What's an Integral, Anyway?

The main goal of this class is to develop a more in-depth understanding of the integral. To support this understanding we will first develop a sophisticated approach to the idea of *measure*, which tells us how large a set is. (You can think of this as a useful generalization of "area" or "volume".)

You probably remember the Riemann integral, from either Analysis I or Calculus II.

Definition 0.1. Let $f : [a, b] \to \mathbb{R}$. We say f is *Riemann Integrable on* $[a, b]$ if there is a number $I \in \mathbb{R}$ so that, for any $\epsilon > 0$, there is a $\delta > 0$ such that, if S is a Riemann sum corresponding to a partition of width less than δ, then $|S - I| < \epsilon$. In this case we say that I is the *Riemann Integral of* f and write $I = \int_a^b f(x)\, dx$.

This definition is perfectly serviceable, but it has a few major issues. One is just that it's incredibly awkward to state, and difficult to use to prove things.

Second, there are a lot of functions that this definition doesn't quite apply to. You may remember so-called "improper" integrals from calculus II: these are integrals either over unbounded sets, like $\int_{-\infty}^{\infty} e^{-x^2}\, dx$, or integrals of unbounded functions like $\int_0^1 \frac{1}{\sqrt{x}}\, dx$. In either case the Riemann integral does not actually converge, and we need to use an awkward limiting process to even define, let alone compute, the integral.

Third, there are many sets we can't integrate over. A Riemann integral can integrate over sets like $[1, 3]$ but not over sets like the rational numbers or the Cantor set. As something of a corollary, we can consistently integrate continuous and nearly-continuous functions, but we can't integrate messy functions like

$$\chi_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}.$$

Fourth, and perhaps most importantly, the Riemann integral doesn't interact well with limits of sequences of functions. If f_n is a sequence of functions, we would like to prove a theorem like

$$\lim_{n \to infty} \int f_n(x)\, dx = \int \lim_{n \to \infty} f(x)\, dx.$$

However, this is unfortunately false. An easy example is to define

$$f_n(x) = \begin{cases} 4n^2 x & 0 \leq x < \frac{1}{2n} \\ 4n - 4n^2 x & \frac{1}{2n} x < \frac{1}{n} \\ 0 & \frac{1}{n} < x \leq 1 \end{cases}.$$

This looks complicated, but the graph is just an isosceles triangle with base $frac1n$ and height $2n$, and thus total area 1. So we know that for each n, $\int_0^1 f_n(x)\,dx = 1$.

However, for any fixed $x \in [0,1]$, it's easy to see that $\lim_{n\to\infty} f_n(x) = 0$. So if f is the pointwise limit of f_n, we have $f = 0$ and $\int_0^1 f(x)\,dx = 0$. Thus

$$\lim_{n\to\infty} \int_0^1 f_n(x)\,dx = 1 \neq 0 = \int_0^1 \lim_{n\to\infty} f(x)\,dx.$$

A new definition of the integral can't fix this example; the triangles under the f_n have area 1, and the pointwise limit is zero, and no amount of redefinition will fix that. But the Lebesgue integral we will define makes it easy to see exactly why this example breaks—and makes it easy to prove that in "most" cases, our desired theorem is actually true.

In the process of building our new and improved approach to the integral, we will develop ideas that help us understand probability better. Tu solve the problems with our integral we will find a way to define the measure or volume of a set. But if we have a collection of possible events, we can treat the probability of something hpapening as the measure of the set of events in which it happens.

If we are rolling a six-sided die, each side appears with probability or measure $\frac{1}{6}$, and we don't need any sophisticated tools to establish this. But if we are choosing a real number between zero and one, how do we describe the probability of getting a rational number? This will require a bit more work. And that work is where the content of this course starts.

1 Euclidean Space

In this section we will review the basic properties of real Euclidean space. Most of them should be familiar to you from Analysis I (Math 310), but I'll collect them here so you can remember the important bits, and also have a useful reference.

1.1 Set Theory

We write $\mathbb{R}$ for the set of real numbers, and $\mathbb{R}^n$ for the set $\{(x_1, \dots, x_n) : x_i \in \mathbb{R}\}$ of n-tuples of real numbers. This is a special case of the *Cartesian product* of sets. If $A_1, \dots, A_n$ are all sets, then

$$\prod_{i=1}^n A_i = A_1 \times \cdots \times A_n = \{(a_1, \dots, a_n) : a_i \in A_i\}.$$

We recall the set operations including union $A \cup B$, intersection $A \cap B$, set complement A^C, and set difference $A \setminus B$. We also have inclusion $A \subset B$ and containment $A \supset B$. If $A \cap B = \varnothing$ the empty set, then A and B are *disjoint*.

In this course we will often want to talk about unions and intersections of many sets. We often use I to stand for an *index set*, such that for each $i \in I$ we have a corresponding set A_i. Then we can write

$$\bigcup_{i \in I} A_i = \{a : \exists i \in I \text{ such that } a \in A_i\};$$
$$\bigcap_{i \in I} A_i = \{a : \forall i \in I, a \in A_i\}.$$

If the index set I is the natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$, instead we often write $\bigcup_{i=1}^\infty A_i$ or $\bigcap_{i=1}^\infty A_i$.

There are a couple of important principles about set intersection and union.

Fact 1.1 (De Morgan's Laws). *Let I be an index set. Then*

$$\left(\bigcup_{i \in I} A_i\right)^C = \bigcap_{i \in I} A_i^C$$
$$\left(\bigcap_{i \in I} A_i\right)^C = \bigcup_{i \in I} A_i^C.$$

Definition 1.2. Let $A_1, A_2, \ldots$ be a sequence of sets. We define

$$\limsup_{k\to\infty} A_k = \bigcap_{j=1}^{\infty}\left(\bigcup_{k=j}^{\infty} A_k\right),$$
$$\liminf_{k\to\infty} A_k = \bigcup_{j=1}^{\infty}\left(\bigcap_{k=j}^{\infty} A_k\right).$$

Proposition 1.3. *Let $A_1, A_2, \ldots,$ be a sequence of sets. Then*

$$\limsup_{k\to\infty} A_k = \{a : a \in A_k \textit{ for infinitely many } k \in \mathbb{N}\}.$$

Proof. If $a \in A_k$ for infinitely many $k \in \mathbb{N}$, then $a \in \bigcup_{k=j}^{\infty} A_k$ for any $j \in \mathbb{N}$. Thus $a \in \bigcap_{j=1}^{\infty}\bigcup_{k=j}^{\infty} A_k = \limsup_{k\to\infty} A_k$.

Conversely, if $a \in A_k$ for only finitely many k, then we can choose some j_0 larger than all of those k and then $a \notin \bigcup_{k=j_0}^{\infty} A_k$. Thus $a \notin \bigcap_{j=1}^{\infty}\bigcup_{k=j}^{\infty} A_k = \limsup_{k\to\infty} A_k$.

□

Exercise 1.4. *State and prove an analogue of Proposition 1.3 for* $\liminf_{k\to\infty} A_k$.

Now recall that we say a set A is *countable* if it is either finite or in bijection with the natural numbers. Informally, A is countable if you can list all of its elements in order. Recall that $\mathbb{N}$ and $\mathbb{Q}$ are countable, but $\mathbb{R}$ is not.

In fact, if $A_1, \ldots, A_n$ are all countable, then $\prod_{i=1}^{n} A_i$ is countable. And if I is a countable index set and A_i is countable for each $i \in I$, then $\bigcup_{i\in I} A_i$ is countable.

1.2 Topology and Metric in Euclidean Space

In order to understand sequences and sets, we need a sense of *topology*: we need to know which sets are "open", which tells us which points are close together. To really do analysis, we need something a bit stronger: we need a *metric*, which tells us how far apart two points are.

In Euclidean space we have something even better: a *norm*.

Definition 1.5. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. We define the norm of x to be $|x| = \sqrt{x_1^2 + \cdots + x_n^2}$.

The norm has the following important properties:

Fact 1.6. *Let $x, y \in \mathbb{R}^n$, and $r \in \mathbb{R}$. Then*

- *(Positive definite)* $|x| \geq 0$, *and* $|x| = 0$ *if and only if* $x = 0 = (0, \ldots, 0)$ *is the zero vector.*
- *(Scalars)* $|rx| = |r||x|$.
- *(Triangle Inequality)* $|x + y| \leq |x| + |y|$.

This norm gives us a *metric*:

Definition 1.7. Let $x, y \in \mathbb{R}^n$. We define the *distance* between x and y to be $d(x, y) = |x-y|$.

Exercise 1.8. *The distance* $d : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ *is a metric. That is, if* $x, y, z \in \mathbb{R}^n$*, then*

- *(Positive definite)* $d(x, y) \geq 0$, *and* $d(x, y) = 0$ *if and only if* $x = y$.
- *(Symmetry)* $d(x, y) = d(y, x)$.
- *(Triangle Inequality)* $d(x, z) \leq d(x, y) + d(y, z)$.

Recall that we can use this metric to define a convergent sequence:

Definition 1.9. Let $x \in \mathbb{R}^n$, and let $x_1, x_2, \ldots$ be a sequence of points in $\mathbb{R}^n$. We say that $\lim_{n\to\infty} x_n = x$ if, for every $\varepsilon > 0$, there is a $N \in \mathbb{N}$ such that if $n > N$ then $d(x_n, x) < \varepsilon$.

From the metric, we can also define open sets.

Definition 1.10. Let $x \in \mathbb{R}^n$ and $0 < r < \infty$. We define the *open ball with radius* r *and center* x to be

$$B_r(x) = B(x, r) = \{y \in \mathbb{R}^n : d(x, y) < r\}.$$

We define the *closed ball with radius* r *and center* x to be

$$\overline{B}_r(x) = \{y \in \mathbb{R}^n : d(x, y) \leq r\}.$$

If $x \in A \subset \mathbb{R}^n$, we say that x is an *interior point* of A if there is some $0 < r < \infty$ such that $B_r(x) \subset A$. We define the *interior* of A, denoted A° or $\mathring{A}$, to be the set of all interior points of A.

We say that A is *open* if every $x \in A$ is an interior point of A. A is open if and only if $A = A^\circ$.

Fact 1.11. *This definition of open sets defines a* topology *on* $\mathbb{R}^n$. *That is:*

- $\varnothing$ *and* $\mathbb{R}^n$ *are open.*

- *The union of* ***any*** *collection of open sets is open.*
- *The intersection of any* ***finite*** *collection of open sets is open.*

Exercise 1.12. *Find a collection of open sets whose intersection is not open.*

Exercise 1.13. *Prove that any open ball is an open set.*

We say that set is *closed* if its complement is open. Then

Fact 1.14.
- *$\varnothing$ and $\mathbb{R}^n$ are closed.*
- *The union of any* ***finite*** *collection of closed sets is closed.*
- *The intersection of* ***any*** *collection of closed sets is closed.*

Remark 1.15. Despite what you might think, "closed" and "open" are not opposites. Some sets are neither open nor closed. (Can you think of one?) Some sets are both open and closed. Topologists call those sets "clopen", because mathematicians have terrible senses of humor.

Definition 1.16. Let $x \in \mathbb{R}^n$ and $A \subset \mathbb{R}^n$. We say that x is a *limit point* of A if for every $r > 0$, there is a point $y \neq x$ such that $y \in A \cap B_r(x)$. That is, any open ball around x contains a point in A that is *not* x.

Fact 1.17. *x is a limit point of A if and only if $B_r(x)$ contains infinitely many points of A for any $r > 0$.*

A set A is closed if and only if it contains all its limit points.

We define the *closure* of A to be the set

$$\overline{A} = \{x : x \in A \text{ or } x \text{ is a limit point of } A\}.$$

We conclude this section with a note on notational convention, quoted directly from Jones:

Convention: Hereafter we shall strive for consistency in denoting open sets with the letter G and closed sets with the letter F. Obviously, any two letters would do, but tradition is on the side of G and F. In German the noun *Gebiet* means region, and in French the adjective *fermé* means closed.

1.3 Compact and Bounded

The definition of compactness is one of the most subtle and important in all of topology.

Definition 1.18. Let $A \subset \mathbb{R}^n$. Suppose that, whenever A is contained in a union of open sets, it is also contained in the union of some finite collection of those sets. Then we say A is *compact*.

We can write this in symbolic notation, which will be clearer in some ways and less clear than others. Suppose that whenever $A \subset \bigcup_{i \in I} G_i$ and each G_i is open, then there exist $i_1, \ldots, i_N \in I$ such that $A \subset \bigcup_{k=1}^N G_{i_k}$. Then A is compact.

We can get a grasp of this definition by seeing how it applies to a few easy cases:

Exercise 1.19. *Prove that:*

- $\varnothing$ *is compact.*
- *Any finite set is compact.*
- *If A, B are compact, then so is $A \cup B$.*
- $B(x, r)$ *is not compact.*
- $\mathbb{R}^n$ *is not compact.*

However, this definition is often unwieldy. Fortunately, in the case of Euclidean space specifically, there is a much easier criterion to check.

Definition 1.20. Let $A \subset \mathbb{R}^n$. We say that A is *bounded* if there is some $x \in \mathbb{R}^n$ and some $r > 0$ so that $A \subset B_r(x)$.

Theorem 1.21 (Heine-Borel)**.** *Let $A \subset \mathbb{R}^n$. Then A is compact if and only if it is closed and bounded.*

Proof. Please don't make me prove this. It's kinda tedious. □

There is one more way to think about compactness: we can relate it to sequence convergence.

Theorem 1.22 (Bolzano-Weierstrass)**.** *Every bounded infinite subset of $\mathbb{R}^n$ has a limit point.*

Proof. Let A be a bounded subset of $\mathbb{R}^n$ with no limit points. We shall prove it is finite.

Since A has no limit points, it contains all its limit points, and thus is closed. Since it is closed and bounded, it is compact.

If $x \in A$, then since x is not a limit point of A, there is some r_x so that $B_{r_x}(x) \cap A = \{x\}$. Now clearly $A \subset \bigcup_{x \in A} B_{r_x}(x)$; and since A is compact, that means there is some finite set I so that $A \subset \bigcup_{x \in I} B_{r_x}(x) = \bigcup_{x \in I}\{x\}$ which is a finite set. Thus A is finite. □

Exercise 1.23. *Every bounded sequence in $\mathbb{R}^n$ has a convergent subsequence.*

Definition 1.24. We say a set A is *sequentially compact* if every sequence in A has a convergent subsequence.

Fact 1.25. *Every compact set is sequentially compact.*

Every sequentially compact subset of $\mathbb{R}^n$ is compact.

Remark 1.26. In a metric space, compactness and sequential compactness are equivalent. Thus for our purposes they are interchangeable. But you can construct a sequentially compact topological space that is not compact.

We have one more family of results we wish to prove about compact sets. When we begin to talk about the volume or measure of a set, we will want to talk about compact subsets of open sets. So we will conclude by proving a couple results about how well we can fit compact sets inside open sets.

Lemma 1.27. *Let K be compact, and G_i be open such that $K \subset \bigcup_{i \in I} G_i$.*

Then there exists an $\varepsilon > 0$ such that: for every $x \in K$ there exists an $i \in I$ such that $B_\varepsilon(x) \subset G_i$.

(The number ε is known as the Lebesgue number *for the covering $\{G_i\}$.)*

Note very importantly that there's a uniformity condition here: we have one ε that works for every x, though each x may work for only one i.

Proof. For each $x \in K$, there is an i_x such that $x \in G_{i_x}$. Since G_{i_x} is open, we can pick an r_x so that $B_{2r_x}(x) \subset G_{i_x}$.

We know that $K \subset \bigcup_{x \in K} B_{r_x}(x)$, and since K is compact and open balls are open, we can pick a finite set $x_1, \dots, x_N \in K$ so that

$$K \subset \bigcup_{j=1}^{N} B_{r_{x_j}}(x_j).$$

Take ε to be the minimum of these r_{x_j}.

Now suppose $x \in K$. Then there is a j such that $x \in B_{r_{x_j}}(x_j)$, and thus $d(x, x_j) < r_{x_j} < 2r_{x_j} - \varepsilon$. Thus $B_\varepsilon(x) \subset B_{2r_{x_j}}(x_j) \subset G_{i_{x_j}}$.

□

Corollary 1.28. *Let K be compact and G be open, with $K \subset G$. Then there is an $\varepsilon > 0$ such that for all $x \in K$, we have $B_\varepsilon(x) \subset G$.*

Corollary 1.29. *Let K be compact and F be closed, with $K \cap F = \varnothing$. Then there exists an $\varepsilon > 0$ such that for every $x \in K$ and $y \in F$ we have $d(x, y) \geq \varepsilon$.*

1.4 Functions and Continuity

Definition 1.30. Let $x_0 \in A \subset \mathbb{R}^n$ and let $f : A \to \mathbb{R}^m$. We say that f is *continuous* at x_0 if for every $\varepsilon > 0$ there is a $\delta > 0$ such that if $x \in A$ and $d(x, x_0) < \delta$ then $d(f(x), f(x_0)) < \varepsilon$.

If f is continuous at x_0 for every $x_0 \in A$ then we say f is *continuous on A*.

It's very important to note that we're only worreid about $x \in A$ for the purposes of this definition. Thus whether a given function f is continuous at a given point x_0 or not can depend on what we give as the domain of f.

Exercise 1.31. *For any fixed $x \in \mathbb{R}^n$, the function $f(y) = d(x, y)$ is continuous on $\mathbb{R}^n$.*

Theorem 1.32. *Let $A \subset \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Then f is continuous if and only if for every open set $G \subset \mathbb{R}^m$, ther is an open set $H \subset \mathbb{R}^n$ such that $f^{-1}(G) = H \cap A$.*

This is basically what it means for a set to be open "in A": thus f is continuous if and only if $f^{-1}(G) \cap A$ is open in A.

In particular, if $f : \mathbb{R}^n \to \mathbb{R}^m$ then f is continuous if and only if $f^{-1}(G) \subset \mathbb{R}^n$ is open for any open $G \subset \mathbb{R}^m$.

Exercise 1.33. *State and prove an equivalent of theorem 1.32 for closed sets.*

A useful set of facts for working with this property:

Fact 1.34. *Let $f : A \to B$, and suppose $X_i \subseteq A, Y_i \subseteq B$. Then we have*

- $f^{-1}\left(\bigcup_{i \in I} Y_i\right) = \bigcup_{i \in I} f^{-1}(Y_i)$
- $f^{-1}\left(\bigcap_{i \in I} Y_i\right) = \bigcap_{i \in I} f^{-1}(Y_i)$

- $f\left(\bigcup_{i\in I} X_i\right) = \bigcup_{i\in I} f(X_i)$
- $f\left(\bigcap_{i\in I} X_i\right) \subseteq \bigcup_{i\in I}(X_i)$.

Exercise 1.35. *Find a function $f : A \to B$ and a family of sets $X_i \subseteq A$ such that $f\left(\bigcap_{i\in I} X_i\right) \neq \bigcup_{i\in I}(X_i)$.*

Exercise 1.36. *Let $f : A \to B$. Then*

- *$f(f^{-1}(Y)) = Y \cap f(A)$ for any $A \subseteq B$.*
- *$f^{-1}(f(X)) \supset X$ for any $X \subseteq A$.*

Proposition 1.37. *Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Let K be a compact subset of A. Then $f(K)$ is compact.*

However, this doesn't work the other way.

Exercise 1.38. *Give an example of a continuous function $f : \mathbb{R} \to \mathbb{R}$ and a compact function $K \subseteq \mathbb{R}$ such that $f^{-1}(K)$ is not compact.*

Corollary 1.39. *Let $K \subseteq \mathbb{R}^n$ be compact and $f : K \to \mathbb{R}$ be continuous. then f attains a maximum and a minimum value on K.*

In math we're often interested in functions which are bijections—which show that two sets are in at least some sense equivalent. To preserve topological equivalence we want those invertible functions to be topologically "nice" functions, which in this cae means continuous.

Definition 1.40. Suppose $A \subseteq \mathbb{R}^n$ and $B \subseteq R^m$, and $f : A \to B$ is a bijection. If f and f^{-1} are continuous functions, we say that f is a *homeomorphism* from A to B.

Exercise 1.41. *Suppose A, B are open and $f : A \to B$ is a homeomorphism. Prove that f gives a bijection between the open subsets of A and the open subsets of B.*

Finally, recall that sometimes we want our functions to be not just continuous, but *uniformly* continuous.

Definition 1.42. Let $A \subseteq \mathbb{R}^n$ and $f : A \to \mathbb{R}^m$. Then f is *uniformly continuous* on A if, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that if $x, y \in A$ with $d(x, y) < \delta$, then $d(f(x), f(y)) < \varepsilon$.

Every uniformly continuous function is continuous, but the converse isn't true. However:

Fact 1.43. *If K is compact and $f : K \to \mathbb{R}^m$ is continouous, then f is uniformly continuous.*

1.5 Distance from a Set

We have a clear definition of the distance between two points, but we often want to know the distance between a point and a set. This turns out to be a bit subtler but not too bad.

Definition 1.44. Let $A \subseteq \mathbb{R}^n$ be nonempty, and let $x \in \mathbb{R}^n$. The *distance* from x to A is the number

$$d(x, A) = \inf\{d(x, y) : y \in A\}.$$

(Since this is a set of real numbers bounded below by 0, it has an infimum.)

Proposition 1.45. *If $A \subseteq \mathbb{R}^n$ is nonempty and $x \in \mathbb{R}^n$, then there is a $x_0 \in \overline{A}$ such that $d(x, A) = d(x, x_0)$.*

Proof. □

Corollary 1.46. *If $A \subseteq \mathbb{R}^n$ is closed and nonempty, and $x \in \mathbb{R}^n$, then there is a $x_0 \in A$ such that $d(x, A) = d(x, x_0)$. That is, there is a closest point to x in A.*

Exercise 1.47. *Let $A \subseteq \mathbb{R}^n$. Then $x \in \overline{A}$ if and only if $d(x, A) = 0$.*

Proposition 1.48. *Assume $A \neq \varnothing$. Then $d(x, A)$ is a continuous function of x.*

Proof. For $x, x' \in \mathbb{R}^n$, for any $y \in A$ we have that

$$d(x, A) \leq d(x, y) \leq d(x, x') + d(x', y).$$

Thus $d(x, A) - d(x, x') \leq d(x', y)$ for any $y \in A$. Thus

$$d(x, A) - d(x, x') \leq d(x', A).d(x, A) - d(x', A) \qquad \leq d(x, x').$$

By symmetry, we also must have

$$d(x', A) - d(x, A) \leq d(x, x')$$
$$|d(x, A) - d(x', A)| \leq d(x, x').$$

□

Theorem 1.49 (Bump Functions)**.** *Assume F is closed, G open, and $F \subseteq G \subseteq \mathbb{R}^n$. Then there is a continuous function $f : \mathbb{R}^n \to \mathbb{R}$ such that*

$$\begin{aligned} 0 \leq f(x) &\leq 1 & \forall x &\in \mathbb{R}^n \\ f(x) &= 1 & \forall x &\in F \\ f(x) &= 0 & \forall x &\in G^c. \end{aligned}$$

That is, for any closed subset of an open set, we can write a continuous function that is 1 on the closed subset and 0 outside of the open set. Such a function is called an *Urysohn function* after Pavel Urysohn.

Proof. This is easy if either F or G^c is empty. Otherwise, we can define

$$f(x) = \frac{d(x, G^c)}{d(x, G^c) + d(x, F)}.$$

We know that f is continuous by proposition 1.48, since the denominator is never zero. □

Definition 1.50. Let $\varnothing \neq A \subseteq \mathbb{R}^n$. The *diameter* of A is

$$\operatorname{diam}(A) = \sup\{d(x, y) : x, y \in A\}.$$

We know that $0 \leq \operatorname{diam}(A) \leq \infty$.

Exercise 1.51. *Let $\varnothing \neq A \subseteq \mathbb{R}^n$. Then*

- $\operatorname{diam}(A) = 0$ *if and only if A contains exactly one point.*
- $\operatorname{diam}(A) < \infty$ *if and only if A is bounded.*
- $\operatorname{diam}(A) = \operatorname{diam}(\overline{A})$.
- $\operatorname{diam}(B_r(x)) = 2r$.

2 The Lebesgue Measure on $\mathbb{R}^n$

2.1 Defining the Lebesgue Measure

2.1.0 The empty set

Define $\lambda(\varnothing) = 0$.

2.1.1 Special rectangles

We can take a closed interval $[a, b] \subset \mathbb{R}$, and then we can take a rectangle or box as

$$I = [a_1, b_1] \times [a_2, b_2] \times \cdots \times [a_n, b_n] \subset \mathbb{R}^n = \{x \in \mathbb{R}^n : a_i \leq x_i \leq b_i\}.$$

Then we define

$$\lambda(I) = (b_1 - a_1) \ldots (b_n - a_n) = \prod_{i=1}^{n} (b_i - a_i).$$

Exercise 2.1. *Let $I \subset \mathbb{R}^n$ be a special rectangle. Prove that the following conditions are equivalent:*

1. $\lambda(I) = 0$
2. $I^\circ = \varnothing$
3. *I is contained in an affine subspace of $\mathbb{R}^n$ having dimension smaller than n. (An affine subspace is a set $\{x_0 + x : x \in E\}$ where E is a subspace and x_0 is a fixed point.)*

We will call these "rectangles" even if they are one-dimensional, or one-hundred-dimensional. This is mostly because the pictures we're going to draw are all two-dimensional, and that's mostly because those are easy to draw.

2.1.2 Special Polygons

A *special polygon* is a finite union of special rectangles, each of which has nonzero measure. All of the sides or edges must be perpendicular to a coordinate axis.

We can define the measure of a special polygon straightforwardly. If $I_1, \ldots, I_N$ are special rectangles with disjoint interiors, and $P = \bigcup_{k=1}^{N} I_k$ is a special polygon, then $\lambda(P) = \sum_{k=1}^{N} \lambda(I_k)$.

This is really the only reasonable definition: if we chop our special polygon into pieces, we want the measure of the pieces to add up to the measure of the polygon.

There are two compatibility conditions we should need to check, but they're tedious and boring and straightforward to check so we'll just state them here.

Fact 2.2.
- *Every special polygon can be expressed as the union of finitely many special rectangles with disjoint interiors.*
- *If P is a special polygon, and $P = \bigcup_{k=1}^n I_k = \bigcup_{\ell=1}^m J_\ell$ are two different ways of writing P as a union of special rectangles with disjoint interiors, then $\sum_{k=1}^n \lambda(I_k) = \sum_{\ell=1}^m J_\ell$.*

Proposition 2.3.
- *If $P_1 \subseteq P_2$ then $\lambda(P_1) \leq \lambda(P_2)$.*
- *If P_1 and P_2 have disjoint interiors, then $\lambda(P_1 \cup P_2) = \lambda(P_1) + \lambda(P_2)$.*

2.1.3 Open Sets

Here we want to define the measure of any open set. We will follow this by defining the measure of compact sets, and then extend to arbitrary sets by squeezing them between open and compact sets.

Definition 2.4. If $\varnothing \neq G \subseteq \mathbb{R}^n$ is an open set, we define

$$\lambda(G) = \sup \{\lambda(P) : P \subseteq G, P \text{ is a special polygon}\}.$$

We know the set we're taking the supremum over is non-empty, since G has some interior and thus contains some rectangle. If the set of polygon measures is bounded then $\lambda(G)$ is a real number; if the set is unbounded, then we write $\lambda(G) = \infty$.

Proposition 2.5. *If G is open and P is a special polygon with $P \subset G$, then there is another special polygon P' with $P \subset P' \subset G$ and $\lambda(P) < \lambda(P')$.*

Proof. Since P is closed and G is open, we have $G \cap P^C$ open. Let $x \in G \cap P^C$; then there is an r such that $x \in B_r(x) \subset G \cap P^C$, and we can let I be a closed special rectangle contained in $B_r(x)$. Then set $P' = P \cup I$; P' is a special rectangle, and clearly $P \subset P' \subset G$. □

Exercise 2.6. *If G is a bounded open set, prove that $\lambda(G) < \infty$.*

Proposition 2.7. *Let $G \subseteq \mathbb{R}^n$ be an open set. Then*

1. $0 \leq \lambda(G) \leq \infty$
2. $\lambda(G) = 0$ *if and only if* $G = \varnothing$.

3. $\lambda(\mathbb{R}^n) = \infty$.

4. *If* $G_1 \subset G_2$ *are open sets, then* $\lambda(G_1) \leq \lambda(G_2)$.

5. *If* G_k *is open for* $k \in \mathbb{N}$, *then*

$$\lambda\left(\bigcup_{k=1}^{\infty} G_k\right) \leq \sum_{k=1}^{\infty} \lambda(G_k).$$

6. *If* G_k *are* disjoint *open sets, then*

$$\lambda\left(\bigsqcup_{k=1}^{\infty} G_k\right) = \sum_{k=1}^{\infty} \lambda(G_k).$$

7. *If* P *is a special polygon, then* $\lambda(P) = \lambda(P^\circ)$.

Proof. 1. By definition.

2. If $G \neq \varnothing$ then G contains some nontrivial special polygon P. Then $\lambda(G) \geq \lambda(P) > 0$.

3. Exercise.

4. This is basically a property of suprema. If $G_1 \subset G_2$, then any special polygon contained in G_1 is also contained in G_2. Thus

$$\{P \subset G_1\} \subset \{P \subset G_2\}$$
$$\{\lambda(P) : P \subset G_1\} \subset \{\lambda(P) : P \subset G_2\}$$
$$\sup\{\lambda(P) : P \subset G_1\} \subset \sup\{\lambda(P) : P \subset G_2\}$$

since any upper bound for the larger set is also an upper bound for the smaller set.

5. This one is trickier than it looks. First note that the union is in fact an open set.

Let P be a special polygon such that $P \subset \bigcup_{k=1}^{\infty} G_k$. Since P is compact, we know there is a Lebesgue number $\varepsilon > 0$ such that, for every $x \in P$, there is a k with $B_\varepsilon(x) \subset G_k$. (See lemma 1.27).

We know P is a union of non-overlapping rectangles; we can always further subdivide those rectangles, and thus we can assume that $P = \bigcup_{j=1}^{n} I_j$ with each I_j a special rectangle of diameter less than 2ε. If we let x_j be the center of the rectangle I_j, then we have $I_j \subset B_\varepsilon(x_j) \subset G_k$ for some k.

Now we can divide our rectangles up according to their open sets. For each k, define P_k to be the union of all I_j such that $I_j \subset G_k$ and $I_J \not\subset G_i$ for $i < k$. (This second condition is just to make sure we don't double-count any rectangle). But every rectangle is contained in at least one of these open sets, so $P = \bigcup_{k=1}^{\infty} P_k$.

Most of these P_k are empty, since there are only finitely many special rectangles running around. But each non-empty P_k is a special polygon, with $P_k \subset G_k$. And we know the P_k have disjoint interiors. Thus we know that

$$\lambda(P) = \sum_{k=1}^{N} \lambda(P_k) \leq \sum_{k=1}^{N} \lambda(G_k) \leq \sum_{k=1}^{\infty} \lambda(G_k).$$

But we've shown that for any special polygon $P \subset \bigcup G_k$, we have $\lambda(P) \leq \sum \lambda(G_k)$. Thus we have

$$\lambda\left(\bigcup_{k=1}^{\infty} G_k\right) = \sup\left\{\lambda(P) : P \subset \bigcup G_k\right\} \leq \sum_{k=1}^{\infty} \lambda(G_k).$$

6. We already know that $\lambda \bigsqcup G_k \leq \sum \lambda(G_k)$ by property 5. So we just need to show the reverse, that $\sum \lambda(G_k) \leq \lambda \bigsqcup G_k$.

   For each k, let P_k be a special polygon with $P_k \subset G_k$. Then the P_k are disjoint, and for any $n \in \mathbb{N}$ we have

$$\sum_{k=1}^{n} \lambda(P_k) = \lambda\left(\bigcup_{k=1}^{n} P_k\right) \leq \lambda\left(\bigsqcup_{k=1}^{\infty} G_k\right)$$

   Since this is true for any special polygons $P_k \subset G_k$, we know that the union is an upper bound for any polygons; thus it's an upper bound for the supremum, and we get

$$\sum_{k=1}^{n} \lambda(G_k) \leq \lambda\left(\bigsqcup_{k=1}^{\infty} G_k\right).$$

   Then this statement is true for any finite sum on the left, so it must still be true in the limit; so we have

$$\sum_{k=1}^{\infty} \lambda(G_k) \leq \lambda\left(\bigsqcup_{k=1}^{\infty} G_k\right).$$

   And this is what we needed to show.

7. It's easy to see that $\lambda(P^\circ) \leq \lambda(P)$ (though not completely trivial). If Q is a special polygon with $Q \subset P^\circ$, then $Q \subset P$ and thus $\lambda(Q) \leq \lambda(P)$. This is true for any Q, and thus we have

$$\lambda(P^\circ) = \sup_{Q \subset P^\circ} \lambda(Q) \leq \lambda(P).$$

Now we need to prove the other direction. We'll start by proving it for special rectangles. If I is a special rectangle, then for any $\varepsilon > 0$ we can find a rectangle $I' \subset I^\circ$ such that $\lambda(I') > \lambda(I) - \varepsilon$ (by simply shrinking each dimension by $\sqrt[n]{\varepsilon}/2$ or something like that).

This tells us that $\lambda(I^\circ) > \lambda(I) - \varepsilon$. But this is true for any $\varepsilon > 0$, so we have $\lambda(I^\circ) \geq \lambda(I)$.

Now if P is a special polygon written as a union of non-overlapping special rectangles I_k, then $\bigcup_{k=1}^n I_k^\circ$ is a disjoint union contained in P°. Thus

$$\lambda(P) = \sum_{k=1}^n \lambda(I_k) \leq \sum_{k=1}^n \lambda(I_k^\circ) \leq \lambda(P)^\circ.$$

□

Exercise 2.8. *Prove that every nonempty open subset of $\mathbb{R}$ can be written as a countable disjoint union of open intervals $G = \bigcup k(a_k, b_k)$, and this expression is unique.*

Then conclude that $\lambda(G) = \sum_k (b_k - a_k)$.

Remark 2.9. In $\mathbb{R}$ we can use this as our construction, but it doesn't really generalize to $\mathbb{R}^n$ easily. You *can* make that work, but it's even more painful.

2.1.4 Compact Sets

If $K \subset \mathbb{R}^n$ is compact, then define

$$\lambda(K) = \inf\{\lambda(G) : K \subset G, G \text{ open}\}.$$

There's something we immediately have to check: if K is a compact special polygon, does this new definition match the old one?

This is a little hard to talk about, so we'll introduce some very temporary notation. We'll use α for the definition of measure we gave in 2.1.2 that applies specifically to special polygons. And we'll use β for the definition that applies to any compact set. We'll prove they're both the same, and then we can go back to calling both of them λ instead.

It's any to see that $\alpha(P) \leq \beta(P)$ for any special polygon P. Whenever $P \subset G$, then $\alpha(P) \leq \lambda(G)$. Therefore, $\alpha(P) \leq \inf\{\lambda(G)\} = \beta(P)$.

Conversely, we want to show that $\beta(P) \leq \alpha(P)$. Suppose $P = \bigcup_{k=1}^n I_k$ is a union of non-overlapping rectangles. For any $\varepsilon > 0$ we can pick special rectangles $I_k' \subset I_k^\circ$ such that $\lambda(I_k') < \lambda(I_k) + \varepsilon/n$.

Then if we set $G = \bigcup_{k=1}^n I_k'^\circ$ we have $P \subset G$, and thus

$$\begin{aligned}\beta(P) \leq \lambda(G) &\leq \sum_{k=1}^n \lambda(I_k'^\circ) \\ &< \sum_{k=1}^n \lambda(I_k) + \varepsilon = \alpha(P) + \varepsilon.\end{aligned}$$

Since this is true for any $\varepsilon > 0$, we have $\beta(P) \leq \alpha(P)$.

We want to prove several properties of the measure of these compact sets. But mostly we can leverage the results we already proved about open sets.

Proposition 2.10. *1.* $0 \leq \lambda(K) < \infty$

2. If $K_1 \subset K_2$ then $\lambda(K_1) \leq \lambda(K_2)$.

Proof. Exercise □

3. $\lambda(K_1 \cup K_2) \leq \lambda(K_1) + \lambda(K_2)$.

Proof. If $K_1 \subset G_1$ and $K_2 \subset G_2$ then $K_1 \cup K_2 \subset G_1 \cup G_2$, and thus

$$\lambda(K_1 \cup K_2) \leq \lambda(G_1 \cup G_2) \leq \lambda(G_1) + \lambda(G_2).$$

Thus

$$\lambda(K_1 \cup K_2) \leq \inf \lambda(G_1) + \lambda(G_2) = \lambda(K_1) + \lambda(K_2).$$

□

4. If K_1 and K_2 are disjoint, then $\lambda(K_1 \cup K_2) = \lambda(K_1) + \lambda(K_2)$.

Proof. Wince K_1 and K_2 are compact, there is a $\varepsilon > 0$ such that for every $x \in K_1, y \in K_2$, then $d(x, y) \geq \varepsilon$. (This is the Lebesgue number again, with the open sets being K_1^C and K_2^C.) Then if we let G be an open set containing $K_1 \cup K_2$, we can write

$$\begin{aligned}G_1 &= G \cap \bigcup_{x \in K_1} B_{\varepsilon/2}(x) \\ G_2 &= G \cap \bigcup_{x \in K_2} B_{\varepsilon/2}(x).\end{aligned}$$

Then we have $K_i \subset G_i$ and $G_1 \cap G_2 = \varnothing$. So we have

$$\lambda(K_1) + \lambda(K_1) \leq \lambda(G_1) + \lambda(G_2) = \lambda(G_1 \cup G_2) \leq \lambda(G).$$

Since this holds for any $G \supset K_1 \cup K_2$, we have $\lambda(K_1) + \lambda(K_2) \leq \lambda(K_1 \cup K_2)$. Since the opposte inequality follows from part (3), that proves equality.

□

Remark 2.11. We didn't try to prove any results about infinite unions of compact sets. Why not?

Here we should mention one very important example: the Cantor set. (It is sometimes known as the ternary Cantor set to distinguish from some generalizations.)

Definition 2.12. We first define a family of open intervals contained in $[0, 1]$. We define $G_1 = (\frac{1}{3}, \frac{2}{3})$; then $[0, 1] \setminus G_1$ is two closed intervals of length one third. We remove the middle third of each of these: we define $G_2 = (\frac{1}{9}, \frac{2}{9}) \cup (\frac{7}{9}, \frac{8}{9})$. Then $[0, 1] \setminus (G_1 \cup G_2) =$ is four closed intervals of length $1/9$. We can iterate this construction to get an infinite sequence of disjoint open sets $G_1, G_2, \dots$.

We define the (ternary) *Cantor set* to be the set

$$C = [0, 1] \setminus \bigcup_{k=1}^{\infty} G_k.$$

Fact 2.13. *The Cantor set C is uncountable.*

Exercise 2.14. *Prove that C is compact. Then prove that $\lambda(C) = 0$.*

2.1.5 Inner and Outer Measure

We would like to extend our definition of measure to cover any set. We don't quite have the ability to do that yet, but we can define two quanties that do apply to any set.

Definition 2.15. Let $A \subseteq \mathbb{R}^n$. Then we define

- The *outer measure* of A

$$\lambda^*(A) = \inf\{\lambda(G) : A \subset G \text{ open}\}$$

- The *inner measure* of A

$$\lambda_*(A) = \sup\{\lambda(K) : A \supset K \text{ compact}\}.$$

Notice that these are basically concepts we've seen before; outer measure is how we defined the measure of a compact set, and inner measure is basically how we defined the measure of an open set. So this entire set of definitions has a sort of push-pull quality.

Proposition 2.16. *1.* $\lambda_*(A) \leq \lambda^*(A)$.

Proof. If $K \subset A \subset G$, then $K \subset G$, and thus $\lambda(K) \leq \lambda(G)$. □

2. If $A \subseteq B$ *then* $\lambda^*(A) \leq \lambda^*(B)$ *and* $\lambda_*(A) \leq \lambda_*(B)$.

3. $\lambda^*\left(\bigcup_{k=1}^\infty A_k\right) \leq \sum_{k=1}^\infty \lambda^*(A_k)$.

Proof. We basically want to cover each A_k with an open set. If $\varepsilon > 0$, then for each k we can find a $G_k \supseteq A_k$ such that $\lambda(G_k) < \lambda^*(A_k) + \varepsilon 2^{-k}$. Then we have

$$\lambda^*\left(\bigcup_{k=1}^\infty A_k\right) \leq \lambda\left(\bigcup_{k=1}^\infty G_k\right) \leq \sum_{k=1}^\infty \lambda(G_k)$$
$$< \sum_{k=1}^\infty (\lambda^*(A_k) + \varepsilon 2^{-k}) = \sum_{k=1}^\infty \lambda^*(A_k) + \varepsilon.$$

Since this holds for any $\varepsilon > 0$, we're done.

□

4. If the A_k *are disjoint, then*

$$\lambda_*\left(\bigcup_{k=1}^\infty A_k\right) \geq \sum_{k=1}^\infty \lambda_*(A_k).$$

Proof. Exercise. □

5. If A *is open or compact, then* $\lambda^*(A) = \lambda_*(A) = \lambda(A)$.

Proof. If A is open, then clearly $\lambda^*(A) = \lambda(A)$. If P is any special polygon with $P \subset A$, then P is compact, so $\lambda(P) \leq \lambda_*(A)$; and thus $\lambda(A) \leq \lambda_*(A)$.

But then $\lambda(A) \leq \lambda_*(A) \leq \lambda^*(A) = \lambda(A)$, so all the numbers are equal.

Now suppose A is compact. Then $\lambda_*(A) = \lambda(A)$ clearly, and $\lambda(A) = \lambda^*(A)$ because that's the definition of $\lambda(A)$. □

2.1.6 Sets with Finite Outer Measure

Recall we want to assign a measure to every possible set. In the last subsubsection 2.1.5 we defined two "measure-like" numbers that apply to any set. But which one should we use?

It turns out that very strange things can happen in general, which we will see later. But all of those strangenesses are avoided if our two measures are in fact the same.

Definition 2.17. Let $A \subseteq \mathbb{R}^n$ be a set with finite outer measure. We say that A is *measurable* and belongs to $\mathcal{L}_0$ if $\lambda^*(A) = \lambda_*(A)$, and in that case we define the *measure* of A to be $\lambda(A) = \lambda^*(A) = \lambda_*(A)$.

Proposition 2.18. *The family $\mathcal{L}_0$ contains all open sets with finite measure and all compact sets. Our new definition of measure belongs to every previous definition of measure we've given.*

Lemma 2.19. *If $A, B \in \mathcal{L}_0$ are disjoint, then $A \cup B \in \mathcal{L}_0$ and $\lambda(A \cup B) = \lambda(A) + \lambda(B)$.*

Proof.

$$\begin{aligned}\lambda^*(A \cup B) &\leq \lambda^*(A) + \lambda^*(B) = \lambda(A) + \lambda(B) \\ &= \lambda_*(A) + \lambda_*(B) \leq \lambda_*(A \cup B) \\ &\leq \lambda^*(A \cup B).\end{aligned}$$

□

We want to be able to tell whether a set is measurable in an easy-to-compute way. The main tool for this is the following theorem on approximation, which says that if we can approximate our set with open sets and compact sets that are "close together" then our set is in $\mathcal{L}_0$.

Theorem 2.20 (Approximation of Measure)**.** *Let $A \subseteq \mathbb{R}^n$ such that $\lambda^*(A) < \infty$. Then $A \in \mathcal{L}_0$ if and only if: for every $\varepsilon > 0$ there is a compact set K and an open G such that $K \subseteq A \subseteq G$ and $\lambda(G \setminus K) < \varepsilon$.*

Proof. If $A \in \L_0$, then that means that the inner measure and outer measure of A are the same. But we can always approximate the outer measure well with an open set, and the inner measure with a compact set. So for any $\varepsilon > 0$ we can find $G \subseteq A, K \subseteq A$ such that

$$\begin{aligned}\lambda(G) &< \lambda^*(A) + \varepsilon/2 = \lambda(A) + \varepsilon/2 \\ \lambda(K) &> \lambda^*(A) - \varepsilon/2 = \lambda(A) - \varepsilon/2.\end{aligned}$$

Since K and $G \setminus K$ are disjoint, we have $\lambda(G) = \lambda(K) + \lambda(G \setminus K)$. Rearranging this gives

$$\begin{aligned}\lambda(G \setminus K) &= \lambda(G) - \lambda(K) \\ &< \lambda(A) + \varepsilon/2 - \lambda(A) + \varepsilon/2 = \varepsilon.\end{aligned}$$

Conversely, suppose that for any $\varepsilon > 0$ there exist $K \subseteq A \subseteq G$ with $\lambda(G \setminus K) < \varepsilon$. Fix an epsilon, and then choose such sets G and K. We have that

$$\begin{aligned}\lambda^*(A) \leq \lambda(G) &= \lambda(K) + \lambda(G \setminus K) \\ &< \lambda(K) + \varepsilon \leq \lambda_*(A) + \varepsilon.\end{aligned}$$

Since this holds for any $\varepsilon > 0$, we conclude that $\lambda^*(A) \leq \lambda_*(A) \leq \lambda^*(A)$. Thus the outer and inner measures are equal, and $A \in \mathcal{L}_0$ by definition.

□

We want to figure out how we can combine $\mathcal{L}_0$ sets to get other $\mathcal{L}_0$ sets. We start by looking at our binary operations, and then we'll figure out how to work with countably many sets at once.

Proposition 2.21. *If $A, B \in \mathcal{L}_0$ then $A \cup B, A \cap B, A \setminus B \in \mathcal{L}_0$ as well.*

Proof. We'll start with the set difference, using the theorem on approximation.

Fix $\varepsilon > 0$, and then we can write $K_1 \subseteq A \subseteq G_1, K_2 \subseteq B \subseteq G_2$ with $\lambda(G_i \setminus K_i) < \varepsilon/2$. Then set $K = K_1 \setminus G_2$ and $G = G_1 \setminus K_2$.

G is clearly open, and K is closed and thus compact. Further, we have $K \subseteq A \setminus B \subseteq G$, and $G \setminus K \subseteq (G_1 \setminus K_1) \cup (G_2 \setminus K_2)$. Thus $\lambda(G \setminus K) < \varepsilon$, and thus $A \setminus B \in \mathcal{L}_0$.

Given this fact, we can prove the other two claims with minimal work. $A \cap B = A \setminus (A \setminus B)$ is a difference of differences of $\mathcal{L}_0$ sets, and thus is in $\mathcal{L}_0$. And $A \cup B = (A \setminus B) \cup B$ is a disjoint union of $\mathcal{L}_0$ sets, and thus is $\mathcal{L}_0$ by lemma 2.19. □

Theorem 2.22 (Countable additivity)**.** *Let $A_k \in \mathcal{L}_0$, and set $A = \bigcup_{k=1}^{\infty} A_k$. Assume $\lambda^*(A) < \infty$. Then $A \in \mathcal{L}_0$, and*

$$\lambda(A) \leq \sum_{k=1}^{\infty} \lambda(A_k).$$

Further, if the A_k are disjoint, then

$$\lambda(A) = \sum_{k=1}^{\infty} \lambda(A_k).$$

Proof. If the A_k are disjoint, this is easy. We know that

$$\begin{aligned} \lambda^*(A) &\leq \sum_{k=1}^{\infty} \lambda^*(A_k) \\ &= \sum_{k=1}^{\infty} \lambda_*(A_k) \leq \lambda_*(A) \leq \lambda^*(A). \end{aligned}$$

If the A_k are not disjoint, we can't do anything this simple. The first inequality holds, but we don't actually have an inequality on the inner measure. But if we can reduce this to a question about a disjoint union, then we can use the previous result and things become much simpler.

Define a new family of sets as follows. We take $B_1 = A_1$, and then for each $k > 1$ we define

$$B_k = A_k \setminus \left(\bigcup_{i=1}^{k-1} A_k \right).$$

Then each $B_k \in \mathcal{L}_0$, and clearly the B_k are disjoint. Each B_k is a subset of the corresponding A_k, and $\bigcup_{k=1}^{\infty} B_k = \bigcup_{k=1}^{\infty} A_k = A$. Then we can use our result on disjoint unions to see that $A \in \mathcal{L}_0$, and

$$\lambda(A) = \sum_{k=1}^{\infty} \lambda(B_k) \leq \sum_{k=1}^{\infty} \lambda(A_k).$$

□

2.1.7 Measurable Sets

Definition 2.23. Let $A \subset \mathbb{R}^n$. We say that A is *(Lebesgue) measurable* if, for every $M \in \mathcal{L}_0$, then $A \cap M \in \mathcal{L}_0$. If A is measurable, we define the *(Lebesgue) measure* of A to be

$$\lambda(A) = \sup\{\lambda(A \cap M) : M \in \mathcal{L}_0\}.$$

We denote the set of all measurable subsets of $\mathbb{R}^n$ with the symbol $\mathcal{L}$.

We now have another (final!) definition of measure; so we need to make sure it's the same as our previous definitions.

Proposition 2.24. *Let $A \subseteq \mathbb{R}^n$ with $\lambda^*(A) < \infty$. Then $A \in \mathcal{L}_0$ if and only if $A \in \mathcal{L}$. And if $A \in \mathcal{L}$, then our two definitions of measure coincide.*

Proof. If $A \in \mathcal{L}_0$, then $A \cap M \in \mathcal{L}_0$ for any $M \in \mathcal{L}_0$, and thus $A \in \mathcal{L}$.

Conversely, suppose $A \in \mathcal{L}$. We know that $B_k(0) \in \mathcal{L}_0$ since it's open, so we know that $A_k = A \cap B_k(0) \in \mathcal{L}_0$. But $A = \bigcup_{k=1}^{\infty} A_k$, and theorem 2.22 tells us that $A \in \mathcal{L}_0$.

Now we need to prove that the measure formulas coincide; for the rest of this proof we'll use λ' for our final definition of measure given in Definition 2.23.

Suppose $A \in \mathcal{L}_0 \subset \mathcal{L}$. Then since for any $M \in \mathcal{L}_0$, we know that $A \cap M \subseteq A$, and so $\lambda(A \cap M) \leq \lambda(A)$, and thus $\lambda'(A) \leq \lambda(A)$. But conversely, $A \in \mathcal{L}_0$, so we must have $\lambda(A) \leq \lambda(A \cap A) \leq \lambda'(A)$. Thus $\lambda' = \lambda$. □

2.2 Basic Properties of the Lebesgue Measure

Now that we have finally given a complete definition of Lebesgue measure, we want to collect all the properties that apply to it. Many of these are properties we've seen already at various earlier stages of the construction, but we need to see they still hold at this completed stage. Some other properties are basically new.

Proposition 2.25. *1. $A \in \mathcal{L}$ if and only if $A^C \in \mathcal{L}$.*

2. If $A_k \in \mathcal{L}$, then $\bigcup_{k=1}^\infty A_k \in \mathcal{L}$ and $\bigcap_{k=1}^\infty A_k \in \mathcal{L}$.

3. If $A, B \in \mathcal{L}$ then $A \setminus B \in \mathcal{L}$.

Proof. 1. For any $M \in \mathcal{L}_0$, we know that $A^c \cap M = M \setminus A = M \setminus (A \cap M)$. This is a difference of $\mathcal{L}_0$ sets, and thus is in $\mathcal{L}_0$. Therefore $A^C \in \mathcal{L}$.

2. If $A_k \in \mathcal{L}$ and $A = \bigcup_{k=1}^\infty A_k$, then for any M we have that $A \cup M = \bigcup_{k=1}^\infty A_k \cap M$. Since $\lambda^*(A \cap M) \leq \lambda(M)$ is finite, theorem 2.22 tells us that $A \cap M \in \mathcal{L}_0$. Thus $A \in \mathcal{L}$.

   The result on intersections follows from De Morgan's Laws: we know that
   $$\bigcap_{k=1}^\infty A_k = \left(\bigcup_{k=1}^\infty A_k^C\right)^C.$$
   Since complements and countable unions preserve measurability, this is a measurable set.

3. $A \setminus B = A \cap B^C$ is measurable.

□

Proposition 2.26. *[Countable Additivity] If A_k are measurable, then*
$$\lambda\left(\bigcup_{k=1}^\infty A_k\right) \leq \sum_{k=1}^\infty \lambda(A_k).$$
If the union is disjoint, then
$$\lambda\left(\bigcup_{k=1}^\infty A_k\right) = \sum_{k=1}^\infty \lambda(A_k).$$

Proof. Let $A = \bigcup_{k=1}^\infty A_k$. Then by theorem 2.22 we know that

$$\lambda(A \cup M) \leq \sum_{k=1}^\infty \lambda(A_k \cap M) \leq \sum_{k=1}^\infty \lambda(A_k).$$

Thus $\sum_{k=1}^\infty \lambda(A_k)$ is an upper bound for $\lambda(A \cap M)$, and so we have $\lambda(A) \leq \sum_{k=1}^\infty \lambda(A_k)$.

Now suppose the sets are disjoint; we just need to prove the opposite inequality. For any $n \in \mathbb{N}$ we can choose sets $M_1, \dots, M_n \in \mathcal{L}_0$, and define $M = \bigcup_{k=1}^n M_k$. Then

$$\begin{aligned} \lambda(A) \geq \lambda(A \cap M) &= \sum_{k=1}^\infty \lambda(A_k \cap M) \\ &\geq \sum_{k=1}^n \lambda(A_k \cap M) \geq \sum_{k=1}^n \lambda(A_k \cap M_k). \end{aligned}$$

Since $\lambda(A_k \cap M_k) \leq \lambda(A_k)$, we conclude that $\lambda(A) \geq \sum_{k=1}^n \lambda(A_k)$. Since this is true for any $n \in \mathbb{N}$, we must have $\lambda(A) \geq \sum_{k=1}^\infty \lambda(A_k)$, as desired.

□

Proposition 2.27. *Suppose $A_1, A_2, \dots$ are measurable sets. Then:*

1. *If $A_1 \subseteq A_2 \subseteq \dots$, then*
$$\lambda\left(\bigcup_{k=1}^\infty A_k\right) = \lim_{k\to\infty} \lambda(A_k).$$
2. *If $A_1 \supseteq A_2 \supseteq \dots$, and further if $\lambda(A_1) < \infty$, then*
$$\lambda\left(\bigcap_{k=1}^\infty A_k\right) = \lim_{k\to\infty} \lambda(A_k).$$

Proof. 1. We can write $\bigcup A_k$ as a disjoint union; in this case this is very easy, since we have
$$\bigcup_{k=1}^\infty A_k = A_1 \cup \bigcup_{k=2}^\infty (A_k \setminus A_{k-1}).$$
Then countable additivity implies that
$$\begin{aligned} \lambda\left(\bigcup_{k=1}^\infty A_k\right) &= \lambda(A_1) + \sum_{k=2}^\infty \lambda(A_k \setminus A_{k-1}) \\ &= \lim_{n\to\infty} \lambda(A_1) + \sum_{k=2}^n \lambda(A_k \setminus A_{k-1}) \\ &= \lim_{n\to\infty} \lambda\left(A_1 \cup \bigcup_{k=2}^n (A_k \setminus A_{k-1})\right) = \lim_{n\to\infty} \lambda(A_n). \end{aligned}$$

2. Exercise.

□

Proposition 2.28. *1. All open sets and all closed sets are measurable.*

2. If $\lambda^*(A) = 0$, then A is measurable and $\lambda(A) = 0$.

Proof. 1. If G is open, then we can write $G = \bigcup_{k=1}^{\infty}(G \cap B_k(0))$ as a countable union of bounded open sets. Each bounded open set has finite outer measure and thus is measurable; and we know a countable union of measurable sets is measurable. Thus G is measurable.

If F is closed, then F^C is open and thus measurable. So F is measurable.

2. We know that $0 \leq \lambda_*(A) = 0 \leq \lambda^*(A) = 0$. Thus $A \in \mathcal{L}_0$ and so A is measurable, and $\lambda(A) = 0$.

□

Proposition 2.29 (Approximation)**.** *Let $A \subseteq \mathbb{R}^n$. Then A is measurable if and only if: for every $\varepsilon > 0$ there exist $F \subseteq A \subseteq G$ such that $\lambda(G \setminus F) < \varepsilon$.*

Proof. First, suppose A has the approximation property as described. We're going to approximate A with a clearly measurable set and then show the remainder is so small that it must also be measurable.

For any $k \in \mathbb{N}$ we can find $F_k \subseteq A \subseteq G_k$, with $\lambda(G_k \setminus F_k) < \frac{1}{k}$. Let $B = \bigcup_{k=1}^{\infty} F_k$. Then B is a countable union of measurable sets and thus measurable, and $B \subseteq A$.

Further, we know that $A \setminus B \subseteq G_k \setminus B \subseteq G_k \setminus F_k$, and thus $\lambda^*(A \setminus B) \leq \lambda(G_k \setminus F_k) < \frac{1}{k}$. Since this holds for each k, we see that $\lambda^*(A \setminus B) = 0$, and thus $A \setminus B$ is measurable. We conclude that $A = B \cup (A \setminus B)$ is a union of measurable sets and thus measurable.

Conversely, suppose A is a measurable subset of $\mathbb{R}^n$. If we take any *finite measure* subset, we know we can approximate it; so we'll build a sequence of these approximations that approximate all of A.

For each k, define $E_k = B_k(0) \setminus B_{k-1}(0)$, which you can visualize like a washer or annulus centered at zero. Since each E_k is bounded, we know that $A \cap E_K \in \mathcal{L}_0$, and thus we can find a compact set K_k and an open set G_k such that $K_k \subseteq A \cap E_k \subseteq G_k$ and $\lambda(G_k \setminus K_k) < \varepsilon 2^{-k}$.

We define $F = \bigcup_{k=1}^{\infty} K_k$ and $G = \bigcup_{k=1}^{\infty} G_k$. It's clear that G_k is open. It's less trivial to see that F is closed, but we can check that it contains all of its limit points; if $x \in \overline{F}$ then x must be a limit point of some finite union $\bigcup_{k=1}^{n} K_k$, and this is a finite union and thus closed, so $x \in \bigcup_{k=1}^{n} K_k \subseteq F$.

So we have $F \subseteq A \subseteq G$ are closed and open respectively. And we can see that

$$G \setminus F = \bigcup_{k=1}^{\infty} (G_k \setminus F) \subseteq \bigcup_{k=1}^{\infty} (G_k \setminus K_k)$$

and so

$$\begin{aligned} \lambda(G \setminus F) &\leq \sum_{k=1}^{\infty} \lambda(G_k \setminus K_k) \\ &< \sum_{k=1}^{\infty} \varepsilon 2^{-k} = \varepsilon. \end{aligned}$$

□

Proposition 2.30. *1. If A is measurable, then $\lambda_*(A) = \lambda^*(A) = \lambda(A)$.*

2. If $A \subseteq B$ and B is measurable, then $\lambda^*(A) + \lambda_*(B \setminus A) = \lambda(B)$.

Proof. 1. If $\lambda^*(A) < \infty$, then this follows from section 2.1.6. So suppose A is measurable, and $\lambda^*(A) = \infty$.

If $\lambda(A) < \infty$, then we could find $F \subseteq A \subseteq G$ with $\lambda(G \setminus F) < 1$, and then we'd have that

$$\lambda(G) = \lambda(G \setminus A) + \lambda(A) \leq \lambda(G \setminus F) + \lambda(A) < 1 + \lambda(A) < \infty$$

which is a contradiction.

Now we just need to show that $\lambda_*(A) = \infty$. We know that $\lambda(A \cap B_k(0)) < \infty$, and for any k we have

$$\lambda(A \cap B_k(0)) = \lambda_*(A \cap B_k(0)) \leq \lambda_*(A).$$

But we know that $\lim_{k \to \infty} \lambda(A \cap B_k(0)) = \lambda(A) = \infty$ since this is a union of an ascending chain. Thus we also must have that $\lambda_*(A) = \infty$.

2. For any open $G \supseteq A$, we know that

$$\begin{aligned} \lambda(G) + \lambda_*(B \setminus A) \geq \lambda(B \cap G) + \lambda_*(B \setminus A) &\geq \lambda(B \cap G) + \lambda_*(B \setminus G) \\ &= \lambda(B \cap G) + \lambda(B \setminus G) = \lambda(B). \end{aligned}$$

This holds for any G, so we have $\lambda(B) \leq \lambda^*(A) + \lambda_*(B \setminus A)$.

Conversely, for any compact $K \subseteq B \setminus A$, we can do basically the same thing:

$$\begin{aligned} \lambda^*(A) + \lambda(K) &\leq \lambda^*(B \setminus K) + \lambda(K) \\ &= \lambda(B \setminus K) + \lambda(K) = \lambda(B). \end{aligned}$$

Thus $\lambda^*(A) + \lambda_*(B \setminus A) \leq \lambda(B)$.

□

Proposition 2.31 (Carathéodory). *A set A is measureable if and only if for every $E \subseteq \mathbb{R}^n$, we have that*

$$\lambda^*(E) = \lambda^*(E \cap A) + \lambda^*(E \cap A^c).$$

Remark 2.32. This proposition provides another way to construct measure; we could have used the outer measure only and avoided inner measure. But this presentation would have been somewhat less concrete, and made some other steps kind of tricky.

Proof. Notice first that this equation is partly cheating. For *any* set A, measurable or not, we know that

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

by the countable subadditivity of outer measure as proven in 2.16. So in either direction we're really just looking at the opposite inequality.

Suppose A is measurable. If $E \subset G$ open, then

$$\lambda(G) = \lambda(G \cap A) + \lambda(G \cap A^c) \geq \lambda^*(E \cap A) + \lambda^*(E \cap A^c).$$

Thus

$$\lambda^*(E) \geq \lambda^*(E \cap A) + \lambda^*(E \cap A^c)$$

by definition of outer measure.

Conversely, suppose that $\lambda^*(E) = \lambda^*(E \cap A) + \lambda^*(E \cap A^c)$ for any E. Then in particular, for any *finitely measurable* $M \in \mathcal{L}_0$ we have

$$\lambda(M) = \lambda^*(M) = \lambda^*(M \cap A) + \lambda^*(M \cap A^c).$$

But we also know that

$$\lambda(M) = \lambda_*(M \cap A) + \lambda^*(M \cap A^c)$$

from proposition 2.30, since we can take $M \cap A^c = M \setminus (M \cap A)$.

But subtracting these equations gives that $0 = \lambda^*(M \cap A) - \lambda_*(M \cap A)$, and thus $\lambda^*(M \cap A) = \lambda_*(M \cap A)$; and this is precisely what it means to say that $M \cap A \in \mathcal{L}_0$. since this holds for any measurable M, then $A \in \mathcal{L}$ by definition. □

2.3 Abstract Measure Spaces

At this point I want to take a moment and discuss which of the properties of the Lebesgue measure generalize, and are necessary for it to be "a measure".

We first want to talk about the properties that measurable sets have to have.

Definition 2.33. Let X be any set. We define an *algebra* of subsets of X to be a subset $\mathcal{M} \subseteq 2^X$ of the power set of X that satisfies the following properties:

- $\emptyset \in \mathcal{M}$
- If $A, B \in \mathcal{M}$ then $A \cup B \in \mathcal{M}$.
- If $A \in \mathcal{M}$ then $A^C = X \setminus A \in \mathcal{M}$.

It's easy to see that an algebra of sets must be closed under any finite unions, and also under finite intersections and under set difference.

All these statements are true of the Lebesgue measurable sets. But the measurable sets have one extra property:

Definition 2.34. Let $\mathcal{M} \subseteq 2^X$ be an algebra. Then it is a σ-*algebra* if it is also closed under countable unions (and thus intersections): if $A_1, A_2, \ldots, \in \mathcal{M}$ then $\bigcup_{k=1}^{\infty} A_k \in \mathcal{M}$.

Example 2.35.

- The power set 2^X is a σ-algebra.
- $\{\emptyset, X\}$ is a σ-algebra. In fact, this is a sub-σ-algebra of any σ-algebra.
- The measurable sets $\mathcal{L} \subset 2^{\mathbb{R}^n}$ are a σ-algebra.
- Let X be any set, ane let $\mathcal{M}_0$ be the set of all sets A such that either A is finite or A^C is finite. Then $\mathcal{M}_0$ is an algebra but not a σ-algebra.
- Let X be any set, ane let $\mathcal{M}_1$ be the set of all sets A such that either A is countable or A^C is countable. Then $\mathcal{M}_0$ is an a σ-algebra.
- Any finite algebra is a σ-algebra for basically dumb reasons.

From this we want to find a way to *build* σ-algebras. There is one lemma which will be very useful for this:

Exercise 2.36. *Let X be a set, and $\mathcal{M}_i \subset 2^X$ be a σ-algebra for each i in some index set I. Prove that $\bigcap_{i \in I} \mathcal{M}_i$ is a σ-algebra.*

Notice this is a little weird. We're not intersecting subsets of X to get a new subset of X; we're intersecting collections of subsets of X to get a new collection of subsets of X.

Now suppose $\mathcal{N} \subset 2^X$ is any collection of subsets—not necessarily an algebra. We can consider the family of σ-algebras that contain $\mathcal{N}$. Clearly there are some such σ-algebras, since 2^X is itself a σ-algebra. If we take the intersection of all these σ-algebras, we will get a new σ-algebra:

$$\mathcal{M} = \bigcap_{\mathcal{P} \supseteq \mathcal{N}} \mathcal{P}.$$

Then $\mathcal{M}$ will contain $\mathcal{N}$, and it is contained in any σ-algebra that contains $\mathcal{N}$, so it is the smallest σ-algebra containing $\mathcal{N}$. We say that $\mathcal{M}$ is the σ-algebra *generated* by $\mathcal{N}$.

An important note is that this is, and essentially must be, non-constructive. There are sets in $\mathcal{M}$ that we can't build by a countable chain of unions or intersections of elements of $\mathcal{N}$. In fact, a set in $\mathcal{M}$ can be a countable union of countable intersections of countable unions of countable intersections of ...

If we want to construct the σ-algebra $\mathcal{M}$ explicitly, we need to do some sort of transfinite induction, which is cumbersome and we just don't want to do it. But it's clear (non-constructively) that $\mathcal{M}$ must exist, and we're satisfied with that.

So far this tells us that we *can* generate σ-algebras, but doens't tell us what we want to do with them, or which σ-algebras we want. But if we want to build a measure, we definitely want to be able to measure all the "reasonable" sets.

Definition 2.37. The class of *Borel sets* in $\mathbb{R}^n$, denoted $\mathcal{B}$, is the σ-algebra generated by the collection of open sets. Clearly $\mathcal{B} \subseteq \mathcal{L}$. We sometimes write $\mathcal{B}_n$ when we need to specify the dimension.

Exercise 2.38. *Prove that the class of Borel sets is* also *the σ-algebra generated by the collection of special rectangles.*

Thus $\mathcal{B}$ is the smallest σ-algebra that contains all the sets we obviously want to be able to measure.

The Borel sets in $\mathbb{R}^n$ are not actually all the Lebesgue measurable sets. But they are close.

Definition 2.39. If $A \subseteq \mathbb{R}^n$ is measurable with $\lambda(A) = 0$, then A is a *null* set. A is null if and only if $\lambda^*(A) = 0$.

Theorem 2.40. *Suppose $A \subseteq \mathbb{R}^n$ is measurable. Then we can write $A = E \cup N$ such that E and N are disjoint, E is a Borel set, and N is a null set.*

Proof. For every $k \in \mathbb{N}$ there is a closed set $F_k \subseteq A$ such that $\lambda(A \setminus F_k) < \frac{1}{k}$, by our approximation property 2.29. Set $E = \bigcup_{k=1}^{\infty} F_k$. Then E is not necessarily closed, but it is certainly Borel since it's a countable union of closed, Borel sets.

Further, $E \subseteq A$. Then for any k, we have

$$\lambda(A \setminus E) \leq \lambda(A \setminus F_k) < \frac{1}{k}.$$

Thus $\lambda(A \setminus E) = 0$ and thus $A \setminus E$ is null. □

In fact, we proved something much stronger than the theorem statement. The set E is not only Borel, it is specifically a countable union of closed sets; we call such sets F_σ sets. Dually, a countable intersection of open sets is called a G_δ set.

Exercise 2.41. *Prove that if $N \subseteq \mathbb{R}^n$ is null, then there is a Borel null set N' such that $N \subseteq N'$. In particular, prove that N' can be chosen to be a G_δ set.*

Theorem 2.42. *Let $E \subseteq \mathbb{R}^n$ be Borel, and let $f : E \to \mathbb{R}^m$ be continuous. If A is Borel in $\mathbb{R}^m$, then $f^{-1}(A)$ is Borel in $\mathbb{R}^n$.*

Proof. This proof has to be a little weird again, because we have to use the universal property of Borel sets; we can't actually study the structure of $f^{-1}(A)$ and see that it's Borel—first because we don't know what it "should" look like, and second because we don't know what A looks like.

So we'll define a class of subsets: let

$$\mathcal{M} = \{A : A \subset \mathbb{R}^m, \quad f^{-1}(A) \in \mathcal{B}_n\}.$$

If we can show that $\mathcal{B}_m \subseteq \mathcal{M}$ then we have proven what we want to prove. But $\mathcal{B}_m$ is the smallest σ-algebra containing all the open sets in $\mathbb{R}^m$; so we want to prove that $\mathcal{M}$ is *a* σ-algebra containing all the open sets in $\mathbb{R}^m$.

First we claim that $\mathcal{M}$ is a σ-algebra. We have to check the three axioms:

1. $f^{-1}(\varnothing) = \varnothing \in \mathcal{B}_n$, so $\varnothing \in \mathcal{M}$.
2. Suppose $A_k \in \mathcal{M}$ for all $k \in \mathbb{N}$. Then for each k we know that $f^{-1}(A_j) \in \mathcal{B}_n$. Thus

$$f^{-1}\left(\bigcup_{k=1}^{\infty} A_k\right) = \bigcup_{k=1}^{\infty} f^{-1}(A_k) \in \mathcal{B}_n$$

since $\mathcal{B}_n$ is a σ-algebra and this is a countable union of $\mathcal{B}_n$ sets. Thus $\bigcup_{k=1}^{\infty} f^{-1}(A_k) \in \mathcal{M}$.

3. Suppose $A \in \mathcal{M}$. Then $f^{-1}(A) \in \mathcal{B}_n$, and thus

$$f^{-1}(A^C) = \{x \in E : f(x) \notin A\} = E \setminus \{x \in E : f(x) \in A\} = E \setminus f^{-1}(A)$$

is a difference of $\mathcal{B}_n$ sets, and thus is in $\mathcal{B}_n$. So $A^C \in \mathcal{M}$.

Thus $\mathcal{M}$ satisfies the three axioms of a σ-algebra: it contains the null set, and is closed under countable unions and under complements. So $\mathcal{M}$ is a σ-algebra.

So now we just need to show that $\mathcal{M}$ contains all the open sets. So let $G \subseteq \mathbb{R}^m$ be an open set. Then we can write $f^{-1}(G) = E \cap H$ where $H \subseteq \mathbb{R}^n$ is open. Thus $H \in \mathcal{B}_n$, and we know $E \in \mathcal{B}_n$, so $f^{-1}(G) = H \cap E \in \mathcal{B}_n$. So $G \in \mathcal{M}$.

□

Remark 2.43. We know that $\mathcal{M}$ contains all the Borel sets; but it might contain far, far more—and whether it does depends on the specific function. In the extreme case where f is constant, then $\mathcal{M}$ is the largest possible σ-algebra, containing every possible subset of $\mathbb{R}^m$.

Corollary 2.44. *Let $E \subseteq \mathbb{R}^n, F \subseteq \mathbb{R}^m$ be Borel, and let $f : E \to F$ be a homeomorphism. Then f gives a bijection between Borel sets in E and in F. That is, If $B \subseteq E$, then $B \in \mathcal{B}_n$ if and only if $f(B) \in \mathcal{B}_m$.*

Proof. This follows because f and f^{-1} are both continuous. The previous theorem shows that if $f(B)$ is Borel, then so is B; considering the function f^{-1} instead shows that if B is Borel, then so is $(f^{-1})^{-1}(B) = f(B)$. □

We still haven't defined an actual measure, though. Clearly we want to use σ-algebras to define the class of measurable sets; but what does an actaul measure look like?

Definition 2.45. A *measure space* consists of three objects:

- A nonempty set X
- A σ-algebra $\mathcal{M} \subseteq 2^X$
- A function $\mu : \mathcal{M} \to [0, \infty]$ such that $\mu(\emptyset) = 0$, and if $A_1, A_2, \dots$ are disjoint then

$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$

We say the function μ is a measure.

Exercise 2.46. *Prove the following facts about abstract measures:*

1. *If $A, B \in \mathcal{M}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$.*

2. *If $A_1, A_2, \dots \in \mathcal{M}$, then*
$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) \leq \sum_{k=1}^{\infty} \mu(A_k).$$

3. *If $A_1 \subseteq A_2 \subseteq \dots$ are in $\mathcal{M}$ then*
$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \lim_{k\to\infty} \mu(A_k).$$

Example 2.47.

- We can take $X = \mathbb{R}^n$, $\mathcal{M} = \mathcal{L}$, and $\mu = \lambda$. This is the Lebesgue measure.
- We can take $X = \mathbb{R}^n$, $\mathcal{M} = \mathcal{B}$ the set of Borel sets, and $\mu = \lambda$. This is the same measure, but allows fewer sets to be measurable. In particular, many sets which are null under the Lebesgue measure are unmeasurable here.
- Take X to be any set, $\mathcal{M} = 2^X$, and $\mu(A) = \infty$ if $A \neq \varnothing$.
- The counting measure: Take X to be a non-empty set, $\mathcal{M} = 2^X$, and
$$\mu(A) = \begin{cases} \#A & A \text{ finite} \\ \infty & A \text{ infinite} \end{cases}$$
- The Dirac measure: let X be any non-empty set and $\mathcal{M} = 2^X$. Fix some $x_0 \in X$ and define $\mu(A) = \chi_A(x_0)$. We usually call this measure the *Dirac measure* and write it δ_{x_0}. It is also sometimes called the Diract delta function, despite not being a function on X.

Most of what we'll prove about the Lebesgue measure is actually true in any abstract measure space; in particular, our definition of integral will work for any measure.

Definition 2.48. Let $X, \mathcal{M}, \mu$ be a measure space. We can define a new measure space called teh *completion* of $(X, \mathcal{M}, \mu)$. We define a σ-algebra $\overline{\mathcal{M}}$ by the property that $A \in \overline{\mathcal{M}}$ if and only if there are $B, C \in \mathcal{M}$ with $B \subseteq A \subseteq C$ and $\mu(C \setminus B) = 0$. Clearly $\mathcal{M} \subseteq \overline{\mathcal{M}}$.

Then in this situation we have $\mu(C) = \mu(B)$, so define $\overline{\mu}(A) = \mu(B)$. It's not too hard to show that $\overline{\mathcal{M}}$ is a σ-algebra and $\overline{\mu}$ is a measure.

Exercise 2.49. *Prove that, if $E \subseteq A \in \overline{\mathcal{M}}$ and $\overline{\mu}(A) = 0$, then $E \in \overline{\mathcal{M}}$ and $\overline{\mu}(E) = 0$.*

Definition 2.50. We say a measure space $X, \mathcal{M}, \mu$ is *complete* if whenever $E \subseteq A \in \mathcal{M}$ and $\mu(A) = 0$, the $E \in \mathcal{M}$.

We observe that the Lebesgue measure is complete; it is in fact the completion of a measure defined on the Borel sets.

Given a measure space, we can find sub-measure-spaces. Suppose $(X, \mathcal{M}, \mu)$ is a measure space, and $B \in \mathcal{M}$ is a measureable subset of X. Then we can define a new measure space $(B, \mathcal{M}_B, \mu_B)$ by taking $\mathcal{M}_B = \{A \cap B : A \in \mathcal{M}\}$, and defining $\mu_B(A) = \mu(A)$.

This just means that A is measurable in B if it's the intersection of a measurable set with B, and the measure is inherited from the larger space.

Example 2.51. We know that $[0,1] \in \mathcal{L}$, so we can define a measure space $([0,1], \mathcal{L}_{[0,1]}, \lambda_{[0,1]})$, where measurable sets are the intersections of Lebesgue measurable sets with the closed interval. This space has total measure one, and does *exactly what you think it should do.*

Example 2.52. If our measure space is $\mathcal{R}^2$, then $\mathbb{R}$ is a Lebesgue-measurable subspace of $\mathcal{R}^2$, so we can look at the measure on $\mathbb{R}$ induced by the measure on $\mathcal{R}^2$. But this isn't really a useful measure! In this case, the induced σ-algebra is exactly the collection of usual Lebesgue measurable subsets of $\mathbb{R}$. But the measure of any set will be 0.

Finally, we are prepared to make some notes on the topic of probability.

Definition 2.53. A *probability space* is a measure space $(\Omega, \mathcal{F}, P)$ (where Ω is a set, $\mathcal{F}$ is a σ-algebra of subsets of Ω, and P is a measure) such that $P(\Omega) = 1$.

We say that the elements of $\mathcal{F}$, which are subsets of Ω, are *events*, and the *probability* of an event $A \in \mathcal{F}$ is $P(A)$.

Example 2.54. The space $[0,1]$ with the (induced) Lebesgue measure is a probability space. In fact, $[0,1] \times [0,1] \times \cdots \times [0,1]$ is a probability space.

Example 2.55. The space $[0,2]$ with the regular Lebesgue measure is a measure space but *not* a probability space, since $\lambda([0,2]) = 2 \neq 1$. But if we define $\mu(A) = \frac{1}{2}\lambda(A)$, then μ is a measure and so $([0,2], \mathcal{L}_{[0,2]}, \mu)$ is a probability space.

Definition 2.56. Suppose $(\Omega, \mathcal{F}, P)$ is a probability space, and $B \in \mathcal{F}$ with $P(B) > 0$. We define the *conditional probability of A given B* by

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

Exercise 2.57. *Prove that $P(A|B)$ is a probability measure on B.*

3 Interesting Sets for the Lebesgue Measure

3.1 Invariance of Lebesgue Measure

Within $\mathbb{R}^n$ there are ways we can move sets around that seem like they either shouldn't change the measure, or should change it in predictable ways.

Definition 3.1. Let $A \subseteq \mathbb{R}^n$ and let $x \in \mathbb{R}^n$. We define the *translation* of A by x to be the set

$$x + A = \{x + a : a \in A\}.$$

Now let $t \in \mathbb{R}^{>0}$. We define the *dilation* of A by t to be the set

$$tA = \{ta : a \in A\}.$$

Lemma 3.2. *Let $A \subseteq \mathbb{R}^n$, let $x \in \mathbb{R}^n$, and let $t \in \mathbb{R}^{>0}$. Then:*

- $\lambda^*(x + A) = \lambda(A)$ *and* $\lambda^*(tA) = t^n \lambda(A)$.
- $\lambda_*(x + A) = \lambda(A)$ *and* $\lambda_*(tA) = t^n \lambda(A)$.
- *If A is measurable, then $x + A$ and tA are measurable, and $\lambda(x + A) = \lambda(A)$ and $\lambda(tA) = t^n \lambda(A)$.*

Proof. We first prove the lemma for special rectangles. If $I = [a_1, b_1] \times \cdots \times [a_n, b_n]$ is a special rectangle, then

$$x + I = [a1 + x1, b1 + x1] \times \cdots \times [a_n + x_n, b_n + x_n]$$

so by definition,

$$\lambda(x + I) = \prod_{i=1}^{n} (b_i + x_i - (a_i + x_i) = \prod_{i=1}^{n} (b_i - a_i) = \lambda(I).$$

Similarly,

$$tI = [ta_1, tb_1] \times \cdots \times [ta_n, tb_n]$$

and so

$$\lambda(tI) = \prod_{i=1}^{n} t(b_i - a_i) = t^n \prod_{i=1}^{n} (b_i - a_i) = t^n \lambda(I).$$

Now we want to extend this to all Lebesgue measurable sets. But this just follows from the steps of the construction of the Lebesgue measure. Clearly the result holds for special polygons; and then the set of special polygons contained in $x + G$ or tG is the set of translations or dilations of special polygons contained in G. Thus the result holds for open sets. Similarly, the result must hold for compact sets, and thus for inner and outer measure. Finally, since the result holds for inner and outer measure, it holds for the measure of measurable sets.

□

We'd like to generalize these two operations a bit further. We want to include translations and dilations, and also some other operations like rotations.

Definition 3.3. Suppose $f : U \to V$ is a function of vector spaces. We say that f is *affine* if

$$f(ax + (1-a)y) = af(x) + (1-a)f(y)$$

for any vectors $x, y \in U$ and scalars $a \in \mathbb{R}$.

This basically tells us that we don't preserve vectors, but we do preserve lines: a point on the line from x to y gets mapped to a point on the line from $f(x)$ to $f(y)$.

Exercise 3.4. *Prove that $f : U \to V$ is affine if and only if there is a linear function $L : U \to V$ and a vector $v \in V$ such that $f(x) = v + L(x)$ for every $x \in U$. Further, this choice of L and v is unique.*

An affine transformation combines a translation and a linear function, but we already understand translations. So let's see what linear functions do to the Lebesgue measure. We wish to prove the following statement:

Theorem 3.5. *Let T be a $n \times n$ matrix, and let $A \subseteq \mathbb{R}^n$. Then*

$$\lambda^*(TA) = |\det T|\lambda^*(A)$$
$$\lambda_*(TA) = |\det T|\lambda_*(A)$$

Further, if A is measurable, then TA is measurable, and

$$\lambda(TA) = |\det T|\lambda(A)$$

sketch. We'll specialize to just proving this for an open set G; once that's proven, we can extend it to the rest of measurable sets. And we can cover G by small cells that we've already understood.

So let $J = [0,1) \times \cdots \times [0,1)$. This is not a special rectangle, but it is a rectangle. Clearly $\lambda(J) = 1$. Then since T is continuous, we can see that $T(J)$ must be measurable. We set $\rho = \frac{\lambda(TJ)}{\lambda(J)} = \lambda(TJ)$. And we claim that $\lambda(TA) = \rho\lambda(A)$.

From here we're essentially going to tile G from the inside with copies of J. We can divide $\mathbb{R}^n$ into translated copies of J of the form $[a_1, a_1+1) \times (a_2, a_2+1) \times \cdots \times [a_n, a_n+1)$. Take all the ones that are inside G. Then tile the remainder with $1/2 \times 1/2 \times \cdots \times 1/2$ rectangles, and then $1/4$, and so on. By following this process we can write $G = \bigcup_{k=1}^\infty J_k$; each J_k is disjoint, and is a translation of a dilation of J.

For any rectangle $J_k = z_k + t_k J$ we see that $\lambda(J_k) = t_k^n \lambda(J)$, and thus

$$\lambda(TJ_k) = \lambda(Tz_k + t_k TJ) = \lambda(t_k TJ) = t_k^n \lambda(TJ) = \rho\lambda(J_k).$$

Then we can see that

$$\begin{aligned}\lambda(TG) &= \lambda\left(\bigcup_{k=1}^\infty TJ_k\right) \\ &= \sum_{k=1}^\infty \lambda(TJ_k) = \sum_{k=1}^\infty \rho\lambda(J_k) \\ &= \rho\sum_{k=1}^\infty \lambda(J_k) = \rho\lambda\left(\bigcup_{k=1}^\infty J_k\right) = \rho\lambda(G).\end{aligned}$$

This proves our formula for open sets; by our sort of standard Lebesgue construction, we can extend this to any Lebesgue measurable set.

To prove the theorem, we have to prove that $\rho = |\det T|$. We can just say this is a theorem of linear algebra: the determinant of a matrix is the volume of the image of the unit cube. But if we want to prove it, we can follow this outline:

If T is invertible, then it's a theorem of linear algebra that T can be written as a product of "elementary" matrices, which correspond to the three row operations. We can show that this result holds for any elementary matrix; since the determinant is multiplicative, that implies that it holds for any invertible matrix.

If T is multiplying one dimension by a scalar, then (without loss of generality) $T(J) = [0,c) \times [0,1) \times \cdots \times [0,1)$, so $\det(T) = c$ and $\lambda(TJ) = |c|$. If T is a row-switching matrix, then $\det T = 1$ and $TJ = J$ so $\lambda(TJ) = 1$.

If T is a row-addition matrix, then $\det T = 1$. Showing that $\lambda(TJ) = 1$ is a bit trickier. But we can carefully choose a set

$$A = \{-cx_2 \leq x_1 \leq 0, 0 \leq x_i \leq 1 \text{ for } i > 1\}$$

and then if we apply the row-adding matrix

$$T = \begin{bmatrix} 1 & c & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

then $T(A)$ is just A reflected across the first coordinate. Thus $\lambda(TA) = \lambda(A)$. Since we know that $\lambda(TA) = \rho\lambda(A)$ that proves that $\rho = 1 = \det T$.

Conversely, if T is invertible, hen the determinant of T is zero, so we want to show that $\rho = 0$, or equivalently, that $\lambda(TA) = 0$. It's sufficient to show that $T\mathbb{R}^n$ has zero measure. But since $\det T = 0$, we know that the kernel is non-trivial, and by the rank-nullity theorem $\dim T(\mathbb{R}^n) < \dim \mathbb{R}^n$. We proved that any proper affine subspace has measure zero, and thus $T(\mathbb{R}^n)$ has measure zero.

(Technically we only proved this if the affine subspace is a special rectangle, but there's nothing really interesting about proving it for the rotated versions.)

□

We'll finish this discussion by mentioning a particularly important class of affine transformations:

Definition 3.6. Suppose V is an inner product space. We say a linear transformation $L : V \to V$ is *orthogonal* if $\langle L(u), L(v)\rangle = \langle u, v\rangle$.

We say a $n \times n$ matrix A is *orthogonal* if A is invertible and $A^{-1} = A^T$ the transpose of A.

Exercise 3.7. *Prove that a matrix is orthogonal if and only if the associated linear transformation is orthogonal.*

Exercise 3.8. *Prove that if L is orthogonal, then $|\det L| = 1$. Hint: use theorem 3.5 and use $A = B(0,1)$.*

This shows that if L is an orthogonal matrix, then $\lambda(A) = \lambda(LA)$ for any measurable set A; that is, orthogonal matrices preserve measure. Since translations also preserve measure, we can generalize just a hair further.

Definition 3.9. Let $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ such that there is a $z \in \mathbb{R}^n$ and an orthogonal matrix L such that $\Phi(x) = z + Lx$ for any $x \in \mathbb{R}^n$. Then we say that Φ is a *rigid motion*. Notice that Φ is an affine transformation.

Remark 3.10. The set of rigid motions on $\mathbb{R}^n$ form a group, known as the *Euclidean group* or the group of rigid motions.

It is equivalent to ask that Φ be an *isometry*, that is, that Φ preserve distances: we say that Φ is an isometry if

$$|\Phi(x) - \Phi(y)| = |x - y|$$

for any $x, y \in \mathbb{R}^n$.

3.2 A non-measurable set

In this section we will construct (after a fashion) a set $E \subseteq \mathbb{R}^n$ that is not measurable.

We begin by looking at the set $\mathbb{Q}^n \subset \mathbb{R}^n$. For any fixed $x \in \mathbb{R}^n$ we can consider the set of translations $x + \mathbb{Q}^n$. It's easy to see that $y \in x + \mathbb{Q}^n$ if and only if $y - x \in \mathbb{Q}^n$.

Exercise 3.11. *Prove that the translates of $\mathbb{Q}^n$ partition $\mathbb{R}^n$. That is, if $x, y \in \mathbb{R}^n$, then either $x + \mathbb{Q}^n = y + \mathbb{Q}^n$ or $(x + \mathbb{Q}^n) \cap (y + \mathbb{Q}^n) = \varnothing$.*

We sometimes might call these translates "cosets" of $\mathbb{Q}^n$.

It's clear that $\mathbb{R}^n = \bigcup_{x \in \mathbb{R}^n} x + \mathbb{Q}^n$. But each set on the right occurs infinitely many times. If we assume the axiom of choice, we can pick exactly one $x \in \mathbb{R}^n$ in each translate of $\mathbb{Q}^n$; let $E \subset \mathbb{R}^n$ be the set of these chosen points. Then $\mathbb{R}^n = \bigcup_{x \in E}(x + \mathbb{Q}^n)$, and this union is disjoint.

But we can also turn this statement around! For every $x \in \mathbb{R}^n$, we have exactly one $y \in E$ such that $x - y \in \mathbb{Q}^n$. But if we write $x - y = z$, we see that there is exactly one $z \in \mathbb{Q}^n$ such that $x - z = y \in E$. So instead we can write a disjoint union

$$\mathbb{R}^n = \bigcup_{z \in \mathbb{Q}^n} z + E.$$

And this union is disjoint.

But this by itself generates a problem. It's easy to see from this that if E is measurable, it must have positive measure. For

$$\lambda^*(\mathbb{R}^n) = \lambda^* \left(\bigcup_{z \in \mathbb{Q}^n} z + E \right) \leq \sum_{z \in \mathbb{Q}^n} \lambda(z + E) = \sum_{z \in \mathbb{Q}^n} \lambda(E).$$

If $\lambda^*(E) = 0$, then, we have $\lambda^*(\mathbb{R}^n) = 0$ which is clearly false.

But we will show that $\lambda_*(E) = 0$. Let K be any compact subset of E; we will show that $\lambda(K) = 0$. Fix $D = B_1(0) \cap \mathbb{Q}^n$ to be the rational points in the unit ball. Then D is a bounded, countably infinite set. We know that

$$\bigcup_{r \in D} r + K \subseteq \bigcup_{r \in D} r + E$$

is a countably infinite disjoint union. We compute that

$$\lambda\left(\bigcup_{r \in D} r + K\right) = \sum_{r \in D} \lambda(r + K) \sum_{r \in D} \lambda(K).$$

If $\lambda(K) > 0$, then this sum is infinite; but since $D and K$ are bounded, the union is bounded and thus has finite measure. Thus we must have $\lambda(K) = 0$. Since this holds for any compact set $K \subseteq E$, this implies that $\lambda_*(E) = 0$.

Then $0 = \lambda_*(E) < \lambda^*(E)$, and so E is not measurable.

Corollary 3.12. *If $A \subseteq \mathbb{R}^n$ is measurable and $\lambda(A) > 0$ then there is a non-measurable subset $B \subseteq A$.*

Proof. Let E be the set we constructed above; then we can write

$$A = \bigcup_{x \in \mathbb{Q}^n} \left((x + E) \cap A\right).$$

Since A has positive measure, and this is a countable union, there is at least one $x_0 \in \mathbb{Q}^n$ such that $(x_0 + E) \cap A$ has positive outer measure. Then set $B = (x_0 + E) \cap A$. By our argument from above, $\lambda_*(B) = 0$, but $\lambda^*(B) > 0$. Thus $B \notin \mathcal{L}$. □

This same logic, with some care, can be used to generate important paradoxical results.

Fact 3.13 (Banach-Tarski)**.** *Let A, B be any two bounded subsets of $\mathbb{R}^3$ with nonempty interior. Then we can write both sets as finite disjoint unions $A = \bigcup_{k=1}^n A_k$, $B = \bigcup_{k=1}^n B_k$, and define rigid motions $\Phi_k : \mathbb{R}^3 \to \mathbb{R}^3$, such that $\Phi_k(A_k) = B_k$.*

This is "paradoxical" because A and B need not have the same measure, but we know the rigid motions Φ_k preserve measure. In the famous example, we take A to be a ball of radius 1, and B to be the disjoint union of two balls of radius 1. Though an explicit construction that uses the axiom of choice, Banach and Tarski showed that you can divide A into five disjoint pieces, and use rigid motions of each piece to produce B.

However, there is no rigid motion such that $\Phi(A) = B$.

Exercise 3.14. *Prove that there are disjoint subsets $A, B \subseteq \mathbb{R}^n$ such that*

$$\lambda^*(A \cup B) < \lambda^*(A) + \lambda^*(B)$$
$$\lambda_*(A \cup B) > \lambda_*(A) + \lambda_*(B).$$

Exercise 3.15. *Let $A, B, C \subseteq \mathbb{R}^n$ such that $A \subseteq C$ and $\lambda(B \cap C) = 0$. Then A, B are not necessarily disjoint but they are separated in a measure theoretic sense. Prove that $\lambda^*(A \cup B) = \lambda^*(A) + \lambda^*(B)$.*

3.3 Cantor Sets and Lebesgue Functions

In this section we're mostly going to stay in $\mathbb{R}$, although there are perfectly reasonable generalizations to $\mathbb{R}^n$ and we'll try to mention them.

We've already seen the Cantor set $C \subseteq \mathbb{R}$ in section 2.1.4. We removed a union of open middle thirds, and saw what was left. Here we can generalize this.

Choose a sequence of positive real number l_k such that $1 > 2l_1 > 4l_2 > \cdots > 2^k l_k > \ldots$. We can start with the closed interval $[0, 1]$ and remove an open interval from the middle of length $1-2l_1$, leaving $[0, l_1] \cup [1-l_1, 1]$ as the remainder. We denote the middle open interval $(l_1, 1 - l_1) = J_{1/2}$.

From each of these closed intervals we can remove a middle bit of length $l_1 - 2l_2$, leaving four intervals of length l_2. We call the removed intervals $J_{1/4} = (l_2, l_1 - l_2)$ and $J_{3/4} = (1 - l_1 + l_2, 1 - l_2)$.

At the kth step of this process, we have remaining 2^k intervals of length l_k, and have removed $2^k - 1$ intervals which we have labeled $J_{i/2^k}$ for $1 \leq i \leq 2^k - 1$.

Let us denote the limiting set

$$A = [0, 1] \setminus \bigcup_{k \in \mathbb{N}, 1 \leq i \leq 2^k - 1} J_{i/2^k}.$$

A is the complement of a union of open intervals, and thus A is closed and hence compact. We see that $\lambda(A) = \lim_{k\to\infty} 2^k l_k$.

We obtain the original Cantor set C by taking $l_k = 3^{-k}$ for each k. Then $\lambda(C) = \lim_{k\to\infty}(2/3)^k = 0$.

The generalized Cantor sets have one more interesting property: they are *nowhere dense*.

Definition 3.16. A set A is *nowhere dense* if its interior is empty. That is, A is nowhere dense if $A^\circ = \varnothing$. Consequently, $\overline{A^\circ} = \varnothing$ as well.

Why is A nowhere dense? if A has non-empty interior, then it must contain an open interval I with positive length r. We can choose a k such that $2^{-k} \leq r$, and then A is contained in a union of disjoint intervals of length 2^{-k}. Thus $I \not\subseteq A$.

You might think that this implies that A has zero measure. Recall we used the original Cantor set to show you can have an uncountable set with zero measure. But we can build "fat" Cantor sets with positive measure. In fact, if we set

$$l_k = \frac{\theta k + 1}{(k+1)2^k}$$

then $\lambda(A) = \theta$. This works for any $\theta \in [0, 1)$. Thus we can have a nowhere dense set of positive measure, and in fact of just about as much measure as we like.

We can now define the *Lebesgue function* associated to A. We'll set $J_0 = (-\infty, 0)$ and $J_1 = (1, \infty)$. Then it's easy to define a function $f : A^C \to [0, 1]$ by $f(x) = r$ for every $x \in J_r$. We know that the interval J_r is entirely to the left of J_s if $r < s$, so f is an increasing function.

Further, f is continuous on A^C. Informally, we can convince ourselves of this because it seems like the function must be locally constant. But there are infinitely many infinitely small sub-intervals, so it's possible something weird is going on.

However, suppose $|x - y| < l_k$. Then if $x \in J_r$ and $y \in J_s$, one of two things must happen. One possibility is that $r = s$, in which case $f(x) = f(y)$. But if $r \neq s$, then the intervals J_r and J_s must be relatively small, and close together. Both r and s will have to have denominators $\geq 2^{-k}$, and thus $|f(x) - f(y)| < 2^{-k}$. Thus f must be continuous, and in fact uniformly continuous.

(You can see Jones p. 88 for a careful proof of this last fact, but it's mostly some careful work with this definition as a limit).

Thus f is continuous on A^C. It turns out that we can extend f to be continuous on the closure of A^C—which is in fact all of $\mathbb{R}$.

Exercise 3.17. *Let $E \subseteq \mathbb{R}^n$ and $f : E \to \mathbb{R}$ be uniformly continuous. Then there is a unique function $F : \overline{E} \to \mathbb{R}$ such that F is continuous and $F(x) = f(x)$ for all $x \in E$.*

In our particular case we will call this extension the *Lebesgue function* corresponding to A. It is a continuous non-decreasing function $f : \mathbb{R} \to [0, 1]$ that has the property that $f(x) = r$ for any $x \in J_r$. By the intermediate value theorem, it is surjective onto $[0, 1]$.

This function is also an almost-bijection between the extended Cantor set A and the open interval $(0, 1)$. First, if $x < y$ then $f(x) < f(y)$, unless $x, y \in \overline{J_r}$ for some r. In particular, if $x, y \in A$ then $f(x) < f(y)$ unless the open interval (x, y) is one of the J_r. Thus x is *almost* strictly increasing on A.

In particular, f is strictly increasing on A except there are two points outputting $i/2^k$ for each i, k. So let $B = \{\inf(J_r)\} \cup \{0\}$ be the set of all the left endpoints of the intervals J_r. Then $f : (A \setminus B) \to (0, 1)$ is strictly increasing surjective function, and thus a bijection.

By standard set theory/cardinality arguments, this means that A has the same cardinality as $(0, 1)$.

Exercise 3.18. *If f is the Lebesgue function associated to some Cantor set A, then $f(1 - x) + f(x) = 1$ for any x.*

3.4 Non-Borel Measurable Sets

In this section we will prove that not every measurable subset of $\mathbb{R}$ is Borel. When we talk about product measures, we'll extend this result to $\mathbb{R}^n$.

Let C be the ternary Cantor set, and let f be the Lebesgue function associated to it. f is strictly increasing on C, but not on $\mathbb{R}$; but we can make it strictly increasing by defining $g(x) = x + f(x)$. Since f is continuous and nondecreasing, g is continuous and strictly increasing. Then g gives us a homeomorphism from $[0, 1]$ onto $[0, 2]$.

We first claim that $g(C)$ has positive measure. Since g is a bijection, we know that

$$g(C) = [0, 2] \setminus g(C^C) = [0, 2] \setminus g\left(\bigcup J_r\right) = [0, 2] \setminus \bigcup g(J_r).$$

But on J_r the function f is constant, so the function g is just given by $g(x) = x + r$. Thus g maps each open interval J to another open interval of the same length, and so $\lambda(g(J_r)) = \lambda(J_r)$.

Then we have

$$\lambda \bigcup g(J_r) = \sum \lambda g(J_r) = \sum \lambda(J_r) = 1$$

since we worked this out when we studied the Cantor set. Thus we have

$$\lambda(g(C)) = \lambda([0, 2]) - \lambda(g(C^C)) = 2 - 1 = 1.$$

So g has already done something strange: it's a homeomorphism between a set of measure zero and a set of measure 1. Somehow it stretches the volume of C infinitely.

But now let's consider this set $g(C)$. It's a closed set of measure 1. And since it has positive measure, by corollary 3.12, there is some set $B \subseteq g(C)$ that is not measurable. Then we define $A = g^{-1}(B)$.

We know that $A \subseteq C$, and thus $\lambda^*(A) \leq \lambda(C) = 0$. Thus A is measurable because the Lebesgue measure is complete. So we just have to prove that A is not Borel. But since g is a homeomorphism, A is Borel if and only if $g(A) = B$ is Borel (by corollary 2.44). But B is not measureable, and so it's definitely not Borel. Thus A isn't Borel either.

So we've constructed a measure zero set which isn't Borel, but is measurable (because it's measure zero). We can easily build a positive measure set that's measurable but not Borel by, like, taking $A \cup [5, 7]$. This will have measure 2, but still not be Borel.

One more observation to make here: we know homeomorphisms preserve Borel sets. But they clearly *don't* preserve measurable sets, since A is measurable and $g(A) = B$ is not.

4 The Integral

4.1 Measurable Functions

Before we can define the integral, we need to spend a bit of time talking about the sort of functions we can integrate.

First, we want to get some notational conventions out of the way. We'll often need to talk about the *extended real number line* $\overline{R} = [-\infty, \infty]$. Most of the algebra with ∞ does what you probably think it should by this point; but it's important to note that sometimes $0 \cdot \infty$ is undefined and other times it's 0.

In order to do integrals, We want to take functions where we can approximate the output in some reasonable sense: if we look at all the values where f takes on a value "near" a, the set we get will be sensible. We thus define:

Definition 4.1. Let X be a set and $\mathcal{M}$ a σ-algebra on X. Let $f : X \to \overline{\mathbb{R}}$. We say that f is $\mathcal{M}$-*measurable* if, for all $t\overline{R}$, the set $f^{-1}([-\infty, t])$ is $\mathcal{M}$-measurable.

Another way of expressing this is that for all $t \in \overline{\mathbb{R}}$, we have $\{x : f(x) \leq t\} \in \mathcal{M}$.

Exercise 4.2. *Let $A \subset X$. Prove that the characteristic function χ_A is $\mathcal{M}$-measurable if and only if $A \in \mathcal{M}$.*

Exercise 4.3. *Let $\mathcal{M} = \{\varnothing, X\}$ and $\mathcal{N} = 2^X$. Describe explicitly the sets of $\mathcal{M}$-measurable functions and of $\mathcal{N}$-measurable functions.*

You might ask why we specficially look at $[-\infty, t]$ and not $[-\infty, t)$ or $(t, \infty]$ or something. The answer is that it doesn't matter.

Proposition 4.4. *Let $\mathcal{M}$ be a σ-algebra and $f : X \to \overline{\mathbb{R}}$. Then the following are equivalent:*

1. *f is measurable*
2. *$f^{-1}([-\infty, t)) \in \mathcal{M}$ for any $t \in (-\infty, \infty]$*
3. *$f^{-1}((t, \infty]) \in \mathcal{M}$ for any $t \in \overline{\mathbb{R}}$*
4. *$f^{-1}((t, \infty,]) \in \mathcal{M}$ for any $t \in [-\infty, \infty)$*
5. *$f^{-1}(\{-\infty\}) \in \mathcal{M}$, $f^{-1}(\{\infty\}) \in \mathcal{M}$, and $f^{-1}(E) \in \mathcal{M}$ for every Borel set $E \subset \mathbb{R}$.*

Proof. □

Proposition 4.5. *Assume $f, g : X \to \mathbb{R}$ are $\mathcal{M}$-measurable, and $\phi : \mathbb{R} \to \mathbb{R}$ is Borel measurable. Then*

1. *$\phi \circ f$ is $\mathcal{M}$-measurable.*

2. *If $f \neq 0$ then $\frac{1}{f}$ is $\mathcal{M}$-measurable.*

3. *If $0 < p < \infty$ then $|f|^p$ is $\mathcal{M}$-measurable.*

4. *$f + g$ is $\mathcal{M}$-measurable.*

5. *fg is $\mathcal{M}$-measurable.*

Proof. 1. If E is a Borel set, then $\phi^{-1}(E)$ is Borel, and thus $f^{-1}(\phi^{-1}(E)) \in \mathcal{M}$.

2. Excercise. Prove that $\phi(t) = \frac{1}{t}$ is Borel measurable and then conclude this result.

3. The function $\phi(t) = |t|^p$ is continuous, and thus Borel measurable.

4. This one takes a small amount of work.

   We know that $f(x) + g(x) < t$ if and only if $f(x) < t - g(x)$, if and only if there is a $r \in \mathbb{Q}$ such that $f(x) < r < t - g(x)$. So we can write

$$(f+g)^{-1}([-\infty, t)) = \bigcup_{r \in \mathbb{Q}} f^{-1}((-\infty, r)) \cap g^{-1}((-\infty, t-r)).$$

Here we use a dumb trick called polarization. We know that $f \cdot f$ is measurable for any measurable f, by (3). So we write

$$fg = \frac{1}{4}(f+g)^2 - \frac{1}{4}(f-g)^2.$$

Since $f + g$ and $f - g$ are measurable, this whole function is measurable. □

Proposition 4.6. *Suppose $f_k : X \to \overline{\mathbb{R}}$ is $\mathcal{M}$-measurable for all $k \in \mathbb{N}$. Then the following functions are all $\mathcal{M}$-measruable:*

- $\sup_k f_k$
- $\inf_k f_k$
- $\limsup_{k \to \infty} f_k$
- $\liminf_{k \to \infty} f_k$

- $\lim_{k\to\infty} f_k$, *if the pointwise limit exists.*

Proof. We can write

$$\{x : \sup f_k(x) \leq t\} = \bigcap_k \{x : f_k(x) \leq t\}.$$

The right-hand side is an intersection of measurable sets since each f_k is measurable, so the left-hand side is measurable. Similarly

$$\{x : \inf f_k(x) \geq t\} = \bigcap_k \{x : f_k(x) \geq t\}.$$

Then we know that $\limsup f_k = \inf \sup f_k$, and $\liminf f_k = \sup \inf f_k$. Since both sup and inf are measurable, so are these.

Finally, if $\lim f_k$ exists, then $\lim f_k = \limsup f_k = \liminf f_k$ is measurable.

□

4.2 Simple Functions

Definition 4.7. A *simple* function from X to $\overline{\mathbb{R}}$ is any function which assumes finitely many values. Thus we can write

$$s = \sum_{k=1}^{m} \alpha_k \chi_{A_k}$$

where the sets A_k are disjoint and the numbers $\alpha_k \in \overline{\mathbb{R}}$ are distinct.

Exercise 4.8. *A simple function s is measurable if and only if each set A_k is measurable.*

Definition 4.9. Let $a \in \overline{\mathbb{R}}$. We define

$$a_+ = \begin{cases} a & a \geq 0 \\ 0 & a < 0 \end{cases}$$

$$a_- = \begin{cases} 0 & a \geq 0 \\ -a & a < 0 \end{cases}$$

We call these the *positive part* and *negative part* of a.

We observe that $a = a_+ - a_-$ and $|a| = a_+ + a_-$. A silly but useful observation is that $a_+ a_- = 0$.

We can extend this definition for functions: if $f : X \to \overline{\mathbb{R}}$ then $f_+(x) = (f(x))_+$ and $f_-(x) = (f(x))_-$.

Exercise 4.10. *If f is $\mathcal{M}$-measurable, then so are f_+ and f_-.*

It's easy to see that the limit of a sequence of simple measurable functions is measurable; this follows directly from proposition 4.6. Much less obvious is that the converse of this statement is also true: every measurable function is the limit of a sequence of simple measurable functions.

That means that the measurable functions are precisely the closure of simple measureable functions under pointwise limits. A function is measurable if and only if it is the limit of $s_k = \sum_{i=1}^{m_k} \alpha_{i,k} \chi_{A_{i,k}}(x)$.

Theorem 4.11. *Suppose* $f : X \to \overline{\mathbb{R}}$ *is* $\mathcal{M}$*-measurable. Then there is a sequence of* $\mathcal{M}$*-measurable simple functions* $s_1, s_2, \ldots$ *that converge pointwise to* f *on* X*. That is,* $\lim_{k\to\infty} s_k(x) = f(x)$ *for every* $x \in X$*.*

If $f \geq 0$*, we may choose the sequence such that* $0 \leq s_1 \leq s_2 \leq \ldots$*. We may always choose the sequence such that* $|s_1| \leq |s_2| \leq \ldots$*.*

Proof. First we prove the case where $f \geq 0$. We define s_k through the following complicated-looking formula:

$$s_k(x) = \begin{cases} \frac{i}{2^k} & \frac{i}{2^k} \leq f(x) < \frac{i+1}{2^k} \leq k \\ k & k \leq f(x) \end{cases}$$

This formula does two things. First, the maximum possible value we give s_k is k, and the only values we allow are those that are integer multiples of $\frac{1}{2^k}$. Thus there are $k2^k+1$ possible values of s_k, so it is simple.

We need to check two things. First, does the sequence s_k converge to f? For large k, we have $|f(x) - s_k(x)| < \frac{1}{2^k}$, so the sequence converges pointwise. (In exercise 4.12 you will prove that this convergence is uniform if the function f is bounded).

Now is each s_k measurable? We have that $s_k(x) = \frac{i}{2^k}$ when $\frac{i}{2^k} \leq f(x) < \frac{i+1}{2^k}$, so

$$s_k^{-1}\left\{\frac{i}{2^k}\right\} = f^{-1}\left(\left[\frac{i}{2^k}, \frac{i+1}{2^k}\right)\right)$$

and the latter set is measurable because f is measurable. The only other possible value of s_k is k, which happens when $k \leq f(x)$; then we have

$$s_k^{-1}\{k\} - f^{-1}\left([k, \infty]\right)$$

and again this set is measurable since f is measurable. Thus f is the pointwise limit of a sequence of simple measurable functions.

For a general function f, we can just leverage the previous result, in a way that we'll use a lot. We have a sequence of functions $0 \leq s_1 \leq s_2 \leq \ldots$ converging to f_+, and a sequence $0 \leq t_1 \leq t_2 \leq \ldots$ converging to f_-. Then the sequence $s_k - t_k$ converges to f.

□

Exercise 4.12. *If $f : X \to \mathbb{R}$ is measurable and bounded, prove that it is the* uniform *limit of a sequence of measurable functions.*

This result has one simple consequence that isn't strictly speaking about measurable functions, but which will be extremely useful to us. Remember we said that we can approximate any Lebesgue measurable set with a Borel set: a Lebesgue measurable set is a Borel set union a set of measure zero. This means that we can approximate a Lebesgue measurable function with a Borel measurable function.

Theorem 4.13. *Suppose $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is Lebesgue measurable. Then there is a Borel measurable function $g : \mathbb{R}^n \to \overline{\mathbb{R}}$ such that $\{x : f(x) \neq g(x)\}$ has measure zero.*

Proof. As usual, start by assuming $f \geq 0$. There is an increasing sequence $0 \leq s_1 \leq s_2 \leq \dots$ of Lebesgue measurable simple functions s_k that converge to f. Then for each k, we can write

$$s_k = \sum_{i=1}^{m_k} \alpha_{i,k} \chi_{A_{i,k}}$$

whre each $A_{i,k}$ is a Lebesgue measurable set. Then there is a Borel set $E_{i,k}$ such that $\lambda(A_{i,k} \setminus E_{i,k}) = 0$. Define

$$t_k = \sum_{i=1}^{m_k} \alpha_{i,k} \chi_{E_{i,k}}.$$

This is a simple, Borel measurable function such that $0 \leq t_k \leq s_k$ and $t_k = s_k$ except on a set N_k of measure zero.

Define $g = \sup_k t_k$; this is Borel measurable since it's the supremum of Borel measurable functions. Then $g(x) = f(x)$ unless $x \in (A_{i,k} \setminus E_{i,k})$ for some i. But

$$\lambda\left(\bigcup_{i=1}^{m_k} A_{i,k} \setminus E_{i,k}\right) = \sum_{i=1}^{m_k} \lambda(A_{i,k} \setminus E_{i,k}) = 0.$$

Now suppose f is any function. We have shown that we can approximate f_+ with some Borel measurable g_+, and can approximate f_- with some Borel measurable g_-. Then $g = g_+ - g_-$ is a Borel measurable function, and $g(x) = f(x)$ except on a set of measure zero. □

And now, with those preliminaries completed, we are ready to start defining the integral.

For the moment, we'll let S be the set of Lebesgue-measurable simple functions $s : \mathbb{R}^n \to [0, \infty)$.

Definition 4.14. Let $s \in S$, with $s = \sum_{k=1}^m \alpha_k \chi_{A_k}$ where the A_k are disjoint measurable sets. Then the *integral* of s is

$$\int s \, d\lambda = \sum_{k=1}^m \alpha_k \lambda(A_k).$$

Here we use the convention that $0 \cdot \infty = 0$. If $\alpha_k = 0$, it doesn't matter if $\lambda(A_k)$ is infinite. And when we allow ∞-valued functions, we'll ignore that as long as it happens on a set of measure 0.

We can always assume that $\bigcup A_k = \mathbb{R}^n$ if that's convenient; if it isn't true, we can always define $A_{m+1} = \left(\bigcup_{k=1}^m A_k\right)^C$ and $\alpha_{m+1} = 0$, and nothing substantive will change.

It's not *immediately* clear that this definition is well-defined; there is more than one way to describe a simple function like this. But we will prove that it is well-defined in the next proposition.

Before we do that, though, it's worth emphasizing the ways this is similar to the Riemann integral. We can look at the Riemann integral as approximating functions below by a series of step functions. So any finite Riemann sum will add up a finite collection of heights-times-widths.

Here the α_k plays the role of the height, and the $\lambda(A_k)$ plays the role of the width. But we get some extra flexibility by not requiring our A_k to all be intervals; this flexibility is given by all the work we did to define the Lebesgue measure in section 2.

Proposition 4.15.

1. *$\int s \, d\lambda$ is well-defined, and doesn't depend on the measurable sets we choose to divide $\mathbb{R}^n$ into.*
2. *$0 \leq \int s \, d\lambda \leq \infty$.*
3. *If $0 \leq c < \infty$ is a constant, then $\int cs \, d\lambda = c \int s \, d\lambda$.*
4. *If $s, t \in S$, then $\int (s + t) \, d\lambda = \int s \, d\lambda + \int t \, d\lambda$.*
5. *If $s, t \in S$ and $s \leq t$, then $\int s \, d\lambda \leq \int t \, d\lambda$.*

Proof. We're going to prove (5) first, and that's going to give us most of the rest for free.

Suppose we have $s, t \in S$ with $s \leq t$. Then we have representations

$$s = \sum_{k=1}^m \alpha_k \chi_{A_k} \qquad t = \sum_{j=1}^n \beta_j \chi_{B_j}.$$

We assume that $\bigcup A_k = \bigcup B_j = \mathbb{R}^n$. So we've partitioned $\mathbb{R}^n$ two ways: into the A_k and into the B_j. We can mutually refine these partitions: the sets $A_k \cap B_j$ are all disjoint, and their union is $\mathbb{R}^n$. Then we can write

$$\int s\,d\lambda = \sum_{k=1}^{m} \alpha_k \lambda(A_k) = \sum_{j,k}^{n,m} \alpha_k \lambda(A_k \cap B_j)$$
$$\int t\,d\lambda = \sum_{k=1}^{n} \beta_j \lambda(B_k) = \sum_{j,k}^{n,m} \beta_j \lambda(A_k \cap B_j).$$

We claim that for each j,k, then $\alpha_k \lambda(A_k \cap B_j) \leq \beta_j \lambda(A_k \cap B_j)$. If $\lambda(A_k \cap B_j) = 0$, then this is trivially true. If $\lambda(A_j \cap B_j) > 0$, then there is some $x \in A_j \cap B_j$. Then $s(x) = \alpha_k$ and $t(x) = \beta_j$, but $s \leq t$, so $\alpha_k \leq \beta_j$, which proves our claim.

But then $\alpha_k \lambda(A_k \cap B_j) \leq \beta_j \lambda(A_k \cap B_j)$ for every j,k, and thus $\int s\,\lambda \leq \int t\,d\lambda$ by definition.

Now this by itself proves that our definition is well-posed. For suppose we have $s = t$ as just two different ways of representing the same underlying function. Then $s \leq t$ and also $t \leq s$, so $\int s\,d\lambda \leq \int t\,d\lambda$ and also $\int t\,d\lambda \leq \int s\,d\lambda$.

Given that the definition is well posed, items (2) and (3) are fairly clear. So we just have to prove (4). But by the logic from above, we have

$$\begin{aligned}
s + t &= \sum_{j,k}^{n,m} (\alpha_k + \beta_j)\chi_{A_k \cap B_j} \\
\int (s+t)\,d\lambda &= \sum_{j,k}^{n,m} (\alpha_k + \beta_j)\lambda(A_k \cap B_j) \\
&= \sum_{j,k}^{n,m} \alpha_k \lambda(A_k \cap B_j) + \sum_{j,k}^{n,m} \beta_j \lambda(A_k \cap B_j) \\
&= \int s\,d\lambda + \int t\,d\lambda.
\end{aligned}$$

□

4.3 The Integral of Non-Negative Functions

We can now integrate simple functions, which are the measure theory analogues of our finite Riemann sums from the Riemann integral. Now we want to extend this as far as possible.

The essential idea is this: we can compute the integrals of simple functions. Since every measurable function is the limit of simple functions, we can define the integral of a measurable function to be the limit of the integrals of the simple functions.

This definition is quite simple, and it's genuinely shocking how well it works.

Definition 4.16. Let $f : \mathbb{R}^n \to [0, \infty]$ be measurable. We define the *(Lebesgue) integral* of f to be

$$\int f\,d\lambda = \sup\left\{\int s\,d\lambda : s \leq f, s \in S\right\}.$$

Exercise 4.17. *Prove that our two definitions of the integral coincide if f is a measurable simple function. In particular, prove that if $f : \mathbb{R}^n \to [0, \infty]$ is a measurable simple function with $0 \leq \alpha_k \leq \infty$, then*

$$\int f\,d\lambda = \sum_{k=1}^{m} \alpha_k \lambda(A_k).$$

We now want to prove an analogue of proposition 4.15 for this more general integral. Most of the statements just follow immediately from the definition:

1. $\int f\,d\lambda$ is well defined (since every set has a supremum);

2. $0 \leq \int f\,d\lambda \leq \infty$

3. $\int cf\,d\lambda = c\int f\,d\lambda$

5. If $f \leq g$ then $\int f\,d\lambda \leq \int g\,d\lambda$.

However, it's highly non-trivial to prove that $\int (f+g)\,d\lambda = \inf f\,d\lambda + \int g\,d\lambda$.

One half of this is easy. We have that

$$\begin{aligned}
\int (f+g)\,d\lambda &= \sup\left\{\int s\,d\lambda : s \leq f+g\right\} \\
&= \sup\left\{\int (s+t)\,d\lambda : s+t \leq f+g\right\} \\
&= \sup\left\{\int s\,d\lambda + \int t\,d\lambda : s+t \leq f+g\right\}.
\end{aligned}$$

But while $s \leq f, t \leq g$ implies that $s+t \leq f+g$, the converse isn't true. So

$$\left\{\int s\,d\lambda + \int t\,d\lambda : s+t \leq f+g\right\} \supsetneq \left\{\int s\,d\lambda + \int t\,d\lambda : s \leq f, t \leq g\right\}$$

and thus

$$\int (f+g)\,d\lambda \geq \int f\,d\lambda + \int g\,d\lambda.$$

This is basically because our definition doesn't apply to *any* sequence of simple functions approaching f, but just sequences approaching from below. (This is similar to a definition of Riemann sum that only uses lower sums.)

There are various ways to prove the converse to this statement, many of which we could work out right now. One example is to show that the supremum over $s \leq f$ is the same as the infimum over $t \geq f$. But there are some major results that we want to prove anyway that will give us this result as a simple corollary.

In particular, one of the primary advantages of the Lebesgue integral formulation is that it allows us to interchange limits and integrals relatively freely.

Proposition 4.18 (Lebesgue Monotone Convergence Theorem). *Let $f_1, f_2, \dots : \mathbb{R}^n \to \overline{\mathbb{R}}$ be measurable such that*

$$0 \leq f_1 \leq f_2 \leq \dots.$$

Then

$$\lim_{k\to\infty} \int f_k \, d\lambda = \int \left(\lim_{k\to\infty} f_k \right) \, d\lambda.$$