

Advanced Statistics - Combinations of Random variables

Ralf Becker

January 24, 2023

1 Introduction

We have, thus far, developed the concept of probability distributions for single continuous or discrete random variables. As we have seen, these distributions assign probabilities to various numerical values of the random variable under consideration. Further, for the distributions most commonly used in practice, mathematical expressions can be obtained for the mean and variance of this random variable in terms of the parameters of the distribution. For example, if X is the number of successes in n independent experiments in which the probability of success is π (that is, X follows a Binomial distribution with parameters n and π), then it was stated that $E[X] = n\pi$ and $Var(X) = n\pi(1 - \pi)$. Clearly, this mean and variance will change with both n and π .

The next section will develop such properties further. This is done first in the context of a single variable, before later examining the extensions of these ideas that are required when we jointly consider two or more variables.

2 Linear Transformations

Before extending the ideas, recall the definitions of mean and variance as expected values.

2.1 Means and variances: a recap

Recall that the mean, or the expected value of a random variable X , is commonly denoted as either $E[X]$ or μ . For a discrete random variable, the mean is defined as

$$E[X] = \mu = \sum_x xp(x).$$

If the variable is continuous, the analogous definition is

$$E[X] = \mu = \int_{-\infty}^{\infty} xf(x)dx$$

The variance of a random variable, which provides a measure of the spread of the possible outcomes around the mean, is also defined as an expected value, namely:

$$\begin{aligned} \text{var}[X] &= E[(X - \mu)^2] \\ &= \sum_x (x - \mu)^2 p(x) \text{ discrete} \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \text{ continuous} \end{aligned}$$

However, for the distributions which were introduced (e.g. the binomial distribution or the normal distribution) we also learned how the mean and variance were related to the parameters that described these distributions.

2.2 Linear transformations of random variables

Often, information is available about a random variable in one form, but we require this to be converted into another form via a linear transformation. Changing units of measurement provides an important example of this: prices may be expressed in £ and converted to Euros at a specific exchange rate, or temperatures can be expressed in fahrenheit or celsius.

In fact we already encountered a linear transformation. That from a normally distributed random variable X (where $X \sim N(\mu, \sigma^2)$) to a standard normal distribution Z (where $Z \sim N(0, 1)$). The way we did transform this was according to

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$$

the last term illustrating that the transformation is a linear one with a multiplicative factor $\frac{1}{\sigma}$ (like a slope coefficient) and a shift factor $-\frac{\mu}{\sigma}$ (like a constant). Let's learn from that example how the properties of a linear transformation of a rv are related to those of the original rv.

Let's consider the following linear transformation

$$Z = c + bX$$

where X is a random variable with $E[X] = \mu$ and $Var[X] = \sigma^2$. c and b are constants. Z is what we call a linear transformation of X .

We have the following results:

1. $E[Z] = E[bX + c] = bE[X] + c$
2. $Var[Z] = Var[bX + c] = b^2Var[X]$

For the discrete random variable case the algebra of the proof is very straightforward:

$$\begin{aligned}
 E[bX + c] &= \sum_x (bx + c)p(x) \\
 &= \sum_x (bxp(x) + cp(x)) \\
 &= b \sum_x xp(x) + c \sum_x p(x) \\
 &= bE[X] + c,
 \end{aligned}$$

since $\sum_x p(x) = 1$ (i.e. all probabilities sum to 1) and $\sum_x xp(x) = E[X]$, by definition of the expected value for a discrete random variable.

We learn that any multiplicative constant, like b here, can be moved outside of the expectations operator, $E[\]$. This applies also to continuous variables, because corresponding operations carry over when integrals replace the summations above.

To calculate the variance of a linearly transformed rv we can ignore any level shift (the addition of c) and we can draw the multiplicative constant (b) outside of the variance operator, $Var[\]$, but need to square it. Here we will not present any proof for that result. It can be derived using similar, but more involved arguments, as for the expected value.

These relationships between the original rv X and its linear transform Z are generally valid, regardless of how the variable X is distributed.

Example. *Let us confirm that these relationships hold when we transform any normally distributed rv X (where $X \sim N(\mu, \sigma^2)$) to a standard normal distribution Z (where $Z \sim N(0, 1)$). As discussed above, the linear transform used is*

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}$$

So, here $c = -\frac{\mu}{\sigma}$ and $b = \frac{1}{\sigma}$. Let us apply the formulae established above:

$$\begin{aligned}
E[Z] &= E[bX + c] \\
&= bE[X] + c \\
&= \frac{1}{\sigma}E[X] + \left(-\frac{\mu}{\sigma}\right) \\
&= \frac{1}{\sigma}\mu + \left(-\frac{\mu}{\sigma}\right) \\
&= \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0
\end{aligned}$$

So indeed the expected value is $E[Z] = 0$.

$$\begin{aligned}
\text{Var}[Z] &= \text{Var}[bX + c] \\
&= b^2\text{Var}[X] \\
&= \left(\frac{1}{\sigma}\right)^2 \text{Var}[X] \\
&= \left(\frac{1}{\sigma}\right)^2 \sigma^2 \\
&= \frac{\sigma^2}{\sigma^2} = 1
\end{aligned}$$

we have therefore established that $\text{Var}[Z] = 1$, confirming the result we had previously just stated.

There is also another element to complement the above result: namely, if X has the form of a normal random variable, then so has Z . Indeed, this is a very powerful result concerning linear transformations of normal variables, which we simply state:

If $X \sim N(\mu, \sigma^2)$ then $Y = bX + c \sim N(b\mu + c, b^2\sigma^2)$ for constants b and c .

This result that the form of the distribution does not change is specific to the normal distribution: for example, a linear transformation of a Binomial variable is not itself (generally) Binomial.

Example. *Playing a game by rolling fair die. If it costs 10p to play, and the player receives (in pence) double the number thrown, then expected winnings per game are*

$$E[2X - 10] = 2 \times (3.5) - 10 = -3.$$

recognising that the expected value of rolling a fair dice is $E[X] = 3.5$. Hence a player is expected to lose 3p every time they play this game.

Example. You flip a fair coin three times and the rv X records how many times the coin shows HEAD. You convince your friend that he should play a game with the following payoff. Every round (equivalent to three coin flips) will cost £1. He will receive £0.5 for every coin showing HEAD. What is the expected value and the variance of the rv $Y = 0.5 \cdot X - 1$.

What type of random variable is X ?

- normally distributed
- binomially distributed
- poisson distributed
- uniformly distributed

Hint: It is the number of successful bernoulli experiments (success rate $\pi = 0.5$ as it is a fair coin) amongst n trials. That is the definition of a Bernoulli rv.

What are the expected value and variance of X ?

$$E[X] = 3 \cdot 0.5 = 1.5$$

$$Var[X] = 3 \cdot 0.5(1 - 0.5) = 0.75$$

Hint: Use the fact that you know that this is a binomial rv. $E(X) = n\pi$ and $Var(X) = n\pi(1-\pi)$ with $n = 3$ and $\pi = 0.5$.

What are the expected value and variance of Y ?

$$E[Y] = 0.5 \cdot E[X] - 1 = -0.25$$

$$Var[Y] = 0.5^2 \cdot Var[X] = \frac{0.75}{4}$$

3 Joint Probability Distributions

You would have already learned about correlations as a descriptive statistics. We discussed how to use correlation and regression to summarise the extent and nature of any linear relationship between two variables. Examples of variables we used in that earlier section were: Covid case rates and health expenditure for countries; height and weight of people; maths and stats grades for students.

Let's take that last example of students' grades in Advanced Mathematics and Statistics. You can think about this as two random variables, M and S for the maths and stats grade respectively.

The objective of statistics is to learn about population characteristics. An experiment is then any process which generates data. Just imagine we go into a second year class and we ask all students what grades they received in Maths and Stats in Year 1. These two pieces of information, generated by the experiment, are considered to be the values of two random variables (M and S) defined on the sample space of an experiment (see the intro to probability section), then the discussion of random variables and probability distributions needs to be extended to the multivariate case. In the first instance to that of two variables, but eventually also to higher dimensions.

Let X and Y be the two random variables: for simplicity, they are considered to be discrete random variables. The outcome of the experiment is a pair of values (x, y) . The probability of this outcome is a joint probability which can be denoted

$$\Pr(X = x \cap Y = y),$$

emphasising the analogy with the probability of a joint event $\Pr(A \cap B)$, or, more commonly, by

$$\Pr(X = x, Y = y).$$

The collection of these probabilities, for all possible combinations of x and y , is the **joint probability distribution** of X and Y , denoted

$$p(x, y) = \Pr(X = x, Y = y).$$

The **Axioms of Probability** discussed in the Introduction to Probability Section carry over to imply

- any possible event has a non-negative probability (but not larger than 1)

$$0 \leq p(x, y) \leq 1,$$

- The sum of all mutually exclusive (x, y) combinations is calculated as the sum of the individual probabilities

$$\sum_x \sum_y p(x, y) = 1,$$

and delivers a probability of 1. This implies that it is certain that one of the possible outcomes does occur.

Example. Let H and W be the random variables representing the population of weekly incomes of husbands and wives, respectively, in some country. There are only three possible weekly

incomes, £0, £100 or £200. The joint probability distribution of H and W is represented as a table:

Probabilities		H		
		£ 0	£ 100	200
Values of W :	£ 0	0.05	0.15	0.10
	£ 100	0.10	0.10	0.30
	£ 200	0.05	0.05	0.10

Then we can read off, for example, that

$$\Pr(H = 0, W = 0) = 0.05,$$

or that in this population, 5% of husbands and wives have each a zero weekly income.

In this example, the nature of the experiment underlying the population data is not explicitly stated, meaning that all we know are these empirical probabilities. Sometimes we say that we do not know the data generating process.

Example. What are the following probabilities? (where we now, for simplicity, represent £100 by 1).

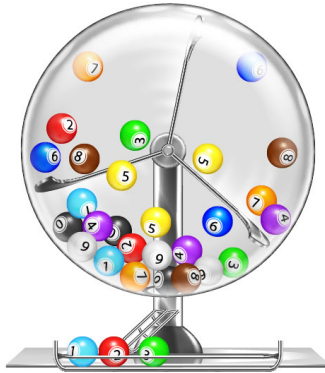
$\Pr(H = 1, W = 0) = 0.15$

$\Pr(H = 2, W = 1) = 0.30$

$\Pr(H = 2, W = 2) = 0.10$

What is the probability that $H = 1$?

$$\Pr(H = 1) = \Pr(H = 1, W = 0) + \Pr(H = 1, W = 1) + \Pr(H = 1, W = 2) = 0.30$$



Example. Consider the following simple version of a lottery. Players in the lottery randomly choose one of 100 tickets which show numbers between 1 and 5, whilst a machine selects the lottery winners by randomly selecting one of five balls (numbered 1 to 5). Any player whose drawn number on a ticket coincides with the number on the ball is a winner. The machine selects one ball at random (so that each ball has an 0.2 chance of selection). The 100 tickets from which the players draw contain the numbers 1 to 5 with the following probabilities:

Number chosen by player	Probability of being chosen
1	0.40
2	0.20
3	0.05
4	0.10
5	0.25
	1.0

Let X denote the number on the player's chosen and Y the number chosen by the machine. If they are assumed to be independent events, then for each possible value of X and Y , we will have

$$\Pr(X \cap Y) = \Pr(X) \Pr(Y).$$

The table above gives the probabilities for X , and $\Pr(Y = y) = 0.2$ for all $y = 1, \dots, 5$, so that a table can be drawn up displaying the joint distribution $p(x, y)$.

For example $p(3, 4) = \Pr(X = 3) \cdot \Pr(Y = 4) = 0.05 \cdot 0.20 = 0.01$.

<i>Probabilities</i>	<i>machine (Y)</i>					<i>Row Total</i>
	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	
<i>player (X)</i>						
<i>1</i>	<i>0.08</i>	<i>0.08</i>	<i>0.08</i>	<i>0.08</i>	<i>0.08</i>	<i>0.40</i>
<i>2</i>	<i>0.04</i>	<i>0.04</i>	<i>0.04</i>	<i>0.04</i>	<i>0.04</i>	<i>0.20</i>
<i>3</i>	<i>0.01</i>	<i>0.01</i>	<i>0.01</i>	<i>0.01</i>	<i>0.01</i>	<i>0.05</i>
<i>4</i>	<i>0.02</i>	<i>0.02</i>	<i>0.02</i>	<i>0.02</i>	<i>0.02</i>	<i>0.10</i>
<i>5</i>	<i>0.05</i>	<i>0.05</i>	<i>0.05</i>	<i>0.05</i>	<i>0.05</i>	<i>0.25</i>
<i>Column Total</i>	<i>0.20</i>	<i>0.20</i>	<i>0.20</i>	<i>0.20</i>	<i>0.20</i>	<i>0.20</i>

You may think that making the probabilities of the numbers on the tickets uneven was somewhat strange. That does mimic a very real issue. When people choose numbers (e.g. for lottery tickets or for passwords) they do not choose random numbers, they choose "lucky" numbers. Lucky numbers are often dates (birthdays, wedding days, etc.). That implies that people choose numbers from 1 - 12 and again numbers from 1 - 31 more often than numbers that are larger than 31. If you do choose such "lucky" numbers you are not less likely to win, but if you win you will have to share the prize with more people!

Example. Make sure you understand how the joint probabilities are calculated?

- $p(2, 1) = Pr(X = 2) \cdot Pr(Y = 1) = 0.20 \cdot 0.20 = 0.04$
- $p(4, 3) = Pr(X = 4) \cdot Pr(Y = 3) = 0.10 \cdot 0.20 = 0.02$
- $p(5, 5) = Pr(X = 5) \cdot Pr(Y = 5) = 0.25 \cdot 0.20 = 0.05$

In the above example we assumed that the random variables X and Y were independent. That allowed us to calculate the joint probabilities according to

$$\Pr(X \cap Y) = \Pr(X) \Pr(Y).$$

A typical problem we may later want to investigate is whether two random variables are in fact independent. This will be discussed later in the section.

3.1 Marginal Probabilities

Let us replicate the earlier table with joint probabilities

Probabilities player (X)	machine (Y)					Row Total
	1	2	3	4	5	
1	0.08	0.08	0.08	0.08	0.08	0.40
2	0.04	0.04	0.04	0.04	0.04	0.20
3	0.01	0.01	0.01	0.01	0.01	0.05
4	0.02	0.02	0.02	0.02	0.02	0.10
5	0.05	0.05	0.05	0.05	0.05	0.25
Column Total	0.20	0.20	0.20	0.20	0.20	0.20

Given a joint probability distribution

$$p(x, y) = \Pr(X = x, Y = y)$$

for the random variables X and Y , a probability of the form $\Pr(X = x)$ or $\Pr(Y = y)$ is called a marginal probability. You can remember the term marginal probabilities as in the above joint probability table these probabilities appeared in the margin of the table. The collection of these probabilities for all values of X is the marginal probability distribution for X (in the above table illustrated by the orange coloured numbers),

$$p_X(x) = \Pr(X = x).$$

If it is clear from the context, write $p_X(x)$ as $p(x)$.

$$\begin{aligned} \Pr(X = x) &= \Pr(X = x, Y = 1) + \Pr(X = x, Y = 2) + \Pr(X = x, Y = 3) + \\ &\quad \Pr(X = x, Y = 4) + \Pr(X = x, Y = 5) \end{aligned}$$

the sum of all the joint probabilities contributing to $X = x$. So, marginal probability distributions are found by summing joint probabilities over all the values of the other variable:

$$p_X(x) = \sum_y p(x, y), \quad p_Y(y) = \sum_x p(x, y).$$

Confirm for yourself that orange marginal probabilities are the sum of the respective row of joint probabilities.

Notice that a marginal probability distribution has to satisfy the usual properties expected of a probability distribution (for a discrete random variable):

$$0 \leq p_X(x) \leq 1, \quad \sum_x p_X(x) = 1,$$

$$0 \leq p_Y(y) \leq 1, \quad \sum_y p_Y(y) = 1.$$

In the above table we can also see the marginal probability for the random variable Y , $p_Y(y)$, this is the probability distribution coloured in blue. Each of these probabilities equal the sum of the respective column of joint probabilities.

The joint probability distribution contains all the information there is about these two random variables. That means we can infer, from that information, the expected values, variances and the correlation between the variables. We will see later how to derive the correlation from the joint probability distribution. Here we will use these marginal probabilities to calculate expected values and variances of the two random variables:

Example. *By calculation, we can find the expected value and variance of W , using the formula established in the discrete random variable section:*

$$E[X] = \sum_{X=1}^5 p_X(x)x = 0.4 \cdot 1 + 0.2 \cdot 2 + 0.05 \cdot 3 + 0.10 \cdot 4 + 0.25 \cdot 5 = 2.60$$

$$Var[X] = E[X^2] - E[X]^2 = (0.4 \cdot 1^2 + 0.2 \cdot 2^2 + 0.05 \cdot 3^2 + 0.10 \cdot 4^2 + 0.25 \cdot 5^2) - 2.6^2 = 2.74.$$

Example. *Find the expected value and variance of Y ,*

$$E[Y] = \sum_{y=1}^5 p_Y(y)y = 0.2 \cdot 1 + 0.2 \cdot 2 + 0.2 \cdot 3 + 0.2 \cdot 4 + 0.2 \cdot 5 = 3.00$$

$$Var[Y] = E[Y^2] - E[Y]^2 = (0.2 \cdot 1^2 + 0.2 \cdot 2^2 + 0.2 \cdot 3^2 + 0.2 \cdot 4^2 + 0.2 \cdot 5^2) - 3^2 = 2.00.$$

4 Independence, Covariance and Correlation

Joint probability distributions have interesting characteristics that are irrelevant (or better not defined) when we look at one random variable only. These characteristics describe how the two individual random variables relate to each other. The relevant concepts are independence, covariance and correlation and will be discussed in turn.

4.1 Independence

Let's first define the following joint probability

$$p(x, y) = \Pr(X = x, Y = y),$$

which represents the probability that simultaneously $X = x$ and $Y = y$. Also $p(x)$ and $p(y)$ are defined as the respective marginal probabilities.

Before discussing the definition of independence it is worth recalling a result of the Bayes' Theorem, that was discussed earlier in the Conditional Probability Section. That Theorem established that a conditional probability was to be calculated according to

$$p(x|y) = \frac{p(x, y)}{p(y)}.$$

From this it follows that the joint probability can be derived from

$$p(x, y) = p(y) p(x|y).$$

The remarkable feature of this result is that it is always valid regardless of how the two random variables X and Y relate to each other. There is one special case of relation between the two random variables, that special case is called *independence*.

Let's think of our little toy example of incomes of two partners in a household, say a husband and wife. Each of these have marginal distributions. So if you just picked a husband in any household, the marginal distribution $p(h)$ tells you all there is to know about the possible incomes for that husband. If the random variables H and W were not independent (i.e. dependent) then knowing what the household's wife's income is would change that probability. For instance it could be that knowing that the wife's income is high (£200) makes it more likely that the husband's income is also high. In that case $p(h|w = 200) \neq p(h)$. If the two random variables were independent then knowing what the wife's income is does not change the probability distribution of the husband's income, $p(h|w = 200) = p(h)$.

The statistical definition of *independence* of the random variables X and Y is that for all values of x and y the following relationship holds:

$$p(x, y) = p(x) p(y).$$

If this relationship indeed holds for all values of x and y , the random variables X and Y are said to be **independent**:

- Independence means that $\Pr(Y = y)$ would not be affected by knowing that $X = x$: knowing the value taken on by one random variable does not affect the probabilities of the outcomes of the other random variable. This is also expressed by the following relationship that *is only valid if X and Y are independent*:

$$p(x, y) = p(x) p(y).$$

Each joint probability is the product of the corresponding marginal probabilities.

- A corollary of this is that if two random variables X and Y are independent, then there can be no relationship of any kind, linear or non-linear, between them.

Example. *The joint probabilities and marginal probabilities (taken from the household incomes example) are*

Probabilities		H :			Row Sums
		£ 0	£ 100	£ 200	$p_W(w)$
Values of W :	£ 0	0.05	0.15	0.10	0.30
	£ 100	0.10	0.10	0.30	0.50
	£ 200	0.05	0.05	0.10	0.20
Column Sums: $p_H(h)$		0.20	0.30	0.50	1.00

If the two variables were independent we should be able to reconstruct the joint probabilities using the formula that is only valid in the case of independence: $p(h, w) = p(h)p(w)$. Let's try that for the outcome $(H = 0, W = 0)$ for which the table shows $p(0, 0) = 0.05$. If H and W were independent this should be

$$p(0, 0) = p(h = 0)p(w = 0) = 0.2 \cdot 0.3 = 0.06 \tag{1}$$

The result is not identical and hence this provides evidence that the variables H and W in the above joint distribution are not independent.

For X and Y to be independent, $p(x, y) = p_X(x)p_Y(y)$ has to hold for all x, y . Finding one pair of values x, y for which this fails is sufficient to conclude that X and Y are not independent.

Example. *Continuing with the above example, what would be the following joint probabilities if H and W were independent (where we now, for simplicity, represent £100 by 1)?*

□ $p(0, 2) = p(h = 0)p(w = 2) = 0.2 \cdot 0.2 = 0.04$

□ $p(2, 1) = p(h = 0)p(w = 2) = 0.5 \cdot 0.5 = 0.25$

$$\square p(1, 2) = p(h = 0)p(w = 2) = 0.3 \cdot 0.2 = 0.06$$

Is any of these probabilities (calculated under the assumption of independence) equal to the respective joint probability? No

4.2 Covariance and Correlation

We previously discussed how the sample covariance and in particular the sample correlation coefficient could be used to establish whether sample data for two variables display a linear relationship. When we know the joint probability population we can calculate the population equivalent of these two measures. Recall that the (population) correlation coefficient is really only a measure of strength of any linear relationship between the random variables.

4.2.1 Covariance

The first step is to define the (population) covariance as a characteristic of the joint probability distribution of X and Y . Let

$$E[X] = \mu_X, \quad E[Y] = \mu_Y.$$

The (population) covariance is defined as

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

Notice that $\text{cov}[X, Y] = \text{cov}[Y, X]$.

There are a number of alternative expressions for the covariance. Let us note that $(X - \mu_X)(Y - \mu_Y)$ can be seen as a function of X and Y , $g(X, Y) = (X - \mu_X)(Y - \mu_Y)$. We can think of $\text{cov}[X, Y] = E[g(X, Y)]$ and then apply the general formula for the expectation of this function of the random variables.

$$\text{cov}[X, Y] = \sum_x \sum_y (x - \mu_X)(y - \mu_Y)p(x, y)$$

We can see from this expression that if enough (x, y) pairs have $x - \mu_X$ and $y - \mu_Y$ values with the same sign, $\text{cov}[X, Y] > 0$, so that large (small) values of $x - \mu_X$ tend to occur with large

(small) values of $y - \mu_Y$. Similarly, if enough (x, y) pairs have $x - \mu_X$ and $y - \mu_Y$ values with different signs, $cov[X, Y] < 0$. Here, large (small) values of $x - \mu_X$ tend to occur with small (large) values of $y - \mu_Y$.

$cov[X, Y] > 0$ represents a positive relationship between X and Y , $cov[X, Y] < 0$ means a negative relationship.

There is a shorthand calculation for covariance, analogous to that given for the variance in the discrete random variable section:

$$\begin{aligned}
 cov[X, Y] &= E[(X - \mu_X)(Y - \mu_Y)] \\
 &= E[XY - X\mu_Y - \mu_X Y + \mu_X \mu_Y] \\
 &= E[XY] - E[X]\mu_Y - \mu_X E[Y] + \mu_X \mu_Y \\
 &= E[XY] - \mu_X \mu_Y - \mu_X \mu_Y + \mu_X \mu_Y \\
 &= E[XY] - \mu_X \mu_Y.
 \end{aligned}$$

As in the earlier section on linear transformations of random variables, the expected value of a sum of terms is equal to the sum of the expected values. Also, constants, like μ_X and μ_Y can be drawn outside the expectations operator, such that $E[X\mu_Y] = E[X]\mu_Y$.

Even with this shorthand method, the calculation of the covariance is rather tedious. To calculate $cov[W, H]$ in the income example, the best approach is to imitate the way in which $E[T]$ was calculated above. Rather than display the values of T (total income), here we display the values of $w \times h$ as we need to calculate $E[WH]$ as part of the variance calculation using the shorthand formula:

		h		
		0	100	200
w	0	(0) 0.05	(0) 0.15	(0) 0.10
	100	(0) 0.10	(10000) 0.10	(20000) 0.30
	200	(0) 0.05	(20000) 0.05	(40000) 0.10

Therefore, using the same strategy of multiplication within cells, and adding up along each row in turn, we find the following

$$\begin{aligned}
 E[WH] &= (0) \times 0.05 + (0) \times 0.15 + (0) \times 0.10 \\
 &\quad + (0) \times 0.10 + (10000) \times 0.10 + (20000) \times 0.30 \\
 &\quad + (0) \times 0.05 + (20000) \times 0.05 + (40000) \times 0.10 \\
 &= 1000 + 6000 + 1000 + 4000 \\
 &= 12000.
 \end{aligned}$$

Previously we found that $E[W] = 90$ and $E[H] = 130$, so that

$$\begin{aligned}
\text{cov}[W, H] &= E[WH] - E[W]E[H] \\
&= 12000 - (90)(130) \\
&= 300.
\end{aligned}$$

Example. *Simulate a small joint probability distribution for R and Q and have students calculate expected values, variances and the covariance.*

For an individual random variable X we previously learned how expectations and variances of linear transformation of that random variable relates to the characteristics of X itself. In particular you learned that

1. $E[Z] = E[bX + c] = bE[X] + c$
2. $\text{Var}[Z] = \text{Var}[bX + c] = b^2\text{Var}[X]$

We can ask the same question about the covariance of two linearly transformed random variables X and Y . Here is the result

$$1. \text{Cov}[Q, Z] = \text{Cov}[(bX + c), (dY + e)] = \text{Cov}[(bX), (dY)] = bd\text{Cov}[X, Y]$$

The additive terms are irrelevant for the covariance (as they are for the variance) and the two factors are drawn outside of the covariance operator.

Example. *You have two random variables X and Y with the following properties: $E[X] = 3$, $\text{Var}[X] = 2$, $E[Y] = -4$, $\text{Var}[Y] = 3$ and $\text{Cov}[X, Y] = 1.22$. Consider the two linear transformations, $Q = 2 + 8X$ and $Z = -6 - 4Y$. What are the following?*

- $E[Q] = 2 + 8E[X] = 26$
- $\text{Var}[Q] = 8^2\text{Var}[X] = 64 \cdot 2 = 128$
- $E[Z] = -6 - 4E[Y] = -6 - 4 \cdot (-4) = -6 + 16 = 10$
- $\text{Var}[Z] = (-4)^2\text{Var}[Y] = 16 \cdot 3 = 48$
- $\text{Cov}[Q, Z] = 8 \cdot (-4)\text{Cov}[X, Y] = -32 \cdot 1.22 = -39.04$

4.2.2 Additional Resources

- The link between covariance and regressions from the Khan Academy.

4.2.3 Strength of Association and Units of Measurement

How does covariance measure the **strength** of a relationship between two random variables? Not well is the answer, because the value of the covariance is dependent on the units of measurement. Take the above example in which both random variables W and H captured the income in £. When we calculated $cov[W, H]$ we needed to calculate terms like 100×100 and hence the units of measurement of the covariance is £^2 which is difficult to interpret.

Perhaps even more worrying, if we had calculated the covariance with 1 and 2 instead of 100 and 200 (and hence measuring income in 100s of pounds rather than pounds) we would have obtained a value of 0.03 (check this yourself by repeating the above calculation). Of course, we just changed the unit of measurement and nothing really has changed in the relationship between the two variables.

In the section on Correlation and Regression as a descriptive statistic we only briefly mentioned covariance for exactly these reasons. There we introduced the correlation statistic as the preferred measure of (linear) association. As discussed there, the correlation statistics is actually based on the covariance statistic, but it does avoid the weaknesses just discussed.

4.2.4 Correlation

What is required is a measure of the strength of association which is invariant to changes in units of measurement. Recall that the (population) **correlation coefficient** between X and Y is defined by

$$\rho_{XY} = \frac{cov[X, Y]}{\sqrt{var[X] var[Y]}}.$$

This is also the correlation between αX and βY as we can see when we apply the rules for linear transformations discussed earlier:

$$\begin{aligned} \rho_{\alpha X, \beta Y} &= \frac{cov[\alpha X, \beta Y]}{\sqrt{var[\alpha X] var[\beta Y]}} \\ &= \frac{\alpha \beta cov[X, Y]}{\sqrt{\alpha^2 \beta^2 var[X] var[Y]}} \\ &= \rho_{XY}, \end{aligned}$$

The correlation coefficient does not depend on the units of measurement.

For the income example we calculated the covariance statistic to be 300. If we calculate the variances, $var[H]$ and $var[W]$ we will get $var[H] = 6100$ and $var[W] = 4900$ respectively when using the \mathcal{L} as the unit of measurement. We then get

$$\begin{aligned}\rho_{WH} &= \frac{300}{\sqrt{(4900)(6100)}} \\ &= 0.0549.\end{aligned}$$

Is this indicative of a strong relationship? Just like the sample correlation coefficient of the Regression and Correlation Section, it can be shown that

- the correlation coefficient ρ_{XY} always satisfies $-1 \leq \rho_{XY} \leq 1$.
- The closer ρ is to 1 or -1 , the stronger the relationship.



Figure 1: Correlation coefficient, strength and direction

So, $\rho_{WH} = 0.0549$ is indicative of a very weak relationship.

It can be shown that if X and Y are exactly linearly related by

$$Y = a + bX \quad \text{with } b > 0$$

then $\rho_{XY} = 1$ - that is, X and Y are perfectly correlated. X and Y are also perfectly correlated if they are exactly linearly related by

$$Y = a + bX \quad \text{with } b < 0,$$

but $\rho_{XY} = -1$. Thus correlation measures the strength of a linear relationship between X and Y . It is important to remember that correlation does not imply causation.

Other notations for the correlation coefficient are

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$\rho_{XY} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sqrt{E[(X - \mu_X)^2] E[(Y - \mu_Y)^2]}}$$

which uses covariance and standard deviation notation, and recognises the definitions of covariances and variances.

Example. You have two random variables X and Y with the following properties: $E[X] = 3$, $Var[X] = 2$, $E[Y] = -4$, $Var[Y] = 3$ and $Cov[X, Y] = 1.22$. Consider the two linear transformations, $Q = 2 + 8X$ and $Z = -6 - 4Y$. What are the following?

- $E[Q] = 2 + 8E[X] = 26$
- $Var[Q] = 8^2 Var[X] = 64 \cdot 2 = 128$
- $E[Z] = -6 - 4E[Y] = -6 - 4 \cdot (-4) = -6 + 16 = 10$
- $Var[Z] = (-4)^2 Var[Y] = 16 \cdot 3 = 48$
- $Cov[Q, Z] = 8 \cdot (-4) Cov[X, Y] = -32 \cdot 1.22 = 39.04$
- $Corr[X, Y] = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}} = \frac{1.22}{\sqrt{2 \cdot 3}} = 0.4981$
- $Corr[Q, Z] = \frac{Cov[Q, Z]}{\sqrt{Var[Q]Var[Z]}} = \frac{39.04}{\sqrt{128 \cdot 48}} = 0.4981$

4.3 The relation between Correlation, Covariance and Independence

Non-zero correlation and covariance between random variables X and Y indicate some linear association between them, whilst independence of X and Y implies no relationship or association of any kind between them. So, it is not surprising that

- **independence** of X and Y implies **zero covariance**: $cov[X, Y] = 0$;
- **independence** of X and Y implies **zero correlation**: $\rho_{XY} = 0$.

The converse is not true, in general:

- **zero covariance or correlation** does not imply independence.

The reason is that there may be a relationship between X and Y which is not linear.

Example. Here is an example of a joint probability distribution.

$p(x, y)$		y		
		1	2	3
x	1	0.1	0.02	0.2
	2	0.08	0.2	0.05
	3	0.25	0.05	0.05

Can you use the information, without going through a full correlation calculation to see whether, here, X and Y are independent or not and whether there is correlation between the two random variables?

The answer is, they are not independent and there is a negative correlation. To illustrate that it may be useful to think graphically about this distribution. Let us firstly graphically display the joint probability distribution.

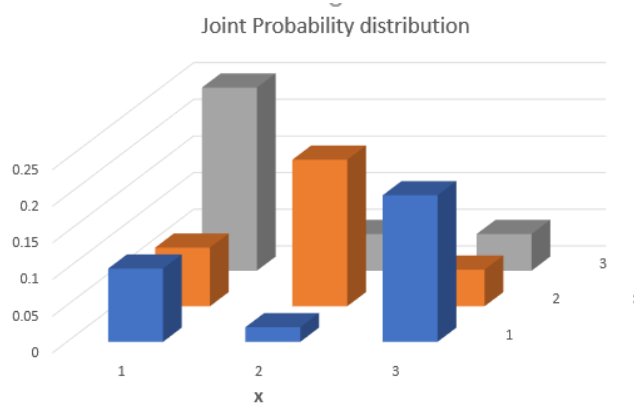


Figure 2: Joint probability distribution

From here you can see that the highest probability is for values where either X is high and Y is low or vice versa. So that indicates that there is a negative correlation. This would be even easier to see if you could see a sample of data randomly drawn from that distribution. In the following scatter diagram you can see just that. 100 observations were drawn from the above distribution. Each point is slightly randomly offset, such that you can see which combinations have been generated more frequently.

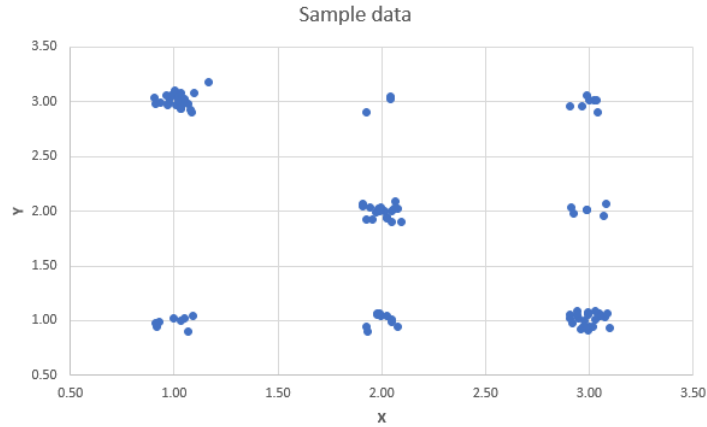


Figure 3: Joint probability mass function.

You can see that a line of best fit would show a negative slope.

Example. Here are three joint probability distributions. What are these distributions properties.

D1		y		
$p(x, y)$		1	2	3
x	1	0.10	0.06	0.04
	2	0.25	0.15	0.10
	3	0.15	0.09	0.06

In this distribution X and Y are independent random variables. **TRUE/FALSE**

Hint: the joint probabilities can be replicated from the marginal probabilities, hence the two variables are independent

In this distribution X and Y are **not**/positively/negatively correlated.

Hint: as the two random variables are independent, the correlation is 0.

D2		y		
$p(x, y)$		1	2	3
x	1	0.25	0.03	0.28
	2	0.04	0.04	0.05
	3	0.03	0.26	0.02

In this distribution X and Y are independent random variables. **TRUE/FALSE**

Hint: the joint probabilities can not be replicated from the marginal probabilities, hence the two variables are dependent

In this distribution X and Y are **not**/positively/negatively correlated.

Hint: While the two variables are not independent (they are dependent), there is no correlation

as the relationship is not a linear one. For low values of X very high and very low values of Y are more likely. You can think of this as a nonlinear (quadratic) relationship.

D3		y		
	$p(x, y)$	1	2	3
x	1	0.25	0.09	0.01
	2	0.08	0.18	0.07
	3	0.01	0.10	0.21

In this distribution X and Y are independent random variables. **TRUE/FALSE**

Hint: Probabilities on the diagonal are higher. This is a sign of two dependent random variables.

In this distribution X and Y are not/**positively**/negatively correlated.

Hint: If X is low (high) lower (higher) values of Y are more likely. This is the sign of a positive linear relationship.

5 Linear Combinations of Several Random Variables

Reconsider the previous example on the incomes of two partners in a household, where the random variables H and W represent the respective salaries. The joint probability distribution of H and W is given here, displaying $p(h, w)$:

		H		
Probabilities		£ 0	£ 100	200
Values of W :	0	0.05	0.15	0.10
	100	0.10	0.10	0.30
	200	0.05	0.05	0.10

You may now be interested in the total household income, T :

$$T = g(H, W) = H + W.$$

This new random variable is a function of H and W , and we can deduce the probability distribution of T from the joint distribution of H and W . You can directly calculate the new variable's expected value and variance from the joint probability distribution $p(h, w)$: the expected value of T is

$$\begin{aligned}
E[T] &= E[g(H, W)] \\
&= \sum_h \sum_w g(h, w) p(h, w) \\
&= 0 \cdot 0.05 + 100 \cdot 0.15 + 200 \cdot 0.10 + \\
&\quad 100 \cdot 0.10 + 200 \cdot 0.10 + 300 \cdot 0.3 + \\
&\quad 200 \cdot 0.05 + 300 \cdot 0.05 + 400 \cdot 0.10 \\
&= 220
\end{aligned}$$

We sum up the product of all joint probabilities and their associated outcome for T .

This is what one may call a pedestrian way of getting the expected value for a function of two (or more) random variables. This will actually work for any type of function of the two random variables, even if that function is a nonlinear one.

If the function $g(X, Y)$ of the two random variables X and Y is a linear one, like the generic

$$V = aX + bY + c$$

then things simplify significantly. In what follows we shall demonstrate how to get the expected value and the variance of a linear combination of random variables directly if we know what $E[X]$, $E[Y]$, $Var[X]$, $Var[Y]$ and $Cov[X, Y]$ are.

5.1 A preliminary result

Let's start with the simplest of all cases and consider

$$V = g(X, Y) = X$$

then it should follow that (for discrete random variables)

$$E[X] = \sum_x \sum_y xp(x, y)$$

This can be shown to apply by summing, for each given x the probabilities for all the possible outcomes for Y , so that

$$E[X] = \sum_x x \left[\sum_y p(x, y) \right]$$

By the definition of the marginal distribution of y , the summation in brackets is the marginal probability $p_X(x)$ that $X = x$, giving

$$E[X] = \sum_x xp_X(x)$$

This shows that the expected value of X is the same whether calculated from the joint probability distribution of X and Y , or from the marginal distribution of X . The same, obviously, applies for $E[Y]$. This is just another way of saying that the joint distribution contains all the information regarding the marginal distributions.

5.2 Expected value of a Linear Combination

The result we need for the expected value of a linear combination is easy to remember: it amounts to saying that

- the expected value of a linear combination is the linear combination of the expected values, or
- more simply, the expected value of a sum is equal to the sum of expected values.

If $V = aX + bY + c$, where a, b, c are constants, then

$$E[V] = E[aX + bY + c] = aE[X] + bE[Y] + c$$

This result is a natural generalisation of the linear transformation result that $E[bX + c] = bE[X] + c$.

Proof: (discrete random variables case). Consider

$$\begin{aligned} E[V] &= E[aX + bY + c] \\ &= \sum_x \sum_y (ax + by + c) \times p(x, y) \\ &= \sum_x \sum_y ax \times p(x, y) + \sum_x \sum_y by \times p(x, y) + \sum_x \sum_y c \times p(x, y) \end{aligned}$$

Examine each of the three terms on the right hand side:

$$\begin{aligned}
\sum_x \sum_y a x p(x, y) &= a \sum_x \sum_y x p(x, y) = a \sum_x x \sum_y p(x, y) = a E[X] \\
\sum_x \sum_y b y p(x, y) &= b \sum_y \sum_x y p(x, y) = b \sum_y y \sum_x p(x, y) = b E[Y] \\
\sum_x \sum_y c p(x, y) &= c \sum_x \sum_y p(x, y) = c \times 1.
\end{aligned}$$

The result that the probabilities in a joint probability distribution must sum to 1 over all x and y combinations is used in the line immediately above. Summing the three terms then gives the earlier result

$$\begin{aligned}
E[V] &= E[aX + bY + c] \\
&= aE[X] + bE[Y] + c
\end{aligned}$$

Notice that nothing need be known about the joint probability distribution $p(x, y)$ of X and Y to prove the result: it applies, for example, irrespective of whether the variables are independent or not. The result is also valid for continuous random variables.

Example. For the joint income of husbands and wives example, we defined $T = W + H$, where $E[W] = 130$, $E[H] = 90$. Therefore, using the expected value of the linear combination gives

$$E[T] = E[W] + E[H] = 220 \tag{2}$$

confirming the result obtained earlier.

Example. Suppose that random variables X and Y have $E[X] = 0.5$ and $E[Y] = 3.5$, and let $V = 5X - Y$. Then,

$$\begin{aligned}
E[V] &= 5E[X] - E[Y] \\
&= 5 \times 0.5 - 3.5 \\
&= -1
\end{aligned}$$

Let X_1, \dots, X_n be random variables with a_1, \dots, a_n, c being constants, and define the random variable W by

$$\begin{aligned}
W &= c + a_1X_1 + \dots + a_nX_n \\
&= c + \sum_{i=1}^n a_iX_i
\end{aligned}$$

Then,

$$E[W] = c + \sum_{i=1}^n a_iE[X_i]$$

This can be proved by repeatedly using the linear combination result for two variables.

Example. *Let's consider the case where we combine $n = 3$ random variables with $E[X_1] = 2$, $E[X_2] = -1$, $E[X_3] = 3$, $W = 2X_1 + 5X_2 - 3X_3 + 4$. What is $E[W]$?*

$$\begin{aligned}
E[W] &= E[2X_1 + 5X_2 - 3X_3 + 4] \\
&= 2E[X_1] + 5E[X_2] - 3E[X_3] + 4 \\
&= 2 \times 2 + 5 \times (-1) - 3 \times 3 + 4 \\
&= -6
\end{aligned}$$

5.3 Variance of a Linear Combination

The result for the mean (or expected value) of a linear combination is fairly intuitive. That for a variance, however, is less so. We start with the two variable case.

5.3.1 Two variables

Let V be the random variable $V = aX + bY + c$. To find $Var[V]$ we start from the alternative form for the variance calculation:

$$\begin{aligned}
Var[V] &= E[V^2] - E[V]^2 \\
&= E[(aX + bY + c)^2] - E[aX + bY + c]^2
\end{aligned}$$

Now we will just apply the above rule for the expectation of a linear combination and multiply out the squared terms. It is a little tedious algebra, but if you do it carefully you should be able to account for all terms.

$$\begin{aligned}
\text{Var}[V] &= \text{Var}[aX + bY + c] \\
&= E[(aX + bY + c)^2] - E[aX + bY + c]^2 \\
&= E[(aX + bY + c)(aX + bY + c)] - (aE[X] + bE[Y] + c)^2 \\
&= E[(aX + bY + c)(aX + bY + c)] - (aE[X] + bE[Y] + c)(aE[X] + bE[Y] + c) \\
&= E[a^2X^2 + abXY + acX + abXY + b^2Y^2 + bcY + acX + bcY + c^2] \\
&\quad - a^2E[X]^2 - abE[X]E[Y] - acE[X] - abE[X]E[Y] - b^2E[Y]^2 - bcE[Y] - acE[X] - bcE[Y] - c^2
\end{aligned}$$

The next step is to apply again the result for the expected value of a linear combination and then rearrange terms in a useful manner, remembering that $\text{Var}[X] = E[X^2] - E[X]^2$ and $\text{Cov}[X, Y] = E[XY] - E[X]E[Y]$:

$$\begin{aligned}
\text{Var}[V] &= a^2E[X^2] + abE[XY] + acE[X] + abE[XY] + b^2E[Y^2] + bcE[Y] + acE[X] + bcE[Y] + c^2 \\
&\quad - a^2E[X]^2 - abE[X]E[Y] - acE[X] - abE[X]E[Y] - b^2E[Y]^2 - bcE[Y] - acE[X] - bcE[Y] - c^2 \\
&= a^2 \underbrace{(E[X^2] - E[X]^2)}_{\text{Var}[X]} + b^2 \underbrace{(E[Y^2] - E[Y]^2)}_{\text{Var}[Y]} \\
&\quad - 2ab \underbrace{(E[XY] - E[X]E[Y])}_{\text{Cov}[X, Y]} \\
&\quad + 2acE[X] + 2bcE[Y] + c^2 - 2acE[X] - 2bcE[Y] - c^2
\end{aligned}$$

Now you can note that all the terms in the last line cancel out, which leaves us with the following very important result:

$$\text{Var}[V] = \text{Var}[aX + bY + c] = a^2\text{Var}[X] + b^2\text{Var}[Y] + 2ab\text{Cov}[X, Y]$$

Summarising, for $V = aX + bY + c$, the following holds:

1. $\text{Var}[V] = a^2\text{Var}[X] + 2ab\text{Cov}[X, Y] + b^2\text{Var}[Y]$
2. If X and Y are independent, then $\text{Cov}[X, Y] = 0$ and

$$\text{Var}[V] = a^2\text{Var}[X] + b^2\text{Var}[Y]$$

3. If X and Y are dependent, but linearly uncorrelated such that $Cov[X, Y] = 0$, then the same result holds.

It was worth going through the details of this calculation as it lies at the heart of modern finance theory (in particular portfolio theory). It is this formula which justifies the benefits of portfolio diversification, the idea that it is better to hold several risky assets rather than just one as long as they are not perfectly correlated. This is true for investors but also for companies which may want to invest in several products rather than just one.

Let's use the following example to illustrate this

Why is this result so important ... diversification



Example. Assume you have two assets you can invest in, asset A which has random annual returns with $E[R_A] = 0.05$ and variance $Var[R_A] = 0.2$. Asset B has $E[R_B] = 0.05$ and variance $Var[R_B] = 0.25$. The correlation between the returns is $Corr[R_A, R_B] = 0.3$, so they are mildly positively correlated.

The variance in asset returns gives you a measure of the risk you carry with that investment. So Asset B has a higher risk at the same expected return. Most of us would, everything else being equal, prefer lower risk.

Which asset would you invest in if you had to invest all your money into either asset A or asset B ? In that case you would choose asset A as it gives you the same expected return but at a lower risk.

Let's say you form a portfolio where you invest 50% of your money into asset A and 50% into asset B . What would be the expected return and the variance of your portfolio?

You need to consider a new random variable $P = 0.5R_A + 0.5R_B$. This is a linear combination of two random variables and we can apply the results we learned above:

$$E[P] = 0.5 E[R_A] + 0.5 E[R_B]$$

$$\text{Var}[P] = \text{Var}[0.5 R_A + 0.5 R_B] = 0.5^2 \text{Var}[R_A] + 0.5^2 \text{Var}[R_B] + 2 \cdot 0.5 \cdot 0.5 \text{Cov}[R_A, R_B]$$

This means that we need $\text{Cov}[R_A, R_B]$ which we have not been provided with. However, we know the correlation and the variances, such that we can use the following relationship

$$\begin{aligned} \text{Corr}[R_A, R_B] &= \frac{\text{Cov}[R_A, R_B]}{\sqrt{\text{Var}[R_A] \cdot \text{Var}[R_B]}} \\ \text{Cov}[R_A, R_B] &= \text{Corr}[R_A, R_B] \cdot \sqrt{\text{Var}[R_A] \cdot \text{Var}[R_B]} \\ &= 0.3 \cdot \sqrt{0.2 \cdot 0.25} = 0.06708 \end{aligned}$$

Now we can calculate the portfolio characteristics:

$$E[P] = 0.5 E[R_A] + 0.5 E[R_B] = 0.5 \cdot 0.05 + 0.5 \cdot 0.05 = 0.05$$

$$\begin{aligned} \text{Var}[P] &= \text{Var}[0.5 R_A + 0.5 R_B] = 0.5^2 \text{Var}[R_A] + 0.5^2 \text{Var}[R_B] + 2 \cdot 0.5 \cdot 0.5 \text{Cov}[R_A, R_B] \\ &= 0.5^2 \cdot 0.2 + 0.5^2 \cdot 0.25 + 2 \cdot 0.5 \cdot 0.5 \cdot 0.06708 = 0.14604 \end{aligned}$$

And this is the wonder of diversification. Your portfolio, consisting of 50% asset A and 50% of asset B has the same expected return, but its variance (risk) is smaller than the individual assets'.

Financial portfolio theory then thinks about what the best possible way of mixing assets A and B would be, i.e. how can you minimise $\text{Var}[P]$ and also how the above relationship generalises to the case where you have more than two assets.

Example. Suppose that X and Y are random variables with $\text{Var}[X] = 0.25$, $\text{Var}[Y] = 2.5$ and $\text{Cov}[X, Y] = 0.5$. Calculate $\text{Var}[V]$ if $V = 2X + 0.5Y$. [Note to Mobius builder, randomise a, b , variances and covariances]

$$\begin{aligned} \text{Var}[V] &= \text{Var}[a X + b Y] = a^2 \text{Var}[X] + b^2 \text{Var}[Y] + 2 \cdot a \cdot b \text{Cov}[X, Y] \\ &= 2^2 \cdot 0.25 + 0.5^2 \cdot 2.5 + 2 \cdot 2 \cdot 0.5 \cdot 0.5 = 2.625 \end{aligned}$$

What is the variance of $W = 2X - 0.5Y$

$$\begin{aligned} \text{Var}[W] &= \text{Var}[a X + b Y] = a^2 \text{Var}[X] + b^2 \text{Var}[Y] + 2 \cdot a \cdot b \text{Cov}[X, Y] \\ &= 2^2 \cdot 0.25 + 0.5^2 \cdot 2.5 + 2 \cdot 2 \cdot (-0.5) \cdot 0.5 = 0.625 \end{aligned}$$

Hint: When you have a negative contribution in the linear combination you got to be careful with the signs. The variance still contributes positively but the covariance term now has a negative sign.

5.4 Variance of linear combinations of many uncorrelated variables

Extend the earlier result to the case of a linear combination of n random variables X_1, \dots, X_n is messy because of the large number of covariance terms involved. But that is exactly what is required to develop portfolio theory. We leave that to finance courses. However, in the special case when X_1, \dots, X_n are uncorrelated random variables then all covariances are equal to zero: $Cov[X_i, X_j] = 0, i \neq j$. And therefore, for X_1, \dots, X_n uncorrelated,

$$Var \left[\sum_{i=1}^n a_i X_i \right] = \sum_{i=1}^n a_i^2 Var[X_i]$$

This also applies when X_1, \dots, X_n are independent random variables as they are, by definition, uncorrelated.

Be careful to remember, however, that this result only applies if the variables are uncorrelated. If they are correlated, then the covariance terms are nonzero and these will enter the computation of the variance of a linear combination.

There is no short-cut to computing the standard deviation of a linear combination. You will have to calculate the variance of the linear combination and then take the square root to get the standard deviation.

6 Normal Random Variables

The results so far on linear combinations relate to characteristics of the probability distribution of a linear combination like

$$V = aX + bY + c$$

not to the probability distribution itself. In particular, we have considered the mean and variance. Indeed, part of the attraction of these results is that they can be obtained without having to find the probability distribution of V . In fact it is, in general, not easily possible to find the probability distribution of V itself, which is why it is so nice that at least we can easily get the mean and the variance.

However, if we know that X and Y have normal distributions then it follows that V itself is normally distributed. This is an extension of the result for the linear transformation of a single normally distributed random variable. The linear combination result is, however, much more powerful. It is also EXTREMELY IMPORTANT! It is also rather unusual: there are not many distributions for which this type of result holds.

If $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, with $V = aX + bY + c$ for constants a, b, c then

$$V \sim N(\mu_V, \sigma_V^2)$$

for which

$$\begin{aligned}\mu_V &= a\mu_X + b\mu_Y + c \\ \sigma_V^2 &= a^2\sigma_X^2 + 2ab\sigma_{XY} + b^2\sigma_Y^2\end{aligned}$$

Note that independence of X and Y has not been assumed and is irrelevant to the general result. - If X_1, \dots, X_n are uncorrelated random variables with $X_i \sim N(\mu_i, \sigma_i^2)$ and $W = \sum_{i=1}^n a_i X_i$, where a_1, \dots, a_n are constants, then

$$W \sim N\left(\sum_{i=1}^n a_i \mu_i, \sum_{i=1}^n a_i^2 \sigma_i^2\right).$$

In this case, the assumption that the variables are uncorrelated plays a role only in specifying the variance of the linear combination. If the variables were to be correlated, the normal distribution also applies to the linear combination with the mean being $\sum_i a_i \mu_i$, but the variance of the sum will change (to include terms involving the covariances - but we did not discuss this case here unless $n = 2$).

Of course, standard normal distribution tables can be used in the usual way to compute probabilities of events involving linear combinations like V and W . This is illustrated in the following.

Example. If X and Y are independent, $X \sim N(20, 5)$, $Y \sim N(30, 11)$, while $D = X - Y$, then $D \sim N(-10, 16)$ as

$$\begin{aligned}E[D] &= 1 \cdot E[X] + (-1) \cdot E[Y] = 1 \cdot 20 + (-1) \cdot 30 = -10 \\ \text{Var}[D] &= 1^2 \cdot \text{Var}[X] + (-1)^2 \cdot \text{Var}[Y] = 1 \cdot 5 + 1 \cdot 11 = 16.\end{aligned}$$

What is the probability that $D > 0$? As D is normally, but not standard normally distributed we shall standardise it to a $Z \sim N(0, 1)$ variable and then apply the standard normal probability table.

$$\begin{aligned}
Pr(D > 0) &= Pr\left(Z > \frac{0 - (-10)}{4}\right) \\
&= Pr(Z > 2.5) \\
&= 0.00621.
\end{aligned}$$

Therefore, there is a 0.621% probability that the random variable D takes a value of larger than 0.

Example. Suppose that X and Y are normally distributed random variables $X \sim N(2, 3)$, $Y \sim N(-3, 10)$ with correlation $Corr[X, Y] = -0.5$. Calculate $Pr(V > 0)$ if $V = 2X + 0.5Y$. [Note to Mobius builder, randomise parameters]

$$\begin{aligned}
Cov[X, Y] &= Corr[X, Y] \cdot \sqrt{Var[X]Var[Y]} = (-0.5) \cdot \sqrt{3 \cdot 10} = -2.7386 \\
E[V] &= 2 \cdot E[X] + 0.5 \cdot E[Y] = 2 \cdot 2 + 0.5 \cdot (-3) = 2.5 \\
Var[V] &= 2^2 \cdot Var[X] + 0.5^2 \cdot Var[Y] + 2 \cdot 2 \cdot 0.5 \cdot Cov[X, Y] \\
&= 4 \cdot 3 + \frac{1}{4} \cdot 10 + 2 \cdot (-2.7386) = 9.0228.
\end{aligned}$$

What is the probability that $V > 0$?

$$\begin{aligned}
Pr(V > 0) &= Pr\left(Z > \frac{0 - 2.5}{\sqrt{9.0228}}\right) \\
&= Pr(Z > -0.83) = 1 - Pr(Z \leq -0.83) \\
&= 1 - 0.2033 = 0.7967.
\end{aligned}$$

Cumulative Area Under the Standard Normal Distribution
Example: $P(Z < -2.54) = 0.0055$, $P(Z \geq -2.54) = 1 - P(Z < -2.54) = 0.9945$

Z	0	1	2	3	4	5	6	7	8	9
-3.00	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.90	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.80	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.70	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.60	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.50	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.40	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.30	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.20	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.10	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.00	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.90	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.80	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.70	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.60	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.50	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.40	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.30	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.20	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.10	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.00	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.90	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.80	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.70	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.60	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.50	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776

Therefore, there is a 79.67% probability that the random variable V takes a value larger than 0.

7 Summary

In this section you learned how to think about two random variables. The joint distribution of two random variables gives you all the information you need to establish whether the two random variables are correlated or not and whether they are independent or not.

Further it was discussed how to calculate expected values of linear combinations of random variables. You learned that you can always calculate the expected value and variance of such linear combinations, but that, in general, you cannot say anything about the distribution of the resulting random variable. An exception to this was where the random variables being combined are normally distributed random variables. In that case, the resulting linear combination is also normally distributed. This, however, is clearly a special case.

Through the discussion of linear combinations you also learned the statistics behind one of the most fundamental insights of finance theory, the value of diversification. This insight is important for all financial but also more generally all business investment. Further, knowing how to deal with multiple random variables and their combinations is crucial when it comes to learning more statistics and econometrics.