly, the power sets of  $A$  and  $B$  are much bigger. For a given set  $A$ , the cardinality of its power set is  $2^{|A|}$ . Therefore,  $\mathcal{P}(A)$  and  $\mathcal{P}(B)$  each contain  $2^{10} = 1024$  different sets.

Finally, the **Cartesian product** of  $A$  and  $B$  is the set denoted by  $A \times B$  whose members are all the pairwise combinations of the elements of  $A$  and  $B$ .

| $A \times B$ | 6        | 7        | ...      | 15       |
|--------------|----------|----------|----------|----------|
| 1            | (1, 6)   | (1, 7)   | ...      | (1, 15)  |
| 2            | (2, 6)   | (2, 7)   | ...      | (2, 15)  |
| $\vdots$     | $\vdots$ | $\vdots$ | $\ddots$ | $\vdots$ |
| 10           | (10, 6)  | (10, 7)  | ...      | (10, 15) |

The cardinality of  $A \times B$  is equal to the product of the cardinalities of  $A$  and  $B$ , i.e.,  $|A \times B| = |A| \times |B|$ .

## Outcomes and Events

In probability theory, a finite **sample space** is a non-empty finite set denoted by  $\Omega$ . The sample space includes all possible outcomes that can occur. A *probability measure* is a function that assigns to each element  $\omega$  of  $\Omega$  a number in  $[0, 1]$  so that

$$\sum_{\omega \in \Omega} P(\omega) = 1.$$

An event  $A$  is a subset of  $\Omega$ , and we define the probability of that event occurring as

$$P(A) = \sum_{\omega \in A} P(\omega). \quad (1)$$

Such a finite probability space is denoted by  $(\Omega, P)$ .

An immediate consequence of (1) is that  $P(\Omega) = 1$ . Furthermore, if  $A$  and  $B$  are disjoint sets of  $\Omega$  we have that

$$\begin{aligned} P(A \cup B) &= \sum_{\omega \in A \cup B} P(\omega) \\ &= \sum_{\omega \in A} P(\omega) + \sum_{\omega \in B} P(\omega) \\ &= P(A) + P(B). \end{aligned}$$

If we denote by  $A^C$  the complement of  $A$  in  $\Omega$ , the last expression implies that  $P(A) + P(A^C) = 1$ . Also, because  $\Omega^C = \emptyset$ , we also have that  $P(\Omega) + P(\emptyset) = 1$ , or  $P(\emptyset) = 0$ .

**Example 1.** If  $\Omega = \{\omega_1, \omega_2, \omega_3\}$ , then

$$\mathcal{P}(\Omega) = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_3\}, \{\omega_1, \omega_2\}, \{\omega_2, \omega_3\}, \{\omega_1, \omega_3\}, \{\omega_1, \omega_2, \omega_3\}\}$$

defines the collection of all possible events that we can measure. As we saw previously, the cardinality of  $\mathcal{P}(\Omega)$  grows exponentially with the size of  $\Omega$ .

The function  $P$  such that  $P(\omega_1) = 1/2$ ,  $P(\omega_2) = 1/4$ , and  $P(\omega_3) = 1/4$  defines a probability measure on  $\Omega$ . For example, we have that  $P(\{\omega_1, \omega_3\}) = 1/2 + 1/4 = 3/4$ .  $\square$

## Random Variables

### Definition

Sets and events tell us *what* can happen, but investment theory is ultimately about *numbers*: a stock's return, a portfolio's payoff, next quarter's dividend. A random variable is the tool that attaches a number to each possible outcome, which is exactly what we need to talk about returns instead of abstract states of the world.

If  $(\Omega, P)$  is a finite probability space, a **random variable** is a real-valued function defined on  $\Omega$ .

**Example 2.** Consider a sample space with four possible outcomes  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ . The table below describes the possible values of three random variables denoted by  $X$ ,  $Y$  and  $Z$ .

| Outcome    | $X$ | $Y$ | $Z$ |
|------------|-----|-----|-----|
| $\omega_1$ | -10 | 20  | 15  |
| $\omega_2$ | -5  | 10  | -10 |
| $\omega_3$ | 5   | 0   | 15  |
| $\omega_4$ | 10  | 0   | -10 |

Observing the values of  $X$  provides perfect information about which event happened. For example, if  $X = 5$  then we know that  $\omega_3$  occurred.

Knowing the values of  $Y$  or  $Z$ , on the other hand, does not provide the same amount of information. If we learn that  $Y = 0$  we only know that either  $\omega_3$  or  $\omega_4$  occurred. If we denote by  $\mathcal{F}_Y$  the set of events that can be generated by  $Y$ , we have that

$$\mathcal{F}_Y = \{\emptyset, \{\omega_1\}, \{\omega_2\}, \{\omega_1, \omega_2\}, \{\omega_3, \omega_4\}, \{\omega_1, \omega_3, \omega_4\}, \{\omega_2, \omega_3, \omega_4\}, \Omega\}.$$

The information set provided by  $Z$  is even smaller, since

$$\mathcal{F}_Z = \{\emptyset, \{\omega_1, \omega_3\}, \{\omega_2, \omega_4\}, \Omega\}.$$

Thus, a random variable does not necessarily provide all the information generated by the probability space  $\Omega$ . □

## Expectation and Variance

If  $X$  is a random variable defined on a finite probability space  $(\Omega, P)$ , the **expectation** or **expected value** of  $X$  is defined to be

$$E(X) = \sum_{\omega \in \Omega} X(\omega) P(\omega),$$

whereas the **variance** of  $X$  is

$$V(X) = E(X - E(X))^2.$$

The **standard deviation** is the square-root of the variance, i.e.,  $\sigma_X = \sqrt{V(X)}$ .

**Example 3.** Consider the sample space  $\Omega = \{\omega_1, \omega_2, \omega_3\}$  in which we define the probability measure  $P$  such that  $P(\omega_1) = 1/2$ ,  $P(\omega_2) = 1/4$ , and  $P(\omega_3) = 1/4$ . There are two random variables  $X$  and  $Y$  that take values in  $\Omega$  according to the table below.

| Outcome    | Probability | $X$ | $Y$ |
|------------|-------------|-----|-----|
| $\omega_1$ | 1/2         | 10  | 2   |
| $\omega_2$ | 1/4         | 8   | 40  |
| $\omega_3$ | 1/4         | 4   | 20  |

Using this information, we can compute the expectation of each random variable.

$$E(X) = \frac{1}{2} \times 10 + \frac{1}{4} \times 8 + \frac{1}{4} \times 4 = 8,$$

$$E(Y) = \frac{1}{2} \times 2 + \frac{1}{4} \times 40 + \frac{1}{4} \times 20 = 16.$$

Having computed the expectations of  $X$  and  $Y$ , we can compute their variances as

$$\begin{aligned} V(X) &= \frac{1}{2} \times (10 - 8)^2 + \frac{1}{4} \times (8 - 8)^2 + \frac{1}{4} \times (4 - 8)^2 = 6, \\ V(Y) &= \frac{1}{2} \times (2 - 16)^2 + \frac{1}{4} \times (40 - 16)^2 + \frac{1}{4} \times (20 - 16)^2 = 246. \end{aligned}$$

Finally, the standard deviations of  $X$  and  $Y$  are  $\sigma_X = \sqrt{6} \approx 2.45$  and  $\sigma_Y = \sqrt{246} \approx 15.68$ , respectively.  $\square$

## Covariance

The **covariance** between two random variables  $X$  and  $Y$  defined on a probability space  $(\Omega, P)$  is defined as

$$\text{Cov}(X, Y) = E(X - E(X))(Y - E(Y)),$$

and their **correlation** is

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}.$$

The correlation between any two random variables is always between -1 and 1.<sup>1</sup>

**Example 4.** Continuing with Example 3, we have that

$$\text{Cov}(X, Y) = \frac{1}{2} \times (10 - 8)(2 - 16) + \frac{1}{4} \times (8 - 8)(40 - 16) + \frac{1}{4} \times (4 - 8)(20 - 16) = -18.$$

---

<sup>1</sup>Define the standardized random variables  $U = \frac{X - E(X)}{\sigma_X}$  and  $V = \frac{Y - E(Y)}{\sigma_Y}$ , so that  $E(U^2) = E(V^2) = 1$  and  $E(UV) = \rho_{X,Y}$ . We can then compute

$$0 \leq E(U - V)^2 = 2 - 2\rho_{X,Y},$$

which implies  $\rho_{X,Y} \leq 1$ . Similarly,

$$0 \leq E(U + V)^2 = 2 + 2\rho_{X,Y},$$

which implies  $\rho_{X,Y} \geq -1$ . Thus, we conclude that

$$-1 \leq \rho_{X,Y} \leq 1.$$

Thus,  $\rho_{X,Y} \approx -0.47$ .

□

The covariance of  $X$  and  $Y$  can also be expressed as<sup>2</sup>

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

## Probability Mass Function

For discrete random variables, the **probability mass function** (or pmf) is a real-valued function that specifies the probability that the random variable  $X$  is equal to a certain value  $x$ , i.e.,

$$p_X(x) = P(\omega \in \Omega : X(\omega) = x).$$

**Example 5.** Suppose we define a probability measure  $P$  to the random variables  $X$  and  $Y$  defined in Example 2 according to the table below.

| Outcome    | P    | $X$ | $Y$ |
|------------|------|-----|-----|
| $\omega_1$ | 0.10 | -10 | 20  |
| $\omega_2$ | 0.30 | -5  | 10  |
| $\omega_3$ | 0.40 | 5   | 0   |
| $\omega_4$ | 0.20 | 10  | 0   |

---

<sup>2</sup>From the definition of covariance and the linearity of expectation we have that

$$\begin{aligned}\text{Cov}(X, Y) &= E(X - E(X))(Y - E(Y)) \\ &= E[X(Y - E(Y))] - E[E(X)(Y - E(Y))] \\ &= E(XY) - E[X E(Y)] - E(X) E(Y - E(Y)) \\ &= E(XY) - E(X) E(Y).\end{aligned}$$

We have that the probability mass function of  $X$  is

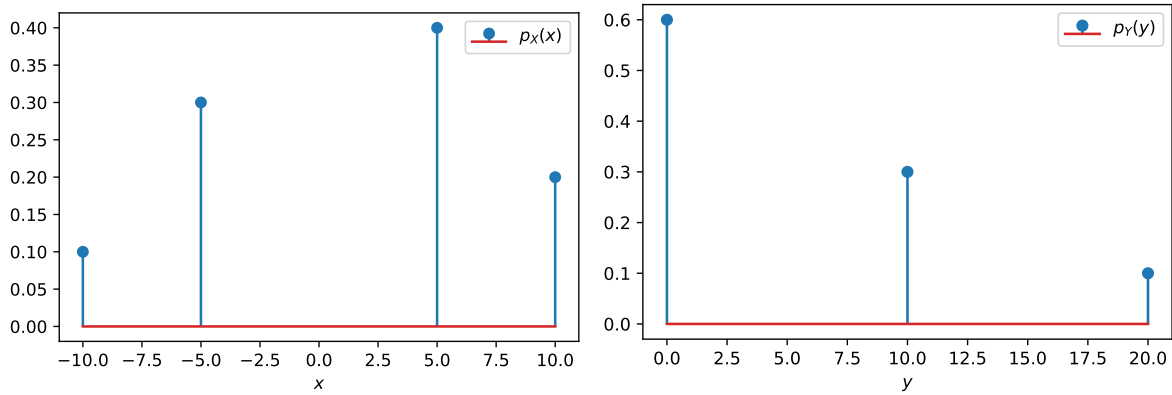
$$p_X(x) = \begin{cases} 0.10 & \text{if } x = -10, \\ 0.30 & \text{if } x = -5, \\ 0.40 & \text{if } x = 5, \\ 0.20 & \text{if } x = 10. \end{cases}$$

The probability mass function of  $Y$  takes only three different values for  $Y$ .

$$p_Y(y) = \begin{cases} 0.60 & \text{if } y = 0, \\ 0.30 & \text{if } y = 10, \\ 0.10 & \text{if } y = 20. \end{cases}$$

□

It is sometimes easier to visualize the probability mass function by plotting the probability of different values of the random variable.



- (a)** The function  $p_X(x)$  defines the probability of  $X$  being equal to  $x = \{-10, -5, 5, 10\}$ . **(b)** The function  $p_Y(y)$  defines the probability of  $Y$  being equal to  $y = \{0, 10, 20\}$ .

**Figure 1:** The figure plots the probability mass function of the random variables  $X$  and  $Y$ .

It is apparent from the pictures that  $p_X(x) = 0$  if  $x \notin \{-10, -5, 5, 10\}$ . Indeed, the set  $\{\omega \in \Omega : X(\omega) = x\}$  is empty for all  $x$  not equal to  $-10, -5, 5$ , or  $10$ . Similarly,  $p_Y(y) = 0$  if  $y \notin \{0, 10, 20\}$ .

To simplify notation, we will often write  $\{X = x\}$  to denote the set  $\{\omega \in \Omega : X(\omega) = x\}$ . Using this notation, we have that  $p_X(x) = P(X = x)$ .

The **support** of  $X$  is the set

$$R_X = \{x \in \mathbb{R} : p_X(x) > 0\},$$

that is, the values  $X$  can actually take with positive probability. Since  $\Omega$  is finite,  $R_X$  is finite as well, say  $R_X = \{x_1, x_2, \dots, x_m\}$ . Using this notation we can re-write the expectation of  $X$  as

$$E(X) = \sum_{x \in R_X} x p_X(x) = \sum_{i=1}^m x_i p_X(x_i), \quad (2)$$

which is commonly used in statistics. Similarly, the variance of  $X$  becomes

$$V(X) = \sum_{x \in R_X} (x - E(X))^2 p_X(x).$$

For two random variables  $X$  and  $Y$  defined in  $(\Omega, P)$ , the set  $\{X = x, Y = y\}$  denotes all outcomes in  $\Omega$  that satisfy  $\{X = x\}$  and  $\{Y = y\}$ . Therefore, we have that

$$\{X = x, Y = y\} = \{X = x\} \cap \{Y = y\}.$$

The function

$$p_{X,Y}(x, y) = P(X = x, Y = y)$$

is called the joint probability mass function of  $X$  and  $Y$ . Similarly to  $R_X$ , the support of  $Y$  is  $R_Y = \{y \in \mathbb{R} : p_Y(y) > 0\}$ , which we write as  $R_Y = \{y_1, y_2, \dots, y_n\}$ .

**Example 6.** The joint pmf of the random variables defined in Example 5 is given in the table below.

| $X \setminus Y$ | 0   | 10  | 20  |
|-----------------|-----|-----|-----|
| -10             | 0   | 0   | 0.1 |
| -5              | 0   | 0.3 | 0   |
| 5               | 0.4 | 0   | 0   |
| 10              | 0.2 | 0   | 0   |

The function  $p_{X,Y}(x, y)$  has many zeros since in Example 5 there are only four outcomes. Any other outcome then has probability zero of occurring.  $\square$

**Example 7.** We can generate any joint pmf for two random variables as long as the sum of all probabilities is equal to one. The table below reports the joint probabilities of a random variable  $X$  taking values in  $[-1, 0, 1]$  and a random variable  $Y$  taking values in  $[0, 1, 2, 3]$ .

| $X \setminus Y$ | 0       | 1       | 2       | 3       |
|-----------------|---------|---------|---------|---------|
| -1              | 0.12500 | 0.09375 | 0.06250 | 0.03125 |
| 0               | 0.06250 | 0.12500 | 0.12500 | 0.06250 |
| 1               | 0.03125 | 0.06250 | 0.09375 | 0.12500 |

In this case the underlying probability space has at least  $3 \times 4 = 12$  possible outcomes. The figure below plots the joint pmf of  $X$  and  $Y$ .  $\square$

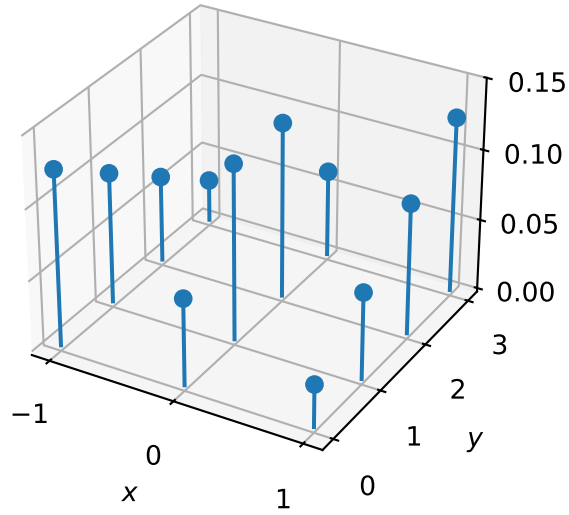
To plot the joint pmf of two random variables we need a three-dimensional graph.

We can use the joint pmf to compute the expectation of a function of two random variables. Indeed, we have that

$$E(f(X, Y)) = \sum_{i=1}^m \sum_{j=1}^n f(x_i, y_j) p_{X,Y}(x_i, y_j). \quad (3)$$

If we write  $\mu_X = E(X)$  and  $\mu_Y = E(Y)$ , equation (3) implies that the covariance of  $X$  and  $Y$  can be computed as

$$\text{Cov}(X, Y) = \sum_{i=1}^m \sum_{j=1}^n (x_i - \mu_X)(y_j - \mu_Y) p_{X,Y}(x_i, y_j).$$



**Figure 2:** The figure plots the joint probability mass function of  $X$  and  $Y$  in Example 7.

The joint pmf contains all the information of  $X$  and  $Y$  since we can recover the individual pmfs of  $X$  and  $Y$  from it. Indeed, we have that

$$p_X(x) = \sum_{j=1}^n p_{X,Y}(x, y_j)$$

and

$$p_Y(y) = \sum_{i=1}^m p_{X,Y}(x_i, y).$$

It is important to note that the joint pmf not only contains the individual information of two random variables but also captures their mutual dependence.

## Independence

Independence formalizes the idea that learning about one random variable tells you nothing about another. As we will see in this section, independence is a notion that is stronger than zero covariance between two random variables.

We say that two events  $A$  and  $B$  are **independent** if  $P(A \cap B) = P(A) P(B)$ .

**Example 8.** Suppose that the weather tomorrow can be either sunny, fair or rainy. In addition, a certain stock tomorrow can either go up or down in price.

We can define

$$W = \{\text{sunny, fair, rainy}\}$$

and

$$S = \{\text{up, down}\}.$$

The set of outcomes can be described as all possible pairwise combinations of weather tomorrow and the stock price movement. The sample space  $\Omega$  is then the cartesian product of  $W$  and  $S$ , i.e.,  $\Omega = W \times S$ .

We can then define the weather events

$$\text{Sunny} = \{(\text{sunny, up}), (\text{sunny, down})\},$$

$$\text{Fair} = \{(\text{fair, up}), (\text{fair, down})\},$$

$$\text{Rainy} = \{(\text{rainy, up}), (\text{rainy, down})\}.$$

The table below describes the probabilities for tomorrow's weather.

| Weather     | Sunny | Fair | Rainy |
|-------------|-------|------|-------|
| Probability | 0.3   | 0.5  | 0.2   |

Similarly, the stock events can be defined as

$$\text{Up} = \{(\text{sunny, up}), (\text{fair, up}), (\text{rainy, up})\},$$

$$\text{Down} = \{(\text{sunny, down}), (\text{fair, down}), (\text{rainy, down})\}.$$

The probabilities of the stock price going up or down are described in the table below.

| Stock       | Up  | Down |
|-------------|-----|------|
| Probability | 0.6 | 0.4  |

If the weather does not affect the likelihood of the stock going up or down, we should expect to see on sunny days 60% of the time the stock going up and 40% of those days the stock going down.

That is, if the weather tomorrow and the stock price movement are independent events, we should expect

$$P(\text{Stock} \cap \text{Weather}) = P(\text{Stock}) P(\text{Weather}),$$

where Stock is either Up or Down, and Weather is either Sunny, Fair, or Rainy.

The table below describes the combined probabilities of the stock price movement and the weather tomorrow that are consistent with the independence of those events.

| Stock\Weather | Sunny | Fair | Rain |
|---------------|-------|------|------|
| Up            | 0.18  | 0.30 | 0.12 |
| Down          | 0.12  | 0.20 | 0.08 |

In the table, the weather does not change the relative proportions of the probabilities for the stock price. □

The previous example shows how to generate independent events out of two finite probability spaces  $(\Omega_1, P_1)$  and  $(\Omega_2, P_2)$ . If we define  $\Omega = \Omega_1 \times \Omega_2$  and let  $P(\omega_1, \omega_2) = P_1(\omega_1) P_2(\omega_2)$  for each  $\omega_1 \in \Omega_1$  and  $\omega_2 \in \Omega_2$ , the pair  $(\Omega, P)$  is a well-defined finite probability space. In this new probability space, the events  $A = \{\omega_1\} \times \Omega_2$  and  $B = \Omega_1 \times \{\omega_2\}$  are independent for any

$\omega_1 \in \Omega_1$  and  $\omega_2 \in \Omega_2$ .<sup>3</sup>

**Example 9.** The sample space  $\Omega$  is always independent from any event  $A \subset \Omega$  since  $P(A \cap \Omega) = P(A) = P(A)P(\Omega)$ . Intuitively, an outcome always happens independently of whether  $A$  happens or not.  $\square$

Two random variables  $X$  and  $Y$  are independent if the events  $\{X = x\}$  and  $\{Y = y\}$  are independent. Thus, if  $X$  and  $Y$  are independent we have that

$$P(X = x, Y = y) = P(X = x) P(Y = y),$$

or equivalently

$$p_{X,Y}(x, y) = p_X(x)p_Y(y).$$

An important consequence of independence is that if  $X$  and  $Y$  are two independent random variables, then<sup>4</sup>

$$E(XY) = E(X) E(Y). \quad (4)$$

---

<sup>3</sup>To prove this statement, we need to show that  $P(A \cap B) = P(A)P(B)$ . Indeed, we have that

$$\begin{aligned} P(A) &= \sum_{\omega_2 \in \Omega_2} P(\omega_1, \omega_2) \\ &= \sum_{\omega_2 \in \Omega_2} P_1(\omega_1) P_2(\omega_2) \\ &= P_1(\omega_1) \sum_{\omega_2 \in \Omega_2} P_2(\omega_2) \\ &= P_1(\omega_1). \end{aligned}$$

Similarly,  $P(B) = P_2(\omega_2)$ . Since  $A \cap B = \{(\omega_1, \omega_2)\}$ , we have that  $P(A \cap B) = P(A)P(B)$ , proving that  $A$  and  $B$  are independent.

<sup>4</sup>If the domains of  $X$  and  $Y$  are  $\{x_1, x_2, \dots, x_m\}$  and  $\{y_1, y_2, \dots, y_n\}$ , respectively, we can then write

$$\begin{aligned} E(X) &= \sum_{i=1}^m x_i p_X(x_i), \\ E(Y) &= \sum_{j=1}^n y_j p_Y(y_j). \end{aligned}$$

Equation (4) implies that if  $X$  and  $Y$  are independent, their covariance is equal to zero. Indeed,

$$\begin{aligned}\text{Cov}(X, Y) &= E(XY) - E(X) E(Y) \\ &= E(X) E(Y) - E(X) E(Y) \\ &= 0.\end{aligned}$$

However, the opposite statement is not true.

**Example 10.** Consider two random variables  $X$  and  $Y$  defined in the table below.

| Outcome    | P    | $X$ | $Y$ |
|------------|------|-----|-----|
| $\omega_1$ | 0.40 | -1  | 0   |
| $\omega_2$ | 0.30 | 1   | 1   |
| $\omega_3$ | 0.30 | 1   | -1  |

We have that

$$\begin{aligned}E(X) &= 0.4 \times (-1) + 0.6 \times 1 = 0.2, \\ E(Y) &= 0.4 \times 0 + 0.3 \times 1 + 0.3 \times (-1) = 0, \\ E(XY) &= 0.4 \times (-1) \times 0 + 0.3 \times 1 \times 1 + 0.3 \times 1 \times (-1) = 0.\end{aligned}$$

Therefore,  $\text{Cov}(X, Y) = 0 - 0.2 \times 0 = 0$ , which shows that  $X$  and  $Y$  are uncorrelated.

However,  $X$  and  $Y$  are not independent. Independence requires  $p_{X,Y}(x, y) = p_X(x)p_Y(y)$  for every pair  $(x, y)$ , but this already fails at  $(x, y) = (-1, 1)$ : both  $p_X(-1) = 0.4$  and  $p_Y(1) =$

Thus,

$$\begin{aligned}E(X) E(Y) &= \left( \sum_{i=1}^m x_i p_X(x_i) \right) \left( \sum_{j=1}^n y_j p_Y(y_j) \right) \\ &= \sum_{i=1}^m \sum_{j=1}^n x_i y_j p_X(x_i) p_Y(y_j) \\ &= \sum_{i=1}^m \sum_{j=1}^n x_i y_j p_{X,Y}(x_i, y_j) \\ &= E(XY).\end{aligned}$$

0.3 are positive, so independence would require

$$p_{X,Y}(-1, 1) = p_X(-1)p_Y(1) = 0.4 \times 0.3 = 0.12,$$

yet no outcome has  $X = -1$  and  $Y = 1$  at the same time, so  $p_{X,Y}(-1, 1) = 0$  instead. Intuitively, this happens because  $Y$  reveals  $X$  exactly:  $Y = 0$  tells us  $X = -1$ , while  $Y = 1$  or  $Y = -1$  each tell us  $X = 1$ . Once we know  $Y = 1$ , the possibility  $X = -1$  is ruled out entirely, and this could not happen if  $X$  and  $Y$  were independent, since then the value of  $Y$  would never change what we think  $X$  can be.  $\square$

The previous example is a special case of the following fact: suppose there is a function  $g : R_Y \rightarrow \mathbb{R}$  such that

$$p_{X,Y}(x, y) > 0 \implies x = g(y)$$

for every  $x \in R_X$  and  $y \in R_Y$  (that is, whenever  $X$  and  $Y$  can jointly take the values  $x$  and  $y$ , we must have  $x = g(y)$ ). Then  $X$  and  $Y$  are independent if and only if  $R_X$  contains a single element, i.e.,  $X$  is constant.<sup>5</sup>

The previous result suggests a general *necessary* condition for independence. If  $X$  and  $Y$  are independent, then  $p_{X,Y}(x, y) = p_X(x) p_Y(y)$  for every  $x \in R_X$  and  $y \in R_Y$ . Since  $p_X(x) > 0$  and  $p_Y(y) > 0$  for every such  $x$  and  $y$ , independence forces  $p_{X,Y}(x, y) > 0$  as well. In other words, the joint pmf must have *full support* on the entire rectangle  $R_X \times R_Y$ : every combination of values that  $X$  and  $Y$  could individually take must also be jointly possible.

This condition is necessary, but not sufficient. Full support only rules out the failure mode from the result above, where some combinations are forbidden outright. It says nothing about whether the probabilities on the combinations that *are* possible are the right ones: independence further requires  $p_{X,Y}(x, y) = p_X(x) p_Y(y)$  for every pair. A joint pmf can have full support on

---

<sup>5</sup>( $\Leftarrow$ ) If  $R_X = \{c\}$ , then both sides of  $p_{X,Y}(x, y) = p_X(x)p_Y(y)$  equal  $p_Y(y)$  when  $x = c$  and 0 otherwise, so the equality holds trivially and  $X$  and  $Y$  are independent.

( $\Rightarrow$ ) Suppose instead that  $X$  and  $Y$  are independent, and fix any  $y \in R_Y$ , so that  $p_Y(y) > 0$ . Since  $p_{X,Y}(x, y) = 0$  for every  $x \neq g(y)$ , independence gives

$$p_X(x) p_Y(y) = 0 \quad \text{for every } x \neq g(y).$$

Because  $p_Y(y) > 0$ , this forces  $p_X(x) = 0$  whenever  $x \neq g(y)$ , i.e.,  $R_X = \{g(y)\}$ . Since this holds for every  $y \in R_Y$ , we conclude that  $R_X$  is the single point  $\{g(y)\}$ , i.e.,  $X$  is constant.

$R_X \times R_Y$  and still violate this equality, in which case  $X$  and  $Y$  remain dependent even though no combination is ruled out.

In Example 10,  $X$  is a (non-injective) function of  $Y$ :  $X = -1$  whenever  $Y = 0$ , and  $X = 1$  whenever  $Y = 1$  or  $Y = -1$ . Since  $X$  is not constant, the previous result tells us  $X$  and  $Y$  must be dependent, confirming what we already found by direct computation. Yet  $\text{Cov}(X, Y) = 0$ , which is why zero covariance, or equivalently zero correlation, never implies that two random variables are unrelated.

By contrast, in Example 8 neither variable is a function of the other: Weather = Sunny occurs together with both Stock = Up and Stock = Down, and symmetrically, Stock = Up occurs together with all three weather outcomes. So the previous result, which only pins down dependence when one variable determines the other, does not rule out independence. Indeed, the two variables were explicitly constructed to be independent via the product-measure recipe described earlier in this section: assigning  $P(w, s) = P_1(w) P_2(s)$  to every pair  $(w, s) \in W \times S$  is precisely what gives the joint pmf full support on  $W \times S$  in the first place.

## Linear Combinations

In investment theory, we usually study linear combinations of random variables of the form  $Z = \alpha X + \beta Y$ . The expectation of  $Z$  is just a linear combination of the expectations of  $X$  and  $Y$ ,<sup>6</sup>

$$E(Z) = \alpha E(X) + \beta E(Y). \quad (5)$$

---

<sup>6</sup>The expectation of  $Z$  is computed as

$$\begin{aligned} E(Z) &= E(\alpha X + \beta Y) \\ &= \sum_{\omega \in \Omega} (\alpha X(\omega) + \beta Y(\omega)) P(\omega) \\ &= \alpha \sum_{\omega \in \Omega} X(\omega) P(\omega) + \beta \sum_{\omega \in \Omega} Y(\omega) P(\omega) \\ &= \alpha E(X) + \beta E(Y). \end{aligned}$$

The variance of  $Z$ , though, includes not only the variances of  $X$  and  $Y$  but also their covariances,<sup>7</sup>

$$V(Z) = \alpha^2 V(X) + \beta^2 V(Y) + 2\alpha\beta \text{Cov}(X, Y). \quad (6)$$

This is an important result which is at the heart of portfolio diversification.

**Example 11.** Suppose  $X$  and  $Y$  are the returns on two stocks with  $V(X) = 0.04$ ,  $V(Y) = 0.09$ , and  $\text{Cov}(X, Y) = -0.01$ . Consider the equally weighted portfolio  $Z = 0.5X + 0.5Y$ . By (6),

$$V(Z) = (0.5)^2(0.04) + (0.5)^2(0.09) + 2(0.5)(0.5)(-0.01) = 0.01 + 0.0225 - 0.005 = 0.0275.$$

Notice that  $V(Z) = 0.0275$  is smaller than the weighted average of the two variances,  $0.5(0.04) + 0.5(0.09) = 0.065$ : combining the two stocks lowers risk below what either stock offers on its own, precisely because their negative covariance means losses in one tend to be offset by gains in the other. This is the mathematical content behind portfolio diversification.  $\square$

More generally, consider the random variables  $X_1, X_2, \dots, X_n$ , and form a new random variable  $X$  such that

$$X = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n,$$

where  $\alpha_i \in \mathbb{R}$  for all  $i \in \{1, 2, \dots, n\}$ .

The expectation of  $X$  is a linear combination of the expectations of  $X_1, X_2, \dots, X_n$ . The variance of  $X$ , though, takes into account of all covariances between  $X_i$  and  $X_j$ , for  $i, j = 1, 2, \dots, n$ .

---

<sup>7</sup>The variance of  $Z$  is computed as

$$\begin{aligned} V(Z) &= V(\alpha X + \beta Y) \\ &= E(\alpha X + \beta Y - E(\alpha X + \beta Y))^2 \\ &= E(\alpha(X - E(X)) + \beta(Y - E(Y)))^2 \\ &= E(\alpha^2(X - E(X))^2 + \beta^2(Y - E(Y))^2 \\ &\quad + 2\alpha\beta(X - E(X))(Y - E(Y))) \\ &= \alpha^2 E(X - E(X))^2 + \beta^2 E(Y - E(Y))^2 \\ &\quad + 2\alpha\beta E(X - E(X))(Y - E(Y)) \\ &= \alpha^2 V(X) + \beta^2 V(Y) + 2\alpha\beta \text{Cov}(X, Y). \end{aligned}$$

Indeed, we have that

$$V(X) = \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \text{Cov}(X_i, X_j). \quad (7)$$

The previous expression can be simplified if the random variables  $X_1, X_2, \dots, X_n$  are independent from each other. In such case, we have that  $\text{Cov}(X_i, X_j) = 0$  for all  $i \neq j$ . Recognizing that  $\text{Cov}(X_i, X_i) = V(X_i)$ , equation (7) implies that

$$V(X) = \sum_{i=1}^n \alpha_i^2 V(X_i). \quad (8)$$

**Example 12.** Suppose that  $X_1, X_2, \dots, X_n$  are independent random variables with the same variance denoted by  $\sigma^2$ . Define  $X$  to be the sum of these random variables so that

$$X = X_1 + X_2 + \dots + X_n.$$

Equation (8) implies that

$$V(X) = \sum_{i=1}^n V(X_i) = n\sigma^2.$$

□

Keeping track of  $n$  weights and  $n(n+1)/2$  distinct covariances quickly becomes cumbersome as  $n$  grows. Matrix notation lets us express the expectation and variance of  $X$  compactly, regardless of  $n$ . To this end, it is convenient to collect the random variables  $X_1, X_2, \dots, X_n$  and the weights  $\alpha_1, \alpha_2, \dots, \alpha_n$  into vectors

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \quad \boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix},$$

so that the linear combination  $X = \alpha_1 X_1 + \dots + \alpha_n X_n$  can be written compactly as  $X = \boldsymbol{\alpha}^\top \mathbf{X}$ .

The expectation of  $\mathbf{X}$  is simply the vector of expectations of each component,

$$\mathbb{E}(\mathbf{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \\ \vdots \\ \mathbb{E}(X_n) \end{pmatrix},$$

and (5) generalizes to  $\mathbb{E}(X) = \boldsymbol{\alpha}^\top \mathbb{E}(\mathbf{X})$ .

The **covariance matrix** of  $\mathbf{X}$ , denoted  $\boldsymbol{\Sigma}$ , is the  $n \times n$  matrix whose  $(i, j)$  entry is  $\text{Cov}(X_i, X_j)$ ,

$$\boldsymbol{\Sigma} = \begin{pmatrix} V(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & V(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & V(X_n) \end{pmatrix}.$$

Since  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$ , the matrix  $\boldsymbol{\Sigma}$  is symmetric. Using  $\boldsymbol{\Sigma}$ , equation (7) can be written as the quadratic form

$$V(X) = \boldsymbol{\alpha}^\top \boldsymbol{\Sigma} \boldsymbol{\alpha}. \quad (9)$$

**Example 13.** Return to the two stocks  $X$  and  $Y$  from Example 11, and collect them into  $\mathbf{X} = (X, Y)^\top$  with weights  $\boldsymbol{\alpha} = (0.5, 0.5)^\top$ . Their covariance matrix is

$$\boldsymbol{\Sigma} = \begin{pmatrix} 0.04 & -0.01 \\ -0.01 & 0.09 \end{pmatrix}.$$

Using (9),

$$V(Z) = \boldsymbol{\alpha}^\top \boldsymbol{\Sigma} \boldsymbol{\alpha} = \begin{pmatrix} 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} 0.04 & -0.01 \\ -0.01 & 0.09 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} = 0.0275,$$

matching the value computed directly in Example 11. The matrix formulation gives the same number, but scales to any number of assets without writing out a double sum.  $\square$

The left-hand side of (9) is the variance of  $X = \boldsymbol{\alpha}^\top \mathbf{X}$ , which can never be negative, regardless of

$\alpha$ . Hence

$$\alpha^T \Sigma \alpha \geq 0 \quad \text{for every } \alpha \in \mathbb{R}^n.$$

A symmetric matrix with this property is called **positive semidefinite**. Thus, every covariance matrix is symmetric and positive semidefinite. If in addition no nonzero linear combination  $\alpha^T X$  has zero variance, then the inequality above is strict for every  $\alpha \neq 0$ , and  $\Sigma$  is **positive definite**.<sup>8</sup>

## Practice Problems

These problems give you a chance to practice the concepts introduced in this chapter. Try to solve each one on your own before expanding the solution.

Solutions to all problems can be found at <https://lorenzonaranjo.com/class-materials/investments/>.

**Problem 1** (Sets). Let  $U = \{1, 2, \dots, 20\}$  be the universe,  $A = \{n \in U : n \text{ is even}\}$ , and  $B = \{n \in U : n \leq 12\}$ . Find  $A \cap B$ ,  $A \cup B$ ,  $A \setminus B$ , and  $|\mathcal{P}(A \cap B)|$ .

**Problem 2** (Events and Probability). A sample space has four equally likely outcomes,  $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ . Let  $A = \{\omega_1, \omega_2\}$  and  $B = \{\omega_2, \omega_3\}$ . Compute  $P(A)$ ,  $P(B)$ ,  $P(A \cup B)$ , and  $P(A^C)$ .

**Problem 3** (Expectation and Variance). A stock's annual return  $X$  takes the values  $-4\%$ ,  $2\%$ , and  $6\%$  with probabilities  $0.2$ ,  $0.5$ , and  $0.3$ , respectively. Compute  $E(X)$ ,  $V(X)$ , and  $\sigma_X$ .

**Problem 4** (Covariance). Two random variables  $X$  and  $Y$  are defined on a sample space with four equally likely outcomes, taking the values below.

---

<sup>8</sup>Positive definiteness of  $\Sigma$  is not just a mathematical curiosity. Once we study how to construct portfolios of risky assets, we will minimize  $\alpha^T \Sigma \alpha$  subject to constraints on the weights  $\alpha$ , and positive definiteness is exactly what guarantees that such a minimum exists and is unique.

| Outcome | $\omega_1$ | $\omega_2$ | $\omega_3$ | $\omega_4$ |
|---------|------------|------------|------------|------------|
| $X$     | -2         | -1         | 1          | 2          |
| $Y$     | 4          | 1          | 1          | 4          |

Compute  $E(X)$ ,  $E(Y)$ , and  $\text{Cov}(X, Y)$ . Are  $X$  and  $Y$  independent?

**Problem 5** (Probability Mass Function). Roll two fair six-sided dice and let  $X$  be the sum of the two dice. What is  $p_X(7)$ ? What is  $p_X(2)$ ?

**Problem 6** (Independence). Suppose that whether it rains tomorrow is independent of whether a particular stock goes up. It rains with probability 0.3, and the stock goes up with probability 0.55. What is the probability that it rains **and** the stock goes up? What is the probability that it does **not** rain **and** the stock goes up?

**Problem 7** (Linear Combinations and Diversification). You invest 50% of your wealth in asset  $A$  and 50% in asset  $B$ . Asset  $A$  has expected return 8% and standard deviation 18%; asset  $B$  has expected return 14% and standard deviation 30%. The correlation between their returns is  $-0.2$ . Compute the expected return and standard deviation of your portfolio.