Probability Basics

These notes introduce basic probability concepts crucial to understanding the modern investment theory approach. Investment theory aims to understand how to allocate resources to different assets whose future payoffs are uncertain. To model the future uncertainty of prices and cash flows, we can rely on well-established mathematical concepts that summarize the expected rewards and risks of investing in a portfolio of financial assets.

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.

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 of 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 answer if 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

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

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

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

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

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

Similarly,
$$B^C = \{0, 1, 2, 3, 4, 5\} \cup \{21, 22, 23, \ldots\}.$$

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 $|A| = |B| = 10$ and $|C| = 3$. 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

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

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 $\mathcal{P}(B)$ contain each $2^{10} = 1024$ 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 Ω. The sample space includes all possible outcomes that can occur. A *probability measure* is a function that assigns to each element ω of Ω a number in $[0, 1]$ so that

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

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

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

Such a finite probability space is denoted by (Ω, P).

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

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 Ω. For example, we have that $P(\{\omega_1, \omega_3\}) = 1/2 + 1/4 = 3/4$. □

Random Variables

Definition

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

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
ω_1	-10	20	15
ω_2	-5	10	-10
ω_3	5	0	15
ω_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 ω_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 ω_3 or ω_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 with all the information generated by the probability space Ω. ☐

Expectation and Variance

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

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

whereas the **variance** of X is

$$\mathsf{V}(X) = \mathsf{E}(X - \mathsf{E}\,X)^2.$$

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

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

Outcome	Probability	X	Y
ω_1	1/2	10	2
ω_2	1/4	8	40
ω_3	1/4	4	20

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

$$\mathsf{E}\,X = \frac{1}{2} \times 10 + \frac{1}{4} \times 8 + \frac{1}{4} \times 4 = 8,$$
$$\mathsf{E}\,Y = \frac{1}{2} \times 2 + \frac{1}{4} \times 40 + \frac{1}{4} \times 20 = 16.$$

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

$$V(X) = \frac{1}{2} \times (10-8)^2 + \frac{1}{4} \times (8-8)^2 + \frac{1}{4} \times (4-8)^2 = 6,$$

$$V(Y) = \frac{1}{2} \times (2-16)^2 + \frac{1}{4} \times (40-16)^2 + \frac{1}{4} \times (20-16)^2 = 246.$$

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. $\qquad\square$

Covariance

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

$$\mathrm{Cov}(X,Y) = E(X - EX)(Y - EY),$$

and their **correlation** is

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

The correlation between any two random variables is always between -1 and 1.

Proof

Let σ_X and σ_Y denote the standard deviations of X and Y, respectively. We can then compute

$$E((X - EX)\sigma_Y + (Y - EY)\sigma_X)^2 = (\sigma_X^2 \sigma_Y^2 + 2\sigma_X \sigma_Y \, \mathrm{Cov}(X,Y) + \sigma_Y^2 \sigma_X^2)$$
$$= 2\sigma_X \sigma_Y (\sigma_X \sigma_Y + \mathrm{Cov}(X,Y)),$$

which implies $\sigma_X \sigma_Y + \mathrm{Cov}(X,Y) \geq 0$ or $-\sigma_X \sigma_Y \leq \mathrm{Cov}(X,Y)$.

Similarly,

$$E((X - EX)\sigma_Y - (Y - EY)\sigma_X)^2 = (\sigma_X^2 \sigma_Y^2 - 2\sigma_X \sigma_Y \, \mathrm{Cov}(X,Y) + \sigma_Y^2 \sigma_X^2)$$
$$= 2\sigma_X \sigma_Y (\sigma_X \sigma_Y - \mathrm{Cov}(X,Y)),$$

which implies $\sigma_X \sigma_Y - \mathrm{Cov}(X,Y) \geq 0$ or $\mathrm{Cov}(X,Y) \leq \sigma_X \sigma_Y$.

Thus, we conclude that
$$-1 \leq \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \leq 1,$$
or equivalently $-1 \leq \rho_{X,Y} \leq 1$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Example 4. Continuing with Example 3, we have that
$$\text{Cov}(X, Y) = \frac{1}{2} \times (10 - 8)(2 - 16) + \frac{1}{4}(8 - 8)(40 - 16) + \frac{1}{4}(4 - 8)(20 - 16) = -18.$$

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

The covariance of X and Y can also be expressed as
$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Proof

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

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

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
ω_1	0.10	-10	20
ω_2	0.30	-5	10
ω_3	0.40	5	0
ω_4	0.20	10	0

We have that the probability mass function of X is

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

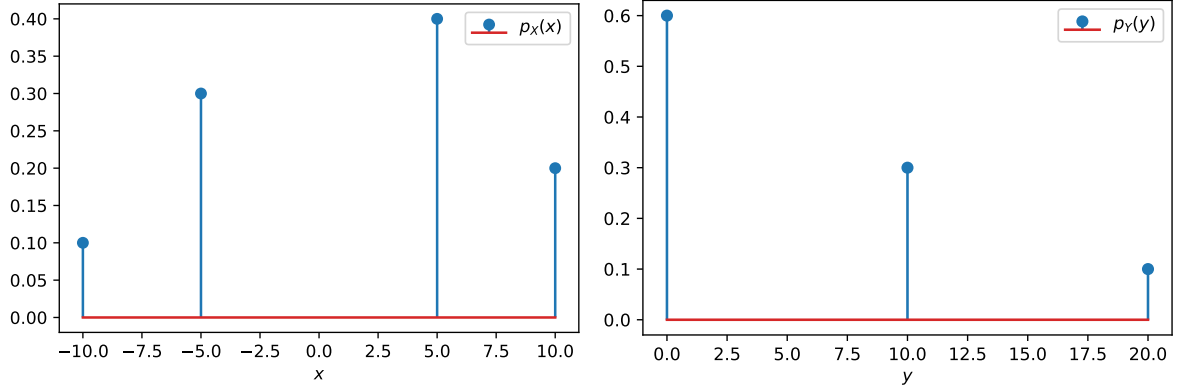The probability mass function of Y only loads on three different values for Y.

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

$\square$

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

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 $p_X(x) = \text{P}(X = x)$.

(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.

If a random variable is defined for m different values of x, we can re-write the expectation of a random variable as

$$\mathsf{E}(X) = \sum_{i=1}^{m} x_i p_X(x_i), \tag{2}$$

which is commonly used in statistics.

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

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

The function

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

is called the joint probability mass function of X and Y.

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

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

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

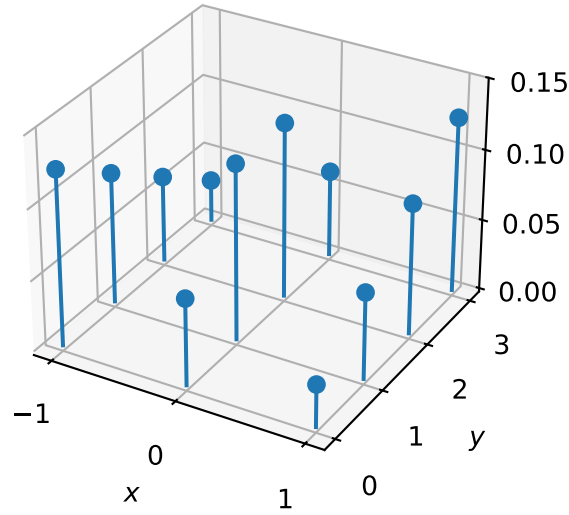$$\text{Cov}(X, Y) = \sum_{x=1}^{m} \sum_{j=1}^{n} (x_i - \mu_X)(y_j - \mu_Y) p_{X,Y}(x_i, y_j).$$

11

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

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.

12

We can define

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

and

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

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

We can then define the weather events

$$\text{Sunny} = \{(\text{sunny}, \text{up}), (\text{sunny}, \text{down})\},$$
$$\text{Fair} = \{(\text{fair}, \text{up}), (\text{fair}, \text{down})\},$$
$$\text{Rainy} = \{(\text{rainy}, \text{up}), (\text{rainy}, \text{down})\}.$$

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

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

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 (Ω_1, P_1) and (Ω_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 (Ω, 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 $\omega_2 \in \Omega_2$.

Proof

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, $P(B) = P_2(w_2)$. 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. □

Example 9. The sample space Ω 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. □

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

$$E(XY) = E(X) E(Y). \tag{4}$$

Proof

If the domains of X and Y are $\{x_1, x_2, \ldots, x_m\}$ and $\{y_1, y_2, \ldots, y_n\}$, respectively, we can then write

$$E(X) = \sum_{i=1}^{m} x_i p_X(x_i),$$

$$E(Y) = \sum_{j=1}^{n} y_j p_Y(y_j).$$

Thus,

$$E(X)E(Y) = \left(\sum_{i=1}^{m} x_i p_X(x_i)\right)\left(\sum_{j=1}^{n} y_j p_Y(y_j)\right)$$

$$= \sum_{i=1}^{m}\sum_{j=1}^{n} x_i y_j p_X(x_i) p_Y(y_j)$$

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

$$= E(XY).$$

☐

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

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

$$= E(X)E(Y) - E(X)E(Y)$$

$$= 0.$$

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
ω_1	0.40	-1	0
ω_2	0.30	1	1
ω_3	0.30	1	-1

We have that

$$E(X) = 0.4 \times (-1) + 0.6 \times 1 = 0.2,$$

$$E(Y) = 0.4 \times 0 + 0.3 \times 1 + 0.3 \times (-1) = 0,$$

$$E(XY) = 0.4 \times (-1) \times 0 + 0.3 \times 1 \times 1 + 0.3 \times 1 \times (-1) = 0.$$

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

However, the two random variables are not independent. If we know that $X = -1$ then we know that $Y = 0$. Similarly, learning that $Y = 1$ tells us that $X = 1$. $\qquad\square$

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,

$$\mathsf{E}\,Z = \alpha\,\mathsf{E}\,X + \beta\,\mathsf{E}\,Y. \tag{5}$$

The variance of Z, though, includes not only the variances of X and Y but also their covariances,

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

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

Proof

The expectation of Z is computed as

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

The variance of Z is computed as

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

$\square$

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

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

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

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

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

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

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

Example 11. Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables with the same variance denoted by σ^2. Define X to be the sum of these random variables so that

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

Equation (8) implies that

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

$\square$