Chapter 2. Length measure on $\mathbb{R}$

The idea here is to develop the idea of the length of a subset of $\mathbb{R}$ in such a way that we can manage to define the Lebesgue integral in more or less the way outlined in Chapter 0.

Ideally we should be able to talk about the length (total length) of any subset of the line, but this turns out to be unworkable. Well, at least lots of things that seem obvious can't work unless we restrict ourselves to subsets of $\mathbb{R}$ that are not too terrible. We'll explain later in precise terms which kinds of sets are going to be allowed.

The approach we take is to start with lengths of intervals and then build up to more complicated sets. If we wanted to cater for double integrals and triple integrals (and more generally integrals of functions of n real variables) we should start with areas of rectangles in $\mathbb{R}^2$, or volumes of boxes in $\mathbb{R}^3$, or n-dimensional volume of n-dimensional boxes in $\mathbb{R}^n$, and build up from there. We will not actually do these more general cases, but the approach we use on $\mathbb{R}$ can be adapted fairly easily to work on $\mathbb{R}^n$.

Contents

2.1 Algebras of subsets of $\mathbb{R}$

As we said, we can't manage with lengths of all subsets of $\mathbb{R}$, that is we will not be able to assign a satisfactory length to each set in the power set $\mathcal{P}(\mathbb{R})$, the collection of all subsets. We do insist that we have our length making sense for a (Boolean) algebra of subsets.

Definition 2.1.1. A collection $\mathcal{A}$ of subsets of $\mathbb{R}$ is called an *algebra of subsets* of $\mathbb{R}$ if

(a) $\emptyset \in \mathcal{A}$

(b) $E \in \mathcal{A} \Rightarrow E^c \in \mathcal{A}$

(c) $E_1, E_2 \in \mathcal{A} \Rightarrow E_1 \cup E_2 \in \mathcal{A}$

(In words, $\mathcal{A}$ contains the empty set, is closed under taking complements and under unions (of two members).)

Lemma 2.1.2. *If $\mathcal{A}$ is an algebra of subsets of $\mathbb{R}$, then*

(i) $\mathbb{R} \in \mathcal{A}$

(ii) $E_1, E_2 \in \mathcal{A} \Rightarrow E_1 \cap E_2 \in \mathcal{A}$

(iii) $E_1, E_2, \ldots, E_k \in \mathcal{A} \Rightarrow \bigcup_{j=1}^k E_j \in \mathcal{A}$ *and* $\bigcap_{j=1}^k E_j \in \mathcal{A}$

Proof. (i) $\mathbb{R} = \emptyset^c \in \mathcal{A}$ by Definition 2.1.1 (a) and (b).

(ii) $E_1, E_2 \in \mathcal{A} \Rightarrow E_1 \cap E_2 = (E_1^c \cup E_2^c)^c$ (by De Morgans laws) $\Rightarrow E_1 \cap E_2 \in \mathcal{A}$ (because from Definition 2.1.1 (b) we have $E_1^c, E_2^c \in \mathcal{A}$, thus $E_1^c \cup E_2^c \in \mathcal{A}$ by Definition 2.1.1 (c), and so $(E_1^c \cup E_2^c)^c \in \mathcal{A}$ by Definition 2.1.1 (b) again).

(iii) This follows by induction on k since $\bigcup_{j=1}^{k+1} E_j = \left(\bigcup_{j=1}^{k} E_j\right) \cup E_k$, and similarly for intersections.

□

Remark 2.1.3. It might help us to know (from topology or the inermediate value theorem) that the intervals in $\mathbb{R}$ are the nonempty connected subsets of the real line. (This fact will not adapt if we were doing rectangles in $\mathbb{R}^2$ or boxes in $\mathbb{R}^n$, however.) A less sophisticated concept that could be used instead (but applies to subsets of the real line and is equivalent to connectedness) is a *between-ness property* of a set $S \subseteq \mathbb{R}$, stated as follows:

$$\text{if } s_1, s_2 \in S \text{ with } s_1 < s_2 \text{ and } x \in \mathbb{R} \text{ with } s_1 < x < s_2 \text{ then } x \in S. \tag{2.1.1}$$

There are quite a few kinds of intervals, finite ones with neither, one or both endpoints included, so

$$(a,b) = \{x \in \mathbb{R} : a < x < b\}, [a,b) = \{x \in \mathbb{R} : a \le x < b\},$$
$$(a,b] = \{x \in \mathbb{R} : a < x \le b\}, [a,b] = \{x \in \mathbb{R} : a \le x \le b\},$$

where we insist that $a < b$ except in the case of $[a,a] = \{a\}$ where we allow the singleton sets as intervals. Then there are infinite ones that are not the whole line with the end point included or nor and the whole line,

$$(a,\infty) = \{x \in \mathbb{R} : a < x\}, [a,\infty) = \{x \in \mathbb{R} : a \le x\},$$

$$(-\infty,a) = \{x \in \mathbb{R} : x < a\}, (-\infty,a] = \{x \in \mathbb{R} : x \le a\},$$

and finally the whole real line $\mathbb{R} = (-\infty, \infty)$.

In addition we can consider the empty set as an interval, though we will say that it has no determined end points, putting it in a category by itself. Allowing $\pm\infty$ as kind of virtual endpoints, all nonempty intervals have a left and a right endpoint (which is unambiguous).

Just for clarity, when we speak of finite intervals, we mean ones with finite endpoints and not ones that are finite sets.

We can show the following.

Lemma 2.1.4. *The subsets $S \subseteq \mathbb{R}$ with the between-ness property (2.1.1) are exactly all the intervals (including the empty subset).*

Proof. If S is bounded and not empty, one can take $a = \inf S$, $b = \sup S$ and use (2.1.1) to show that $(a,b) \subseteq S$. In more detail if $a < x < b$, then $x > \inf S$ implies there is $s_1 \in S$ with $s_1 < x$. (So $a \le s_1 < b$.) Similarly $x < b = \sup S$ implies there is $s_2 \in S$ with $x < s_2$ (and this s_2 must have $s_2 \le b$). By (2.1.1) we get $x \in S$. This way we show $(a,b) \subseteq S$ must hold. Then there are 4 cases to consider depending on $a \in S$ or $a \notin S$, and $b \in S$ or $b \notin S$.

If, say, S is bounded below but not above we can do something similar to show $(a, \infty) \subset S$ where $a = \inf S$ and then there are two cases depending on whether $a \in S$ or not. If S is bounded above but not below we take $b = \sup S$ and show $(-\infty, b) \subseteq S$ and then we we have two cases depending on whether or not $b \in S$.

If S is bounded neither above nor below we show $\mathbb{R} = (-\infty, \infty) \subseteq S \Rightarrow S = \mathbb{R}$. The last case is $S = \emptyset$. □

Lemma 2.1.5. *If $I_1, I_2 \subseteq \mathbb{R}$ are intervals, then either $I_1 \cup I_2$ is also an interval or there is $t \in \mathbb{R}$ satisfying one of*

(i) $x < t$ for each $x \in I_1$ and $t < x$ for each $x \in I_2$, or

(ii) $x < t$ for each $x \in I_2$ and $t < x$ for each $x \in I_1$.

A formal proof is possible based on the negation of the between-ness property (2.1.1). If $I_1 \cup I_2$ is not an interval, then there is an x that fails to satisfy (2.1.1) for $S = I_1 \cup I_2$. Then argue that we must have one of the cases stated.

If $I_1 = \emptyset$ or $I_2 = \emptyset$, then we have a collapsing case where $I_1 \cup I_2$ is just one of the intervals and we don't need to do more. If not, then we can use (2.1.1) to show that I_1 must be on one side of x and I_2 on the other.

A less formal (longer) argument would be to look at all the (many) possible casess of how two nonempty intervals can be situated with respect to each other. If I_1 and I_2 actually overlap, so that there is $c \in I_1 \cap I_2$, it is fairly easy to see that $I_1 \cup I_2$ must be an interval. Another case is where there is a common endpoint between I_1 and I_2 and that endopoint is in one of the intervals. An example is $[0, 1) \cup (1, 3) = [0, 3)$.

We won't give the formal proof (or the longer one).

Definition 2.1.6. We define the *(full) interval algebra* $\mathscr{J}$ to be the collection of all **finite unions** of intervals. We include the empty set in $\mathscr{J}$ (as the union of the empty collection of such intervals!).[1]

Lemma 2.1.7. *$\mathscr{J}$ is an algebra (of subsets of $\mathbb{R}$).*

Proof. This is almost obvious, but a formal check is perhaps a good idea. If we take the union of two members of $\mathscr{J}$, say of two finite unions $E_1 = \bigcup_{j=1}^n I_j$ and $E_2 = \bigcup_{k=1}^m I'_k$ (where the I_j and I'_k are intervals), then

$$E_1 \cup E_2 = \bigcup_{j=1}^n I_j \cup \bigcup_{k=1}^m I'_k$$

is again a finite union of intervals and so $E_1 \cup E_2 \in \mathscr{J}$.

[1]In previous versions of these notes, I used a smaller algebra that was denoted $\mathscr{I}$ and it only included finite unions of some intervals, those where the right hand end point is included when it is finite and the left-hand end point is not included. Starting with $\mathscr{I}$, some things are easier to check right away but there is then a price to pay later. In the end, it makes no big difference but this time we use the full interval algebra $\mathscr{J}$.

Considering the intersection $E_1 \cap E_2$ (instead of the union) we have

$$E_1 \cap E_2 = \bigcup_{j=1}^{n} (I_j \cap E_2) = \bigcup_{j=1}^{n} \bigcup_{k=1}^{m} (I_j \cap I'_k)$$

and to see that $E_1 \cap E_2 \in \mathscr{I}$ notice that each $I_j \cap I'_j$ is either empty or another interval. By induction, we have then that finite intersections of sets in $\mathscr{I}$ will be in $\mathscr{I}$.

The complement of one interval is either another interval or the union of two such, except in the case of the whole line where we have $\mathbb{R}^c = (-\infty, \infty)^c = \emptyset$. For instance

$$(a, b)^c = (-\infty, a] \cup [b, \infty), (a, b]^c = (-\infty, a] \cup (b, \infty), [a, b)^c = (-\infty, a) \cup [b, \infty),$$

(for $a < b$),

$$[a, b]^c = (-\infty, a) \cup (b, \infty), \qquad (a \leq b)$$

and

$$(-\infty, a]^c = (a, \infty), (-\infty, a)^c = [a, \infty), (a, \infty)^c = (-\infty, a], [a, \infty)^c = (-\infty, a),$$

The complement of the empty set is $\phi^c = \mathbb{R} = (-\infty, \infty) = \mathbb{R}$. What we need is that if I is any kind of interval, inclduing the case $I = \emptyset$, then $I^c \in \mathscr{I}$.

By De Morgans laws then, if $E \in \mathscr{I}$ with $E = \bigcup_{j=1}^{n} I_j$, we have

$$E^c = \bigcap_{j=1}^{n} I_j^c$$

and we have just noted that each $I_j^c \in \mathscr{I}$. So $E^c \in \mathscr{I}$ (using what we just showed about finite intersections). □

Remark 2.1.8 (Standard Form). One thing to note is that if two intervals overlap or share an endpoint that belongs in one of the intervals, then their union is another interval. Or, more precisely, we have Lemma 2.1.5.

By using induction on the smallest number of intervals required to express E, we can write every set $E = \mathscr{I}$ uniquely as one of the following

$$E = \emptyset,$$

or there is $n \geq 1$ so that E is a disjoint union of $n \geq 1$ nonempty intervals

$$E = I_1 \cup I_2 \cup \cdots \cup I_n$$

in such a way that (if $n > 1$) there are points $t_1, \ldots, t_{n-1} \in \mathbb{R}$ such that $x < t_j$ holds for each $x \in I_j$ and $t_j < x$ holds for each $x \in I_{j+1}$ (for $1 \leq j < n$).

So we order the intervals left to right and the existence of the t_j ensures that we cannot merge any pair of consecutive intervals (or any pair at all) into a single interval.

The idea for the induction would start with E as a union of (2 or more) nonempty intervals and if any pair of the intervals has a union that is an interval, then we can reduce the number of intervals we need by 1. Then we cn apply our induction hypothesis. So assume we cannot find any pair of the intervals that merge into a single interval. Then by Lemma 2.1.5, we can order each pair of the intervals left to right. In this way we can find a leftmost one. Then apply the induction hypothesis in that case. If not there must be a leftmost interval and a gap betwen that and the others.

We call this way of writing $E \in \mathscr{I}$ the *Standard Form* of E.

We treat the standard form for $E = \emptyset$ to have $n = 0$ and it is just considered as the empty union (of no intervals).

Example 2.1.9. To clarify, here are examples. Say $E = (1, 4) \cup (3, 6)$, then we can write it in a simpler way as $E = (1, 6)$.

If $E = (1, 4) \cup [4, 6) \cup (-\infty, 0)$, then we would write $E = (-\infty, 0) \cup (1, 6)$.

We could say in this last example that there are two connected components of E. To make the representation totally unique (and not just unique up to the order of writing down the union, we are ordering them left to right).

It might help to think for youself of a few examples that are a little more involved.

Remark 2.1.10. We want to define a length function on subsets of $\mathbb{R}$, certainly including the sets in $\mathscr{I}$ for a start. But some of the intervals above have infinite length and so we are led to allow ∞ as a possible length.

That means we introduce an extended system of nonnegative numbers by adding an extra symbol ∞ to the usual $[0, \infty)$. We regard ∞ as bigger than every ordinary $x \in [0, \infty)$, that is we extend the definition of $x_1 < x_2$ for $x_1, x_2 \geq 0$ to have $x < \infty$ for all $x \in [0, \infty)$. Then we have $[0, \infty]$ including ∞.

We will also want to allow most arithmetic operations to be done where ∞ is allowed. So we will extend the definition of addition and multiplication of positive numbers so that

$$x + \infty = \infty + x = \infty + \infty = \infty \quad (0 \leq x < \infty),$$

$$x\infty = \infty x = \infty\infty = \infty \quad (0 < x < \infty),$$

$$0\infty = \infty 0 = 0$$

and sometimes we will have subtraction (mostly where we end up with a positive result) but we will **not define** $\infty - \infty$.

So we will have $\infty - x = \infty$ for $0 \leq x < \infty$.

We could go further and also introduce $-\infty$ (smaller than all $x \in \mathbb{R}$) and then get an extended real line $[-\infty, \infty]$, but we will not need that (until later).

Note that the rule $0\infty = \infty 0 = 0$ is not always good to allow, for instance in the case of limits, but it will be appropriate in the ways that we will encounter it for integration.

Definition 2.1.11. We define a length function (or 'measure') $m \colon \mathscr{I} \to [0, \infty]$ by taking

$$m((a, b)) = m([a, b)) = m((a, b]) = m([a, b]) = b - a,$$

$$m((-\infty, a)) = m((-\infty, a]) = m((a, \infty)) = m([a, \infty)) = m((-\infty, \infty)) = \infty$$

and defining $m(E)$ for general $E \in \mathscr{I}$ as the sum of the lengths of the intervals in the unique representation mentioned above in Remark 2.1.8.

(For $m(\emptyset)$ we are defining $m(\emptyset) = 0$, for example because we have $n = 0$ in the 'Standard Form' (Remark 2.1.8) for that case and then the empty sup should be considered formally to mean 0. Or just add $m(\emptyset) = 0$ as one more rule. In the case of nonempty E it will have a standard form $E = \bigcup_{j=1}^n I_j$ and we are defining $m(E) = \sum_{j=1}^n m(I_j)$.)

Example 2.1.12. Looking at one of the earlier examples $E = (1, 4) \cup (3, 6)$, we know $E \in \mathscr{J}$. To compute $m(E)$ from the definition we need E in its standard form $E = (1, 6)$. Then we have $m(E) = 6 - 1 = 5$ (the usual total length of E).

If $E = (1, 4) \cup (3, 6) \cup [-2, -1]$, the standard form would be $E = [-2, -1] \cup (1, 6)$ and $m(E) = (-1 - (-2)) + (6 - 1) = 1 + 5 = 6$.

Terminology 2.1.13. We say that a function $m\colon \mathcal{A} \to [0, \infty]$ is *finitely additive* if whenever $E_1, E_2, \ldots, E_n \in \mathcal{A}$ are disjoint (that is $E_j \cap E_k = \emptyset$ for $j \neq k$), then

$$m(E_1 \cup E_2 \cup \cdots \cup E_n) = m(E_1) + m(E_2) + \cdots + m(E_n).$$

By induction, to check finite additivity we only need to check it for $n = 2$.

We say that m is *finitely subadditive* if whenever $E_1, E_2, \ldots, E_n \in \mathcal{A}$, then

$$m(E_1 \cup E_2 \cup \cdots \cup E_n) \leq m(E_1) + m(E_2) + \cdots + m(E_n).$$

Again the case $n = 2$ implies the general case.

Lemma 2.1.14. *The length function m we have defined on $\mathscr{J}$ has the properties*

(i) $E_1, E_2 \in \mathscr{J}$*,* $E_1 \cap E_2 = \emptyset \Rightarrow m(E_1 \cup E_2) = m(E_1) + m(E_2)$ *(which implies finite additivity);*

(ii) $E_1, E_2 \in \mathscr{J}$*,* $E_1 \subseteq E_2 \Rightarrow m(E_1) \leq m(E_2)$ *(monotonicity);*

(iii) m *is finitely subadditive.*

Proof. Most of this is quite easy, almost obvious perhaps, but maybe not so easy to organise into a proper proof.

(i) It is sufficient to establish this in the case when $E_1 \cup E_2$ is an interval. The reason is this. We know $E_1 \cup E_2 \in \mathscr{J}$ and so we can write it in its 'standard form' (Remark 2.1.8) $E_1 \cup E_2 = \bigcup_{j=1}^n I_j$ where the I_j are intervals ordered from left to right (and no pair of the intervals overlaps and can only share an end point that is not included in either interval) Since $E_1 \subseteq E_1 \cup E_2$, it must be that when we write E_1 in the form we mentioned in Remark 2.1.8, then the intervals used for E_1 must all be contained in one of the intervals I_j. So it is clear then that $m(E_1) = \sum_{j=1}^n m(E_1 \cap I_j)$ and similarly $m(E_2) = \sum_{j=1}^n m(E_2 \cap I_j)$. Now $I_j = (E_1 \cap I_j) \cup (E_2 \cap I_j)$ (since $I_j \subseteq E_1 \cup E_2$) and that justifies the simplification.

Assuming then that $E_1 \cup E_2 = I$ is a single interval, the intervals that make up E_1 and E_2 must fit together to fill up I, with the intervals for E_1 alternating with those for E_2 (because $E_1 \cap E_2 = \emptyset$ and there must be gaps between the intervals mking up the standard form of E_1, also gaps between the intervals making up the standard form of E_2). It is then pretty obvious that the sum of the lengths of the intervals for E_1 together with those of E_2 must be $m(I)$. So $m(I) = m(E_1) + m(E_2)$, or $m(E_1 \cup E_2) = m(E_1) + m(E_2)$.

As noted before, this implies finite additivity.

(ii) If $E_1 \subseteq E_2$, then $E_2 = E_1 \cup (E_2 \setminus E_1)$, a disjoint union, and so (by (i))

$$m(E_2) = m(E_1 \cup (E_2 \setminus E_1)) = m(E_1) + m(E_2 \setminus E_1) \geq m(E_1).$$

(iii) For $E_1, E_2 \in \mathscr{I}$, we get $E_1 \cup E_2 = E_1 \cup (E_2 \setminus E_1)$, a disjoint union and so

$$m(E_1 \cup E_2) = m(E_1 \cup (E_2 \setminus E_1)) = m(E_1) + m(E_2 \setminus E_1) \leq m(E_1) + m(E_2)$$

using (i) and (ii). As noted before, this implies finite subadditivity. □

Remark 2.1.15. Looking back at the example with $y = \sin(1/x)$ in §0.3 we already indicated that we need to be able to deal with more than just lengths of finite unions of intervals. We need to be able at least to deal with sets that are made up of countably many intervals, and the obvious way to define their total length is to add up all the lengths.

This brings us to a more general kind of additivity, called 'countable additivity'.

Before we get to that, notice that there is a nice thing about series $\sum_{n=1}^\infty t_n$ of terms $t_n \in [0, \infty]$. If there is any term where $t_{n_0} = \infty$, then $\sum_{n=1}^N t_n = \infty$ whenever $N \geq n_0$ and so it seems perfectly logical to say that $\sum_{n=1}^\infty t_n = \infty$. On the other hand, if $t_n < \infty$ always then the sequence of partial sums $s_N = \sum_{n=1}^N t_n$ is a monotone increasing sequence. If the sequence of partial sums is bounded above (by a finite number), then $\lim_{N\to\infty} s_N = \sup_{N\geq 1} s_N$ and this number is $\sum_{n=1}^\infty t_n$ (in the ordinary sense). If the sequence of partial sums is not bounded above by a finite quantity, we define $\sum_{n=1}^\infty t_n = \infty$.

So every series $\sum_{n=1}^\infty t_n$ of terms t_n in $[0, \infty]$ has a sum.

Another good thing is that the sum remains the same if we change the order of the terms. That follows from the fact that the sum is the supremum (in $[0, \infty]$) of the partial sums.

Terminology 2.1.16. If $\mathcal{A}$ is an algebra and $m \colon \mathcal{A} \to [0, \infty]$, then we say that m is *countably additive* if whenever $E_1, E_2, E_3, \ldots$ is an infinite sequence of disjoint sets in $\mathcal{A}$ such that the union $\bigcup_{n=1}^\infty E_n \in \mathcal{A}$, then

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

What we want is to get to a situation where we consider only algebras $\mathcal{A}$ that are closed under taking countable unions, but the algebra $\mathscr{I}$ does not have that property. We will need one more result about the length function we have defined on $\mathscr{I}$.

Proposition 2.1.17. *The length function on $\mathscr{I}$ is countably additive.*

Before embarking on the proof, here are two little facts that we will need:

Lemma 2.1.18. *If $I \subseteq \mathbb{R}$ is a finite length nonempty interval (one with $m(I) < \infty$) and $\varepsilon > 0$ is arbitrary, then there is a finite closed interval $[c, d] \subseteq I$ with $d - c > m(I) - \varepsilon$.*

Proof. We could have $m(I) = 0$ in which case $I = [a, a]$ for some $a \in \mathbb{R}$ (as we are not allowing $I = \emptyset$). In that case take $c = d = a$.

Otherwise I has endpoints $a < b$ (which are finite). Take a number $\delta > 0$ so that $2\delta < \varepsilon$ and $2\delta < b - a$. Then put $c = a + \delta$, $d = b - \delta$. □

Lemma 2.1.19. *If $I \subseteq \mathbb{R}$ is a finite length interval (one with $m(I) < \infty$) and $\varepsilon > 0$ is arbitrary, then there is a finite open interval $(e, f) \supseteq I$ with $f - e < m(I) + \varepsilon$.*

Proof. In the case $I = \emptyset$ we can take $\delta = \varepsilon/3$ and then $e = -\delta$, $f = \delta$.

Otherwise I has endpoints $a < b$ (which are finite). Take $\delta = \varepsilon/3$, $e = a - \delta$, $f = b + \delta$. □

Now for the proof of Proposition 2.1.17.

Proof. As in the proof of finite additivity, it is enough to consider the case where an interval $I \in \mathscr{J}$ is a countable union $I = \bigcup_{n=1}^{\infty} E_n$ of disjoint $E_n \in \mathscr{I}$. Furthermore, we can write each $E_n = \bigcup_{j=1}^{j_n} I_{n,j}$ in its 'standard form' as a finite disjoint union of intervals (where $m(E_n) = \sum_{j=1}^{j_n} m(I_{n,j})$) and so we can reduce to the case where $I = \bigcup_{n=1}^{\infty} I_n$ is a disjoint union of intervals I_n. [To spell it out more, the idea is that $E = E_1 \cup E_2 \cup \cdots = I_{1,1} \cup I_{1,2} \cup \cdots I_{1,n_1} \cup I_{2,1} \cup \cdots \cup I_{2,n_2} \cup I_{3,1} \cup \cdots$ and we number all these intervals as $I_1, I_2, \ldots$.]

Since $\bigcup_{n=1}^{N} I_n \subseteq I$ for N finite, we can apply finite additivity and monotonicity to get

$$\sum_{n=1}^{N} m(I_n) = m\left(\bigcup_{n=1}^{N} I_n\right) \leq m(I)$$

for each N. Letting $N \to \infty$ gives half of what we want to show

$$\sum_{n=1}^{\infty} m(I_n) \leq m(I).$$

It remains to prove $\geq$ holds also.

If any of the intervals I_n is of infinite length, then it is clear that $I \supseteq I_n$ must be infinite also and so $m(I) = \sum_{n=1}^{\infty} I_n$ in this case.

So we assume each I_n is a finite (length) interval. Fix $\varepsilon > 0$ arbitrarily small and choose a finite closed interval $[a_0, b_0] \subset I$. If $m(I) < \infty$, choose $a_0 < b_0$ with $b_0 - a_0 > m(I) - \varepsilon$ (using Lemma 2.1.18). If I is infinite take $[a_0, b_0]$ a finite interval contained it of any large length.

Now $[a_0, b_0] \subseteq I \subseteq \bigcup_{n=1}^{\infty} I_n$. Using Lemma 2.1.19 we can find open intervals J_n with $I_n \subseteq J_n$ and $m(J_n) < m(I_n) + \varepsilon/2^n$. We then have an open cover $[a_0, b_0] \subseteq \bigcup_{n=1}^{\infty} I_n \subseteq \bigcup_{n=1}^{\infty} J_n$ of the compact interval $[a_0, b_0]$ by the open intervals J_n. By the Heine Borel theorem, there must be a finite subcover

$$[a_0, b_0] \subseteq \bigcup_{n=1}^{N} J_n$$

(for some $N < \infty$). By subadditivity of m

$$\begin{aligned} b_0 - a_0 = m([a_0, b_0] &\leq \sum_{j=1}^{N} m(J_n) \\ &\leq \sum_{j=1}^{N} (m(I_n) + \varepsilon/2^n) \\ &\leq \sum_{j=1}^{N} m(I_n) + \sum_{n=1}^{N} \varepsilon/2^n \\ &\leq \sum_{j=1}^{\infty} m(I_n) + \sum_{n=1}^{\infty} \varepsilon/2^n \\ &= \varepsilon + \sum_{j=1}^{\infty} m(I_n) \end{aligned}$$

In the case where $m(I) < \infty$ then we get

$$m(I) - \varepsilon \leq \varepsilon + \sum_{n=1}^{\infty} m(I_n) \Rightarrow m(I) \leq 2\varepsilon + \sum_{n=1}^{\infty} m(I_n) + 2\varepsilon.2$$

But this is true no matter how small $\varepsilon > 0$ is and so we have $m(I) \leq \sum_{n=1}^{\infty} m(I_n)$ if $m(I) < \infty$.

If $m(I) = \infty$ we can choose $b_0 - a_0$ as big as we like (with $\varepsilon = 1$ say) to get $\sum_{n=1}^{\infty} m(I_n) = \infty = m(I)$. □

Corollary 2.1.20. *The length function on $\mathscr{J}$ is countably subadditive — in the sense that whenever $E \in \mathscr{J}$ and $E_1, E_2, E_3, \ldots$ is an infinite sequence of sets in $\mathscr{J}$ such that the union $E \subseteq \bigcup_{n=1}^{\infty} E_n$ (think of the E_n as a countable cover of E), then*

$$m(E) \leq \sum_{n=1}^{\infty} m(E_n).$$

Proof. We can deduce this from countable additivity (Proposition 2.1.17) and monotonicty as follows. First we deal with the case where $E = \bigcup_{n=1}^{\infty} E_n \in \mathscr{J}$.

We generate a sequence of disjoint sets $F_1, F_2, \ldots \in \mathscr{I}$ such that $F_n \subseteq E_n$, $\bigcup_{j=1}^{n} F_j = \bigcup_{j=1}^{n} E_j$ for each n and $\bigcup_{n=1}^{\infty} F_n = E$. To do that let $F_1 = E_1$ and inductively define

$$F_n = E_n \setminus \left(\bigcup_{j=1}^{n-1} E_j \right) = E_n \cap \left(\bigcup_{j=1}^{n-1} E_j \right)^c \in \mathscr{I}$$

for $n > 1$.

The properties claimed for $(F_n)_{n=1}^\infty$ are rather easy to check (by induction) and the by countable addiivity of m we have

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

(because $F_n \subseteq E_n \Rightarrow m(F_n) \leq m(F_n)$ using monotonicty of m).

For the general case where $E \subsetneq \bigcup_{n=1}^\infty E_n$, note that

$$\bigcup_{n=1}^\infty (E_n \cap E) = \left(\bigcup_{n=1}^\infty E_n\right) \cap E = E$$

and so by the first case (since $E_n \cap E \in \mathscr{I}$ for all n)

$$m(E) \leq \sum_{n=1}^\infty m(E_n \cap E) \leq \sum_{n=1}^\infty m(E_n)$$

(using monotonicity of m to get $m(E_n \cap E) \leq m(E_n)$). $\square$

Notation 2.1.21. If $E \subseteq \mathbb{R}$ and $x_0 \in \mathbb{R}$ we define

$$x_0 + E = \{x_0 + x : x \in E\}$$

(and refer to $x_0 + E$ as the *translate* of E by x_0).

Proposition 2.1.22 (Translation invariance)**.** *For $E \in \mathscr{J}$ and $x_0 \in \mathbb{R}$ we have $x_0 + E \in \mathscr{J}$ and $m(x_0 + E) = m(E)$.*

Proof. Exercise. [Hint: Let $T(x) = x + x_0$, so that T translates x by x_0, and show that if I is an interval with endpoints $a < b$ (and length $m(I) = b - a$)), then $I + x_0$ has endpoints $a + x_0 < b + x_0$ (and the same length $(b + x_0) - (a + x_0) = b - a$). Look at E in standard form.] $\square$

2.2 Outer measure

Definition 2.2.1 (Outer measure)**.** We now define the *outer measure* of an arbitrary subset $S \subseteq \mathbb{R}$ to be

$$m^*(S) = \inf\left\{\sum_{n=1}^\infty m(E_n) : E_1, E_2, \ldots \in \mathscr{J} \text{ with } S \subseteq \bigcup_{n=1}^\infty E_n\right\}$$

(with the understanding that if $\sum_{n=1}^\infty m(E_n) = \infty$ always, then $m^*(S) = \infty$).

Remark 2.2.2. Outer measure is defined on the algebra $\mathcal{P}(\mathbb{R})$ of all subsets of $\mathbb{R}$, but it fails to be even finitely additive. We will identify a smaller algebra, the algebra of Lebesgue measurable subsets, on which m^* is countably additive. In the meantime, we will give some properties that m^* does have.

It should be clear soon that m^* is defined to be as large as it can be subject to two requirements (a) countably subadditive and (b) agrees with m on $\mathscr{I}$.

In the arguments that follow, you may like to notice any time we delve into what $\mathscr{I}$ is. We will be making use of rather general facts — the fact that $\mathscr{I}$ is an algebra, countable (sub) additivity of m, monotonicity of m, $m(\emptyset) = 0$ and translation invariance of m. But we won't be delving into the details of m or $\mathscr{I}$.

Since $\mathscr{I}$ is an algebra, for any $S \subseteq \mathbb{R}$ we can take $E_1 = \mathbb{R}$, $E_n = \emptyset$ for $n = 2, 3 \ldots$ and so there is always at least one possible choice for the sequence $(E_n)_{n=1}^\infty$ used in the definition of $m^*(S)$.

Proposition 2.2.3. *(i)* $m^*(\emptyset) = 0$

(ii) $S_1 \subseteq S_2 \subseteq \mathbb{R} \Rightarrow m^*(S_1) \leq m^*(S_2)$ *(monotonicity)*

(iii) If $S_1, S_2, \ldots \subseteq \mathbb{R}$*, then*

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

(recall that this is called countable subadditivity*)*

(iv) If $E \in \mathscr{I}$*, then* $m^*(E) = m(E)$.

(v) For $S \subseteq \mathbb{R}$ *and* $x_0 \in \mathbb{R}$, $m^*(x_0 + S) = m^*(S)$ *(translation invariance of outer measure).*

Proof. (i) We can take $E_n = \emptyset \in \mathscr{I}$ for each $n \geq 1$ and then $\emptyset \subseteq \bigcup_{n=1}^\infty E_n$ while $\sum_{n=1}^\infty m(E_n) = 0$. So $m^*(\emptyset) \leq 0$. Hence $m^*(\emptyset) = 0$.

(ii) If $m^*(S_2) = \infty$ then certainly the inequality is true. For the case $m^*(S_2) < \infty$, if $E_1, E_2, \ldots \in \mathscr{I}$ with $S_2 \subseteq \bigcup_{n=1}^\infty E_n$, then certainly also $S_1 \subseteq \bigcup_{n=1}^\infty E_n$ and $m^*(S_1) \leq \sum_{n=1}^\infty m(E_n)$. But this sum can be arbitrarily close to $m^*(S_2)$ and so $m^*(S_1) \leq m^*(S_2)$.

(iii) If $\sum_{n=1}^\infty m^*(S_n) = \infty$, then the inequality is true and so we consider the case $\sum_{n=1}^\infty m^*(S_n) < \infty$. Fix $\varepsilon > 0$ and for each n, choose $E_{n,1}, E_{n,2}, \ldots \in \mathscr{I}$ with

$$S_n \subseteq \bigcup_{j=1}^\infty E_{n,j} \text{ and } \sum_{j=1}^\infty m(E_{n,j}) \leq m^*(S_n) + \frac{\varepsilon}{2^n}.$$

Then, taking $S = \bigcup_{n=1}^\infty S_n$ we have

$$S \subseteq \bigcup_{n=1}^\infty \bigcup_{j=1}^\infty E_{n,j}$$

(a countable number of sets in $\mathscr{I}$ that can be arranged as a single sequence) and so

$$m^*(S) \leq \sum_{n=1}^\infty \sum_{j=1}^\infty m(E_{n,j}) \leq \sum_{n=1}^\infty \left(m^*(S_n) + \frac{\varepsilon}{2^n}\right) = \sum_{n=1}^\infty m^*(S_n) + \varepsilon.$$

Since we can take $\varepsilon > 0$ arbitrarily small we have the result.

(iv) For $E \in \mathscr{J}$ we can take $E_1 = E$ and $E_n = \emptyset$ for $n = 2, 3, \ldots$. That gives $E_1, E_2, \ldots \in \mathscr{J}$ with $E \subseteq \bigcup_{n=1}^\infty E_n$ and shows that $m^*(E) \leq \sum_{n=1}^\infty m(E_n) = m(E) + 0 = m(E)$.

On the other hand if we have any $E_1, E_2, \ldots \in \mathscr{J}$ with $E \subseteq \bigcup_{n=1}^\infty E_n$, then we can say that

$$E = E \cap \left(\bigcup_{n=1}^\infty E_n \right) = \bigcup_{n=1}^\infty E \cap E_n \in \mathscr{J}$$

and apply countable subadditivity of m (Corollary 2.1.20) to deduce

$$m(E) \leq \sum_{n=1}^\infty m(E \cap E_n) \leq \sum_{n=1}^\infty m(E_n).$$

In this way we conclude that $m(E) \leq m^*(E)$. Combining with the reverse inequality obtained first we get equality, as required.

(v) We show $m^*(x_0 + S) \leq m^*(S)$ first. Of course if $m^*(S) = \infty$, this is true. If $m^*(S) < \infty$, for any $\varepsilon > 0$ we must be able to find $E_1, E_2, \ldots \in \mathscr{J}$ so that $S \subseteq \bigcup_{n=1}^\infty E_n$ and $\sum_{n=1}^\infty m(E_n) \leq m^*(S) + \varepsilon$. Then $x_0 + S \subseteq \bigcup_{n=1}^\infty (x_0 + E_n)$, where we know $x_0 + E_n \in \mathscr{J}$ with $m(x_0 + E_n) = m(E_n)$ (by translation invariance of m), and so we have

$$m^*(x_0 + S) \leq \sum_{n=1}^\infty m(x_0 + E_n) = \sum_{n=1}^\infty m(E_n) \leq m^*(S) + \varepsilon.$$

As $\varepsilon > 0$ can be arbitrary, we conclude $m^*(x_0 + S) \leq m^*(S)$.

The reverse inequality follows by applying the same fact to $-x_0$ and $\tilde{S} = x_0 + S$ to get $m^*((-x_0) + \tilde{S}) \leq m^*(\tilde{S})$, which means $m^*(S) \leq m^*(x_0 + S)$. So we have equality. □

Corollary 2.2.4. *m^* is finitely subadditive. That is, if $S_1, S_2, \ldots, S_n \subseteq \mathbb{R}$, then*

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

Proof. Exercise. □

2.3 Lebesgue measurable sets

Although we have not justified the claim that m^* fails to be finitely additive, we use it as a way to distinguish the well-behaved sets.

Definition 2.3.1. We say that a subset $F \subset \mathbb{R}$ is *Lebesgue measurable* (or measurable with respect to the outer measure m^*) if for every subset $S \subset \mathbb{R}$,

$$m^*(S) = m^*(S \cap F) + m^*(S \cap F^c).$$

We denote the collection of Lebesgue measurable sets by $\mathscr{L}$.

The criterion used to define $\mathscr{L}$ is called the Carathéodory criterion. Our aim now is to show that this class $\mathscr{L}$ is a well-behaved algebra, in fact a σ-algebra and that the restriction of m^* to $\mathscr{L}$ is also well-behaved. We need some definitions so we can express things concisely.

Notice that the Carathéodory criterion says that F and F^c divides every subset $S \subseteq \mathbb{R}$ additively for m^*. Here is a small simplification of the criterion.

Lemma 2.3.2. *A set $F \subset \mathbb{R}$ is in $\mathscr{L}$ if it satisfies*

$$m^*(S) \geq m^*(S \cap F) + m^*(S \cap F^c).$$

for each $S \subseteq \mathbb{R}$.

Proof. By subadditivity of m^*, we always know

$$m^*(S) \leq m^*(S \cap F) + m^*(S \cap F^c).$$

and combining with the assumption in the lemma we get equality, □

Example 2.3.3. If $F \subset \mathbb{R}$ has $m^*(F) = 0$, then $F \in \mathscr{L}$.

Proof. From Lemma 2.3.2 what we need to do is consider and arbitrary $S \subseteq \mathbb{R}$ and show

$$m^*(S) \geq m^*(S \cap F) + m^*(S \cap F^c).$$

Since $S \cap F \subseteq F$, we have $m^*(S \cap F) \leq m^*(F) = 0$ (as m^* is monotone) and so $m^*(S \cap F) = 0$. Hence what we need to show is

$$m^*(S) \geq 0 + m^*(S \cap F^c) = m^*(S \cap F^c).$$

But $S \cap F^c \subseteq S \Rightarrow m^*(S \cap F^c) \leq m^*(S)$ and that shows $F \in \mathscr{L}$. □

Definition 2.3.4. An algebra $\mathcal{A}$ (of subsets of $\mathbb{R}$) is called a σ-algebra ('sigma-algebra') if whenever $E_1, E_2, \ldots \in \mathcal{A}$, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{A}$. [In words: $\mathcal{A}$ is closed under the operation of taking countable unions.]

Definition 2.3.5. If Σ is a σ-algebra (of subsets of $\mathbb{R}$) and $\mu\colon \Sigma \to [0, \infty]$ is a function, then we call μ a *measure* on Σ if it satisfies

(a) $\mu(\emptyset) = 0$

(b) μ is countably additive, that is whenever $E_1, E_2, \ldots \in \Sigma$ are disjoint, then $\mu\left(\bigcup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mu(E_n)$.

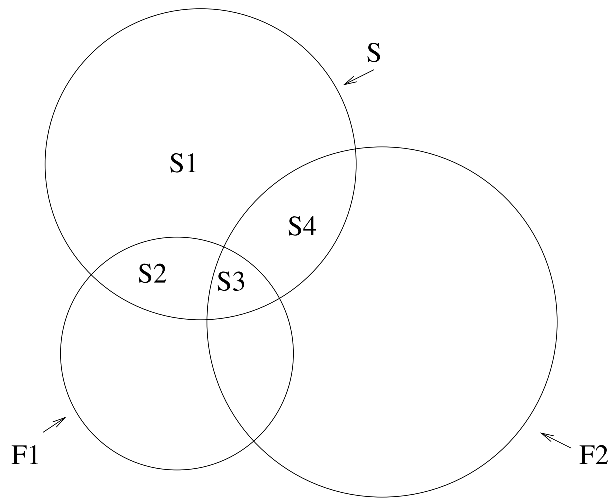
Theorem 2.3.6 (Carathéodory)**.** *$\mathscr{L}$ is a σ-algebra, $\mathscr{J} \subset \mathscr{L}$ and the restriction of outer measure m^* to $\mathscr{L}$ defines a measure on $\mathscr{L}$.*

Proof. **Step 1:** $\mathscr{L}$ is an algebra.

(i) $\emptyset \in \mathscr{L}$
For any $S \subseteq \mathbb{R}$, $m^*(S \cap \emptyset) + m^*(S \cap \emptyset^c) = m^*(\emptyset) + m^*(S) = 0 + m^*(S) = m^*(S)$. (We could also use the more general result from Example 2.3.3.)

(ii) $F \in \mathscr{L} \Rightarrow F^c \in \mathscr{L}$
For any $S \subseteq \mathbb{R}$, $m^*(S \cap F^c) + m^*(S \cap (F^c)^c) = m^*(S \cap F^c) + m^*(S \cap F) = m^*(S \cap F) + m^*(S \cap F^c) = m^*(S)$ by measurability of F.

(iii) $F_1, F_2 \in \mathscr{L} \Rightarrow F_1 \cup F_2 \in \mathscr{L}$.
We first show $F_1 \cap F_2 \in \mathscr{L}$ (for $F_1, F_2 \in \mathscr{L}$). Take $S \subseteq \mathbb{R}$. It may help to look at this Venn diagram to follow the argument.



Notice that

$$\begin{aligned} m^*(S) &= m^*(S \cap F_1) + m^*(S \cap F_1^c) \\ &= m^*(S \cap F_1 \cap F_2) + m^*(S \cap F_1 \cap F_2^c) + m^*(S \cap F_1^c) \end{aligned}$$

(using first $F_1 \in \mathscr{L}$, then $F_2 \in \mathscr{L}$). But also

$$\begin{aligned} m^*(S \cap (F_1 \cap F_2)^c) &= m^*(S \cap (F_1 \cap F_2)^c \cap F_1) + m^*(S \cap (F_1 \cap F_2)^c \cap F_1^c) \\ &\qquad (F_1 \text{ and } F_1^c \text{ split } S \cap (F_1 \cap F_2)^c \text{ additively}) \\ &= m^*(S \cap (F_1^c \cup F_2^c) \cap F_1) + m^*(S \cap (F_1^c \cup F_2^c) \cap F_1^c) \\ &= m^*(S \cap F_2^c \cap F_1) + m^*(S \cap F_1^c) \\ &= m^*(S \cap F_1 \cap F_2^c) + m^*(S \cap F_1^c) \end{aligned}$$

and so we get

$$m^*(S) = m^*(S \cap F_1 \cap F_2) + m^*(S \cap (F_1 \cap F_2)^c).$$

This shows $F_1 \cap F_2 \in \mathscr{L}$
As $F_1 \cup F_2 = (F_1^c \cap F_2^c)^c$, we get $F_1 \cup F_2 \in \mathscr{L}$ because $F_1^c, F_2^c \in \mathscr{L} \Rightarrow F_1^c \cap F_2^c \in \mathscr{L} \Rightarrow (F_1^c \cap F_2^c)^c \in \mathscr{L}$, or $F_1 \cup F_2 \in \mathscr{L}$.

Step 2: $\mathscr{L}$ is a σ-algebra and m^* is countably additive on $\mathscr{L}$.

If $F_1, F_2, \ldots \in \mathscr{L}$ we aim to show that $F = \bigcup_{n=1}^\infty F_n \in \mathscr{L}$. It is enough to do that when $F_1, F_2, \ldots$ are disjoint, since we can replace $F_1, F_2, \ldots$ by $\tilde{F}_1 = F_1$, $\tilde{F}_2 = F_2 \cap F_1^c$, $\ldots$, $\tilde{F}_n = F_n \cap \left(\bigcup_{j=1}^{n-1} F_j\right)^c$, thus getting a disjoint sequence in $\mathscr{L}$ with the same union $F = \bigcup_{n=1}^\infty \tilde{F}_n$.

So we assume that $F_1, F_2, \ldots$ are disjoint, $F = \bigcup_{n=1}^\infty F_n$.

Take any $S \subseteq \mathbb{R}$. We claim that $m^*(S) \geq m^*(S \cap F) + m^*(S \cap F^c)$ always (which shows $F \in \mathscr{L}$ via Lemma 2.3.2).

Take $G_n = \bigcup_{j=1}^n F_n$ (which we know is in $\mathscr{L}$ as $\mathscr{L}$ is an algebra) and notice that

$$m^*(S) \geq m^*(S \cap G_n) + m^*(S \cap G_n^c).$$

Since $F_n \in \mathscr{L}$ we have

$$m^*(S \cap G_n) = m^*(S \cap G_n \cap F_n) + m^*(S \cap G_n \cap F_n^c) = m^*(S \cap F_n) + m^*(S \cap G_{n-1})$$

(for $n > 1$). Since $F_1 = G_1$ we then can see by induction on n that

$$m^*(S \cap G_n) = \sum_{j=1}^n m^*(S \cap F_j),$$

and so

$$m^*(S) \geq \sum_{j=1}^n m^*(S \cap F_j) + m^*(S \cap G_n^c).$$

However $G_n \subset F$ implies $G_n^c \supset F^c$ and so (by monotonicity of m^*) we get

$$m^*(S) \geq \sum_{j=1}^n m^*(S \cap F_j) + m^*(S \cap F^c).$$

Let $n \to \infty$ to get

$$m^*(S) \geq \sum_{n=1}^\infty m^*(S \cap F_n) + m^*(S \cap F^c). \tag{2.3.1}$$

From subadditivity of m^*, this inequality implies

$$m^*(S) \geq m^*\left(\bigcup_{n=1}^\infty S \cap F_n\right) + m^*(S \cap F^c) = m^*(S \cap F) + m^*(S \cap F^c).$$

So $F \in \mathscr{L}$ is verified.

Using (2.3.1) for $S = F$ we get

$$m^*(F) = m^*\left(\bigcup_{n=1}^{\infty} F_n\right) \geq \sum_{n=1}^{\infty} m^*(F_n)$$

and subadditivity of m^* gives the opposite inequality. That shows countable additivity of m^* on $\mathscr{L}$.

Step 3: $\mathscr{J} \subseteq \mathscr{L}$.

Fix $E \in \mathscr{J}$. Take any $S \subseteq \mathbb{R}$. We claim that $m^*(S) \geq m^*(S \cap E) + m^*(S \cap E^c)$ always (which shows $E \in \mathscr{L}$ via Lemma 2.3.2). If $m^*(S) = \infty$ there is nothing to do.

For any $\varepsilon > 0$ (fixed) and then there must be $E_1, E_2, \ldots \in \mathscr{J}$ with $S \subseteq \bigcup_{n=1}^{\infty} E_n$ and $\sum_{n=1}^{\infty} m(E_n) < m^*(S) + \varepsilon$.

We then have $S \cap E \subseteq \bigcup_{n=1}^{\infty} E_n \cap E$, and $E_n \cap E \in \mathscr{J}$. So

$$m^*(S \cap E) \leq \sum_{n=1}^{\infty} m(E_n \cap E).$$

Similarly, as $E^c \in \mathscr{J}$,

$$m^*(S \cap E^c) \leq \sum_{n=1}^{\infty} m(E_n \cap E^c).$$

But $m(E_n) = m(E_n \cap E) + m(E_n \cap E^c)$ (finite additivity of m) and so

$$\begin{aligned} m^*(S \cap E) + m^*(S \cap E^c) &\leq \sum_{n=1}^{\infty} m(E_n \cap E) + \sum_{n=1}^{\infty} m(E_n \cap E^c) \\ &= \sum_{n=1}^{\infty} (m(E_n \cap E) + m(E_n \cap E^c)) \\ &= \sum_{n=1}^{\infty} m(E_n) \\ &\leq m^*(S) + \varepsilon. \end{aligned}$$

As this holds for each $\varepsilon > 0$ we have $m^*(S \cap E) + m^*(S \cap E^c) \leq m^*(S)$ and that completes the argument. □

Remark 2.3.7. It may be a good idea now to recap a little on where we have reached.

We began with lengths of finite intervals I with endpoints $a < b$ (or even $a = b$ in the case of one point intervals), where we assigned the length $m(I) = b - a$, together with infinite intervals to which we assigned infinite length. Just to make things more convenient we worked with finite unions of intervals of these types so that we would have an algebra of sets to work with.

For lengths of sets in that algebra, the interval algebra $\mathscr{J}$, we showed that certain nice properties worked.

We then defined outer measure for arbitrary subsets of $\mathbb{R}$, basically by making $m^*(E)$ be the biggest thing it could be if we wanted it to agree with $m(E)$ when E is in the interval algebra $\mathscr{J}$ and to be countably subadditive.

Finally we identified a σ-algebra $\mathscr{L}$ containing $\mathscr{J}$ and on which m^* is well-behaved in the sense of being countably additive.

We have pretty much what we need to start looking at integration, but first we take stock a bit of this σ-algebra $\mathscr{L}$.

Proposition 2.3.8. *$\mathscr{L}$ is translation invariant. That is if $F \in \mathscr{L}$ and $x_0 \in \mathbb{R}$, then $x_0 + F \in \mathscr{L}$.*

Proof. Exercise. (It is not that hard because we already know $m^*(x_0 + E) = m^*(E)$ for every $E \subseteq \mathbb{R}$. See Proposition 2.2.3 (v).) $\square$

Proposition 2.3.9. *If $\mathcal{C}$ is a set of subsets of $\mathbb{R}$ (so that means $\mathcal{C} \subseteq \mathcal{P}(\mathbb{R})$ is a subset of the power set of $\mathbb{R}$, the collection of all subsets of $\mathbb{R}$), then $\mathcal{C}$ is a σ-algebra if and only if it satisfies*

(i) $\emptyset \in \mathcal{C}$

(ii) $E \in \mathcal{C} \Rightarrow E^c \in \mathcal{C}$

(iii) $E_1, E_2, \ldots \in \mathcal{C} \Rightarrow \bigcup_{n=1}^\infty E_n \in \mathcal{C}$.

Proof. We defined a σ-algebra to be an algebra with the additional property of being closed under taking countable unions. So every σ-algebra $\mathcal{C}$ has these three properties.

For the converse, suppose $\mathcal{C}$ has these three properties. We are just missing the property of an algebra that

$$E_1, E_2 \in \mathcal{C} \Rightarrow E_1 \cup E_2 \in \mathcal{C}.$$

However, if $E_1, E_2 \in \mathcal{C}$ we can make an infinite sequence by defining $E_n = \emptyset$ for $n \geq 3$. Now $E_1, E_2, \ldots \in \mathcal{C}$ and so we have $\bigcup_{n=1}^\infty E_n \in \mathcal{C}$. But $\bigcup_{n=1}^\infty E_n = E_1 \cup E_2$ because the other sets in the sequence are all empty. So we do have $E_1 \cup E_2 \in \mathcal{C}$. $\square$

Proposition 2.3.10. *For any set $\mathcal{S} \subset \mathcal{P}(\mathbb{R})$ of subsets of $\mathbb{R}$, there is a smallest σ-algebra (of subsets of $\mathbb{R}$) that contains $\mathcal{S}$.*

We call it the σ-algebra generated by $\mathcal{S}$ and denote it sometimes by $\Sigma_{\mathcal{S}}$.

Proof. This is sort of easy, but in an abstract way. There is certainly one possible σ-algebra containing $\mathcal{S}$, that is $\mathcal{P}(\mathbb{R})$.

What we do is look at all possible σ-algebra, the set

$$\mathscr{S} = \{\Sigma : \Sigma \subseteq \mathcal{P}(\mathbb{R}) \text{ a } \sigma\text{-algebra and } \mathcal{S} \subset \Sigma\}.$$

Then we take their intersection

$$\Sigma_{\mathcal{S}} = \bigcap_{\Sigma \in \mathscr{S}} \Sigma$$

and argue that $\Sigma_{\mathcal{S}}$ is still a σ-algebra. In fact that $\Sigma_{\mathcal{S}} \in \mathscr{S}$ and is the contained in every $\Sigma \in \mathscr{S}$ by the way we defined it.

We check

(i) $\emptyset \in \Sigma_{\mathcal{S}}$

because $\emptyset \in \Sigma$ for each $\Sigma \in \mathscr{S}$ (and $\mathscr{S}$ is not empty — there is at least one Σ in it).

(ii) $E \in \Sigma_{\mathcal{S}} \Rightarrow E^c \in \Sigma_{\mathcal{S}}$

because $E \in \Sigma_{\mathcal{S}} \Rightarrow E \in \Sigma$ for each $\Sigma \in \mathscr{S}$, and so $E^c \in \Sigma$ for each $\Sigma \in \mathscr{S}$. Thus $E^c \in \Sigma_{\mathcal{S}}$.

(iii) $E_1, E_2, \ldots \in \Sigma_{\mathcal{S}} \Rightarrow \bigcup_{n=1}^{\infty} E_n \in \Sigma_{\mathcal{S}}$

If $E_n \in \Sigma_{\mathcal{S}}$ for $n = 1, 2, \ldots$, then we have $E_1, E_2, \ldots \in \Sigma$ for each $\Sigma \in \mathscr{S}$. So $\bigcup_{n=1}^{\infty} E_n \in \Sigma$ for each $\Sigma \in \mathscr{S}$, and hence $\bigcup_{n=1}^{\infty} E_n \in \Sigma_{\mathcal{S}}$.

Finally $\mathcal{S} \subset \Sigma_{\mathcal{S}}$ holds because $\mathcal{S} \subset \Sigma$ for each $\Sigma \in \mathscr{S}$. □

Lemma 2.3.11. *If Σ is a σ-algebra (of subsets of $\mathbb{R}$) and $E_1, E_2, \ldots \in \Sigma$, then $\bigcap_{n=1}^{\infty} E_n \in \Sigma$. (So σ-algebras are closed under countable intersections as well as under countable unions.)*

Proof. This is because De Morgans laws allow us to turn intersections unto unions via taking complements:

$$\bigcap_{n=1}^{\infty} E_n = \left(\bigcup_{n=1}^{\infty} E_n^c \right)^c$$

We have $E_n^c \in \Sigma$ for all n, hence $\bigcup_{n=1}^{\infty} E_n^c \in \Sigma$, hence $\left(\bigcup_{n=1}^{\infty} E_n^c\right)^c \in \Sigma$. □

Theorem 2.3.12. *The following σ-algebras of subsets of $\mathbb{R}$ are all the same σ-algebra (and usually called the* Borel σ-algebra *on $\mathbb{R}$)*

1. $\Sigma_1 =$ the σ-algebra generated by $\mathscr{I}$ (the interval algebra)

2. $\Sigma_2 =$ the σ-algebra generated by the collection of all finite open intervals (a, b) (with $a < b$)

3. $\Sigma_3 =$ the σ-algebra generated by the collection of all finite closed intervals $[a, b]$ (with $a \leq b$)

4. $\Sigma_4 =$ the σ-algebra generated by the collection of all open subsets of $\mathbb{R}$

5. $\Sigma_5 =$ the σ-algebra generated by the collection of all closed subsets of $\mathbb{R}$

Once we know they are all equal we denote the σ-algebra by Σ_{Borel}. Sets in Σ_{Borel} are called Borel subsets of $\mathbb{R}$.

Proof. It helps to notice first that one point sets are included in each of these σ-algebras.

Starting with the interval algebra $\mathscr{I}$ and any $x_0 \in \mathbb{R}$ we have $\{x_0\} = [x_0, x_0] \in \mathscr{I}$. Thus $\{x_0\} \in \Sigma_1$.

For open intervals, we can say instead that $\{x_0\} = \bigcap_{n=1}^{\infty} \left(x_0 - \frac{1}{n}, x_0 + \frac{1}{n}\right)$. So we have $\{x_0\} \in \Sigma_2$.

That also covers Σ_4 (open intervals are open sets).

Since single point sets are closed, we have no bother with $\{x_0\} \in \Sigma_5$ (for each $x_0 \in \mathbb{R}$). Since single point sets are closed intervals also, because $[a, a] = \{a\}$, the fact that $\{x_0\} \in \Sigma_3$ is also easy.

We can see immediately that any σ-algebra that contains $\mathscr{J}$ (in particular Σ_1) must contain finite open intervals (a, b). It follows that $\Sigma_1 \supseteq \Sigma_2$.

To show $\Sigma_2 \supseteq \Sigma_1$ we need to show Σ_2 contains all the kinds of intervals that generate the algebra $\mathscr{J}$, namely all intervals. We have $(a, b] = (a, b) \cup \{b\} \in \Sigma_2$, $[a, b] = (a, b) \cup \{a\} \in \Sigma_2$, and $[a, b] = (a, b) \cup \{a\} \cup \{b\} \in \Sigma_2$ if $a < b$. Since we also have one point (closed) intervals $[a, a] \in \Sigma_2$, that covers all the finite ones. Since $(a, \infty) = \bigcup_{n=1}^{\infty}(a + n - 1, a + n] \in \Sigma_2$ (countable union of sets in Σ_2 already), $[a, \infty) = \{a\} \cup (a, \infty) \in \Sigma_2$, $(-\infty, a) = [a, \infty)^c \in \Sigma_2$, $(-\infty, a] = \{a\} \cup (a, \infty)^c \in \Sigma_2$, and $(-\infty, \infty) = \emptyset^c \in \Sigma_2$ (true for any σ-algebra). It follows that all intervals are in Σ_2, thus $\mathscr{J} \subset \Sigma_2$ and so $\Sigma_1 = \Sigma_{\mathscr{J}} \subseteq \Sigma_2$.

$\Sigma_2 \supseteq \Sigma_3$ since $[a, b] = (a, b) \cup \{a\} \cup \{b\}$ and $\Sigma_3 \supseteq \Sigma_2$ also because $(a, b) = [a, b] \cap (\{a\} \cup \{b\})^c$. So $\Sigma_2 = \Sigma_3$.

To show $\Sigma_2 \supseteq \Sigma_4$ we show that every open subset $U \subseteq \mathbb{R}$ is a countable union of open intervals. For each $x \in U$, there is $\delta > 0$ with $(x - \delta, x + \delta) \subset U$. (That is what it means for U to be an open set.) Now there are rational numbers q_1, q_2 with $x - \delta < q_1 < x < q_2 < x + \delta$ and that means $x \in (q_1, q_2) \subset U$. It follows that U is the union of all open intervals (q_1, q_2) with rational endpoints $q_1 < q_2$ that are contained in U. But there are fewer such intervals than there are pairs $(q_1, q_2) \in \mathbb{Q} \times \mathbb{Q} = \mathbb{Q}^2$. Since $\mathbb{Q}^2$ is a countable set, the collection of these rational intervals contained in U is countable. So U is a countable union of open intervals, hence $U \in \Sigma_2$. This shows $\Sigma_2 \supseteq \Sigma_4$.

Since the finite open intervals (a, b) are all open sets in $\mathbb{R}$, it is clear that they are all in Σ_4 and $\Sigma_4 \supseteq \Sigma_2$. Combining we have $\Sigma_2 = \Sigma_4$.

So we now have $\Sigma_1 = \Sigma_2 = \Sigma_3 = \Sigma_4$.

Finally $\Sigma_4 = \Sigma_5$ because closed sets are exactly the complements of open sets. □

Corollary 2.3.13. *The Borel σ-algebra (of subsets of $\mathbb{R}$) is the σ-algebra generated by the intervals $(-\infty, b]$ with $b \in \mathbb{R}$.*

Proof. We have

$$(a, b] = (-\infty, b] \cap ((-\infty, a])^c.$$

So any σ-algebra that contains the intervals $(-\infty, b]$ must also contain all half-open intervals of the form $(a, b]$. Then it must contain single point sets as $\{b\} = \bigcap_{n=1}^{\infty}(b - 1/n, b]$. So it contains all closed intervals $[a, b] = \{a\} \cup (a, b]$.

Thus it must contain the Borel σ-algebra (as the Borel σ-algebra is Σ_3 above).

The reverse inclusion, that Borel σ-algebra must contain the σ-algebra generated by the intervals $(-\infty, b]$ is easy because each $(-\infty, b] \in \mathscr{J}$. □

Remark 2.3.14. We have not proved that there any sets not included in $\mathscr{L}$. This is 'true', though the proof relies on the axiom of choice, something that cannot be proved based on the usual axioms of mathematical logic.

The idea for the proof is to define an equivalence relation on $\mathbb{R}$ by $x \sim y \iff x - y \in \mathbb{Q}$. Then choose one $x \in [0,1]$ belonging to each equivalence class and let E be the set of those x. Now

$$[0,1] \subseteq \bigcup_{q \in \mathbb{Q} \cap [-1,1]} (q + E) \subseteq [-1,2]$$

We know $m^*(q + E) = m^*(E)$ for each q. If E was in $\mathscr{L}$, then so would $q + E$ for each q, and countable additivity of m^* on $\mathscr{L}$ would give a contradiction from

$$1 = m^*([0,1]) \leq m^* \left(\bigcup_{q \in \mathbb{Q} \cap [-1,1]} (q + E) \right) = \sum_{q \in \mathbb{Q} \cap [-1,1]} m^*(q + E) \leq m^*([-1,2]) = 3.$$

(Here we use that the sets $q + E$ are disjoint for different $q \in \mathbb{Q}$.) If $m^*(E) > 0$, the sum would be ∞, while if $m^*(E) = 0$, the sum would be 0. Neither possibility is between 1 and 3. So $E \notin \mathscr{L}$.

Remark 2.3.15. We have also not proved that there are fewer Borel sets than there are Lebesgue measurable sets, but that is also true (and does not need the axiom of choice — it needs just a bit more about how to distinguish different sizes of uncountable sets than we have looked into).

It is not too hard to show that for each $E \in \mathscr{L}$, there is a set B is in the Borel σ-algebra (that is a Borel set B) with $m^*(E \Delta B) = 0$ (where $E \Delta B = (E \setminus B) \cup (B \setminus E)$ is the symmetric difference). I think we can omit that proof. The opposite is relatively easy to prove: that if B is in the Borel σ-algebra and $E \subset \mathbb{R}$ has $m^*(E \Delta B) = 0$, then $E \in \mathscr{L}$.

So, in a way, the difference between the Borel σ-algebra and $\mathscr{L}$ comes from the fact that $\mathscr{L}$ includes all sets E with $m^*(E) = 0$.