Probability with Measure

Dr Nic Freeman

September 16, 2020

Contents

Chapter 0

Introduction

OMITTED

0.1 Preliminaries

This section contains lots of definitions, from earlier courses, that we will use in MAS350. Most of the material here should be familiar to you. There may be one or two minor extensions of ideas you have seen before.

1. *Set Theory.*

    Let S be a set and $A, B, C, \ldots$ be subsets.

    A^c is the complement of A in S so that

    $$A^c = \{x \in S; x \notin A\}.$$

    Union $A \cup B = \{x \in S; x \in A \text{ or } x \in B\}$.

    Intersection $A \cap B = \{x \in S; x \in A \text{ and } x \in B\}$.

    Set theoretic difference: $A - B = A \cap B^c$.

    We have finite and infinite unions and intersections so if $A_1, A_2, \ldots, A_n$ are subsets of S.

    $$\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \ldots \cup A_n.$$
    $$\bigcap_{i=1}^{n} A_i = A_1 \cap A_2 \cap \ldots \cap A_n.$$

    We will also need *infinite* unions and intersections. So let (A_n) be a sequence of subsets in S.

    Let $x \in S$. We say that $x \in \bigcup_{i=1}^{\infty} A_i$ if $x \in A_i$ for at least one value of i. We say that $x \in \bigcap_{i=1}^{\infty} A_i$ if $x \in A_i$ for all values of i.

    Note that *de Morgan's laws* hold in this context:

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

    The *Cartesian product* $S \times T$ of sets S and T is

    $$S \times T = \{(s, t); s \in s, t \in T\}.$$

2. *Sets of Numbers*

    - Natural numbers $\mathbb{N} = \{1, 2, 3, \ldots\}$.
    - Non-negative integers $\mathbb{Z}_+ = \mathbb{N} \cup \{0\} = \{0, 1, 2, 3, \ldots\}$.
    - Integers $\mathbb{Z}$.
    - Rational numbers $\mathbb{Q}$.
    - Real numbers $\mathbb{R}$.

- Complex numbers $\mathbb{C}$.

A set X is *countable* if there exists an injection between X and $\mathbb{N}$. A set is *uncountable* if it fails to be countable. $\mathbb{N}, \mathbb{Z}_+, \mathbb{Z}$ and $\mathbb{Q}$ are countable. $\mathbb{R}$ and $\mathbb{C}$ are uncountable. All finite sets are countable.

3. *Images and Preimages.*

Suppose that S_1 and S_2 are two sets and that $f : S_1 \to S_2$ is a mapping (or function). Suppose that $A \subseteq S_1$. The *image* of A under f is the set $f(A) \subseteq S_2$ defined by

$$f(A) = \{y \in S_2; y = f(x) \text{ for some } x \in S_1\}.$$

If $B \subseteq S_2$ the *inverse image* of B under f is the set $f^{-1}(B) \subseteq S_1$ defined by

$$f^{-1}(B) = \{x \in S_1; f(x) \in B\}.$$

Note that $f^{-1}(B)$ makes sense irrespective of whether the mapping f is invertible.

Key properties are, with $A, A_1, A_2 \subseteq S_1$ and $B, B_1, B_2 \subseteq S_2$:

$$f^{-1}(B_1 \cup B_2) = f^{-1}(B_1) \cup f^{-1}(B_2),$$
$$f^{-1}(B_1 \cap B_2) = f^{-1}(B_1) \cap f^{-1}(B_2),$$
$$f^{-1}(A^c) = f^{-1}(A)^c,$$
$$f(A_1 \cup A_2) = f(A_1) \cup f(A_2),$$
$$f(A_1 \cap A_2) \subseteq f(A_1) \cap f(A_2),$$
$$f(f^{-1}(B)) \subseteq B$$
$$(f \circ g)^{-1}(A) = g^{-1}(f^{-1}(A))$$
$$A \subseteq f^{-1}(f(A)),$$
$$A \subseteq B \quad \Rightarrow \quad f^{-1}(A) \subseteq f^{-1}(B).$$

4. *Extended Real Numbers*

We will often find it convenient to work with ∞ and $-\infty$. These are *not* real numbers, but we find it convenient to treat them a bit like real numbers. To do so we specify the extra arithmetic rules, for all $x \in \mathbb{R}$,

$$\infty + x = x + \infty = \infty,$$

$$x - \infty = -\infty + x = -\infty,$$

$$\infty.x = x.\infty = \infty \text{ for } x > 0,$$

$$\infty.x = x.\infty = -\infty \text{ for } x < 0,$$

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

Note that $\infty - \infty$, $\infty.\infty$ and $\frac{\infty}{\infty}$ are undefined. We also specify that, for all $x \in \mathbb{R}$,

$$-\infty < x < \infty.$$

We write $\mathbb{R}^* = \{-\infty\} \cup \mathbb{R} \cup \{\infty\}$, which is known as the *extended* real numbers.

5. *Analysis.*

- sup and inf. If A is a bounded set of real numbers, we write $\sup(A)$ and $\inf(A)$ for the real numbers that are their least upper bounds and greatest lower bounds (respectively.) If A fails to be bounded above, we write $\sup(A) = \infty$ and if A fails to be bounded below we write $\inf(A) = -\infty$. Note that $\inf(A) = -\sup(-A)$ where $-A = \{-x; x \in A\}$. If $f : S \to \mathbb{R}$ is a mapping, we write $\sup_{x \in S} f(x) = \sup\{f(x); x \in S\}$. A very useful inequality is

$$\sup_{x \in S} |f(x) + g(x)| \leq \sup_{x \in S} |f(x)| + \sup_{x \in S} |g(x)|.$$

- Sequences and Limits. Let $(a_n) = (a_1, a_2, a_3, \ldots)$ be a sequence of real numbers. It *converges* to the real number a if given any $\epsilon > 0$ there exists a natural number N so that whenever $n > N$ we have $|a - a_n| < \epsilon$. We then write $a = \lim_{n \to \infty} a_n$.

  A sequence (a_n) which is *monotonic increasing* (i.e. $a_n \leq a_{n+1}$ for all $n \in \mathbb{N}$) and *bounded above* (i.e. there exists $K > 0$ so that $a_n \leq K$ for all $n \in \mathbb{N}$) converges to $\sup_{n \in \mathbb{N}} a_n$.

  A sequence (a_n) which is *monotonic decreasing* (i.e. $a_{n+1} \leq a_n$ for all $n \in \mathbb{N}$) and *bounded below* (i.e. there exists $L > 0$ so that $a_n \geq L$ for all $n \in \mathbb{N}$) converges to $\inf_{n \in \mathbb{N}} a_n$.

  A *subsequence* of a sequence (a_n) is itself a sequence of the form (a_{n_k}) where $n_{k_1} < n_{k_2}$ when $k_1 < k_2$.

- Series. If the sequence (s_n) converges to a limit s where $s_n = a_1 + a_2 + \cdots + a_n$ we write $s = \sum_{n=1}^{\infty} a_n$ and call it the *sum of the series*. If each $a_n \geq 0$ then the sequence (s_n) is either convergent to a limit or properly divergent to infinity. In the latter case we write $s = \infty$ and interpret this in the sense of extended real numbers.

- Continuity. A function $f : \mathbb{R} \to \mathbb{R}$ is *continuous* at $a \in \mathbb{R}$ if given any $\epsilon > 0$ there exists $\delta > 0$ so that $|x - a| < \delta \Rightarrow |f(x) - f(a)| < \epsilon$. Equivalently f is continuous at a if given any sequence (a_n) that converges to a, the sequence $(f(a_n))$ converges to $f(a)$.

  f is a *continuous function* if it is continuous at every $a \in \mathbb{R}$.

Chapter 1

Measure Spaces and Measure

1.1   What is Measure?

Measure theory is the abstract mathematical theory that underlies all models of measurement in the real world. This includes measurement of length, area and volume, mass but also chance/probability. Measure theory is on the one hand a branch of pure mathematics, but it also plays a key role in many applied areas such as physics and economics. In particular it provides a foundation for both the modern theory of integration and also the theory of probability. It is one of the milestones of modern analysis and is an invaluable tool for functional analysis.

To motivate the key definitions, suppose that we want to measure the lengths of several line segments. We represent these as closed intervals of the real number line $\mathbb{R}$ so a typical line segment is $[a, b]$ where $b > a$. We all agree that its length is $b - a$. We write this as

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

and interpret this as telling us that the measure m of length of the line segment $[a, b]$ is the number $b - a$. We might also agree that if $[a_1, b_1]$ and $[a_2, b_2]$ are two non-overlapping line segments and we want to measure their combined length then we want to apply m to the set-theoretic union $[a_1, b_1] \cup [a_2, b_2]$ and

$$m([a_1, b_1] \cup [a_2, b_2]) \quad = (b_2 - a_2) + (b_1 - a_1) = m([a_1, b_1]) + m([a_2, b_2]).$$

$$(1.1)$$

An isolated point c has zero length and so

$$m(\{c\}) = 0.$$

and if we consider the whole real line in its entirety then it has infinite length, i.e.

$$m(\mathbb{R}) = \infty.$$

We have learned so far that if we try to abstract the notion of a measure of length, then we should regard it as a mapping m defined on subsets of the real line and taking values in the extended non-negative real numbers $[0, \infty]$.

Question Does it make sense to consider m on **all** subsets of $\mathbb{R}$?

Example 1.1.1 (The Cantor Set.) Start with the interval $[0,1]$ and remove the middle third to create the set $C_1 = [0,1/3) \cup (2/3,1]$. Now remove the middle third of each remaining piece to get $C_2 = [0,1/9) \cup (2/9,1/3) \cup (2/3,7/9) \cup (8/9,1]$. Iterate this process so for $n > 2$, C_n is obtained from C_{n-1} by removing the middle third of each set within that union. The *Cantor set* is $C = \bigcap_{n=1}^{\infty} C_n$. It turns out that C is uncountable. Does $m(C)$ make sense?

We'll see later that $m(C)$ does make sense and is a finite number (can you guess what it is?). But it turns out that there are even wilder sets in $\mathbb{R}$ than C which have no length. These are quite difficult to construct (they require the axiom of choice) so we won't try to describe them here.

Conclusion. The set of all subsets of $\mathbb{R}$ is its power set $\mathcal{P}(\mathbb{R})$. We've just learned that the power set is too large to support a good theory of measure of length. So we need to find a smaller class of subsets that we can work with.

1.2 Sigma Fields

So far we have only discussed length but now we want to be more ambitious. Let S be an arbitrary set. We want to define mappings from subsets of S to $[0, \infty]$ which we will continue to denote by m. These will be called measures and they will share some of the properties that we've just been looking at for measures of length. Now on what type of subset of S can m be defined? The power set of S is $\mathcal{P}(S)$ and we have just argued that this could be too large for our purposes as it may contain sets that can't be measured.

Suppose that A and B are subsets of S that we can measure. Then we should surely be able to measure the complement A^c, the union $A \cup B$ and the whole set S. Note that we can then also measure $A \cap B = (A^c \cup B^c)^c$. This leads to a definition

Definition 1.2.1 Let S be a set. A *Boolean algebra* **B** is a set of subsets of S that has the following properties

B(i) $S \in \mathbf{B}$,

B(ii) If $A, B \in \mathbf{B}$ then $A \cup B \in \mathbf{B}$,

B(iii) If $A \in \mathbf{B}$ then $A^c \in \mathbf{B}$.

Note that **B** is a set, and each element of **B** is a subset of S. In other words, **B** is a subset of the power set $\mathcal{P}(S)$. In this course we will frequently work with sets, whose elements are sets. It's important to get used to working with these objects; don't forget the difference between $\{\{1\}, \{2\}\}$ and $\{1, 2\}$.

Boolean algebras are named after the British mathematician George Boole (1815-1864) who introduced them in his book *The Laws of Thought* published in 1854. They are well studied mathematical objects that are extremely useful in logic and digital electronics. It turns out that they are inadequate for our own purposes – we need a little more sophistication.

If we use induction on B(ii) then we can show that, if $A_1, A_2, \ldots A_n \in \mathbf{B}$ then $A_1 \cup A_2 \cup \cdots \cup A_n \in \mathbf{B}$. This is left for you to prove, in Problem **1.1**. But we need to be able to do analysis and this requires us to be able to handle infinite unions. The next definition gives us what we need:

Definition 1.2.2 Let S be a set. A σ-field Σ is a set of subsets of S that has the following properties

S(i) $S \in \Sigma$,

S(ii) If (A_n) is a sequence of sets with $A_n \in \Sigma$ for all $n \in \mathbb{N}$ then $\bigcup_{n=1}^{\infty} A_n \in \Sigma$,

S(iii) If $A \in \Sigma$ then $A^c \in \Sigma$.

The terms σ-field and σ-algebra have the same meaning (this is an unfortunate accident of history!). Often you will find that 'σ-field' is used in advanced texts and 'σ-algebra' is used within lecture courses. I prefer σ-field, you may use either.

Lastly, a piece of terminology.

Definition 1.2.3 Given a σ-field Σ on S, we say that a set $A \subset S$ is *measurable* if $A \in \Sigma$

Facts about σ-fields

- By S(i) and S(iii), $\emptyset = S^c \in \Sigma$.

- We have seen in S(ii) that infinite unions of sets in Σ are themselves in Σ. The same is true of finite unions. To see this let $A_1, \ldots, A_m \in \Sigma$ and define the sequence (A'_n) by
$$A'_n = \begin{cases} A_n \text{ if } 1 \le n \le m \\ \emptyset \text{ if } n > m \end{cases}$$ Now apply S(ii) to get the result. We can deduce from this that every σ-field is a Boolean algebra.

- Σ is also closed under infinite (or finite) intersections. To see this use de Morgan's law to write
$$\bigcap_{i=1}^{\infty} A_i = \left(\bigcup_{i=1}^{\infty} A_i^c \right)^c.$$

- Σ is closed under set theoretic differences $A - B$, since (by definition) $A - B = A \cap B^c$.

Examples of σ-fields

1. $\mathcal{P}(S)$ is a σ-field. If S is finite with n elements then $\mathcal{P}(S)$ has 2^n elements (Problem **1.2**).

2. For any set S, $\{\emptyset, S\}$ is a σ-field which is called the *trivial* σ-field. It is the basic tool for modelling logic circuitry where $\emptyset$ corresponds to "OFF" and S to "ON".

3. If S is any set and $A \subset S$ then $\{\emptyset, A, A^c, S\}$ is a σ-field .

4. The most important σ-field for studying the measure of length is the *Borel σ-field* of $\mathbb{R}$ which is denoted $\mathcal{B}(\mathbb{R})$. It is named after the French mathematican Emile Borel (1871-1956) who was one of the founders of measure theory. It is defined rather indirectly and we postpone this definition until after the next section.

A pair (S, Σ) where S is a set and Σ is a σ-field of subsets of S is called a *measurable space* There are typically many possible choices of Σ to attach to S. For example we can always take Σ to be trivial or the power set. The choice of Σ is determined by what we want to measure.

1.3 Measure

Definition 1.3.1 Let (S, Σ) be a measurable space. A measure on (S, Σ) is a mapping $m : \Sigma \to [0, \infty]$ which satisfies

M(i) $m(\emptyset) = 0$,

M(ii) (σ-*additivity*) If $(A_n)_{n \in \mathbb{N}}$ is a sequence of sets where each $A_n \in \Sigma$ and if these sets are mutually disjoint, i.e. $A_n \cap A_m = \emptyset$ if $m \neq n$, then

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

M(ii) may appear to be rather strong. Our earlier discussion about length led us to $m(A \cup B) = m(A) + m(B)$ and straightforward induction then extends this to *finite additivity*: $m(A_1 \cup A_2 \cup \cdots \cup A_n) = m(A_1) + m(A_2) + \cdots + m(A_n)$ but if we were to replace M(ii) by this weaker finite additivity condition, we would not have an adequate tool for use in analysis, and this would make our theory much less powerful.

The key point here is, of course, limits. Limits are how we rigorously justify that approximations work – consequently we need them, if we are to create a theory that will, ultimately, be useful to experimentalists and modellers.

Basic Properties of Measures

1. (Finite additivity) If $A_1, A_2, \ldots, A_r \in \Sigma$ and are mutually disjoint then

$$m(A_1 \cup A_2 \cup \cdots \cup A_r) = m(A_1) + m(A_2) + \cdots + m(A_r).$$

To see this define the sequence (A'_n) by $A'_n = \begin{cases} A_n \text{ if } 1 \leq n \leq r \\ \emptyset \text{ if } n > r \end{cases}$ Then

$$m \left(\bigcup_{i=1}^{r} A_i \right) = m \left(\bigcup_{i=1}^{\infty} A'_i \right) = \sum_{i=1}^{\infty} m(A'_i) = \sum_{i=1}^{r} m(A_i),$$

where we used M(ii) and then M(i) to get the last two expressions.

2. If $A, B \in \Sigma$ with $B \subseteq A$ and either $m(A) < \infty$, or $m(A) = \infty$ but $m(B) < \infty$, then

$$m(A - B) = m(A) - m(B). \tag{1.2}$$

To prove this write the disjoint union $A = (A - B) \cup B$ and then use the result of (1) (with $r = 2$).

3. (Monotonicity) If $A, B \in \Sigma$ with $B \subseteq A$ then $m(B) \leq m(A)$.

If $m(A) < \infty$ this follows from (1.2) using the fact that $m(A - B) \geq 0$. If $m(A) = \infty$, the result is immediate.

4. If $A, B \in \Sigma$ are arbitrary (i.e. not necessarily disjoint) then

$$m(A \cup B) + m(A \cap B) = m(A) + m(B). \tag{1.3}$$

The proof of this is Problem **??** part (a). Note that if $m(A \cap B) < \infty$ we have

$$m(A \cup B) = m(A) + m(B) - m(A \cap B).$$

10

Now some concepts and definitions. First, let us define the setting that we will work in for all of Chapters 1-**??**.

Definition 1.3.2 A triple (S, Σ, m) where S is a set, Σ is a σ-field on S, and $m : \Sigma \to [0, \infty)$ is a measure is called a *measure space.*

The extended real number $m(S)$ is called the *total mass* of m. The measure m is said to be *finite* if $m(S) < \infty$.

We will start to think about probability in Chapter 3. A finite measure is called a *probability measure* if $m(S) = 1$. When we have a probability measure, we use a slightly different notation.

We write Ω instead of S and call it a *sample space.*
We write $\mathcal{F}$ instead of Σ. Elements of $\mathcal{F}$ are called *events.*
We use $\mathbb{P}$ instead of m.
The triple $(\Omega, \mathcal{F}, \mathbb{P})$ is called a *probability space.*

Examples of Measures

1. **Counting Measure**

   Let S be a finite set and take $\Sigma = \mathcal{P}(S)$. For each $A \subseteq S$ define

   $$m(A) = \#(A) \text{ i.e. the number of elements in } A.$$

2. **Dirac Measure**

   This measure is named after the famous British physicist Paul Dirac (1902-84). Let (S, Σ) be an arbitrary measurable space and fix $x \in S$. The Dirac measure δ_x at x is defined by

   $$\delta_x(A) = \left\{ \begin{array}{c} 1 \text{ if } x \in A \\ 0 \text{ if } x \notin A \end{array} \right.$$

   Note that we can write counting measure in terms of Dirac measure, so if S is finite and $A \subseteq S$,

   $$\#(A) = \sum_{x \in S} \delta_x(A).$$

3. **Discrete Probability Measures**

   Let Ω be a countable set and take $\mathcal{F} = \mathcal{P}(\Omega)$. Let $\{p_\omega, \omega \in \Omega\}$ be a set of real numbers which satisfies the conditions

   $$p_\omega \geq 0 \text{ for all } \omega \in \Omega \text{ and } \sum_{\omega \in \Omega} p_\omega = 1.$$

   Now define the discrete probability measure P by

   $$P(A) = \sum_{\omega \in A} p_\omega = \sum_{\omega \in \Omega} p_\omega \delta_\omega(A),$$

   for each $A \in \mathcal{F}$.

   For example if $\#(\Omega) = n + 1$ and $0 < p < 1$ we can obtain the *binomial distribution* as a probability measure by taking $p_r = \binom{n}{r} p^r (1-p)^{n-r}$ for $r = 0, 1, \ldots, n$.

4. **Measures via Integration**

   Let (S, Σ, m) be an arbitrary measure space and $f : S \to [0, \infty)$ be a function that takes non-negative values. In Chapter 3, we will meet a powerful integration theory that allows us to cook up a new measure I_f from m and f (provided that f is suitably well-behaved, which will think about in Chapter 2) by the prescription:

   $$I_f(A) = \int_A f(x) m(dx),$$

   for all $A \in \Sigma$.

12

1.4   The Borel σ-field and Lebesgue Measure

In this section we take S to be the real number line $\mathbb{R}$. We want to describe a measure λ that captures the notion of length as we discussed at the beginning of this chapter. So we should have $\lambda((a,b)) = b - a$. The first question is - which σ-field should we use? We have already argued that the power set $\mathcal{P}(\mathbb{R})$ is too big. Our σ-field should contain open intervals, and also unions, intersections and complements of these.

Definition 1.4.1 The *Borel σ-field* of $\mathbb{R}$ to be denoted $\mathcal{B}(\mathbb{R})$ is the smallest σ-field that contains all open intervals (a,b) where $-\infty \le a < b \le \infty$. Sets in $\mathcal{B}(\mathbb{R})$ are called *Borel sets*.

Note that $\mathcal{B}(\mathbb{R})$ also contains isolated points $\{a\}$ where $a \in \mathbb{R}$. To see this first observe that $(a, \infty) \in \mathcal{B}(\mathbb{R})$ and also $(-\infty, a) \in \mathcal{B}(\mathbb{R})$. Now by S(iii),$(-\infty, a] = (a, \infty)^c \in \mathcal{B}(\mathbb{R})$ and $[a, \infty) = (-\infty, a)^c \in \mathcal{B}(\mathbb{R})$. Finally as σ-fields are closed under intersections, $\{a\} = [a, \infty) \cap (-\infty, a] \in \mathcal{B}(\mathbb{R})$. You can show that $\mathcal{B}(\mathbb{R})$ also contains all closed intervals (see Problem **??**).

We make two observations:

1. $\mathcal{B}(\mathbb{R})$ is defined quite indirectly and there is no "formula" that can be used to give the most general element in it. However it is very hard to find a subset of $\mathbb{R}$ that isn't in $\mathcal{B}(\mathbb{R})$ – we will give an example of one in Section 1.5.

2. $\mathcal{B}(S)$ makes sense on any set S for which there are subsets that can be called "open" in a sensible way. In particular this works for metric spaces. The most general type of S for which you can form $\mathcal{B}(S)$ is a *topological space*.

The measure that precisely captures the notion of length is called *Lebesgue measure* in honour of the French mathematician Henri Lebesgue (1875-1941), who founded the modern theory of integration. We will denote it by λ. First we need a definition.

Let $A \in \mathcal{B}(\mathbb{R})$ be arbitrary. A *covering* of A is a finite or countable collection of open intervals $\{(a_n, b_n), n \in \mathbb{N}\}$ so that

$$A \subseteq \bigcup_{n=1}^{\infty} (a_n, b_n).$$

Definition 1.4.2 Let $\mathcal{C}_A$ be the set of all coverings of the set $A \in \mathcal{B}(\mathbb{R})$. The Lebesgue measure λ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is defined by the formula:

$$\lambda(A) = \inf_{\mathcal{C}_A} \sum_{n=1}^{\infty} (b_n - a_n), \tag{1.4}$$

where the inf is taken over all possible coverings of A.

It would take a long time to prove that λ really is a measure, and it wouldn't help us understand λ any better if we did it, so we'll omit that from the course. For the proof, see the standard text books e.g. Cohn, Schilling or Tao.

Let's check that the definition (1.4) agrees with our intuitive ideas about length.

1. If $A = (a,b)$ then $\lambda((a,b)) = b - a$ as expected, since (a,b) is a covering of itself and any other cover will have greater length.

2. If $A = \{a\}$ then choose any $\epsilon > 0$. Then $(a - \epsilon/2, a + \epsilon/2)$ is a cover of a and so $\lambda(\{a\}) \leq (a + \epsilon/2) - (a - \epsilon/2) = \epsilon$. But ϵ is arbitrary and so we conclude that $\lambda(\{a\}) = 0$.

   From (1) and (2), and using M(ii), we deduce that for $a < b$,

   $$\lambda([a, b)) = \lambda(\{a\} \cup (a, b)) = \lambda(\{a\}) + \lambda((a, b)) = b - a.$$

3. If $A = [0, \infty)$ write $A = \bigcup_{n=1}^{\infty}[n - 1, n)$. Then by M(ii), $\lambda([0, \infty)) = \infty$. By a similar argument, $\lambda((-\infty, 0)) = \infty$ and so $\lambda(\mathbb{R}) = \lambda((-\infty, 0)) + \lambda([0, \infty)) = \infty$.

4. If $A \in \mathcal{B}(\mathbb{R})$, and for some $x \in \mathbb{R}$ we define $A_x = \{x + a \,;\, a \in A\}$, then $\lambda(A) = \lambda(A_x)$.

   In words, if we take a set A and translate it (by x), we do not change its measure. This is easily seen from (1.4), because any cover of A can be translated by x to be a cover of A_x.

In simple practical examples on Lebesgue measure, it is generally best not to try to use (1.4), but to just apply the properties (1) to (4) above:

e.g. to find $\lambda((-3, 10) - (-1, 4))$, use (1.2) to obtain

$$\lambda((-3, 10) - (-1, 4)) = \lambda((-3, 10)) - \lambda((-1, 4)))$$
$$= (10 - (-3)) - (4 - (-1)) = 13 - 5 = 8.$$

If I is a closed interval (or in fact any Borel set) in $\mathbb{R}$ we can similarly define $\mathcal{B}(I)$, the Borel σ-field of I, to be the smallest σ-field containing all open intervals in I. Then Lebesgue measure on $(I, \mathcal{B}(I))$ is obtained by restricting the sets A in (1.4) to be in $\mathcal{B}(I)$.

Sets of measure zero play an important role in measure theory. Here are some interesting examples of quite "large" sets that have Lebesgue measure zero

1. **Countable Subsets of $\mathbb{R}$ have Lebesgue Measure Zero**

   Let $A \subset \mathbb{R}$ be countable. Write $A = \{a_1, a_2, \ldots\} = \bigcup_{n=1}^{\infty}\{a_n\}$. Since A is an infinite union of point sets, it is in $\mathcal{B}(\mathbb{R})$. Then

   $$\lambda(A) = \lambda\left(\bigcup_{n=1}^{\infty}\{a_n\}\right) = \sum_{n=1}^{\infty} \lambda(\{a_n\}) = 0.$$

   It follows that

   $$\lambda(\mathbb{N}) = \lambda(\mathbb{Z}) = \lambda(\mathbb{Q}) = 0.$$

   The last of these is particularly intriguing as it tells us that the only contribution to length of sets of real numbers comes from the irrationals.

2. **The Cantor Set has Lebesgue Measure Zero**

   Recall the construction of the Cantor set $C = \bigcap_{n=1}^{\infty} C_n$ given earlier in this chapter. Recall also that the C_n are decreasing, that is $C_{n+1} \subseteq C_n$, and hence also $C \subseteq C_n$ for all n.

   Since C_n is a union of intervals, $C_n \in \mathcal{B}(\mathbb{R})$ for all $n \in \mathbb{N}$. Hence $C \in \mathcal{B}(\mathbb{R})$. We easily see that $\lambda(C_1) = 1 - \frac{1}{3}$ and $\lambda(C_2) = 1 - \frac{1}{3} - \frac{2}{9}$. Iterating, we deduce that $\lambda(C_n) = 1 - \sum_{r=1}^{n} \frac{2^{r-1}}{3^r}$ and since $\lambda(C) \leq \lambda(C_n)$ we thus have

   $$\lambda(C) \leq \lambda(C_n) = 1 - \sum_{r=1}^{n} \frac{2^{r-1}}{3^r}.$$

   Letting $n \to \infty$, and using that limits preserve weak inequalities, we obtain $\lambda(C) \leq 0$. But by definition of a measure we have $\lambda(C) \geq 0$. Hence $\lambda(C) = 0$.

1.5   An example of a non-measurable set $(\star)$

Note that this section has a $(\star)$, meaning that it is off-syllabus. It is included for interest.

We might wonder, why go to all the trouble of defining the Borel σ-field? In other words, why can't we measure (the 'size' of) every possible subset of $\mathbb{R}$? We will answer these questions by constructing a strange looking set $\mathscr{V} \subseteq \mathbb{R}$; we will then show that it is not possible to define the Lebesgue measure of $\mathscr{V}$.

As usual, let $\mathbb{Q}$ denote the rational numbers. For any $x \in \mathbb{R}$ we define

$$\mathbb{Q}_x = \{x + q \,;\, q \in \mathbb{Q}\}. \tag{1.5}$$

Note that different x values may give the same $\mathbb{Q}_x$. For example, an exercise for you is to prove that $\mathbb{Q}_{\sqrt{2}} = \mathbb{Q}_{1+\sqrt{2}}$. You can think of $\mathbb{Q}_x$ as the set $\mathbb{Q}$ translated by x.

It is easily seen that $\mathbb{Q}_x \cap [0,1]$ is non-empty; just pick some rational q that is slightly less than x and note that $x + (-q) \in \mathbb{Q}_x \cap [0,1]$. Now, for each set $\mathbb{Q}_x$, we pick precisely one element $r \in \mathbb{Q}_x \cap [0,1]$ (it does not matter which element we pick). We write this number r as $r(\mathbb{Q}_x)$. Define

$$\mathscr{V} = \{r(\mathbb{Q}_x) \,;\, x \in \mathbb{R}\},$$

which is a subset of $[0,1]$. For each $q \in \mathbb{Q}$ define

$$\mathscr{V}_q = \{q + m \,;\, m \in \mathscr{V}\}.$$

Clearly $\mathscr{V} = \mathscr{V}_0$, and $\mathscr{V}_q$ is precisely the set $\mathscr{V}$ translated by q. Now, let us record some facts about $\mathscr{V}_q$.

Lemma 1.5.1 *It holds that*

1. *If $q_1 \neq q_2$ then $\mathscr{V}_{q_1} \cap \mathscr{V}_{q_2} = \emptyset$.*

2. *$\mathbb{R} = \bigcup_{q \in \mathbb{Q}} \mathscr{V}_q$.*

3. *$[0,1] \subseteq \bigcup_{q \in \mathbb{Q} \cap [-1,1]} \mathscr{V}_q \subseteq [-1,2]$.*

Before we prove this lemma, let us use it to show that $\mathscr{V}$ cannot have a Lebesgue measure. We will do this by contradiction: assume that $\lambda(\mathscr{V})$ is defined.

Since $\mathscr{V}$ and $\mathscr{V}_q$ are translations of each other, they must have the same Lebesgue measure. We write $c = \lambda(\mathscr{V}) = \lambda(\mathscr{V}_q)$, which does not depend on q. Let us write $\mathbb{Q} \cap [-1,1] = \{q_1, q_2, \ldots, \}$, which we may do because $\mathbb{Q}$ is countable. By parts (1) and (3) of Lemma 1.5.1 and property M(ii) we have

$$\lambda \left(\bigcup_{q \in \mathbb{Q} \cap [-1,1]} \mathscr{V}_q \right) = \sum_{i=1}^{\infty} \lambda(\mathscr{V}_{q_i}) = \sum_{i=1}^{\infty} c.$$

Using the monotonicity property of measures (see Section 1.3) and part (3) of Lemma 1.5.1 we thus have

$$1 \leq \sum_{i=1}^{\infty} c \leq 3.$$

However, there is no value of c which can satisfy this equation! So it is not possible to make sense of the Lebesgue measure of $\mathscr{V}$.

The set $\mathscr{V}$ is known as a *Vitali set*. In higher dimensions even stranger things can happen with non-measurable sets; you might like to investigate the *Banach-Tarski paradox*.

PROOF: [Of Lemma 1.5.1.] We prove the three claims in turn.

(1) Let $q_1, q_2 \in \mathbb{Q}$ be unequal. Suppose that some $x \in \mathscr{V}_{q_1} \cap \mathscr{V}_{q_2}$ exists – and we now look for a contradiction. By definition of $\mathscr{V}_q$ we have

$$x = q_1 + r(\mathbb{Q}_{x_1}) = q_2 + r(\mathbb{Q}_{x_2}). \tag{1.6}$$

By definition of $\mathbb{Q}_x$ we may write $r(\mathbb{Q}_{x_1}) = x_1 + q_1'$ for some $q_1' \in \mathbb{Q}$, and similarly for x_2, so we obtain $x = q_1 + x_1 + q_1' = q_2 + x_2 + q_2'$ where $q, q' \in \mathbb{Q}$. Hence, setting $q = q_2 - q_1 + q_2' - q_1' \in \mathbb{Q}$, we have $x_1 + q = x_2$, which by (1.5) means that $\mathbb{Q}_{x_1} = \mathbb{Q}_{x_2}$. Thus $r(\mathbb{Q}_{x_1}) = r(\mathbb{Q}_{x_2})$, so going back to (1.6) we obtain that $q_1 = q_2$. But this contradicts our assumption that $q_1 \neq q_2$. Hence x does not exist and $\mathscr{V}_{q_1} \cap \mathscr{V}_{q_2} = \emptyset$.

(2) We will show $\supseteq$ and $\subseteq$. The first is easy: since $\mathscr{V}_q \subseteq \mathbb{R}$ it is immediate that $\mathbb{R} \supseteq \bigcup_{q \in \mathbb{Q}} \mathscr{V}_q$.

Now take some $x \in \mathbb{R}$. Since we may take $q = 0$ in (1.5) we have $x \in \mathbb{Q}_x$. By definition of $r(\mathbb{Q}_x)$ we have $r(\mathbb{Q}_x) = x + q'$ for some $q' \in \mathbb{Q}$. By definition of $\mathscr{V}$ we have $r(\mathbb{Q}_x) \in \mathscr{V}$ and since $x = r(\mathbb{Q}_x) - q'$ we have $x \in \mathscr{V}_{-q'}$. Hence $x \in \bigcup_{q \in \mathbb{Q}} \mathscr{V}_q$.

(3) Since $\mathscr{V} \subseteq [0, 1]$, we have $\mathscr{V}_q \cap [0, 1] = \emptyset$ whenever $q \notin [-1, 1]$. Hence, from part (2) and set algebra we have

$$\mathbb{R} \cap [0, 1] \;=\; \left(\bigcup_{q \in \mathbb{Q}} \mathscr{V}_q \right) \cap [0, 1] \;=\; \bigcup_{q \in \mathbb{Q}} \mathscr{V}_q \cap [0, 1] \;=\; \bigcup_{q \in \mathbb{Q} \cap [-1, 1]} \mathscr{V}_q \cap [0, 1] \;\subseteq\; \bigcup_{q \in \mathbb{Q} \cap [-1, 1]} \mathscr{V}_q.$$

This proves the first $\subseteq$ of (3). For the second simply note that $\mathscr{V} \subseteq [0, 1]$ so $\mathscr{V}_q \subseteq [-1, 2]$ whenever $q \in [-1, 1]$. ∎

Remark 1.5.2 *We used the axiom of choice to define the function $r(\cdot)$.*

1.6   Exercises

1.1 Give a careful proof by induction of the fact that if **B** is a Boolean algebra and $A_1, A_2, \ldots, A_n \in$ **B**, then $A_1 \cup A_2 \cup \cdots \cup A_n \in$ **B**.

1.2 Show that if S is a set containing n elements, then the power set $\mathcal{P}(S)$ contains 2^n elements.

Hint: How many subsets are there of size r, for a fixed $1 \leq r \leq n$? The binomial theorem may also be of some use.

1.3 OMITTED

Chapter 2

Measurable Functions

OMITTED

Chapter 3

Probability and Measure

OMITTED

Chapter 4

Product Measures and Fubini's Theorem (△)

OMITTED

Appendix A

Solutions to exercises

OMITTED