

2

Measure

2.1 Null sets

The idea of a ‘negligible’ set relates to one of the limitations of the Riemann integral, as we saw in the previous chapter. Since the function $f = \mathbf{1}_{\mathbb{Q}}$ takes a non-zero value only on \mathbb{Q} , and equals 1 there, the ‘area under its graph’ (if such makes sense) must be very closely linked to the ‘length’ of the set \mathbb{Q} . This is why it turns out that we cannot integrate f in the Riemann sense: the sets \mathbb{Q} and $\mathbb{R} \setminus \mathbb{Q}$ are so different from intervals that it is not clear how we should measure their ‘lengths’ and it is clear that the ‘integral’ of f over $[0, 1]$ should equal the ‘length’ of the set of rationals in $[0, 1]$. So how *should* we define this concept for more general sets?

The obvious way of defining the ‘length’ of a set is to start with intervals nonetheless. Suppose that I is a bounded interval of any kind, i.e. $I = [a, b]$, $I = [a, b)$, $I = (a, b]$ or $I = (a, b)$. We simply define the length of I as $l(I) = b - a$ in each case.

As a particular case we have $l(\{a\}) = l([a, a]) = 0$. It is then natural to say that a one-element set is ‘null’. Before we extend this idea to more general sets, first consider the length of a finite set. A finite set is not an interval but since a single point has length 0, adding finitely many such lengths together should still give 0. The underlying concept here is that if we decompose a set into a finite number of disjoint intervals, we compute the length of this set by adding the lengths of the pieces.

As we have seen, in general it may not be always possible to decompose a set

into non-trivial intervals. Therefore, we consider systems of intervals that cover a given set. We shall generalise the above idea by allowing a countable number of covering intervals. Thus we arrive at the following more general definition of sets of ‘zero length’:

Definition 2.1

A *null set* $A \subseteq \mathbb{R}$ is a set that may be covered by a sequence of intervals of arbitrarily small total length, i.e. given any $\varepsilon > 0$ we can find a sequence $\{I_n : n \geq 1\}$ of intervals such that

$$A \subseteq \bigcup_{n=1}^{\infty} I_n$$

and

$$\sum_{n=1}^{\infty} l(I_n) < \varepsilon.$$

(We also say simply that ‘ A is null’.)

Exercise 2.1

Show that we get an equivalent notion if in the above definition we replace the word ‘intervals’ by any of these: ‘open intervals’, ‘closed intervals’, ‘the intervals of the form $(a, b]$ ’, ‘the intervals of the form $[a, b)$ ’.

Note that the intervals do not need to be disjoint. It follows at once from the definition that the empty set is null.

Next, any one-element set $\{x\}$ is a null set. For, let $\varepsilon > 0$ and take $I_1 = (x - \frac{\varepsilon}{4}, x + \frac{\varepsilon}{4})$, $I_n = [0, 0]$ for $n \geq 2$. (Why take $I_n = [0, 0]$ for $n \geq 2$? Well, why not! We could equally have taken $I_n = (0, 0) = \emptyset$, of course!) Now

$$\sum_{n=1}^{\infty} l(I_n) = l(I_1) = \frac{\varepsilon}{2} < \varepsilon.$$

More generally, any countable set $A = \{x_1, x_2, \dots\}$ is null. The simplest way to show this is to take $I_n = [x_n, x_n]$, for all n . However, as a gentle introduction to the next theorem we will cover A by open intervals. This way it is more fun.

For, let $\varepsilon > 0$ and cover A with the following sequence of intervals:

$$\begin{aligned} I_1 &= (x_1 - \frac{\varepsilon}{8}, x_1 + \frac{\varepsilon}{8}) & l(I_1) &= \frac{1}{2}\varepsilon \cdot \frac{1}{2^1} \\ I_2 &= (x_2 - \frac{\varepsilon}{16}, x_2 + \frac{\varepsilon}{16}) & l(I_2) &= \frac{1}{2}\varepsilon \cdot \frac{1}{2^2} \\ I_3 &= (x_3 - \frac{\varepsilon}{32}, x_3 + \frac{\varepsilon}{32}) & l(I_3) &= \frac{1}{2}\varepsilon \cdot \frac{1}{2^3} \\ &\dots & \dots & \\ I_n &= (x_n - \frac{\varepsilon}{2 \cdot 2^n}, x_n + \frac{\varepsilon}{2 \cdot 2^n}) & l(I_n) &= \frac{1}{2}\varepsilon \cdot \frac{1}{2^n} \end{aligned}$$

Since $\sum_{n=1}^{\infty} \frac{1}{2^n} = 1$,

$$\sum_{n=1}^{\infty} l(I_n) = \frac{\varepsilon}{2} < \varepsilon,$$

as needed.

Here we have the following situation: A is the union of countably many one-element sets. Each of them is null and A turns out to be null as well.

We can generalise this simple observation:

Theorem 2.2

If $(N_n)_{n \geq 1}$ is a sequence of null sets, then their union

$$N = \bigcup_{n=1}^{\infty} N_n$$

is also null.

Proof

We assume that all N_n , $n \geq 1$, are null and to show that the same is true for N we take any $\varepsilon > 0$. Our goal is to cover the set N by countably many intervals with total length less than ε .

The proof goes in three steps, each being a little bit tricky.

Step 1. We carefully cover each N_n by intervals.

‘Carefully’ means that the lengths have to be small. ‘Small’ means that we are going to add them up later to end up with a small number (and ‘small’ here means less than ε).

Since N_1 is null, there exist intervals I_k^1 , $k \geq 1$, such that

$$\sum_{k=1}^{\infty} l(I_k^1) < \frac{\varepsilon}{2}, \quad N_1 \subseteq \bigcup_{k=1}^{\infty} I_k^1.$$

For N_2 we find a system of intervals I_k^2 , $k \geq 1$, with

$$\sum_{k=1}^{\infty} l(I_k^2) < \frac{\varepsilon}{4}, \quad N_2 \subseteq \bigcup_{k=1}^{\infty} I_k^2.$$

You can see a cunning plan of making the total lengths smaller at each step at a geometric rate. In general, we cover N_n with intervals I_k^n , $k \geq 1$, whose total length is less than $\frac{\varepsilon}{2^n}$:

$$\sum_{k=1}^{\infty} l(I_k^n) < \frac{\varepsilon}{2^n}, \quad N_n \subseteq \bigcup_{k=1}^{\infty} I_k^n.$$

Step 2. The intervals I_k^n are formed into a sequence.

We arrange the countable family of intervals $\{I_k^n\}_{k \geq 1, n \geq 1}$ into a sequence J_j , $j \geq 1$. For instance we put $J_1 = I_1^1$, $J_2 = I_2^1$, $J_3 = I_1^2$, $J_4 = I_3^1$, etc. so that none of the I_k^n are skipped. The union of the new system of intervals is the same as the union of the old one and so

$$N = \bigcup_{n=1}^{\infty} N_n \subseteq \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{\infty} I_k^n = \bigcup_{j=1}^{\infty} J_j.$$

Step 3. Compute the total length of J_j .

This is tricky because we have a series of numbers with two indices:

$$\sum_{j=1}^{\infty} l(J_j) = \sum_{n=1, k=1}^{\infty} l(I_k^n).$$

Now we wish to write this as a series of numbers, each being the sum of a series. We can rearrange the double sum because the components are non-negative (a fact from elementary calculus)

$$\sum_{n=1, k=1}^{\infty} l(I_k^n) = \sum_{n=1}^{\infty} \left(\sum_{k=1}^{\infty} l(I_k^n) \right) < \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} = \varepsilon,$$

which completes the proof. \square

Thus any countable set is null, and null sets appear to be closely related to countable sets – this is no surprise as any proper interval is uncountable, so any countable subset is quite ‘sparse’ when compared with an interval, hence makes no real contribution to its ‘length’. (You may also have noticed the

similarity between Step 2 in the above proof and the ‘diagonal argument’ which is commonly used to show that \mathbb{Q} is a countable set.)

However, uncountable sets *can* be null, provided their points are sufficiently ‘sparsely distributed’, as the following famous example, due to Cantor, shows:

1. Start with the interval $[0, 1]$, remove the ‘middle third’, i.e. the interval $(\frac{1}{3}, \frac{2}{3})$, obtaining the set C_1 , which consists of the two intervals $[0, \frac{1}{3}]$ and $[\frac{2}{3}, 1]$.
2. Next remove the middle third of each of these two intervals, leaving C_2 , consisting of four intervals, each of length $\frac{1}{9}$, etc. (See Figure 2.1.)
3. At the n th stage we have a set C_n , consisting of 2^n disjoint closed intervals, each of length $\frac{1}{3^n}$. Thus the total length of C_n is $(\frac{2}{3})^n$.



Figure 2.1 Cantor set construction (C_3)

We call

$$C = \bigcap_{n=1}^{\infty} C_n$$

the *Cantor set*.

Now we show that C is null as promised.

Given any $\varepsilon > 0$, choose n so large that $(\frac{2}{3})^n < \varepsilon$. Since $C \subseteq C_n$, and C_n consists of a (finite) sequence of intervals of total length less than ε , we see that C is a null set.

All that remains is to check that C is an uncountable set. This is left for you as

Exercise 2.2

Prove that C is uncountable.

Hint Adapt the proof of the uncountability of \mathbb{R} : begin by expressing each x in $[0, 1]$ in ternary form:

$$x = \sum_{k=1}^{\infty} \frac{a_k}{3^k} = 0.a_1a_2\dots$$

with $a_k = 0, 1$ or 2 . Note that $x \in C$ iff all its a_k equal 0 or 2 .

Why is the Cantor set null, even though it is uncountable? Clearly it is the distribution of its points, the fact that it is ‘spread out’ all over $[0,1]$, which causes the trouble. This makes it the source of many examples which show that intuitively ‘obvious’ things are not always true! For example, we can use the Cantor set to define a function, due to Lebesgue, with very odd properties:

If $x \in [0,1]$ has ternary expansion (a_n) , i.e. $x = 0.a_1a_2\dots$ with $a_n = 0, 1$ or 2 , define N as the first index n for which $a_n = 1$, and set $N = \infty$ if none of the a_n are 1 (i.e. when $x \in C$). Now set $b_n = \frac{a_n}{2}$ for $n < N$ and $b_N = 1$, and let $F(x) = \sum_{n=1}^N \frac{b_n}{2^n}$ for each $x \in [0,1]$. Clearly, this function is monotone increasing and has $F(0) = 0$, $F(1) = 1$. Yet it is constant on the middle thirds (i.e. the complement of C), so all its increase occurs on the Cantor set. Since we have shown that C is a null set, F ‘grows’ from 0 to 1 entirely on a ‘negligible’ set. The following exercise shows that it has no jumps!

Exercise 2.3

Prove that the Lebesgue function F is continuous and sketch its partial graph.

2.2 Outer measure

The simple concept of null sets provides the key to our idea of length, since it tells us what we can ‘ignore’. A quite general notion of ‘length’ is now provided by:

Definition 2.3

The (Lebesgue) *outer measure* of any set $A \subseteq \mathbb{R}$ is given by

$$m^*(A) = \inf Z_A$$

where

$$Z_A = \left\{ \sum_{n=1}^{\infty} l(I_n) : I_n \text{ are intervals, } A \subseteq \bigcup_{n=1}^{\infty} I_n \right\}.$$

We say the $(I_n)_{n \geq 1}$ *cover* the set A . So the outer measure is the infimum of lengths of all possible covers of A . (Note again that some of the I_n may be empty; this avoids having to worry whether the sequence (I_n) has finitely or infinitely many different members.)

Clearly $m^*(A) \geq 0$ for any $A \subseteq \mathbb{R}$. For some sets A , the series $\sum_{n=1}^{\infty} l(I_n)$ may diverge for any covering of A , so $m^*(A)$ may be equal to ∞ . Since we wish to be able to add the outer measures of various sets we have to adopt a convention to deal with infinity. An obvious choice is $a + \infty = \infty$, $\infty + \infty = \infty$ and a less obvious but quite practical assumption is $0 \times \infty = 0$, as we have already seen.

The set Z_A is bounded from below by 0, so the infimum always exists. If $r \in Z_A$, then $[r, +\infty] \subseteq Z_A$ (clearly, we may expand the first interval of any cover to increase the total length by any number). This shows that Z_A is either $\{+\infty\}$ or the interval $(x, +\infty]$ or $[x, +\infty]$ for some real number x . So the infimum of Z_A is just x .

First we show that the concept of null set is consistent with that of outer measure:

Theorem 2.4

$A \subseteq \mathbb{R}$ is a null set if and only if $m^*(A) = 0$.

Proof

Suppose that A is a null set. We wish to show that $\inf Z_A = 0$. To this end we show that for any $\varepsilon > 0$ we can find an element $z \in Z_A$ such that $z < \varepsilon$.

By the definition of null set we can find a sequence (I_n) of intervals covering A with $\sum_{n=1}^{\infty} l(I_n) < \varepsilon$ and so $\sum_{n=1}^{\infty} l(I_n)$ is the required element z of Z_A .

Conversely, if $A \subseteq \mathbb{R}$ has $m^*(A) = 0$, then by the definition of \inf , given any $\varepsilon > 0$, there is $z \in Z_A$, $z < \varepsilon$. But a member of Z_A is the total length of some covering of A . That is, there is a covering (I_n) of A with total length less than ε , so A is null. \square

This combines our general outer measure with the special case of ‘zero measure’. Note that $m^*(\emptyset) = 0$, $m^*(\{x\}) = 0$ for any $x \in \mathbb{R}$, and $m^*(\mathbb{Q}) = 0$ (and in fact, for any countable X , $m^*(X) = 0$).

Next we observe that m^* is monotone: the bigger the set, the greater its outer measure.

Proposition 2.5

If $A \subset B$ then $m^*(A) \leq m^*(B)$.

Hint Show that $Z_B \subset Z_A$ and use the definition of \inf .

The second step is to relate outer measure to the length of an interval. This innocent result contains the crux of the theory, since it shows that the formal definition of m^* , which is applicable to *all* subsets of \mathbb{R} , coincides with the intuitive idea for intervals, where our thought processes began. We must therefore expect the proof to contain some hidden depths, and we have to tackle these in stages: the hard work lies in showing that the length of the interval cannot be greater than its outer measure: for this we need to appeal to the famous Heine–Borel theorem, which states that every closed, bounded subset B of \mathbb{R} is *compact*: given any collection of open sets O_α covering B (i.e. $B \subset \bigcup_\alpha O_\alpha$), there is a finite subcollection $(O_{\alpha_i})_{i \leq n}$ which still covers B , i.e. $B \subset \bigcup_{i=1}^n O_{\alpha_i}$ (for a proof see [1]).

Theorem 2.6

The outer measure of an interval equals its length.

Proof

If I is unbounded, then it is clear that it cannot be covered by a system of intervals with finite total length. This shows that $m^*(I) = \infty$ and so $m^*(I) = l(I) = \infty$.

So we restrict ourselves to bounded intervals.

Step 1. $m^*(I) \leq l(I)$.

We claim that $l(I) \in Z_I$. Take the following sequence of intervals: $I_1 = I$, $I_n = [0, 0]$ for $n \geq 2$. This sequence covers the set I , and the total length is equal to the length of I , hence $l(I) \in Z_I$. This is sufficient since the infimum of Z_I cannot exceed any of its elements.

Step 2. $l(I) \leq m^*(I)$.

(i) $I = [a, b]$. We shall show that for any $\varepsilon > 0$,

$$l([a, b]) \leq m^*([a, b]) + \varepsilon. \quad (2.1)$$

This is sufficient since we may obtain the required inequality passing to the limit, $\varepsilon \rightarrow 0$. (Note that if $x, y \in \mathbb{R}$ and $y > x$ then there is an $\varepsilon > 0$ with $y > x + \varepsilon$, e.g. $\varepsilon = \frac{1}{2}(y - x)$.)

So we take an arbitrary $\varepsilon > 0$. By the definition of outer measure we can find a sequence of intervals I_n covering $[a, b]$ such that

$$\sum_{n=1}^{\infty} l(I_n) \leq m^*([a, b]) + \frac{\varepsilon}{2}. \quad (2.2)$$

We shall slightly increase each of the intervals to an open one. Let the endpoints of I_n be a_n, b_n , and we take

$$J_n = \left(a_n - \frac{\varepsilon}{2^{n+2}}, b_n + \frac{\varepsilon}{2^{n+2}}\right).$$

It is clear that

$$l(I_n) = l(J_n) - \frac{\varepsilon}{2^{n+1}},$$

so that

$$\sum_{n=1}^{\infty} l(I_n) = \sum_{n=1}^{\infty} l(J_n) - \frac{\varepsilon}{2}.$$

We insert this in (2.2) and we have

$$\sum_{n=1}^{\infty} l(J_n) \leq m^*([a, b]) + \varepsilon. \quad (2.3)$$

The new sequence of intervals of course covers $[a, b]$, so by the Heine–Borel theorem we can choose a finite number of J_n to cover $[a, b]$ (the set $[a, b]$ is compact in \mathbb{R}). We can add some intervals to this finite family to form an initial segment of the sequence (J_n) – just for simplicity of notation. So for some finite index m we have

$$[a, b] \subseteq \bigcup_{n=1}^m J_n. \quad (2.4)$$

Let $J_n = (c_n, d_n)$. Put $c = \min\{c_1, \dots, c_m\}$, $d = \max\{d_1, \dots, d_m\}$. The covering (2.4) means that $c < a$ and $b < d$, hence $l([a, b]) < d - c$.

Next, the number $d - c$ is certainly smaller than the total length of J_n , $n = 1, \dots, m$ (some overlapping takes place) and

$$l([a, b]) < d - c < \sum_{n=1}^m l(J_n). \quad (2.5)$$

Now it is sufficient to put (2.3) and (2.5) together in order to deduce (2.1) (the finite sum is less than or equal to the sum of the series since all terms are non-negative).

(ii) $I = (a, b)$. As before, it is sufficient to show (2.1). Let us fix any $\varepsilon > 0$.

$$\begin{aligned} l((a, b)) &= l\left([a + \frac{\varepsilon}{2}, b - \frac{\varepsilon}{2}]\right) + \varepsilon \\ &\leq m^*\left([a + \frac{\varepsilon}{2}, b - \frac{\varepsilon}{2}]\right) + \varepsilon \quad (\text{by (1)}) \\ &\leq m^*((a, b)) + \varepsilon \quad (\text{by Proposition 2.5}). \end{aligned}$$

(iii) $I = [a, b)$ or $I = (a, b]$.

$$\begin{aligned} l(I) = l((a, b)) &\leq m^*((a, b)) \quad (\text{by (2)}) \\ &\leq m^*(I) \quad (\text{by Proposition 2.5}), \end{aligned}$$

which completes the proof. \square

Having shown that outer measure coincides with the natural concept of length for intervals, we now need to investigate its properties. The next theorem gives us an important technical tool which will be used in many proofs.

Theorem 2.7

Outer measure is countably subadditive, i.e. for any sequence of sets $\{E_n\}$,

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} m^*(E_n).$$

(Note that both sides may be infinite here.)

Proof (a warm-up)

Let us prove first a simpler statement:

$$m^*(E_1 \cup E_2) \leq m^*(E_1) + m^*(E_2).$$

Take an $\varepsilon > 0$ and we show an even easier inequality

$$m^*(E_1 \cup E_2) \leq m^*(E_1) + m^*(E_2) + \varepsilon.$$

This is however sufficient because taking $\varepsilon = \frac{1}{n}$ and letting $n \rightarrow \infty$ we get what we need.

So for any $\varepsilon > 0$ we find covering sequences $(I_k^1)_{k \geq 1}$ of E_1 and $(I_k^2)_{k \geq 1}$ of E_2 such that

$$\begin{aligned} \sum_{k=1}^{\infty} l(I_k^1) &\leq m^*(E_1) + \frac{\varepsilon}{2}, \\ \sum_{k=1}^{\infty} l(I_k^2) &\leq m^*(E_2) + \frac{\varepsilon}{2}; \end{aligned}$$

hence, adding up,

$$\sum_{k=1}^{\infty} l(I_k^1) + \sum_{k=1}^{\infty} l(I_k^2) \leq m^*(E_1) + m^*(E_2) + \varepsilon.$$

The sequence of intervals $(I_1^1, I_1^2, I_2^1, I_2^2, I_3^1, I_3^2, \dots)$ covers $E_1 \cup E_2$, hence

$$m^*(E_1 \cup E_2) \leq \sum_{k=1}^{\infty} l(I_k^1) + \sum_{k=1}^{\infty} l(I_k^2),$$

which combined with the previous inequality gives the result. \square

Proof of the theorem

If the right-hand side is infinite, then the inequality is of course true. So, suppose that $\sum_{n=1}^{\infty} m^*(E_n) < \infty$. For each given $\varepsilon > 0$ and $n \geq 1$ find a covering sequence $(I_k^n)_{k \geq 1}$ of E_n with

$$\sum_{k=1}^{\infty} l(I_k^n) \leq m^*(E_n) + \frac{\varepsilon}{2^n}.$$

The iterated series converges:

$$\sum_{n=1}^{\infty} \left(\sum_{k=1}^{\infty} l(I_k^n) \right) \leq \sum_{n=1}^{\infty} m^*(E_n) + \varepsilon < \infty$$

and since all its terms are non-negative,

$$\sum_{n=1}^{\infty} \left(\sum_{k=1}^{\infty} l(I_k^n) \right) = \sum_{n,k=1}^{\infty} l(I_k^n).$$

The system of intervals $(I_k^n)_{k,n \geq 1}$ covers $\bigcup_{n=1}^{\infty} E_n$, hence

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) \leq \sum_{n,k=1}^{\infty} l(I_k^n) \leq \sum_{n=1}^{\infty} m^*(E_n) + \varepsilon.$$

To complete the proof we let $\varepsilon \rightarrow 0$. \square

A similar result is of course true for a finite family $(E_n)_{n=1}^m$:

$$m^*\left(\bigcup_{n=1}^m E_n\right) \leq \sum_{n=1}^m m^*(E_n).$$

It is a corollary to Theorem 2.7 with $E_k = \emptyset$ for $k > m$.

Exercise 2.4

Prove that if $m^*(A) = 0$ then for each B , $m^*(A \cup B) = m^*(B)$.

Hint Employ both monotonicity and subadditivity of outer measure.

Exercise 2.5

Prove that if $m^*(A \Delta B) = 0$, then $m^*(A) = m^*(B)$.

Hint Note that $A \subseteq B \cup (A \Delta B)$.

We conclude this section with a simple and intuitive property of outer measure. Note that the length of an interval does not change if we shift it along the real line: $l([a, b]) = l([a + t, b + t]) = b - a$ for example. Since the outer measure is defined in terms of the lengths of intervals, it is natural to expect it to share this property. For $A \subset \mathbb{R}$ and $t \in \mathbb{R}$ we put $A + t = \{a + t : a \in A\}$.

Proposition 2.8

Outer measure is translation-invariant, i.e.

$$m^*(A) = m^*(A + t)$$

for each A and t .

Hint Combine two facts: the length of interval does not change when the interval is shifted and outer measure is determined by the length of the coverings.

2.3 Lebesgue-measurable sets and Lebesgue measure

With outer measure, subadditivity (as in Theorem 2.7) is as far as we can get. We wish, however, to ensure that if sets (E_n) are pairwise disjoint (i.e. $E_i \cap E_j = \emptyset$ if $i \neq j$), then the inequality in Theorem 2.7 becomes an equality. It turns out that this will not in general be true for outer measure, although examples where it fails are quite difficult to construct (we give such examples in the Appendix). But our wish is an entirely reasonable one: any ‘length function’ should at least be finitely additive, since decomposing a set into finitely many disjoint pieces should not alter its length. Moreover, since we constructed our

length function via approximation of complicated sets by ‘simpler’ sets (i.e. intervals) it seems fair to demand a *continuity property*: if pairwise disjoint (E_n) have union E , then the lengths of the sets $B_n = E \setminus \bigcup_{k=1}^n E_k$ may be expected to decrease to 0 as $n \rightarrow \infty$. Combining this with finite additivity leads quite naturally to the demand that ‘length’ should be *countably additive*, i.e. that

$$m^* \left(\bigcup_{n=1}^{\infty} E_n \right) = \sum_{n=1}^{\infty} m^*(E_n) \quad \text{when } E_i \cap E_j = \emptyset \text{ for } i \neq j.$$

We therefore turn to the task of finding the class of sets in \mathbb{R} which have this property. It turns out that it is also the key property of the abstract concept of measure, and we will use it to provide mathematical foundations for probability.

In order to define the ‘good’ sets which have this property, it also seems plausible that such a set should apportion the outer measure of *every* set in \mathbb{R} properly, as we state in Definition 2.9 below. Remarkably, this simple demand will suffice to guarantee that our ‘good’ sets have all the properties we demand of them!

Definition 2.9

A set $E \subseteq \mathbb{R}$ is (Lebesgue-) *measurable* if for every set $A \subseteq \mathbb{R}$ we have

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c) \quad (2.6)$$

where $E^c = \mathbb{R} \setminus E$, and we write $E \in \mathcal{M}$.

We obviously have $A = (A \cap E) \cup (A \cap E^c)$, hence by Theorem 2.7 we have

$$m^*(A) \leq m^*(A \cap E) + m^*(A \cap E^c)$$

for any A and E . So our future task of verifying (2.6) has simplified: $E \in \mathcal{M}$ if and only if the following inequality holds:

$$m^*(A) \geq m^*(A \cap E) + m^*(A \cap E^c) \text{ for all } A \subseteq \mathbb{R}. \quad (2.7)$$

Now we give some examples of measurable sets.

Theorem 2.10

- (i) Any null set is measurable.
- (ii) Any interval is measurable.

Proof

(i) If N is a null set, then (Proposition 2.4) $m^*(N) = 0$. So for any $A \subset \mathbb{R}$ we have

$$\begin{aligned} m^*(A \cap N) &\leq m^*(N) = 0 \quad \text{since } A \cap N \subseteq N \\ m^*(A \cap N^c) &\leq m^*(A) \quad \text{since } A \cap N^c \subseteq A \end{aligned}$$

and adding together we have proved (2.7).

(ii) Let $E = I$ be an interval. Suppose, for example, that $I = [a, b]$. Take any $A \subseteq \mathbb{R}$ and $\varepsilon > 0$. Find a covering of A with

$$m^*(A) \leq \sum_{n=1}^{\infty} l(I_n) \leq m^*(A) + \varepsilon.$$

Clearly the intervals $I'_n = I_n \cap [a, b]$ cover $A \cap [a, b]$ hence

$$m^*(A \cap [a, b]) \leq \sum_{n=1}^{\infty} l(I'_n).$$

The intervals $I''_n = I_n \cap (-\infty, a)$, $I'''_n = I_n \cap (b, +\infty)$ cover $A \cap [a, b]^c$ so

$$m^*(A \cap [a, b]^c) \leq \sum_{n=1}^{\infty} l(I''_n) + \sum_{n=1}^{\infty} l(I'''_n).$$

Putting the above three inequalities together we obtain (2.7).

If I is unbounded, $I = [a, \infty)$ say, then the proof is even simpler since it is sufficient to consider $I'_n = I_n \cap [a, \infty)$ and $I''_n = I_n \cap (-\infty, a)$. \square

The fundamental properties of the class \mathcal{M} of all Lebesgue-measurable subsets of \mathbb{R} can now be proved. They fall into two categories: first we show that certain set operations on sets in \mathcal{M} again produce sets in \mathcal{M} (these are what we call ‘closure properties’) and second we prove that for sets in \mathcal{M} the outer measure m^* has the property of countable additivity announced above.

Theorem 2.11

- (i) $\mathbb{R} \in \mathcal{M}$,
- (ii) if $E \in \mathcal{M}$ then $E^c \in \mathcal{M}$,
- (iii) if $E_n \in \mathcal{M}$ for all $n = 1, 2, \dots$ then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{M}$.

Moreover, if $E_n \in \mathcal{M}$, $n = 1, 2, \dots$ and $E_j \cap E_k = \emptyset$ for $j \neq k$, then

$$m^*\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} m^*(E_n). \quad (2.8)$$

Remark 2.12

This result is the most important theorem in this chapter and provides the basis for all that follows. It also allows us to give names to the quantities under discussion.

Conditions (i)–(iii) mean that \mathcal{M} is a σ -field. In other words, we say that a family of sets is a σ -field if it contains the base set and is closed under complements and countable unions. A $[0, \infty]$ -valued function defined on a σ -field is called a *measure* if it satisfies (2.8) for pairwise disjoint sets, i.e. it is *countably additive*.

An alternative, rather more abstract and general, approach to measure theory is to begin with the above properties as *axioms*, i.e. to call a triple $(\Omega, \mathcal{F}, \mu)$ a *measure space* if Ω is an abstractly given set, \mathcal{F} is a σ -field of subsets of Ω , and $\mu : \mathcal{F} \mapsto [0, \infty]$ is a function satisfying (2.8) (with μ instead of m^*). The task of defining Lebesgue measure on \mathbb{R} then becomes that of verifying, with \mathcal{M} and $m = m^*$ on \mathcal{M} defined as above, that the triple $(\mathbb{R}, \mathcal{M}, m)$ satisfies these axioms, i.e. becomes a measure space.

Although the requirements of probability theory will mean that we have to consider such general measure spaces in due course, we have chosen our more concrete approach to the fundamental example of Lebesgue measure in order to demonstrate how this important measure space arises quite naturally from considerations of the ‘lengths’ of sets in \mathbb{R} , and leads to a theory of integration which greatly extends that of Riemann. It is also sufficient to allow us to develop most of the important examples of probability distributions.

Proof of the theorem

(i) Let $A \subseteq \mathbb{R}$. Note that $A \cap \mathbb{R} = A$, $\mathbb{R}^c = \emptyset$, so that $A \cap \mathbb{R}^c = \emptyset$. Now (2.6) reads $m^*(A) = m^*(A) + m^*(\emptyset)$ and is of course true since $m^*(\emptyset) = 0$.

(ii) Suppose $E \in \mathcal{M}$ and take any $A \subseteq \mathbb{R}$. We have to show (2.6) for E^c , i.e.

$$m^*(A) = m^*(A \cap E^c) + m^*(A \cap (E^c)^c),$$

but since $(E^c)^c = E$ this reduces to the condition for E which holds by hypothesis.

We split the proof of (iii) into several steps. But first:

A warm-up Suppose that $E_1 \cap E_2 = \emptyset$, $E_1, E_2 \in \mathcal{M}$. We shall show that $E_1 \cup E_2 \in \mathcal{M}$ and $m^*(E_1 \cup E_2) = m^*(E_1) + m^*(E_2)$.

Let $A \subseteq \mathbb{R}$. We have the condition for E_1 :

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c). \quad (2.9)$$

Now, apply (2.6) for E_2 with $A \cap E_1^c$ in place of A :

$$\begin{aligned} m^*(A \cap E_1^c) &= m^*((A \cap E_1^c) \cap E_2) + m^*((A \cap E_1^c) \cap E_2^c). \\ &= m^*(A \cap (E_1^c \cap E_2)) + m^*(A \cap (E_1^c \cap E_2^c)) \end{aligned}$$

(the situation is depicted in Figure 2.2).

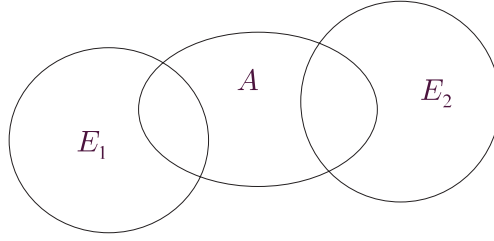


Figure 2.2 The sets A , E_1 , E_2

Since E_1 and E_2 are disjoint, $E_1^c \cap E_2 = E_2$. By de Morgan's law $E_1^c \cap E_2^c = (E_1 \cup E_2)^c$. We substitute and we have

$$m^*(A \cap E_1^c) = m^*(A \cap E_2) + m^*(A \cap (E_1 \cup E_2)^c).$$

Inserting this into (2.9) we get

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_2) + m^*(A \cap (E_1 \cup E_2)^c). \quad (2.10)$$

Now by the subadditivity property of m^* we have

$$\begin{aligned} m^*(A \cap E_1) + m^*(A \cap E_2) &\geq m^*((A \cap E_1) \cup (A \cap E_2)) \\ &= m^*(A \cap (E_1 \cup E_2)), \end{aligned}$$

so (2.10) gives

$$m^*(A) \geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap (E_1 \cup E_2)^c),$$

which is sufficient for $E_1 \cup E_2$ to belong to \mathcal{M} (the inverse inequality is always true, as observed before (2.7)).

Finally, put $A = E_1 \cup E_2$ in (2.10) to get $m^*(E_1 \cup E_2) = m^*(E_1) + m^*(E_2)$, which completes the argument. \square

We return to the proof of the theorem.

Step 1. If pairwise disjoint E_k , $k = 1, 2, \dots$, are in \mathcal{M} then their union is in \mathcal{M} and (2.8) holds.

We begin as in the proof of the warm-up and we have

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c)$$

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_2) + m^*(A \cap (E_1 \cup E_2)^c)$$

(see (2.10)) and after n steps we expect

$$m^*(A) = \sum_{k=1}^n m^*(A \cap E_k) + m^*(A \cap (\bigcup_{k=1}^n E_k)^c). \quad (2.11)$$

Let us demonstrate this by induction. The case $n = 1$ is the first line above. Suppose that

$$m^*(A) = \sum_{k=1}^{n-1} m^*(A \cap E_k) + m^*(A \cap (\bigcup_{k=1}^{n-1} E_k)^c). \quad (2.12)$$

Since $E_n \in \mathcal{M}$, we may apply (2.6) with $A \cap (\bigcup_{k=1}^{n-1} E_k)^c$ in place of A :

$$m^*(A \cap (\bigcup_{k=1}^{n-1} E_k)^c) = m^*(A \cap (\bigcup_{k=1}^{n-1} E_k)^c \cap E_n) + m^*(A \cap (\bigcup_{k=1}^{n-1} E_k)^c \cap E_n^c). \quad (2.13)$$

Now we make the same observations as in the warm-up:

$$(\bigcup_{k=1}^{n-1} E_k)^c \cap E_n = E_n \quad (E_i \text{ are pairwise disjoint}),$$

$$(\bigcup_{k=1}^{n-1} E_k)^c \cap E_n^c = (\bigcup_{k=1}^n E_k)^c \quad (\text{by de Morgan's law}).$$

Inserting these into (2.13) we get

$$m^*(A \cap (\bigcup_{k=1}^{n-1} E_k)^c) = m^*(A \cap E_n) + m^*(A \cap (\bigcup_{k=1}^n E_k)^c),$$

and inserting this into the induction hypothesis (2.12) we get

$$m^*(A) = \sum_{k=1}^{n-1} m^*(A \cap E_k) + m^*(A \cap E_n) + m^*(A \cap (\bigcup_{k=1}^n E_k)^c)$$

as required to complete the induction step. Thus (2.11) holds for all n by induction.

As will be seen at the next step the fact that E_k are pairwise disjoint is not necessary in order to ensure that their union belongs to \mathcal{M} . However, with this assumption we have equality in (2.11) which does not hold otherwise. This equality will allow us to prove countable additivity (2.8).

Since

$$\left(\bigcup_{k=1}^n E_k\right)^c \supseteq \left(\bigcup_{k=1}^{\infty} E_k\right)^c,$$

from (2.11) by monotonicity (Proposition 2.5) we get

$$m^*(A) \geq \sum_{k=1}^n m^*(A \cap E_k) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right).$$

The inequality remains true after we pass to the limit $n \rightarrow \infty$:

$$m^*(A) \geq \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right). \quad (2.14)$$

By countable subadditivity (Theorem 2.7)

$$\sum_{k=1}^{\infty} m^*(A \cap E_k) \geq m^*\left(A \cap \bigcup_{k=1}^{\infty} E_k\right)$$

and so

$$m^*(A) \geq m^*\left(A \cap \bigcup_{k=1}^{\infty} E_k\right) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right) \quad (2.15)$$

as required. So we have shown that $\bigcup_{k=1}^{\infty} E_k \in \mathcal{M}$ and hence the two sides of (2.15) are equal. The right-hand side of (2.14) is squeezed between the left and right of (2.15), which yields

$$m^*(A) = \sum_{k=1}^{\infty} m^*(A \cap E_k) + m^*\left(A \cap \left(\bigcup_{k=1}^{\infty} E_k\right)^c\right). \quad (2.16)$$

The equality here is a consequence of the assumption that E_k are pairwise disjoint. It holds for any set A , so we may insert $A = \bigcup_{j=1}^{\infty} E_j$. The last term on the right is zero because we have $m^*(\emptyset)$. Next $(\bigcup_{j=1}^{\infty} E_j) \cap E_n = E_n$ and so we have (2.8).

Step 2. If $E_1, E_2 \in \mathcal{M}$, then $E_1 \cup E_2 \in \mathcal{M}$ (not necessarily disjoint).

Again we begin as in the warm-up:

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c). \quad (2.17)$$

Next, applying (2.6) to E_2 with $A \cap E_1^c$ in place of A we get

$$m^*(A \cap E_1^c) = m^*(A \cap E_1^c \cap E_2) + m^*(A \cap E_1^c \cap E_2^c).$$

We insert this into (2.17) to get

$$m^*(A) = m^*(A \cap E_1) + m^*(A \cap E_1^c \cap E_2) + m^*(A \cap E_1^c \cap E_2^c). \quad (2.18)$$

By de Morgan's law, $E_1^c \cap E_2^c = (E_1 \cup E_2)^c$, so (as before)

$$m^*(A \cap E_1^c \cap E_2^c) = m^*(A \cap (E_1 \cup E_2)^c). \quad (2.19)$$

By subadditivity of m^* we have

$$m^*(A \cap E_1) + m^*(A \cap E_1^c \cap E_2) \geq m^*(A \cap (E_1 \cup E_2)). \quad (2.20)$$

Inserting (2.19) and (2.20) into (2.18) we get

$$m^*(A) \geq m^*(A \cap (E_1 \cup E_2)) + m^*(A \cap (E_1 \cup E_2)^c)$$

as required.

Step 3. If $E_k \in \mathcal{M}$, $k = 1, \dots, n$, then $E_1 \cup \dots \cup E_n \in \mathcal{M}$ (not necessarily disjoint).

We argue by induction. There is nothing to prove for $n = 1$. Suppose the claim is true for $n - 1$. Then

$$E_1 \cup \dots \cup E_n = (E_1 \cup \dots \cup E_{n-1}) \cup E_n$$

so that the result follows from Step 2.

Step 4. If $E_1, E_2 \in \mathcal{M}$, then $E_1 \cap E_2 \in \mathcal{M}$.

We have $E_1^c, E_2^c \in \mathcal{M}$ by (ii), $E_1^c \cup E_2^c \in \mathcal{M}$ by Step 2, $(E_1^c \cup E_2^c)^c \in \mathcal{M}$ by (ii) again, but by de Morgan's law the last set is equal to $E_1 \cap E_2$.

Step 5. The general case: if E_1, E_2, \dots are in \mathcal{M} , then so is $\bigcup_{k=1}^{\infty} E_k$.

Let $E_k \in \mathcal{M}$, $k = 1, 2, \dots$. We define an auxiliary sequence of pairwise disjoint sets F_k with the same union as E_k :

$$\begin{aligned} F_1 &= E_1 \\ F_2 &= E_2 \setminus E_1 = E_2 \cap E_1^c \\ F_3 &= E_3 \setminus (E_1 \cup E_2) = E_3 \cap (E_1 \cup E_2)^c \\ &\dots \\ F_k &= E_k \setminus (E_1 \cup \dots \cup E_{k-1}) = E_k \cap (E_1 \cup \dots \cup E_{k-1})^c; \end{aligned}$$

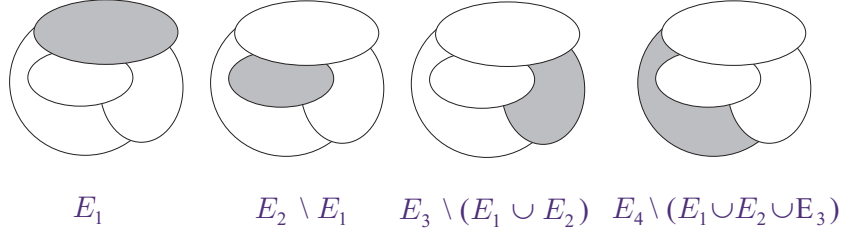


Figure 2.3 The sets F_k

(see Figure 2.3).

By Steps 3 and 4 we know that all F_k are in \mathcal{M} . By the very construction they are pairwise disjoint, so by Step 1 their union is in \mathcal{M} . We shall show that

$$\bigcup_{k=1}^{\infty} F_k = \bigcup_{k=1}^{\infty} E_k.$$

This will complete the proof since the latter is now in \mathcal{M} . The inclusion

$$\bigcup_{k=1}^{\infty} F_k \subseteq \bigcup_{k=1}^{\infty} E_k$$

is obvious since for each k , $F_k \subseteq E_k$ by definition. For the inverse let $a \in \bigcup_{k=1}^{\infty} E_k$. Put $S = \{n \in \mathbb{N} : a \in E_n\}$ which is non-empty since a belongs to the union. Let $n_0 = \min S \in S$. If $n_0 = 1$, then $a \in E_1 = F_1$. Suppose $n_0 > 1$. So $a \in E_{n_0}$ and, by the definition of n_0 , $a \notin E_1, \dots, a \notin E_{n_0-1}$. By the definition of F_{n_0} this means that $a \in F_{n_0}$, so a is in $\bigcup_{k=1}^{\infty} F_k$. \square

Using de Morgan's laws you should easily verify an additional property of \mathcal{M} .

Proposition 2.13

If $E_k \in \mathcal{M}$, $k = 1, 2, \dots$, then

$$E = \bigcap_{k=1}^{\infty} E_k \in \mathcal{M}.$$

We can therefore summarise the properties of the family \mathcal{M} of Lebesgue-measurable sets as follows:

\mathcal{M} is closed under countable unions, countable intersections, and complements. It contains intervals and all null sets.

Definition 2.14

We shall write $m(E)$ instead of $m^*(E)$ for any E in \mathcal{M} and call $m(E)$ the *Lebesgue measure* of the set E .

Thus Theorems 2.11 and 2.6 now read as follows, and describe the construction which we have laboured so hard to establish:

Lebesgue measure $m : \mathcal{M} \rightarrow [0, \infty]$ is a countably additive set function defined on the σ -field \mathcal{M} of measurable sets. Lebesgue measure of an interval is equal to its length. Lebesgue measure of a null set is zero.

2.4 Basic properties of Lebesgue measure

Since Lebesgue measure is nothing else than the outer measure restricted to a special class of sets, some properties of the outer measure are automatically inherited by Lebesgue measure:

Proposition 2.15

Suppose that $A, B \in \mathcal{M}$.

- (i) If $A \subset B$ then $m(A) \leq m(B)$.
- (ii) If $A \subset B$ and $m(A)$ is finite then $m(B \setminus A) = m(B) - m(A)$.
- (iii) m is translation-invariant.

Since $\emptyset \in \mathcal{M}$ we can take $E_i = \emptyset$ for all $i > n$ in (2.8) to conclude that Lebesgue measure is additive: if $E_i \in \mathcal{M}$ are pairwise disjoint, then

$$m\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n m(E_i).$$

Exercise 2.6

Find a formula describing $m(A \cup B)$ and $m(A \cup B \cup C)$ in terms of measures of the individual sets and their intersections (we do not assume that the sets are pairwise disjoint).

Recalling that the *symmetric difference* $A\Delta B$ of two sets is defined by $A\Delta B = (A \setminus B) \cup (B \setminus A)$ the following result is also easy to check:

Proposition 2.16

If $A \in \mathcal{M}$, $m(A\Delta B) = 0$, then $B \in \mathcal{M}$ and $m(A) = m(B)$.

Hint Recall that null sets belong to \mathcal{M} and that subsets of null sets are null.

As we noted in Chapter 1, every open set in \mathbb{R} can be expressed as the union of a countable number of open intervals. This ensures that open sets in \mathbb{R} are Lebesgue-measurable, since \mathcal{M} contains intervals and is closed under countable unions. We can approximate the Lebesgue measure of any $A \in \mathcal{M}$ from above by the measures of a sequence of open sets containing A . This is clear from the following result:

Theorem 2.17

(i) For any $\varepsilon > 0$, $A \subset \mathbb{R}$ we can find an open set O such that

$$A \subset O, \quad m(O) \leq m^*(A) + \varepsilon.$$

Consequently, for any $E \in \mathcal{M}$ we can find an open set O containing E such that $m(O \setminus E) < \varepsilon$.

(ii) For any $A \subset \mathbb{R}$ we can find a sequence of open sets O_n such that

$$A \subset \bigcap_n O_n, \quad m\left(\bigcap_n O_n\right) = m^*(A).$$

Proof

(i) By definition of $m^*(A)$ we can find a sequence (I_n) of intervals with $A \subset \bigcup_n I_n$ and

$$\sum_{n=1}^{\infty} l(I_n) - \frac{\varepsilon}{2} \leq m^*(A).$$

Each I_n is contained in an open interval whose length is very close to that of I_n ; if the left and right endpoints of I_n are a_n and b_n respectively let $J_n = (a_n - \frac{\varepsilon}{2^{n+2}}, b_n + \frac{\varepsilon}{2^{n+2}})$. Set $O = \bigcup_n J_n$, which is open. Then $A \subset O$ and

$$m(O) \leq \sum_{n=1}^{\infty} l(J_n) \leq \sum_{n=1}^{\infty} l(I_n) + \frac{\varepsilon}{2} \leq m^*(A) + \varepsilon.$$

When $m(E) < \infty$ the final statement follows at once from (ii) in Proposition 2.15, since then $m(O \setminus E) = m(O) - m(E) \leq \varepsilon$. When $m(E) = \infty$ we first write \mathbb{R} as a countable union of finite intervals: $\mathbb{R} = \bigcup_n (-n, n)$. Now $E_n = E \cap (-n, n)$ has finite measure, so we can find an open $O_n \supset E_n$ with $m(O_n \setminus E_n) \leq \frac{\varepsilon}{2^n}$. The set $O = \bigcup_n O_n$ is open and contains E . Now

$$O \setminus E = \left(\bigcup_n O_n \right) \setminus \left(\bigcup_n E_n \right) \subset \bigcup_n (O_n \setminus E_n),$$

so that $m(O \setminus E) \leq \sum_n m(O_n \setminus E_n) \leq \varepsilon$ as required.

(ii) In (i) use $\varepsilon = \frac{1}{n}$ and let O_n be the open set so obtained. With $E = \bigcap_n O_n$ we obtain a measurable set containing A such that $m(E) < m(O_n) \leq m^*(A) + \frac{1}{n}$ for each n , hence the result follows. \square

Remark 2.18

Theorem 2.17 shows how the freedom of movement allowed by the closure properties of the σ -field \mathcal{M} can be exploited by producing, for any set $A \subset \mathbb{R}$, a measurable set $O \supset A$ which is obtained from open intervals with two operations (countable unions followed by countable intersections) and whose measure equals the outer measure of A .

Finally we show that monotone sequences of measurable sets behave as one would expect with respect to m .

Theorem 2.19

Suppose that $A_n \in \mathcal{M}$ for all $n \geq 1$. Then we have:

(i) if $A_n \subset A_{n+1}$ for all n , then

$$m\left(\bigcup_n A_n\right) = \lim_{n \rightarrow \infty} m(A_n),$$

(ii) if $A_n \supset A_{n+1}$ for all n and $m(A_1) < \infty$, then

$$m\left(\bigcap_n A_n\right) = \lim_{n \rightarrow \infty} m(A_n).$$

Proof

(i) Let $B_1 = A_1$, $B_i = A_i - A_{i-1}$ for $i > 1$. Then $\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$ and the $B_i \in \mathcal{M}$ are pairwise disjoint, so that

$$\begin{aligned}
 m\left(\bigcup_i A_i\right) &= m\left(\bigcup_i B_i\right) \\
 &= \sum_{i=1}^{\infty} m(B_i) \quad (\text{by countable additivity}) \\
 &= \lim_{n \rightarrow \infty} \sum_{i=1}^n m(B_i) \\
 &= \lim_{n \rightarrow \infty} m\left(\bigcup_{i=1}^n B_i\right) \quad (\text{by additivity}) \\
 &= \lim_{n \rightarrow \infty} m(A_n),
 \end{aligned}$$

since $A_n = \bigcup_{i=1}^n B_i$ by construction – see Figure 2.4.

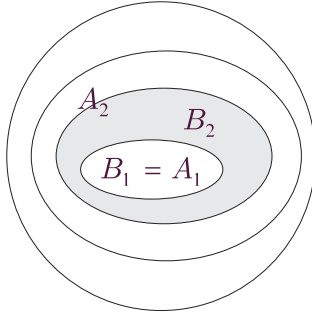


Figure 2.4 Sets A_n, B_n

(ii) $A_1 \setminus A_1 = \emptyset \subset A_1 \setminus A_2 \subset \cdots \subset A_1 \setminus A_n \subset \cdots$ for all n , so that by (i)

$$m\left(\bigcup_n (A_1 \setminus A_n)\right) = \lim_{n \rightarrow \infty} m(A_1 \setminus A_n)$$

and since $m(A_1)$ is finite, $m(A_1 \setminus A_n) = m(A_1) - m(A_n)$. On the other hand, $\bigcup_n (A_1 \setminus A_n) = A_1 \setminus \bigcap_n A_n$, so that

$$m\left(\bigcup_n (A_1 \setminus A_n)\right) = m(A_1) - m\left(\bigcap_n A_n\right) = m(A_1) - \lim_{n \rightarrow \infty} m(A_n).$$

The result follows. \square

Remark 2.20

The proof of Theorem 2.19 simply relies on the countable additivity of m and on the definition of the sum of a series in $[0, \infty]$, i.e.

$$\sum_{i=1}^{\infty} m(A_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n m(A_i).$$

Consequently the result is true, not only for the set function m we have constructed on \mathcal{M} , but for any countably additive set function defined on a σ -field. It also leads us to the following claim, which, though we consider it here only for m , actually characterizes countably additive set functions.

Theorem 2.21

The set function m satisfies:

- (i) m is finitely additive, i.e. for pairwise disjoint sets (A_i) we have

$$m\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n m(A_i)$$

for each n ;

- (ii) m is continuous at \emptyset , i.e. if (B_n) decreases to \emptyset , then $m(B_n)$ decreases to 0.

Proof

To prove this claim, recall that $m : \mathcal{M} \mapsto [0, \infty]$ is countably additive. This implies (i), as we have already seen. To prove (ii), consider a sequence (B_n) in \mathcal{M} which decreases to \emptyset . Then $A_n = B_n \setminus B_{n+1}$ defines a disjoint sequence in \mathcal{M} , and $\bigcup_n A_n = B_1$. We may assume that B_1 is bounded, so that $m(B_n)$ is finite for all n , so that, by Proposition 2.15 (ii), $m(A_n) = m(B_n) - m(B_{n+1}) \geq 0$ and hence we have

$$\begin{aligned} m(B_1) &= \sum_{n=1}^{\infty} m(A_n) \\ &= \lim_{k \rightarrow \infty} \sum_{n=1}^k [m(B_n) - m(B_{n+1})] \\ &= m(B_1) - \lim_{n \rightarrow \infty} m(B_n) \end{aligned}$$

which shows that $m(B_n) \rightarrow 0$, as required. \square

2.5 Borel sets

The definition of \mathcal{M} does not easily lend itself to verification that a particular set belongs to \mathcal{M} ; in our proofs we have had to work quite hard to show that \mathcal{M} is closed under various operations. It is therefore useful to add another construction to our armoury, one which shows more directly how open sets (and indeed open intervals) and the structure of σ -fields lie at the heart of many of the concepts we have developed.

We begin with an auxiliary construction enabling us to produce new σ -fields.

Theorem 2.22

The intersection of a family of σ -fields is a σ -field.

Proof

Let \mathcal{F}_α be σ -fields for $\alpha \in \Lambda$ (the index set Λ can be arbitrary). Put

$$\mathcal{F} = \bigcap_{\alpha \in \Lambda} \mathcal{F}_\alpha.$$

We verify the conditions of the definition.

- (i) $\mathbb{R} \in \mathcal{F}_\alpha$ for all $\alpha \in \Lambda$, so $\mathbb{R} \in \mathcal{F}$.
- (ii) If $E \in \mathcal{F}$, then $E \in \mathcal{F}_\alpha$ for all $\alpha \in \Lambda$. Since the \mathcal{F}_α are σ -fields, $E^c \in \mathcal{F}_\alpha$ and so $E^c \in \mathcal{F}$.
- (iii) If $E_k \in \mathcal{F}$ for $k = 1, 2, \dots$, then $E_k \in \mathcal{F}_\alpha$, all α, k , hence $\bigcup_{k=1}^{\infty} E_k \in \mathcal{F}_\alpha$, all α , and so $\bigcup_{k=1}^{\infty} E_k \in \mathcal{F}$. \square

Definition 2.23

Put

$$\mathcal{B} = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field containing all intervals} \}.$$

We say that \mathcal{B} is the σ -field generated by all intervals and we call the elements of \mathcal{B} *Borel sets* (after Emile Borel, 1871–1956). It is obviously the smallest σ -field containing all intervals. In general, we say that \mathcal{G} is the σ -field generated by a family of sets \mathcal{A} if $\mathcal{G} = \bigcap \{ \mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-field such that } \mathcal{F} \supset \mathcal{A} \}$.

Example 2.24

(Borel sets) The following examples illustrate how the closure properties of the σ -field \mathcal{B} may be used to verify that most familiar sets in \mathbb{R} belong to \mathcal{B} .

- (i) By construction, all intervals belong to \mathcal{B} , and since \mathcal{B} is a σ -field, all open sets must belong to \mathcal{B} , as any open set is a countable union of (open) intervals.
- (ii) Countable sets are Borel sets, since each is a countable union of closed intervals of the form $[a, a]$; in particular \mathbb{N} and \mathbb{Q} are Borel sets. Hence, as the complement of a Borel set, the set of irrational numbers is also Borel. Similarly, finite and cofinite sets are Borel sets.

The definition of \mathcal{B} is also very flexible – as long as we start with *all* intervals of a particular type, these collections generate the same Borel σ -field:

Theorem 2.25

If instead of the family of all intervals we take all open intervals, all closed intervals, all intervals of the form (a, ∞) (or of the form $[a, \infty)$, $(-\infty, b)$, or $(-\infty, b]$), all open sets, or all closed sets, then the σ -field generated by them is the same as \mathcal{B} .

Proof

Consider for example the σ -field generated by the family of open intervals \mathcal{J} and denote it by \mathcal{C} :

$$\mathcal{C} = \bigcap \{ \mathcal{F} \supset \mathcal{J}, \mathcal{F} \text{ is a } \sigma\text{-field} \}.$$

We have to show that $\mathcal{B} = \mathcal{C}$. Since open intervals are intervals, $\mathcal{J} \subset I$ (the family of all intervals), then

$$\{ \mathcal{F} \supset I \} \subset \{ \mathcal{F} \supset \mathcal{J} \}$$

i.e. the collection of all σ -fields \mathcal{F} which contain I is smaller than the collection of all σ -fields which contain the smaller family \mathcal{J} , since it is a more demanding requirement to contain a bigger family, so there are fewer such objects. The inclusion is reversed after we take the intersection on both sides, thus $\mathcal{C} \subset \mathcal{B}$ (the intersection of a smaller family is bigger, as the requirement of belonging to each of its members is a less stringent one).

We shall show that \mathcal{C} contains all intervals. This will be sufficient, since \mathcal{B} is the intersection of such σ -fields, so it is contained in each, so $\mathcal{B} \subset \mathcal{C}$.

To this end consider intervals $[a, b)$, $[a, b]$, $(a, b]$ (the intervals of the form (a, b) are in \mathcal{C} by definition):

$$[a, b) = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, b),$$

$$[a, b] = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, b + \frac{1}{n}\right),$$

$$(a, b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right).$$

\mathcal{C} as a σ -field is closed with respect to countable intersection, so it contains the sets on the right. The argument for unbounded intervals is similar. The proof is complete. \square

Exercise 2.7

Show that the family of intervals of the form $(a, b]$ also generates the σ -field of Borel sets. Show that the same is true for the family of all intervals $[a, b)$.

Remark 2.26

Since \mathcal{M} is a σ -field containing all intervals, and \mathcal{B} is the smallest such σ -field, we have the inclusion $\mathcal{B} \subset \mathcal{M}$, i.e. every Borel set in \mathbb{R} is Lebesgue-measurable. The question therefore arises whether these σ -fields might be the same. In fact the inclusion is proper. It is not altogether straightforward to construct a set in $\mathcal{M} \setminus \mathcal{B}$, and we shall not attempt this here (but see the Appendix). However, by Theorem 2.17 (ii), given any $E \in \mathcal{M}$ we can find a Borel set $B \supset E$ of the form $B = \bigcap_n O_n$, where the (O_n) are open sets, and such that $m(E) = m(B)$. In particular,

$$m(B \Delta E) = m(B \setminus E) = 0.$$

Hence m cannot distinguish between the measurable set E and the Borel set B we have constructed.

Thus, given a Lebesgue-measurable set E we can find a Borel set B such that their symmetric difference $E \Delta B$ is a null set. Now we know that $E \Delta B \in \mathcal{M}$, and it is obvious that subsets of null sets are also null, and hence in \mathcal{M} . However, we cannot conclude that every null set will be a Borel set (if \mathcal{B} did contain all null sets then by Theorem 2.17 (ii) we would have $\mathcal{B} = \mathcal{M}$), and this points to an ‘incompleteness’ in \mathcal{B} which explains why, even if we begin by defining m on intervals and then extend the definition to Borel sets, we would also need to extend it further in order to be able to identify precisely which sets are ‘negligible’ for our purposes. On the other hand, extension of the measure m to the σ -field \mathcal{M} will suffice, since \mathcal{M} does contain all m -null sets and all subsets of null sets also belong to \mathcal{M} .

We show that \mathcal{M} is the smallest σ -field on \mathbb{R} with this property, and we say that \mathcal{M} is the *completion* of \mathcal{B} relative to m and $(\mathbb{R}, \mathcal{M}, m)$ is complete (whereas the measure space $(\mathbb{R}, \mathcal{B}, m)$ is not complete). More precisely, a measure space (X, \mathcal{F}, μ) is *complete* if for all $F \in \mathcal{F}$ with $\mu(F) = 0$, for all $N \subset F$ we have $N \in \mathcal{F}$ (and so $\mu(N) = 0$).

The *completion* of a σ -field \mathcal{G} , relative to a given measure μ , is defined as the smallest σ -field \mathcal{F} containing \mathcal{G} such that, if $N \subset G \in \mathcal{G}$ and $\mu(G) = 0$, then $N \in \mathcal{F}$.

Proposition 2.27

The completion of \mathcal{G} is of the form $\{G \cup N : G \in \mathcal{G}, N \subset F \in \mathcal{F} \text{ with } \mu(F) = 0\}$.

This allows us to extend the measure μ uniquely to a measure $\bar{\mu}$ on \mathcal{F} by setting $\bar{\mu}(G \cup N) = \mu(G)$ for $G \in \mathcal{G}$.

Theorem 2.28

\mathcal{M} is the completion of \mathcal{B} .

Proof

We show first that \mathcal{M} contains all subsets of null sets in \mathcal{B} : so let $N \subset B \in \mathcal{B}$, B null, and suppose $A \subset \mathbb{R}$. To show that $N \in \mathcal{M}$ we need to show that

$$m^*(A) \geq m^*(A \cap N) + m^*(A \cap N^c).$$

First note that $m^*(A \cap N) \leq m^*(N) \leq m^*(B) = 0$. So it remains to show that $m^*(A) \geq m^*(A \cap N^c)$ but this follows at once from monotonicity of m^* .

Thus we have shown that $N \in \mathcal{M}$. Since \mathcal{M} is a complete σ -field containing \mathcal{B} , this means that \mathcal{M} also contains the completion \mathcal{C} of \mathcal{B} .

Finally, we show that \mathcal{M} is the minimal such σ -field, i.e. that $\mathcal{M} \subset \mathcal{C}$: first consider $E \in \mathcal{M}$ with $m^*(E) < \infty$, and choose $B = \bigcap_n O_n \in \mathcal{B}$ as described above such that $B \supset E$, $m(B) = m^*(E)$. (We reserve the use of m for sets in \mathcal{B} throughout this argument.)

Consider $N = B \setminus E$, which is in \mathcal{M} and has $m^*(N) = 0$, since m^* is additive on \mathcal{M} . By Theorem 2.17 (ii) we can find $L \supset N$, $L \in \mathcal{B}$ and $m(L) = 0$. In other words, N is a subset of a null set in \mathcal{B} , and therefore $E = B \setminus N$ belongs to the completion \mathcal{C} of \mathcal{B} . For $E \in \mathcal{M}$ with $m^*(E) = \infty$, apply the above to $E_n = E \cap [-n, n]$ for each $n \in \mathbb{N}$. Each $m^*(E_n)$ is finite, so the E_n all belong to \mathcal{C} and hence so does their countable union E . Thus $\mathcal{M} \subset \mathcal{C}$ and so they are equal. \square

Despite these technical differences, measurable sets are never far from ‘nice’ sets, and, in addition to approximations from above by open sets, as observed in Theorem 2.17, we can approximate the measure of any $E \in \mathcal{M}$ from below by those of closed subsets.

Theorem 2.29

If $E \in \mathcal{M}$ then for given $\varepsilon > 0$ there exists a closed set $F \subset E$ such that $m(E \setminus F) < \varepsilon$. Hence there exists $B \subset E$ in the form $B = \bigcup_n F_n$, where all the F_n are closed sets, and $m(E \setminus B) = 0$.

Proof

The complement E^c is measurable and by Theorem 2.17 we can find an open set O containing E^c such that $m(O \setminus E^c) \leq \varepsilon$. But $O \setminus E^c = O \cap E = E \setminus O^c$, and $F = O^c$ is closed and contained in E . Hence this F is what we need. The final part is similar to Theorem 2.17 (ii), and the proof is left to the reader. \square

Exercise 2.8

Show that each of the following two statements is equivalent to saying that $E \in \mathcal{M}$:

- (a) given $\varepsilon > 0$ there is an open set $O \supset E$ with $m^*(O \setminus E) < \varepsilon$,
- (b) given $\varepsilon > 0$ there is a closed set $F \subset E$ with $m^*(E \setminus F) < \varepsilon$.

Remark 2.30

The two statements in the above Exercise are the key to a considerable generalization, linking the ideas of measure theory to those of topology:

A non-negative countably additive set function μ defined on \mathcal{B} is called a *regular Borel measure* if for every Borel set B we have:

$$\begin{aligned}\mu(B) &= \inf\{\mu(O) : O \text{ open, } O \supset B\}, \\ \mu(B) &= \sup\{\mu(F) : F \text{ closed, } F \subset B\}.\end{aligned}$$

In Theorems 2.17 and 2.29 we have verified these relations for Lebesgue measure. We shall consider other concrete examples of regular Borel measures later.

2.6 Probability

The ideas which led to Lebesgue measure may be adapted to construct measures generally on arbitrary sets: any set Ω carrying an outer measure (i.e. a mapping from $P(\Omega)$ to $[0, \infty]$, monotone and countably subadditive) can be equipped with a measure μ defined on an appropriate σ -field \mathcal{F} of its subsets. The resulting triple $(\Omega, \mathcal{F}, \mu)$ is then called a measure space, as observed in Remark 2.12. Note that in the construction of Lebesgue measure we only used the properties, not the particular form of the outer measure.

For the present, however, we shall be content with noting simply how to restrict Lebesgue measure to any Lebesgue-measurable subset B of \mathbb{R} with $m(B) > 0$:

Given Lebesgue measure m on the Lebesgue σ -field \mathcal{M} , let

$$\mathcal{M}_B = \{A \cap B : A \in \mathcal{M}\}$$

and for $A \in \mathcal{M}_B$ write

$$m_B(A) = m(A).$$

Proposition 2.31

(B, \mathcal{M}_B, m_B) is a complete measure space.

Hint $\bigcup_i (A_i \cap B) = (\bigcup_i A_i) \cap B$ and $(A_1 \cap B) \setminus (A_2 \cap B) = (A_1 \setminus A_2) \cap B$.

We can finally state precisely what we mean by ‘selecting a number from $[0, 1]$ at random’: restrict Lebesgue measure m to the interval $B = [0, 1]$ and consider the σ -field of $\mathcal{M}_{[0,1]}$ of measurable subsets of $[0, 1]$. Then $m_{[0,1]}$ is a measure on $\mathcal{M}_{[0,1]}$ with ‘total mass’ 1. Since all subintervals of $[0, 1]$ with the same length have equal measure, the ‘mass’ of $m_{[0,1]}$ is spread uniformly over $[0, 1]$, so that, for example, the ‘probability’ of choosing a number from $[0, \frac{1}{10})$ is the same as that of choosing a number from $[\frac{6}{10}, \frac{7}{10})$, namely $\frac{1}{10}$. Thus all numerals are equally likely to appear as first digits of the decimal expansion of the chosen number. On the other hand, with this measure, the probability that the chosen number will be rational is 0, as is the probability of drawing an element of the Cantor set C .

We now have the basis for some probability theory, although a general development still requires the extension of the concept of measure from \mathbb{R} to abstract sets. Nonetheless the building blocks are already evident in the detailed development of the example of Lebesgue measure. The main idea in providing a mathematical foundation for probability theory is to use the concept of measure

to provide the mathematical model of the intuitive notion of probability. The distinguishing feature of probability is the concept of *independence*, which we introduce below. We begin by defining the general framework.

2.6.1 Probability space

Definition 2.32

A *probability space* is a triple (Ω, \mathcal{F}, P) where Ω is an arbitrary set, \mathcal{F} is a σ -field of subsets of Ω , and P is a measure on \mathcal{F} such that

$$P(\Omega) = 1,$$

called *probability measure* or, briefly, *probability*.

Remark 2.33

The original definition, given by Kolmogorov in 1932, is a variant of the above (see Theorem 2.21): (Ω, \mathcal{F}, P) is a probability space if (Ω, \mathcal{F}) are given as above, and P is a finitely additive set function with $P(\emptyset) = 0$ and $P(\Omega) = 1$ such that $P(B_n) \searrow 0$ whenever (B_n) in \mathcal{F} decreases to \emptyset .

Example 2.34

We see at once that Lebesgue measure restricted to $[0, 1]$ is a probability measure. More generally: suppose we are given an arbitrary Lebesgue-measurable set $\Omega \subset \mathbb{R}$, with $m(\Omega) > 0$. Then $P = c \cdot m_\Omega$, where $c = \frac{1}{m(\Omega)}$, and $m = m_\Omega$ denotes the restriction of Lebesgue measure to measurable subsets of Ω , provides a probability measure on Ω , since P is complete and $P(\Omega) = 1$.

For example, if $\Omega = [a, b]$, we obtain $c = \frac{1}{b-a}$, and P becomes the ‘uniform distribution’ over $[a, b]$. However, we can also use less familiar sets for our base space; for example, $\Omega = [a, b] \cap (\mathbb{R} \setminus \mathbb{Q})$, $c = \frac{1}{b-a}$ gives the same distribution over the irrationals in $[a, b]$.

2.6.2 Events: conditioning and independence

The word ‘event’ is used to indicate that something is happening. In probability a typical event is to draw elements from a set and then the event is concerned with the outcome belonging to a particular subset. So, as described above, if $\Omega = [0, 1]$ we may be interested in the fact that a number drawn at random

from $[0, 1]$ belongs to some $A \subset [0, 1]$. We want to estimate the probability of this happening, and in the mathematical setup this is the number $P(A)$, here $m_{[0,1]}(A)$. So it is natural to require that A should belong to $\mathcal{M}_{[0,1]}$, since these are the sets we may measure. By a slight abuse of the language, probabilists tend to identify the actual ‘event’ with the set A which features in the event. The next definition simply confirms this abuse of language.

Definition 2.35

Given a probability space (Ω, \mathcal{F}, P) we say that the elements of \mathcal{F} are *events*.

Suppose next that a number has been drawn from $[0, 1]$ but has not been revealed yet. We would like to bet on it being in $[0, \frac{1}{4}]$ and we get a tip that it certainly belongs to $[0, \frac{1}{2}]$. Clearly, given this ‘inside information’, the probability of success is now $\frac{1}{2}$ rather than $\frac{1}{4}$. This motivates the following general definition.

Definition 2.36

Suppose that $P(B) > 0$. Then the number

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

is called the *conditional probability of A given B*.

Proposition 2.37

The mapping $A \mapsto P(A|B)$ is countably additive on the σ -field \mathcal{F}_B .

Hint Use the fact that $A \mapsto P(A \cap B)$ is countably additive on \mathcal{F} .

A classical application of the conditional probability is the total probability formula which enables the computation of the probability of an event by means of conditional probabilities given some disjoint hypotheses:

Exercise 2.9

Prove that if H_i are pairwise disjoint events such that $\bigcup_{i=1}^{\infty} H_i = \Omega$, $P(H_i) \neq 0$, then

$$P(A) = \sum_{i=1}^{\infty} P(A|H_i)P(H_i).$$

It is natural to say that the event A is *independent of* B if the fact that B takes place has no influence on the chances of A , i.e.

$$P(A|B) = P(A).$$

By definition of $P(A|B)$ this immediately implies the relation

$$P(A \cap B) = P(A)P(B),$$

which is usually taken as the definition of independence. The advantage of this practice is that we may dispose of the assumption $P(B) > 0$.

Definition 2.38

The events A, B are *independent* if

$$P(A \cap B) = P(A) \cdot P(B).$$

Exercise 2.10

Suppose that A and B are independent events. Show that A^c and B are also independent.

The exercise indicates that if A and B are independent events, then all elements of the σ -fields they generate are mutually independent, since these σ -fields are simply the collections $\mathcal{F}_A = \{\emptyset, A, A^c, \Omega\}$ and $\mathcal{F}_B = \{\emptyset, B, B^c, \Omega\}$ respectively. This leads us to a natural extension of the definition: two σ -fields \mathcal{F}_1 and \mathcal{F}_2 are *independent* if for any choice of sets $A_1 \in \mathcal{F}_1$ and $A_2 \in \mathcal{F}_2$ we have $P(A_1 \cap A_2) = P(A_1)P(A_2)$.

However, the extension of these definitions to three or more events (or several σ -fields) needs a little care, as the following simple examples show:

Example 2.39

Let $\Omega = [0, 1]$, $A = [0, \frac{1}{4}]$ as before; then A is independent of $B = [\frac{1}{8}, \frac{5}{8}]$ and of $C = [\frac{1}{8}, \frac{3}{8}] \cup [\frac{3}{4}, 1]$. In addition, B and C are independent. However,

$$P(A \cap B \cap C) \neq P(A)P(B)P(C).$$

Thus, given three events, the pairwise independence of each of the three possible pairs does *not* suffice for the extension of ‘independence’ to all three events.

On the other hand, with $A = [0, \frac{1}{4}]$, $B = C = [0, \frac{1}{16}] \cup [\frac{1}{4}, \frac{11}{16}]$ (or alternatively with $C = [0, \frac{1}{16}] \cup [\frac{9}{16}, 1]$),

$$P(A \cap B \cap C) = P(A)P(B)P(C) \quad (2.21)$$

but none of the pairs make independent events.

This confirms further that we need to demand rather more if we wish to extend the above definition – pairwise independence is not enough, nor is (2.21); therefore we need to require both conditions to be satisfied together. Extending this to n events leads to:

Definition 2.40

The events A_1, \dots, A_n are *independent* if for all $k \leq n$ for each choice of k events, the probability of their intersection is the product of the probabilities.

Again there is a powerful counterpart for σ -fields (which can be extended to sequences, and even arbitrary families):

Definition 2.41

The σ -fields $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_n$ defined on a given probability space (Ω, \mathcal{F}, P) are *independent* if, for all choices of distinct indices i_1, i_2, \dots, i_k from $\{1, 2, \dots, n\}$ and all choices of sets $F_{i_n} \in \mathcal{F}_{i_n}$, we have

$$P(F_{i_1} \cap F_{i_2} \cap \dots \cap F_{i_k}) = P(F_{i_1})P(F_{i_2}) \dots P(F_{i_k}).$$

The issue of independence will be revisited in the subsequent chapters where we develop some more tools to calculate probabilities.

2.6.3 Applications to mathematical finance

As indicated in the Preface, we will explore briefly how the ideas developed in each chapter can be applied in the rapidly growing field of mathematical finance. This is not intended as an introduction to this subject, but hopefully it will demonstrate how a consistent mathematical formulation can help to clarify ideas central to many disciplines. Readers who are unfamiliar with mathematical finance should consult texts such as [4], [5], [7] for definitions and a discussion of the main ideas of the subject.

Probabilistic modelling in finance centres on the analysis of models for the evolution of the value of traded assets, such as *stocks* or *bonds*, and seeks to identify trends in their future behaviour. Much of the modern theory is concerned with evaluating *derivative securities* such as *options*, whose value is determined by the (random) future values of some underlying security, such as a stock.

We illustrate the above probability ideas on a classical model of stock prices, namely the *binomial tree*. This model is based on finitely many time instants at which the prices may change, and the changes are of a very simple nature. Suppose that the number of steps is N , and denote the price at the k -th step by $S(k)$, $0 \leq k \leq N$. At each step the stock price changes in the following way: the price at a given step is the price at the previous step multiplied by U with probability p or D with probability $q = 1 - p$, where $0 < D < U$. Therefore the final price depends on the sequence $\omega = (\omega_1, \omega_2, \dots, \omega_N)$ where $\omega_i = 1$ indicates the application of the factor U or $\omega_i = 0$, which indicates application of the factor D . Such a sequence is called a *path* and we take Ω to consist of all possible paths. In other words,

$$S(k) = S(0) \prod_{i=1}^k \eta(i),$$

where

$$\eta(k) = \begin{cases} U & \text{with probability } p \\ D & \text{with probability } q. \end{cases}$$

Exercise 2.11

Suppose $N = 5$, $U = 1.2$, $D = 0.9$, and $S(0) = 500$. Find the number of all paths. How many paths lead to the price $S(5) = 524.88$? What is the probability that $S(5) > 900$ if the probability of going up in a single step is 0.5?

In general, the total number of paths is clearly 2^N and at step k there are $k + 1$ possible prices.

We construct a probability space by equipping Ω with the σ -field 2^Ω of all subsets of Ω , and the probability defined on single-element sets by $P(\{\omega\}) = p^k q^{N-k}$, where $k = \sum_{i=1}^N \omega_i$.

As time progresses we gather information about stock prices, or, what amounts to the same, about paths. This means that having observed some prices, the range of possible future developments is restricted. Our information increases with time and this idea can be captured by the following family of σ -fields.

Fix $m < N$ and let \mathcal{F}_m to be the σ -field generated by the following family of sets $\{A : \omega, \omega' \in A \implies \omega_1 = \omega'_1, \omega_2 = \omega'_2, \dots, \omega_m = \omega'_m\}$. So all paths from a particular set A in this σ -field have identical initial segments while the remaining coordinates are arbitrary. Note that

$$\mathcal{F}_0 = \{\Omega, \emptyset\},$$

$\mathcal{F}_1 = \{A_1, A_1^c, \Omega, \emptyset\}$, where $A_1 = \{\omega : \omega_1 = 1\}$, i.e. $S(1) = S(0)U$, and $A_1^c = \{\omega : \omega_1 = 0\}$ i.e. $S(1) = S(0)D$.

Exercise 2.12

Prove that \mathcal{F}_m has 2^{2^m} elements.

Exercise 2.13

Prove that the sequence \mathcal{F}_m is increasing.

This sequence is an example of a *filtration* (the identifying features are that the σ -fields should be contained in \mathcal{F} and form an increasing chain), a concept which we shall revisit later on.

The consecutive choices of stock prices are closely related to coin-tossing. Intuition tells us that the latter are independent. This can be formally seen by introducing another σ -field describing the fact that at a particular step we have a particular outcome. Suppose ω is such that $\omega_k = 1$. Then we can identify the set of all paths with this property $A_k = \{\omega : \omega_k = 1\}$ and extend to a σ -field: $\mathcal{G}_k = \{A_k, A_k^c, \Omega, \emptyset\}$. In fact, $A_k^c = \{\omega : \omega_k = 0\}$.

Exercise 2.14

Prove that \mathcal{G}_m and \mathcal{G}_k are independent if $m \neq k$.

2.7 Proofs of propositions

Proof (of Proposition 2.5)

If the intervals I_n cover B , then they also cover A : $A \subset B \subset \bigcup_n I_n$, hence $Z_B \subset Z_A$. The infimum of a larger set cannot be greater than the infimum of a smaller set (trivial illustration: $\inf\{0, 1, 2\} < \inf\{1, 2\}$, $\inf\{0, 1, 2\} = \inf\{0, 2\}$), hence the result. \square

Proof (of Proposition 2.8)

If a system I_n of intervals covers A then the intervals $I_n + t$ cover $A + t$. Conversely, if J_n cover $A + t$ then $J_n - t$ cover A . Moreover, the total length of a family of intervals does not change when we shift each by a number. So we

have a one-one correspondence between the interval coverings of A and $A + t$ and this correspondence preserves the total length of the covering. This implies that the sets Z_A and Z_{A+t} are the same, so their infima are equal. \square

Proof (of Proposition 2.13)

By de Morgan's law

$$\bigcap_{k=1}^{\infty} E_k = \left(\bigcup_{k=1}^{\infty} E_k^c \right)^c.$$

By Theorem 2.11 (ii) all E_k^c are in \mathcal{M} , hence by (iii) the same can be said about the union $\bigcup_{k=1}^{\infty} E_k^c$. Finally, by (ii) again, the complement of this union is in \mathcal{M} , and so the intersection $\bigcap_{k=1}^{\infty} E_k$ is in \mathcal{M} . \square

Proof (of Proposition 2.15)

- (i) Proposition 2.5 tells us that the outer measure is monotone, but since m is just the restriction of m^* to \mathcal{M} , then the same is true for m : $A \subset B$ implies $m(A) = m^*(A) \leq m^*(B) = m(B)$.
- (ii) We write B as a disjoint union $B = A \cup (B \setminus A)$ and then by additivity of m we have $m(B) = m(A) + m(B \setminus A)$. Subtracting $m(A)$ (here it is important that $m(A)$ is finite) we get the result.
- (iii) Translation invariance of m follows at once from translation invariance of the outer measure in the same way as in (i) above. \square

Proof (of Proposition 2.16)

The set $A \Delta B$ is null, hence so are its subsets $A \setminus B$ and $B \setminus A$. Thus these sets are measurable, and so is $A \cap B = A \setminus (A \setminus B)$, and therefore also $B = (A \cap B) \cup (B \setminus A) \in \mathcal{M}$. Now $m(B) = m(A \cap B) + m(B \setminus A)$ as the sets on the right are disjoint. But $m(B \setminus A) = 0 = m(A \setminus B)$, so $m(B) = m(A \cap B) = m(A \cap B) + m(A \setminus B) = m((A \cap B) \cup (A \setminus B)) = m(A)$. \square

Proof (of Proposition 2.27)

The family $\mathcal{G} = \{G \cup N : G \in \mathcal{F}, N \subset F \in \mathcal{F} \text{ with } \mu(F) = 0\}$ contains the set X since $X \in \mathcal{F}$. If $G_i \cup N_i \in \mathcal{G}$, $N_i \subset F_i$, $\mu(F_i) = 0$, then $\bigcup G_i \cup \bigcup N_i = \bigcup G_i \cup \bigcup N_i$ is in \mathcal{G} since the first set on the right is in \mathcal{F} and the second is a subset of a null set $\bigcup F_i \in \mathcal{F}$. If $G \cup N \in \mathcal{G}$, $N \subset F$, then $(G \cup N)^c = (G \cup F)^c \cup ((F \setminus N) \cap G^c)$, which is also in \mathcal{G} . Thus \mathcal{G} is a σ -field. Consider any other σ -field \mathcal{H} containing

\mathcal{F} and all subsets of null sets. Since \mathcal{H} is closed with respect to the unions, it contains \mathcal{G} and so \mathcal{G} is the smallest σ -field with this property. \square

Proof (of Proposition 2.31)

It follows at once from the definitions and the hint that \mathcal{M}_B is a σ -field. To see that m_B is a measure we check countable additivity: with $C_i = A_i \cap B$ pairwise disjoint in \mathcal{M}_B , we have

$$m_B\left(\bigcup_i C_i\right) = m\left(\bigcup_i (A_i \cap B)\right) = \sum_i m(A_i \cap B) = \sum_i m(C_i) = \sum_i m_B(C_i).$$

Therefore (B, \mathcal{M}_B, m_B) is a measure space. It is complete, since subsets of null sets contained in B are by definition m_B -measurable. \square

Proof (of Proposition 2.37)

Assume that A_n are measurable and pairwise disjoint. By the definition of conditional probability

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n | B\right) &= \frac{1}{P(B)} P\left(\left(\bigcup_{n=1}^{\infty} A_n\right) \cap B\right) \\ &= \frac{1}{P(B)} P\left(\bigcup_{n=1}^{\infty} (A_n \cap B)\right) \\ &= \frac{1}{P(B)} \sum_{n=1}^{\infty} P(A_n \cap B) \\ &= \sum_{n=1}^{\infty} P(A_n | B) \end{aligned}$$

since $A_n \cap B$ are also pairwise disjoint and P is countably additive. \square

Measure, Integral and Probability

Capinski, M.; Kopp, P.E.

2004, XV, 311 p. 3 illus., Softcover

ISBN: 978-1-85233-781-0