Chapter 5

Probability

Investors think about the future and try to take advantage of the uncertainties of the future. The study of uncertainty and randomness is part of a branch of mathematics known as probability theory. It is important to distinguish between *statistics* and *probability*. Statistics (the analysis of patterns in data) is founded upon probability, but probability can exist without statistics (data). To be more specific, *probability* is part of the theory of *distributions* and *measures* and is really *pure mathematics*.

Probability is usually thought of as a practical experimental science. However, this is far from the truth. The mathematical theory of probability can be based on axioms without any reference to statistics or data. And in the world of finance, it turns out that the rules of probability are counterintuitive as they apply to derivatives.

5.1 Probability

Since we are taking the *pure mathematics* approach, we have to develop the probability theory on the basis on axioms as the foundations. We can then construct the further theory on these foundations.

5.1.1 Probability axioms

In general, suppose we have n outcomes of a random event, where n is finite or *countably infinite*[1] and we express these events in terms of *sets*:

$$A_1.A_2, A_3, \ldots, A_n \qquad . \tag{5.1}$$

If these events cover all possible outcomes, then the union of all these sets covers the entire *sample space*:

$$\Omega = \bigcup_{i=1}^{n} A_i \quad . \tag{5.2}$$

We say the set is countably infinite and *exhaustive*.

Then we can define a *probability* (or measure) of an *event* happening in terms of a probability measure P as:

$$P(A_i) \quad . \tag{5.3}$$

The definition of a *probability measure* is such that:

[1] An example of a countable infinity is the (infinite) set of integers. Any set having a one-to-one correspondence with the set of integers would mean this would also be a countable infinity.

1.
$$P : A \mapsto p \quad \{p \in \mathbb{R} : 0 \leq p \leq 1\} \tag{5.4}$$

Probability is simply a method of converting a set of events (the *domain* variables) into a number between 0 and 1 (the *range* variable). In mathematical terms we say that probability is a *measure*, a general type of *mapping* that relates a *set* A to a real number in the closed interval $[0, 1]$.

When the probability is 1 we say the event is *certain* to occur.

2.
$$P(\Omega) = 1 \qquad . \tag{5.5}$$

3. Two events are *mutually exclusive* if
$$A_i \cap A_j = \emptyset \quad , \tag{5.6}$$
in which case we also have that:
$$P(A_i \cup A_j) = P(A_i) + P(A_j) \qquad . \tag{5.7}$$
This is the law of addition.

4. The law of addition extends to countable additivity, if m events are all mutually exclusive:
$$A_i \cap A_j = \emptyset \quad , \quad i \neq j \quad , \quad i, j = 1, 2, \ldots, m \quad , \tag{5.8}$$
then
$$P\left(\bigcup_{i=1}^{m} A_i\right) = \sum_{i=1}^{m} P\left(A_i\right) \qquad . \tag{5.9}$$

Clearly the symbol P is not a function in the conventional sense, like $A^2 - 2$ or e^{-2A}, since we cannot even define A^2, when it applies to a set, for example. This is why it is termed a *measure*. Obviously the closer the probability gets to 1, the more certain it is to occur, while the closer it gets to 0, the more unlikely it is to occur. If we have two events A and B and
$$P(A) > P(B) \quad , \tag{5.10}$$
then we say event A is more probable than event B.

5.1.2 Addition

Following on from these axioms one can build the theory of probability. A few important *lemmas* and *theorems* will be mentioned here, and referred to later. We note that the *negation* of an event (equivalent to the complement of a set) is such that:
$$P(A^c) = 1 - P(A) \quad . \tag{5.11}$$

Sometimes, if we want to calculate the probability of event A, it may be more convenient to find $P(A^c)$, and then use this relation. Clearly:
$$A \cup A^c = \Omega \quad , \quad A \cap A^c = \emptyset \quad . \tag{5.12}$$

The general law of addition is, for any two events A and B
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad . \tag{5.13}$$

5.1.3 Sets of measure zero

An impossible event has probability zero. However, there are also events which are not impossible but which also have a zero probability. We call the events (or sets): *sets of measure zero*.

By the same token we have events which are certain and thus have probability 1. There also exist events that have measure 1 but are *not* certain. In this case you will often see, in the text books, the additional abbreviation

$$a.s. \quad \text{meaning} \quad \text{almost} \quad \text{surely} \quad .$$

Thus an event is almost sure to occur if it is certain except for sets of measure zero. These notes do not, in any way, purport to be a rigorous approach to *measure theory*, the concept is simply mentioned at this point.

5.2 Conditional Probability

So far, we have arrived at an algebra of addition (and the inverse operation - subtraction) of probabilities. The next section shows how we extend our algebra to devise rules for multiplication, and its inverse, division. We start with the analogue of division, *conditional probability*.

The *conditional probability* is denoted as $P(A|B)$, which is read as 'the probability of A *given* B' and is defined as:

$$\boxed{P(A|B) \equiv \frac{P(A \cap B)}{P(B)}} \quad , \quad P(B) \neq 0 \quad . \tag{5.14}$$

It is said to be the probability that A occurs *given* that B occurred. This is our *rule of division*, or rather the *conditional probability* is just a notation (definition) for division.

5.3 Multiplication Rule

Since conditional probability is an expression for division, and division is the inverse of multiplication, simple rearrangement gives the *multiplication rule*:

$$\boxed{P(A \cap B) = P(A|B)P(B)} \tag{5.15}$$

This provides a viable algebra which can be used as the basis for further study or probability theory.

Conditional probability obeys the axioms, and associated lemmas, of unconditional probability.

Example: In analogy to $P(A^c) = 1 - P(A)$, show that: $P(A^c|B) = 1 - P(A|B)$.

$$P(A^c|B) = \frac{P(A^c \cap B)}{P(B)} \quad ,$$

but since, $B = (A \cap B) \cup (B \cap A^c)$, then

$$P(B) = P(A \cap B) + P(B \cap A^c) \Rightarrow P(A^c \cap B) = P(B) - P(A \cap B) \quad ,$$

$$\Rightarrow P(A^c|B) = \frac{P(B) - P(A \cap B)}{P(B)} = 1 - P(A|B).$$

5.4 Partition Rule

A family of sets (events) $B_1, B_2, B_3, \ldots, B_n$ is a *partition* of Ω if

$$B_i \cap B_j = \emptyset \text{ for all } \; i \neq j \text{ and } \; \bigcup_{i=1}^{n} B_i = \Omega.$$

That is all members are *mutually exclusive* and *exhaustive* of the sample space. That is, the subsets do not overlap, but they (as a whole) cover the entire sample space.

For *any* events A, B, such that $P(A), P(B) > 0$, the *partition rule* states that:

$$\boxed{P(A) = P(A|B)P(B) + P(A|B^c)P(B^c)} \quad . \tag{5.16}$$

Proof We have a partition $\{B, B^c\}$ (disjoint and exhaustive subsets of the event space). Then the intersections of the set A with the two elements of the partition are disjoint, that is:

$$A = (A \cap B) \cup (A \cap B^c) \qquad , \tag{5.17}$$

in which:

$$(A \cap B) \cap (A \cap B^c) = A \cap (B \cap B^c) = A \cap \emptyset = \emptyset \qquad . \tag{5.18}$$

And since this is the union of disjoint sets (figure 5.2) then

$$\begin{aligned} P(A) &= P((A \cap B) \cup (A \cap B^c)) \\ &= P(A \cap B) + P(A \cap B^c) \\ &= P(A|B)P(B) + P(A|B^c)P(B^c). \end{aligned}$$

This can be illustrated by the Venn diagram (figure 5.2).

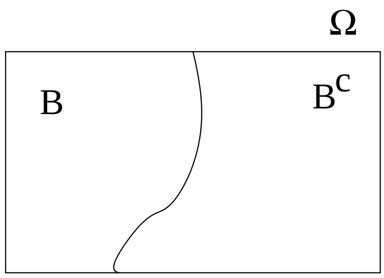

Figure 5.1: A partition of the event space: B and B^c. The subsets B and B^c are, by definition, disjoint subsets of Ω.

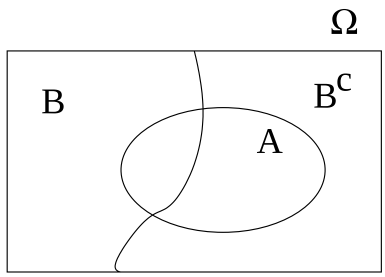

Figure 5.2: This illustrates the division of A across the partition. The set A is thus divided into two disjoint subsets: $A \cap B$ and $A \cap B^c$ according to equation (5.17).

More generally for any *partition* $\{B_1, B_2, \ldots, B_n\}$, we have the union of disjoint subsets:

$$A = \bigcup_{i=1}^{n} (A \cap B_i) \quad .$$

Then it follows that:

$$P(A) = \sum_{i=1}^{n} P(A \cap B_i)$$

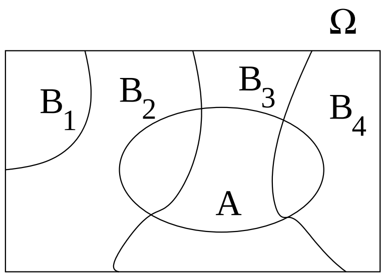

Figure 5.3: The visual representation of the *partition rule* expressed by equation (5.19). This illustration considers the set A divided between the partition sets: $\{B_1, B_2, B_3, B_4\}$.

and using the multiplication rule (5.15), we arrive at the *partition rule*:

$$\boxed{P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i).} \tag{5.19}$$

An illustration of the partition rule is given in figure 5.3.

5.5 Bayes' theorem

In general, we have,

$$P(A \cap B) = P(A|B)P(B) \quad , \quad P(A \cap B) = P(B|A)P(A) \quad , \tag{5.20}$$

that is,

$$P(A|B)P(B) = P(B|A)P(A) \quad , \tag{5.21}$$

or equivalently, if $P(B) \neq 0$,

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}} \quad . \tag{5.22}$$

which is *Bayes' theorem.* It is a slight exaggeration, but only slight, to say that Bayes' theorem is the foundation of all *statistical inference.*

Consider one simple application. Suppose we have a collection of events, a partition: $\{A_1, A_2, \ldots, A_n\}$ and an event B, such that $P(B) \neq 0$. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i) \tag{5.23}$$

by the partition rule. Then, we have a common expression of Bayes' theorem;

$$P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)} \quad . \tag{5.24}$$

and in particular:

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}} \quad . \tag{5.25}$$

Example Consider two urns containing a mixture of white and blue balls. Urn 1 contains 2 white and 7 blue balls. Urn 2 has 5 white and 6 blue balls. We flip a fair coin and draw a ball from urn 1 if we get H, and draw from urn 2 if T occurs.

What is the probability that the outcome of the toss was H *given* that a W occurred (that is a white ball was selected)?

This problem can be solved by using Bayes' theorem. We seek $P(H|W)$, and this can be expressed as:

$$P(H|W) = \frac{P(W|H)P(H)}{P(W)} = \frac{P(W|H)P(H)}{P(W|H)P(H) + P(W|T)P(T)}$$

$$P(H|W) = \frac{(\frac{2}{9}) \times (\frac{1}{2})}{(\frac{2}{9})(\frac{1}{2}) + (\frac{5}{11})(\frac{1}{2})} = \frac{22}{67} \qquad .$$

5.6 Independent Events

We call two events, A and B, *independent* if the occurrence of one does not affect the probability that the other occurs.

Thus if $P(A), P(B) > 0$ then A and B are independent if

$$P(A|B) = P(A) \qquad \text{and} \qquad P(B|A) = P(B) \qquad .$$

Then since,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = P(A)$$

it follows that:

$$\boxed{P(A \cap B) = P(A)P(B) \qquad .} \tag{5.26}$$

Thus an equivalent definition, and the more commonly used expression for *independent events* is,

$$P(A \cap B) = P(A)P(B) \qquad \Leftrightarrow \qquad A \text{ and } B \text{ independent} \quad . \tag{5.27}$$

5.7 Discrete random variables

Suppose we have a variable X that is random but that it can take on a limited number of discrete values (*outcomes* or *events*): $X \in \{x_1, x_2, \ldots, x_m\}$, and none of these values are equal to any others in the set. Then X is called a *discrete random variable.* The classic example of such a variable would be the result of a coin toss, in which only HEADS or TAILS would be possible outcomes. In our application we will be considering X as the price change in the value of an asset, say from day to day, or from hour to hour.

Let us assume that the probability of X having each of the allowed values is known, for every value. This *probability mass function* we denote as follows:

$$P(X = x_i) = f_X(x_i) \quad , \qquad i = 1, 2, \ldots, m \quad . \tag{5.28}$$

However, X can have one (and *only* one) of the values $\{x_1, x_2, \ldots, x_m\}$. Each of these *events* is *mutually exclusive*: one cannot have any two events occurring simultaneously. This can be expressed as follows:

$$P(X = x_i \text{ and } X = x_j) = 0 \quad , \quad i \neq j \quad . \tag{5.29}$$

Moreover:

$$P(X = x_i \text{ or } X = x_j) = P(X = x_i) + P(X = x_j) \quad , \quad i \neq j \quad , \tag{5.30}$$

in such a case. It then follows that:

$$P(X = x_1 \text{ or } X = x_2, \ldots, X = x_{m-1} \text{ or } X = x_m) =$$
$$P(X = x_1) + P(X = x_2) + \cdots + P(X = x_{m-1}) + P(X = x_m) \quad .$$

This can be expressed in terms of the probability mass function as:

$$\sum_{i=1}^{m} P(X = x_i) = \sum_{i=1}^{m} f_X(x_i) = 1 \qquad . \tag{5.31}$$

This is called the *addition law* or the *law of total probability.* It is nothing more than a lemma that follows from axioms (5.2) and (5.5).

We define the *mean*, also called *expected value*, average or *expectation value*, of X as follows:

$$\mathbb{E}(X) \equiv \mu \equiv \sum_{i=1}^{m} x_i f_X(x_i) \qquad . \tag{5.32}$$

The symbol μ is commonly used for *expected value.* Again, knowing the probability mass $f_X(x_i)$ this is easy to calculate. Note that this is the true mean *not* the sample mean. The mean is a mathematical quantity and not based on data or statistical estimation: we will come to that shortly.

Another useful quantity is the *variance* of X. This is defined as:

$$\mathrm{var}(X) \equiv \mathbb{E}\left(X^2\right) - \mu^2 \quad . \tag{5.33}$$

A simple bit of manipulation shows that:

$$\mathrm{var}(X) = \mathbb{E}\left((X - \mu)^2\right) \quad . \tag{5.34}$$

We denote/define the *standard deviation*, σ, of the variable as the square-root of the variance:

$$\sigma \equiv \sqrt{\mathrm{var}(X)} \qquad . \tag{5.35}$$

5.7.1 Markov inequality

Consider a (positive) discrete random variable $X \geq 0$. Then, the following result, the *Markov inequality*, is true for any $a > 0$

$$\boxed{P(X \geq a) \leq \frac{\mu}{a}} \qquad . \tag{5.36}$$

Proof: To begin with, we note that,

$$P(X \geq a) \equiv \sum_{x_i \geq a} f_X(x_i) \qquad . \tag{5.37}$$

Then consider the mean (expected value):

$$\mu = \sum_{x_i} x_i f_X(x_i) = \sum_{x_i < a} x_i f_X(x_i) + \sum_{x_i \geq a} x_i f_X(x_i) \qquad . \tag{5.38}$$

This leads to the expression:

$$\mu - \sum_{x_i \geq a} x_i f_X(x_i) = \sum_{x_i < a} x_i f_X(x_i) \qquad . \tag{5.39}$$

Since the right-hand-side is clearly non-negative, one can write:

$$\mu - \sum_{x_i \geq a} x_i f_X(x_i) \geq 0 \qquad . \tag{5.40}$$

This can be written as, subtracting the same terms from either side:

$$\mu - \sum_{x_i \geq a} a f_X(x_i) \geq \sum_{x_i \geq a} (x_i - a) f_X(x_i) \qquad . \tag{5.41}$$

And again the right-hand side is non-negative and thus:

$$\mu \geq a \sum_{x_i \geq a} f_X(x_i) \qquad . \tag{5.42}$$

From this, and referring to (5.37), it immediately follows that:

$$P(X \geq a) \leq \frac{\mu}{a} \qquad , \tag{5.43}$$

which proves the result.

A corollary of this result is Chebyshev's inequality:

$$\boxed{P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}} \qquad . \tag{5.44}$$

This follows by setting $X \to (X - \mu)^2$ in the Markov inequality and noting that:

$$P(|X - \mu| \geq a) = P((X - \mu)^2 \geq a^2) \qquad . \tag{5.45}$$

5.7.2 Convergence theorems

Consider a set of n discrete random variables:

$$\{X_1, X_2, \ldots, X_n\} \qquad . \tag{5.46}$$

all of which are *independent* and *identically distributed*. That is

$$\mathbb{E}(X_i X_j) = \mathbb{E}(X_i)\,\mathbb{E}(X_j) \quad , \qquad i \neq j \quad , \tag{5.47}$$

and each X_i has the same probability mass function. That is, each X has an identical mean and variance:

$$\mathbb{E}(X_i) = \mu \qquad , \qquad \mathrm{var}(X_i) = \sigma^2 \qquad , \tag{5.48}$$

for all $i = 1, 2, \ldots, n$. Then we have,

$$\mathrm{var}(X_1 + X_2) = \mathrm{var}(X_1) + \mathrm{var}(X_2) \qquad , \tag{5.49}$$

and in general,

$$\mathrm{var}(X_1 + X_2 + \cdots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n) = n\sigma^2 \qquad , \tag{5.50}$$

Let us consider the n values of X as a *sample*. Then the *sample sum* can be defined as:

$$S_n \equiv X_1 + X_2 + \cdots + X_n \tag{5.51}$$

Then the *sample mean* is defined as:

$$\frac{S_n}{n} = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{5.52}$$

which itself is random. Nonetheless

$$\mathbb{E}\left(\frac{S_n}{n}\right) = \frac{1}{n}\left[\mathbb{E}\left(X_1\right) + \mathbb{E}\left(X_2\right) + \cdots + \mathbb{E}\left(X_n\right)\right] = \mu \tag{5.53}$$

That is - the expected value of the sample mean is the true mean. Furthermore:

$$\mathrm{var}\left[\frac{S_n}{n}\right] = \frac{1}{n^2}\left[\mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n)\right] = \frac{\sigma^2}{n} \qquad . \tag{5.54}$$

5.8 The law of large numbers and random walks

One of the cornerstones of statistics is that the sample mean converges to the true mean with certainty as the sample size increases. This is expressed formally by the law of large numbers, originally obtained by Jacob Bernoulli in (1713), and which in its *weak form* can be written as:

$$\boxed{\lim_{n\to\infty} P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) = 0} \qquad , \tag{5.55}$$

for any $\varepsilon > 0$. By letting ε be vanishingly small, we can assert that:

$$\frac{S_n}{n} \approx \mu \qquad , \tag{5.56}$$

for very large n.

That is the sample mean converges to the true mean, in the limit of an infinitely large sample size. Sometimes this result is incorrectly said to be the law of averages.

Proof: Consider the Chebyshev inequality (5.44) and replacing the variable for $X \to S_n/n$. This gives us:

$$X - \mathbb{E}\left(X\right) \to \frac{S_n}{n} - \frac{\mathbb{E}\left(S_n\right)}{n} = \frac{S_n}{n} - \mu \qquad , \tag{5.57}$$

Moreover, according to (5.54):

$$\mathrm{var}\left[\frac{S_n}{n} - \mu\right] = \frac{\sigma^2}{n} \qquad . \tag{5.58}$$

This then leads to,

$$P\left(\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right) \leq \frac{\sigma^2}{n\epsilon} \qquad . \tag{5.59}$$

Then in the limit $n \to \infty$, with ε finite and non-zero, we arrive at the result (5.55).

The formula (5.56) can be misunderstood. It seems to suggest that the random variable S_n is, in some sense predictable. However, this is incorrect. The variance of S_n/n does indeed go to zero as σ^2/n. However, the variance of S_n therefore increases linearly with n. Thus, on average the difference between S_n and its mean increases with $\sqrt{n}$.

Example: Consider an experiment in which a sequence of n tosses of a fair coin is conducted. For a single toss, the theoretical probability of heads is 0.5 and tails 0.5. So the expected value μ for the number of heads in one toss is 0.5.

The law states that, in the experiment S_n/n, that is the total number of heads as a fraction of the total number of tosses, n, will tend towards 50% as n increases. But consider that as n increases there is more variation in the total number of heads. The equation (5.56) does not mean that the *number* of heads is always half the total. In fact:

$$\lim_{n\to\infty} S_n - n\mu \not\to 0 \qquad .$$

In fact, the probability that after N tosses, there are exactly $N/2$ heads and $N/2$ tails is given by

$$P(N/2) = \left(\begin{array}{c} N \\ N/2 \end{array} \right) \left(\frac{1}{2}\right)^N \approx \sqrt{\frac{2}{N\pi}},$$

where we have used the approximation that for large N

$$N! \to \sqrt{2\pi N} N^N e^{-N}.$$

This is closely related to another problem in mathematics called the *random walk.* A walker starts at the position $x = 0$ and then tosses a coin. If it is heads then the walker takes a step right of length 1m, whereas if the coin comes up tails then the walker steps left 1m. So the sequence of coin tosses can be followed by watching the random walk, and predicting the future value of an asset price, and this is equivalent to guessing where the walker will end up in the future. Even if the coin is fair, if the duration of the walk is long $n \to \infty$ the chances that the walker will have returned to the starting point $x = 0$ (that is the probability that the total number of heads matches that of tails) becomes increasingly unlikely the longer the walk.

We will see a clarification of this law in a more precise form, *the central-limit theorem*, later in the course.

5.9 Asset price as a discrete random variable

An experiment for which the *outcomes* are taken randomly from a finite set of values is said to be a *discrete random variable.* Suppose the time now $t = 0$ is the present. Then the current asset price (spot price) S_0 is known with certainty, as are all the asset prices for the previous (past) $t < 0$. We *assert* that, at some time $t = T$ in the future, the corresponding asset price, S_T, can have one of the (discrete) values:

$$S_T \in \{s_1, s_2, \ldots, s_m\} \tag{5.60}$$

but we don't know which one. Thus S_T is a discrete random variable, and in this case the outcomes are mutually exclusive, and thus form a partition of S_T.

For this random variable we define a *probability function*:

$$0 \leq P(S_T = s_i) \leq 1 \quad , \quad i = 1, \ldots, m \qquad . \tag{5.61}$$

If we know this function then we are indeed fortunate since this *completely* describes the randomness of the variable. We cannot be better informed about a random variable than knowing its probability distribution.

As already discussed at the start of chapter 4, we can describe such a discrete random variable through a *probability mass function*, $f_{S_T}(s)$:

$$P(S_T = s_i) = f_{S_T}(s_i) \qquad , \tag{5.62}$$

The probability is zero everywhere except at certain precise values. This is consistent with financial markets where the price is restricted to discrete variables, separated by the *tick size.*

Again, the probabilities are always between the limits (5.61), and the total probability must add to one:

$$\sum_{i=1}^{m} f_{S_T}(s_i) = 1 \qquad . \tag{5.63}$$

As discussed in chapter 4, from the probability mass function, we can calculate various properties of S_T such as its expectation value and its variance.

5.10 Implied Probability

As discussed above, there are *true probabilities* - random events for which the randomness is exactly known. However, in financial markets there are also *implied probabilities* - these are not explicitly known but they are implicit in the pricing of certain financial instruments.

A simple example of this is the kind of speculation called gambling. A betting market is one in which a speculator bets money on a certain outcome, and wins money if the outcome is correct, or loses if it is not correct.

Given that we associate uncertainty with probability, this is reflected in the *betting odds* which are in effect like the *price* of an asset.

5.10.1 Implied Probability in a market

In contrast to flipping a coin, in most practical cases, it is very difficult to estimate the probability of an event happening. However this does not prevent us from making rough guesses based on our perception. Examples commonly arise in financial markets, in which the prices express the probabilities *implied* by the market.

Consider a tennis match between player A and player B, there are only two outcomes (excluding the cancellation of the match); player A loses or player A wins.

In a gambling market the probabilities of these events are expressed by the 'odds' assigned to each event by the bookmakers. The odds are expressed as a ratio of two numbers in the form a/b (or equivalently a–b)

For example, the odds for a match were given as follows: player A wins 1/4, player B wins 11/4. In the language of gambling, this is stated as *four-to-one on* and *eleven-to-four against*. What this means is that, if A wins and you bet £1 on this event you receive a profit of £1/4= £0.25. While if you bet £1 on B and this is successful, you receive £11/4 = £2.75 profit.

Mathematically, this ratio a/b can be ascribed the equivalent *implied probability*:

$$P = \frac{b}{(a+b)} \qquad .$$

If we consider the event W, as A winning, then W^c, is the event B wins. Thus

$$P(W) = \frac{4}{5} = 0.800 \qquad P(W^c) = \frac{4}{15} \approx 0.267 \qquad .$$

Implied probabilities of this type do not add to 1! For the example above we see that:

$$P(W) + P(W^c) = 1.067 \qquad .$$

There is a very good reason for this, it helps ensure the bookmaker makes a profit. If, however, the bookmaker set the odds such that:

$$P(W) + P(W^c) < 1 \quad ,$$

then he/she has made a serious mistake. In these circumstances we can ensure a profit (and that the bookmaker loses), whatever the outcome by betting on both outcomes - see below.

Bookmaker's odds reflect the amount of money wagered by the public on the event. **The probabilities are based on the public perception (market) of the likelihood of each event. These probabilities are not based on the true (mathematical) probabilities.** When the market (implied) probabilities are approximately equal to the true (mathematical) probabilities, we say the that market is 'efficient': that is the market is a reliable estimate of the true probabilities.

5.10.2 Arbitrage

Let's consider an example of a tennis match. There are only two outcomes: a player wins or he/she loses. There are no draws. Suppose the betting odds (*implied probabilities*) are as follows: We see

Table 5.1: Betting odds: arbitrage not possible

Player	Winning odds	implied probability
A	2/9	9/11 = 0.818
B	11/4	4/15 = 0.267

that in this case, the implied probabilities add up to 1.085 - that is more that a true probability which would add to 1.000. So clearly the market has incorrect pricing - but in this case, it is arranged by the betting company so that the 'error' is in its favour. The company is using *arbitrage* to aim at a guaranteed profit.

Suppose instead that the betting market shows the odds in Table 5.2: The total is 0.910.

Table 5.2: Betting odds for a tennis match - where arbitrage is possible

Player	Winning odds	Implied probability
A	5/9	9/14 = 0.643
B	11/4	4/15 = 0.267

Then, even though the outcome is uncertain, one can make a certain profit, by simultaneously betting on player A and player B. In terms of the market one buys A and at the same times sells A for a higher price.

If p_A denotes the probability (price) of A, and p_B the probability (price) of B. Then

$$p_A + p_B < 1 \qquad . \tag{5.64}$$

So, the arbitrageur bets an amount x_A on A and an amount x_B on B, so that the total investment equals $x_A + x_B$. Then if A wins (B loses) the *profit* is:

$$\varepsilon_A = \text{winnings} - \text{investment} = x_A/p_A - x_A - x_B \quad . \tag{5.65}$$

if B wins, the *profit* is:

$$\varepsilon_B = \text{winnings} - \text{investment} = x_B/p_B - x_A - x_B \quad . \tag{5.66}$$

But, it is possible to choose x_A and x_B such that a profit is certain in wither case. We seek the values of x_A and x_B such that:

$$\varepsilon_A = \varepsilon_B > 0 \quad . \tag{5.67}$$

Thus

$$x_A/p_A - x_A - x_B = x_B/p_B - x_A - x_B \tag{5.68}$$

so that the arbitrage strategy should be that the bets need to be in the ratio:

$$x_A/p_A = x_B/p_B. \tag{5.69}$$

For a total amount to wager Y, we have that

$$x_A = Y\frac{p_A}{p_A + p_B}; x_B = Y\frac{p_B}{p_A + p_B}.$$

This then gives a profit

$$\varepsilon = x_B/p_B - x_A - x_B = \frac{x_A}{p_A}(1 - p_A - p_B) = Y\left(\frac{1}{p_A + p_B} - 1\right) > 0 \quad . \tag{5.70}$$

This explains why, when the condition (5.64) occurs, *arbitrage* is possible. That is the arbitrageur has eliminated the risk and will make a profit whatever the outcome of the tennis match, and in fact the value of the profit is known in advance of the match. That is the key aspect of arbitrage.

There is one further nuanced version of arbitrage. In the example above, the arbitrageur required the initial investment $x_A + x_B$ to do the deal. If an arbitrageur can find a *self-financing* deal then that requires minimal (or even no) initial investment.