NOTES ON MEASURE THEORY AND THE LEBESGUE INTEGRAL MAA5229, SPRING 2019

1. σ-ALGEBRAS

Notation: Let X be a set, and let 2^X denote the set of all subsets of X. Let E^c denote the complement of E in X, and for $E, F \subset X$, write $E \setminus F = E \cap F^c$. Let $E\Delta F$ denote the *symmetric difference* of E and F:

$$E\Delta F := (E \setminus F) \cup (F \setminus E) = (E \cup F) \setminus (E \cap F).$$

Definition 1.1. Let X be a set. A *Boolean algebra* is a nonempty collection $\mathscr{A} \subset 2^X$ which is closed under finite unions and complements. A *σ-algebra* is a Boolean algebra which is also closed under countable unions.

If $\mathscr{M} \subset \mathscr{N} \subset 2^X$ are σ-algebras, we say that $\mathscr{M}$ is *coarser* than $\mathscr{N}$; likewise $\mathscr{N}$ is *finer* than $\mathscr{M}$. ◁

Remark 1.2. If $\{E_\alpha\}$ is any collection of sets in X, then

$$\left(\bigcup_\alpha E_\alpha^c\right)^c = \bigcap_\alpha E_\alpha. \tag{1}$$

Hence a Boolean algebra (resp. σ-algebra) is automatically closed under finite (resp. countable) intersections. It follows that a Boolean algebra (and a σ-algebra) on X always contains $\varnothing$ and X. (Proof: $X = E \cup E^c$ and $\varnothing = E \cap E^c$.) ◇

Definition 1.3. A *measurable space* is a pair $(X, \mathscr{M})$ where $\mathscr{M} \subset 2^X$ is a σ-algebra. A function $f : X \to Y$ from one measurable space $(X, \mathscr{M})$ to another $(Y, \mathscr{N})$ is called *measurable* if $f^{-1}(E) \in \mathscr{M}$ whenever $E \in \mathscr{N}$. ◁

Definition 1.4. A topological space $X = (X, \tau)$ consists of a set X and a subset τ of 2^X such that

(i) $\emptyset, X \in \tau$;
(ii) τ is closed under finite intersections;
(iii) τ is closed under arbitrary unions.

The set τ is a *topology* on X.

(a) Elements of τ are called *open sets*;
(b) A subset S of X is *closed* if $X \setminus S$ is open;
(c) S is a G_δ if $S = \cap_{j=1}^\infty O_j$ for open sets O_j;
(d) S is an F_σ if it is an (at most) countable union of closed sets;

Date: March 12, 2019.

(e) A subset C of X is *compact*, if for any collection $\mathcal{F} \subset \tau$ such that $C \subset \cup\{T : T \in \mathcal{F}\}$ there exist a finite subset $\mathcal{G} \subset \mathcal{F}$ such that $C \subset \cup\{T : T \in \mathcal{G}\}$;
(f) If τ and σ are both topologies on X, and $\sigma \subset \tau$, then we say that τ is *finer* than σ (and σ is *coarser* than τ) and
(g) If (X, τ) and (Y, σ) are topological spaces, a function $f : X \to Y$ is *continuous* if $S \in \sigma$ implies $f^{-1}(S) \in \tau$.

◁

Example 1.5. If (X, d) is a metric space, then the collection τ of open sets (in the metric space sense) is a topology on X. There are important topologies in analysis that are not metrizable (do not come from a metric). △

Example 1.6. Let X be a nonempty set.

(a) The power set 2^X is the finest σ-algebra on X.
(b) At the other extreme, the set $\{\varnothing, X\}$ is the coarsest σ-algebra on X.
(c) Let X be an uncountable set. The collection

$$\mathscr{M} = \{E \subset X : E \text{ is at most countable or } X \setminus E \text{ is at most countable }\} \tag{2}$$

is a σ-algebra (the proof is left as an exercise).
(d) If $\mathscr{M} \subset 2^X$ a σ-algebra, and E is any nonempty subset of X, then

$$\mathscr{M}_E := \{A \cap E : A \in \mathscr{M}\} \subset 2^E$$

is a σ-algebra on E (exercise).
(e) If $\{\mathscr{M}_\alpha : \alpha \in A\}$ is a collection of σ-algebras on X, then their intersection $\cap_{\alpha \in A} \mathscr{M}_\alpha$ is also a σ-algebra (checking this statement is a simple exercise). Hence given any set $\mathscr{E} \subset 2^X$, we can define the σ-algebra

$$\mathscr{M}(\mathscr{E}) = \bigcap\{\mathscr{M} : \mathscr{M} \text{ is a } \sigma\text{-algebra and } \mathscr{E} \subset \mathscr{M}\}. \tag{3}$$

Note that the intersection is over a nonempty collection since $\mathscr{E}$ is a subset of the σ-algebra 2^X. We call $\mathscr{M}(\mathscr{E})$ the *σ-algebra generated by $\mathscr{E}$*.
(f) An important instance of the construction in (d) is when X is a topological space and $\mathscr{E}$ is the collection of open sets of X. In this case the σ-algebra generated by $\mathscr{E}$ is called the *Borel σ-algebra* and is denoted $\mathscr{B}_X$. The Borel σ-algebra over $\mathbb{R}$ is studied more closely later.

△

The following proposition is trivial but useful.

Proposition 1.7. *If $\mathscr{M} \subset 2^X$ is a σ-algebra and $\mathscr{E} \subset \mathscr{M}$, then $\mathscr{M}(\mathscr{E}) \subset \mathscr{M}$.* †

The proposition is used in the following way: suppose we want to prove that a particular statement is true for every set in some σ-algebra $\mathscr{M}$ (say, the Borel σ-algebra $\mathscr{B}_X$), which we know is generated by a collection of sets $\mathscr{E}$ (say, the open sets of X). Then it suffices to prove that 1) the statement is true for every set in $\mathscr{E}$, and 2) the collection of sets for which the statement is true forms a σ-algebra.

A function $f : X \to Y$ between topological spaces is said to be *Borel measurable* if it is measurable when X and Y are equipped with their respective Borel σ-algebras.

Proposition 1.8. *If X and Y are topological spaces, then every continuous function $f : X \to Y$ is Borel measurable.* †

Proof. The proof is left as an exercise. (Hint: follow the strategy described after Proposition 1.7.) □

We will defer further discussion of measurable functions.

1.1. The Borel σ-algebra over $\mathbb{R}$.

Before going further, we take a closer look at the Borel σ-algebra over $\mathbb{R}$. We begin with a useful lemma on the structure of open subsets of $\mathbb{R}$:

Lemma 1.9. *Every nonempty open subset $U \subset \mathbb{R}$ is an (at most countable) disjoint union of open intervals.* †

Here we allow the "degenerate" intervals $(-\infty, a), (a, +\infty), (-\infty, +\infty)$.

Proof outline. First verify that if I and J are intervals and $I \cap J \neq \emptyset$, then $I \cup J$ is an interval. Given $x \in U$, let

$$\begin{aligned} \alpha_x &= \sup\{a : [x, a) \subset U\} \\ \beta_x &= \inf\{b : (b, x] \subset U\} \end{aligned}$$

and let $I_x = (\alpha_x, \beta_x)$. Verify that, for $x, y \in U$ either $I_x = I_y$ or $I_x \cap I_y = \emptyset$. Indeed, $x \sim y$ if $I_x = I_y$ is an equivalence relation on U. Hence, $U = \cup_{x \in U} I_x$ expresses U as a disjoint union of nonempty intervals, say $U = \cup_{p \in P} I_p$ where P is an index set and the I_p are nonempty intervals. For each $q \in \mathbb{Q} \cap U$ there exists a unique p_q such that $q \in I_{p_q}$. On the other hand, for each $p \in P$ there is a $q \in \mathbb{Q} \cap U$ such that $q \in I_p$. Thus, the mapping from $\mathbb{Q} \cap U$ to P defined by $q \mapsto p_q$ is onto. It follows that P is at most countable. □

Proposition 1.10 (Generators of $\mathscr{B}_{\mathbb{R}}$). *Each of the following collections of sets $\mathscr{E} \subset 2^{\mathbb{R}}$ generates the Borel σ-algebra $\mathscr{B}_{\mathbb{R}}$:*

- *(i) the open intervals $\mathscr{E}_1 = \{(a, b) : a, b \in \mathbb{R}\}$*
- *(ii) the closed intervals $\mathscr{E}_2 = \{[a, b] : a, b \in \mathbb{R}\}$*
- *(iii) the (left or right) half-open intervals $\mathscr{E}_3 = \{[a, b) : a, b \in \mathbb{R}\}$ or $\mathscr{E}_4 = \{[a, b) : a, b \in \mathbb{R}\}$*
- *(iv) the (left or right) open rays $\mathscr{E}_5 = \{(-\infty, a) : a \in \mathbb{R}\}$ or $\mathscr{E}_6 = \{(a, +\infty) : a \in \mathbb{R}\}$*
- *(v) the (left or right) closed rays $\mathscr{E}_7 = \{(-\infty, a] : a \in \mathbb{R}\}$ or $\mathscr{E}_8 = \{[a, +\infty) : a \in \mathbb{R}\}$*

†

Proof. We prove only the open and closed interval cases, the rest are similar and left as exercises. The proof makes repeated use of Proposition 1.7. To prove $\mathscr{M}(\mathscr{E}_1) = \mathscr{B}_{\mathbb{R}}$, first note that since each interval (a,b) is open, $\mathscr{M}(\mathscr{E}_1) \subset \mathscr{B}_{\mathbb{R}}$ by Proposition 1.7. Conversely, each open set $U \subset \mathbb{R}$ is a countable union of open intervals, so $\mathscr{M}(\mathscr{E}_1)$ contains all the open sets of $\mathbb{R}$, and since the open sets generate $\mathscr{B}_{\mathbb{R}}$ by definition, Proposition 1.7 implies $\mathscr{B}_{\mathbb{R}} \subset \mathscr{M}(\mathscr{E}_1)$. Thus $\mathscr{M}(\mathscr{E}_1) = \mathscr{B}_{\mathbb{R}}$.

For the closed intervals $\mathscr{E}_2$, first note that each closed set is a Borel set, since it is the complement of an open set; thus $\mathscr{E}_2 \subset \mathscr{B}_{\mathbb{R}}$ so $\mathscr{M}(\mathscr{E}_2) \subset \mathscr{B}_{\mathbb{R}}$ by Proposition 1.7. Conversely, each open interval (a,b) is a countable union of closed intervals $[a+\frac{1}{n}, b-\frac{1}{n}]$. More precisely, since $a<b$ we can choose an integer N such that $a+\frac{1}{N} < b - \frac{1}{N}$; then

$$(a,b) = \bigcup_{n=N}^{\infty} [a+\frac{1}{n}, b-\frac{1}{n}].$$

It follows that $\mathscr{E}_1 \subset \mathscr{M}(\mathscr{E}_2)$, so by Proposition 1.7 and the first part of the proof,

$$\mathscr{B}_{\mathbb{R}} = \mathscr{M}(\mathscr{E}_1) \subset \mathscr{M}(\mathscr{E}_2).$$

□

2. Measures

Definition 2.1. Let X be a set and $\mathscr{M}$ a σ-algebra on X. A *measure* on $\mathscr{M}$ is a function $\mu : \mathscr{M} \to [0,+\infty]$ such that

(i) $\mu(\varnothing) = 0$,
(ii) If $\{E_j\}_{j=1}^{\infty}$ is a sequence of disjoints sets in $\mathscr{M}$, then

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

If $\mu(X) < \infty$, then μ is called *finite*; if $X = \cup_{j=1}^{\infty} X_j$ with $\mu(X_j) < \infty$ for each j then μ is called *σ-finite*.

Almost all of the measures of importance in analysis (and certainly all of the measures we will work with) are σ-finite.

A triple $(X, \mathscr{M}, \mu)$ where X is a set, $\mathscr{M}$ is a σ-algebra and μ a measure on $\mathscr{M}$, is called a *measure space*. ◁

Here are some simple measures and some procedures for producing new measures from old. Non-trivial examples of measures—such as Lebesgue measure on $\mathbb{R}$— will have to wait for the Caratheodory and Hahn-Kolmogorov theorems in the following sections.

Example 2.2. (a) Let X be any set, and for $E \subset X$ let $|E|$ denote the cardinality of E. The function $\mu : 2^X \to [0,+\infty]$ defined by

$$\mu(E) = \begin{cases} |E| & \text{if } E \text{ is finite} \\ +\infty & \text{if } E \text{ is infinite} \end{cases}$$

is a measure on $(X, 2^X)$, called *counting measure*. It is finite if and only if X is finite, and σ-finite if and only if X is (at most) countable.

(b) Let X be an uncountable set and $\mathscr{M}$ the σ-algebra of (at most) countable and co-countable sets (Example 1.6(b)). For $E \in \mathscr{M}$ define $\mu(E) = 0$ if E is countable and $\mu(E) = +\infty$ if E is co-countable. Then μ is a measure.

(c) Let $(X, \mathscr{M}, \mu)$ be a measure space and $E \in \mathscr{M}$. Recall the σ-algebra $\mathscr{M}_E$ from Example 1.6(c). The function $\mu_E(A) := \mu(A \cap E)$ is a measure on $(E, \mathscr{M}_E)$. (Why did we assume $E \in \mathscr{M}$?)

(d) (Linear combinations) If μ is a measure on $\mathscr{M}$ and $c > 0$, then $(c\mu)(E) := c \cdot \mu(E)$ is a measure, and if $\mu_1, \dots \mu_n$ are measures on the same $\mathscr{M}$, then

$$(\mu_1 + \cdots \mu_n)(E) := \mu_1(E) + \cdots \mu_n(E)$$

is a measure. Likewise an infinite sum of measures $\sum_{n=1}^{\infty} \mu_n$ is a measure. (The proof of this last fact requires a small amount of care.)

(e) (Point masses) Let $X = \mathbb{R}$ and $\mathscr{M} = \mathscr{B}_{\mathbb{R}}$ (or $\mathscr{M} = 2^{\mathbb{R}}$). Fix a point $x \in \mathbb{R}$ and Define for each $E \in \mathscr{M}$

$$\delta_x(E) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{otherwise} \end{cases}$$

Then δ_x is a measure, called the *point mass at x* (or *Dirac measure*, or, much worse, *Dirac delta function*).

△

One can also define products and pull-backs of measures, compatible with the constructions of product and pull-back σ-algebras; these examples will be postponed until we have built up some more of the machinery of measurable functions.

Our next goal is to describe some basic properties of measures, but before doing so make note of the disjointification trick, which is often useful.

Proposition 2.3 (Disjointification). *If $\mathscr{M} \subset 2^X$ is a σ-algebra if and $G_1, G_2, \dots$ is a sequence of sets from $\mathscr{M}$, then there exists a sequence $F_1, F_2, \dots$ of pairwise disjoint sets from $\mathscr{M}$ such that*

$$\bigcup_{j=1}^{n} F_j = \bigcup_{j=1}^{n} G_j$$

for n either a positive integer or ∞.

If $\emptyset \neq \mathscr{M} \subset 2^X$ is closed with respect to complements, finite intersections and countable disjoint unions, then $\mathscr{M}$ is a σ-algebra. †

Proof. The proof amounts to the observation that if $\{G_n\}$ is a sequence of subsets of X, then the sets

$$F_n = G_n \setminus \left(\bigcup_{k=1}^{n-1} G_k \right) \tag{4}$$

are disjoint and in $\mathscr{M}$, and $\bigcup_{j=1}^{n} F_j = \bigcup_{j=1}^{n} G_j$ for all N (and thus $\bigcup_{n=1}^{\infty} F_n = \bigcup_{n=1}^{\infty} G_n$).

To prove the second part of the Proposition, given a sequence (G_n) from $\mathscr{M}$ use the disjointification trick to obtain a sequence of disjoint sets $F_n \in \mathscr{M}$ such that $\cup G_n = \cup F_n$. □

Theorem 2.4 (Basic properties of measures)**.** *Let $(X, \mathscr{M}, \mu)$ be a measure space.*

(a) (Monotonicity) If $E, F \in \mathscr{M}$ and $E \subset F$, then $\mu(F) = \mu(F \setminus E) + \mu(E)$. In particular, $\mu(E) \leq \mu(F)$ and if $\mu(E) < \infty$, then $\mu(F \setminus E) = \mu(F) - \mu(E)$.
(b) (Subadditivity) If $\{E_j\}_{j=1}^\infty \subset \mathscr{M}$, then $\mu(\bigcup_{j=1}^\infty E_j) \leq \sum_{j=1}^\infty \mu(E_j)$
(c) (Monotone convergence for sets) If $\{E_j\}_{j=1}^\infty \subset \mathscr{M}$ and $E_j \subset E_{j+1}$ for all j, then $\lim \mu(E_j)$ exists and moreover $\mu(\cup E_j) = \lim \mu(E_j)$.
(d) (Dominated convergence for sets) If $\{E_j\}_{j=1}^\infty \subset \mathscr{M}$ and $E_j \supset E_{j+1}$ for all j, and $\mu(E_1) < \infty$, then $\lim \mu(E_j)$ exists and moreover $\mu(\cap E_j) = \lim \mu(E_j)$.

Proof. a) By additivity, $\mu(F) = \mu(F \setminus E) + \mu(E) \geq \mu(E)$.

b) We use the disjointification trick (see Proposition 2.3): for each $j \geq 1$ let

$$F_j = E_j \setminus \left(\bigcup_{k=1}^{j-1} E_k \right).$$

Then the F_j are disjoint, and $F_j \subset E_j$ for all j, so by countable additivity and (a)

$$\mu\left(\bigcup_{j=1}^\infty E_j\right) = \mu\left(\bigcup_{j=1}^\infty F_j\right) = \sum_{j=1}^\infty \mu(F_j) \leq \sum_{j=1}^\infty \mu(E_j).$$

c) The sets $F_j = E_j \setminus E_{j-1}$ are disjoint sets whose union is $\bigcup_{j=1}^\infty E_j$, and for each j, $\bigcup_{k=1}^j F_k = E_j$. So by countable additivity,

$$\begin{aligned}
\mu\left(\bigcup_{j=1}^\infty E_j\right) &= \mu\left(\bigcup_{j=1}^\infty F_j\right) \\
&= \sum_{k=1}^\infty \mu(F_k) \\
&= \lim_{j\to\infty} \sum_{k=1}^j \mu(F_k) \\
&= \lim_{j\to\infty} \mu\left(\bigcup_{k=1}^j F_k\right) \\
&= \lim_{j\to\infty} \mu(E_j).
\end{aligned}$$

d) The sequence $\mu(E_j)$ is decreasing (by (a)) and bounded below, so $\lim \mu(E_j)$ exists. Let $F_j = E_1 \setminus E_j$. Then $F_j \subset F_{j+1}$ for all j, and $\bigcup_{j=1}^\infty F_j = E_1 \setminus \bigcap_{j=1}^\infty E_j$. So by (c) applied to the F_j, and since $\mu(E_1) < \infty$,

$$\begin{aligned}\mu(E_1) - \mu(\bigcap_{j=1}^{\infty} E_j) &= \mu(E_1 \setminus \bigcap_{j=1}^{\infty} E_j) \\ &= \lim \mu(F_j) \\ &= \lim(\mu(E_1) - \mu(E_j)) \\ &= \mu(E_1) - \lim \mu(E_j).\end{aligned}$$

Again since $\mu(E_1) < \infty$, it can be subtracted from both sides. □

Remark 2.5. Note that in item (d) of Theorem 2.4, the hypothesis "$\mu(E_1) < \infty$" can be replaced by "$\mu(E_j) < \infty$ for some j". However the finiteness hypothesis cannot be removed entirely. For instance, consider $(\mathbb{N}, 2^{\mathbb{N}})$ equipped with counting measure, and let $E_j = \{k : k \geq j\}$. Then $\mu(E_j) = \infty$ for all j but $\mu(\bigcap_{j=1}^{\infty} E_j) = \mu(\varnothing) = 0$. ◇

For any set X and subset $E \subset X$, there is a function $\mathbf{1}_E : X \to \{0, 1\}$ defined by

$$\mathbf{1}_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E \end{cases},$$

called the *characteristic function* or *indicator function* of E. For a sequence of subsets (E_n) of X, say (E_n) *converges to E pointwise* if $\mathbf{1}_{E_n} \to \mathbf{1}_E$ pointwise[1]. This notion allows the formulation of a more refined version of the dominated convergence theorem for sets, which foreshadows (and is a special case of) the dominated convergence theorem for the Lebesgue integral.

Definition 2.6. Let $(X, \mathscr{M}, \mu)$ be a measure space. A *null set* is a set $E \in \mathscr{M}$ with $\mu(E) = 0$. If $E \in \mathscr{M}$ and E^c is null, we say E has *full measure.* ◁

It follows immediately from countable subadditivity that a countable union of null sets is null. The contrapositive of this statement is a measure-theoretic version of the pigeonhole principle:

Proposition 2.7 (Pigeonhole principle for measures)**.** *If $(E_n)_{n=1}^{\infty}$ is a sequence of sets in $\mathscr{M}$ and $\mu(\cup E_n) > 0$, then $\mu(E_n) > 0$ for some n.* †

It will often be tempting to assert that if $\mu(E) = 0$ and $F \subset E$, then $\mu(F) = 0$, but one must be careful: F need not be a measurable set. This caveat is not a big deal in practice, however, because we can always enlarge the σ-algebra on which a measure is defined so as to contain all subsets of null sets, and it will usually be convenient to do so.

Definition 2.8. If $(X, \mathscr{M}, \mu)$ has the property that $F \in \mathscr{M}$ whenever $E \in \mathscr{M}$, $\mu(E) = 0$, and $F \subset E$, then μ is *complete.* ◁

[1]What would happen if we asked for uniform convergence?

Theorem 2.9. *Let $(X, \mathscr{M}, \mu)$ be a measure space. Let $\mathscr{N} := \{N \in \mathscr{M} | \mu(N) = 0\}$. Define*

$$\overline{\mathscr{M}} := \{E \cup F | E \in \mathscr{M}, F \subset N \text{ for some } N \in \mathscr{N}\}$$

Then $\overline{\mathscr{M}}$ is a σ-algebra, and

$$\overline{\mu}(E \cup F) := \mu(E)$$

is a well-defined function from $\overline{\mathscr{M}}$ to $[0, \infty]$ and is a complete measure on $\overline{\mathscr{M}}$ such that $\overline{\mu}|_{\mathscr{M}} = \mu$.

Proof. First note that $\mathscr{M}$ and $\mathscr{N}$ are both closed under countable unions, so $\overline{\mathscr{M}}$ is as well. To see that $\overline{\mathscr{M}}$ is closed under complements, consider $E \cup F \in \overline{\mathscr{M}}$ with $E \in \mathscr{M}, F \subset N \in \mathscr{N}$. Using, $F^c = N^c \cup (N \setminus F)$ gives

$$(E \cup F)^c = E^c \cap F^c = (E \cap N)^c \cup (N \cap F^c \cap E^c).$$

The first set on the right hand side is in $\mathscr{M}$ and the second is a subset of N. Thus the union is in $\overline{\mathscr{M}}$ as desired.

To prove that $\overline{\mu}$ is well defined, suppose $G = E \cup F = E' \cup F'$ for $E, E' \in \mathscr{M}$ and $F, F' \in \mathscr{N}$. In particular, there exists μ-null sets $N, N' \in \mathscr{M}$ with $F \subset N$ and $F' \subset N'$. Observe that

$$E \setminus E' \subset G \setminus E' \subset F' \subset N'.$$

Thus $\mu(E \setminus E') = 0$. By symmetry, $\mu(E' \setminus E) = 0$. On the other hand,

$$E = (E \cap E') \cup (E \setminus E').$$

Thus, $\mu(E) = \mu(E \cap E')$. By symmetry, $\mu(E') = \mu(E' \cap E)$.

The proof that $\overline{\mu}$ is a complete measure on $\overline{\mathscr{M}}$ which extends μ, is left as an exercise. □

3. Outer measures and the Caratheodory Extension Theorem

3.1. **Outline.**
The point of the construction of Lebesgue measure on the real line is to extend the naive notion of length for intervals to a suitably large family of subsets of $\mathbb{R}$. We will accomplish this objective via the Caratheodory Extension Theorem (Theorem 3.3)

Definition 3.1. Let X be a nonempty set. A function $\mu^* : 2^X \to [0, +\infty]$ is an *outer measure* if:

(i) $\mu^*(\varnothing) = 0$,
(ii) (Monotonicity) $\mu^*(A) \leq \mu^*(B)$ whenever $A \subset B$,
(iii) (Subadditivity) if $\{A_j\}_{j=1}^{\infty} \subset 2^X$, then

$$\mu^*\left(\bigcup_{j=1}^{\infty} A_j\right) \leq \sum_{j=1}^{\infty} \mu^*(A_j).$$

◁

Definition 3.2. If μ^* is an outer measure on X, then a set $E \subset X$ is *outer measurable* (or *μ^*-measurable, measurable with respect to μ^**, or just *measurable*) if

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c) \tag{5}$$

for *every* $A \subset X$. ◁

The significance of outer measures and (outer) measurable sets stems from the following theorem:

Theorem 3.3 (Caratheodory Extension Theorem)**.** *If μ^* is an outer measure on X, then the collection $\mathscr{M}$ of outer measurable sets is a σ-algebra, and the restriction of μ^* to $\mathscr{M}$ is a complete measure.*

The proof will be given later in this section; we first explain how to construct outer measures. In fact, all of the outer measures we consider will be constructed using the following proposition.

Proposition 3.4. *Let $\mathscr{E} \subset 2^X$ and $\mu_0 : \mathscr{E} \to [0, +\infty]$ be such that $\varnothing, X \in \mathscr{E}$ and $\mu_0(\varnothing) = 0$. The function $\mu^* : 2^X \to [0, \infty]$ defined by*

$$\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \mu_0(E_n) : E_n \in \mathscr{E} \text{ and } A \subset \bigcup_{n=1}^{\infty} E_n \right\} \tag{6}$$

is an outer measure. †

Note that we have assumed $X \in \mathscr{E}$, so there is at least one covering of A by sets in $\mathscr{E}$ (take $E_1 = X$ and all other E_j empty), so the definition (6) makes sense. To prove the Proposition we need the following lemma on double sums, whose proof is left as an (important!) exercise.

Lemma 3.5. *Let $(a_{m,n})_{m,n=1}^{\infty}$ be a doubly indexed sequence of nonnegative real numbers. Suppose there is a real number C so that for every finite set $F \subset \mathbb{N} \times \mathbb{N}$,*

$$\sum_{(m,n)\in F} a_{m,n} \leq C. \tag{7}$$

Then for each n and each m, the sums $\sum_{m=1}^{\infty} a_{m,n}, \sum_{n=1}^{\infty} a_{m,n}$ are convergent, and

$$\sum_{m=1}^{\infty} \sum_{n=1}^{\infty} a_{m,n} = \sum_{n=1}^{\infty} \sum_{m=1}^{\infty} a_{m,n} \leq C.$$

†

Proof of Proposition 3.4. It is immediate from the definition that $\mu^*(\varnothing) = 0$ (cover the empty set by empty sets) and that $\mu^*(A) \leq \mu^*(B)$ whenever $A \subset B$ (any covering of B is also a covering of A). To prove countable subadditivity, we make our first use of the "$\epsilon/2^n$" trick. Let (A_n) be a sequence in 2^X and let $\epsilon > 0$ be given. If $\sum_{n=1}^{\infty} \mu^*(A_n) = +\infty$ there is nothing to prove, so we assume this sum is convergent. In particular $\mu^*(A_n) < +\infty$ for each n. It follows from the definition of μ^* that for each

$n \geq 1$ there exists a countable collection of sets $(E_{n,k})_{k=1}^{\infty}$ in $\mathscr{E}$ such that $A_n \subset \bigcup_{k=1}^{\infty} E_{n,k}$ and

$$\sum_{k=1}^{\infty} \mu_0(E_{n,k}) < \mu^*(A_n) + \epsilon 2^{-n}.$$

One may now verify that the numbers $a_{n,k} = \mu_0(E_{n,k})$ satisfy the hypothesis of Lemma 3.5. We now have that the countable collection $(E_{n,k})_{n,k=1}^{\infty}$ covers $A = \bigcup_{n=1}^{\infty} A_n$, and using Lemma 3.5

$$\mu^*(A) \leq \sum_{k,n=1}^{\infty} \mu_0(E_{n,k}) < \sum_{n=1}^{\infty} (\mu^*(A_n) + \epsilon 2^{-n}) = \epsilon + \sum_{n=1}^{\infty} \mu^*(A_n).$$

Since $\epsilon > 0$ was arbitrary, we conclude $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$. □

Example 3.6. [Lebesgue outer measure] Let $\mathscr{E} \subset 2^{\mathbb{R}}$ be the collection of all open intervals $(a,b) \subset \mathbb{R}$, with $-\infty < a < b < +\infty$, together with $\varnothing$ and $\mathbb{R}$. Define $m_0((a,b)) = b - a$, the length of the interval; $\mu_0(\varnothing) = 0$; and $\mu_0(\mathbb{R}) = +\infty$. The corresponding outer measure is Lebesgue outer measure and it is the mapping $m^* : 2^{\mathbb{R}} \to [0,\infty]$ defined, for $A \in 2^{\mathbb{R}}$, by

$$m^*(A) = \inf \left\{ \sum_{n=1}^{\infty} (b_n - a_n) : A \subset \bigcup_{n=1}^{\infty} (a_n, b_n) \right\} \tag{8}$$

where we allow the *degenerate* intervals $\mathbb{R} = (-\infty, +\infty)$ and $\emptyset$. The value $m^*(A)$ is the *Lebesgue outer measure of* A. In the next section we will construct Lebesgue measure from m^* via the Caratheodory Extension Theorem; the main issues will be to show that the outer measure of an interval is equal to its length, and that every Borel subset of $\mathbb{R}$ is outer measurable. The other desirable properties of Lebesgue measure (such as translation invariance) will follow from this construction. △

Before proving Theorem 3.3 we make one observation, which will be used repeatedly. Namely, if μ^* is an outer measure on a set X, to prove that a subset $E \subset X$ is outer measurable, it suffices to prove that

$$\mu^*(A) \geq \mu^*(A \cap E) + \mu^*(A \setminus E)$$

for all $A \subset X$, since the opposite inequality for all A is immediate from the subadditivity of μ^*.

We also need one lemma, which will be used to obtain completeness. A set $E \subset X$ is called μ^*-*null* if $\mu^*(E) = 0$.

Lemma 3.7. *Every μ^*-null set is μ^*-measurable.* †

Proof. Let E be μ^*-null and $A \subset X$. By monotonicity, $A \cap E$ is also μ^*-null, so by monotonicity again,

$$\mu^*(A) \geq \mu^*(A \setminus E) = \mu^*(A \cap E) + \mu^*(A \setminus E).$$

Thus the lemma follows from the observation immediately preceding the lemma. □

Proof of Theorem 3.3. We first show that $\mathscr{M}$ is a σ-algebra. It is immediate from Definition 3.2 that $\mathscr{M}$ contains $\varnothing$ and X, and since (5) is symmetric with respect to E and E^c, $\mathscr{M}$ is also closed under complementation. Next we check that $\mathscr{M}$ is closed under finite unions (which will prove that $\mathscr{M}$ is a Boolean algebra). So, let $E, F \in \mathscr{M}$ and fix an arbitrary $A \subset X$. Since F is μ^*-measurable,

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

By subadditivity and the set equality $A \cap (E \cup F) = (A \cap E) \cup (A \cap (F \cap E^c))$,

$$\mu^*(A \cap (E \cup F)) \leq \mu^*(A \cap E) + \mu^*(A \cap (F \cap E^c)).$$

Using the last two displayed equations and the outer-measurability of E,

$$\begin{aligned}\mu^*(A \cap (E \cup F)) &+ \mu^*(A \cap ((E \cup F)^c)) \\ &\leq \mu^*(A \cap E) + \mu^*(A \cap (F \cap E^c)) + \mu^*(A \cap (F^c \cap E^c)) \\ &= \mu^*(A \cap E) + \mu^*(A \cap E^c) \\ &= \mu^*(A).\end{aligned}$$

Hence $E \cup F$ is outer-measurable. By induction, $\mathscr{M}$ is closed under finite unions.

It remains to prove that, for any disjoint sequence E_n from $\mathscr{M}$ and any $A \subset X$,

$$\mu^*(A) \geq \mu^*(A \cap \bigcup_{n=1}^{\infty} E_n) + \mu^*(A \setminus \bigcup_{n=1}^{\infty} E_n).$$

For each $N \geq 1$, we have already proved that $\bigcup_{n=1}^{N} E_n$ is outer measurable, and therefore

$$\mu^*(A) \geq \mu^*(A \cap \bigcup_{n=1}^{N} E_n) + \mu^*(A \setminus \bigcup_{n=1}^{N} E_n).$$

By monotonicity, we also know that $\mu^*(A \setminus \bigcup_{n=1}^{N} E_n) \geq \mu^*(A \setminus \bigcup_{n=1}^{\infty} E_n)$, so it suffices to prove that

$$\lim_{N\to\infty} \mu^*(A \cap \bigcup_{n=1}^{N} E_n) \geq \mu^*(A \cap \bigcup_{n=1}^{\infty} E_n). \tag{9}$$

(By monotonicity of the outer measure, the sequence on the left is increasing, and so the limit exists as an extended real number.) By the outer measurability of $\bigcup_{n=1}^{N} E_n$

$$\mu^*(A \cap \bigcup_{n=1}^{N+1} E_n) = \mu^*(A \cap \bigcup_{n=1}^{N} E_n) + \mu^*(A \cap E_{N+1})$$

for all $N \geq 1$ (the disjointness of the E_n was used here). Iterating this identity gives

$$\mu^*(A \cap \bigcup_{n=1}^{N+1} E_n) = \sum_{n=1}^{N+1} \mu^*(A \cap E_n) \tag{10}$$

and taking limits,

$$\lim_{N\to\infty} \mu^*(A \cap \bigcup_{n=1}^{N} E_n) = \sum_{n=1}^{\infty} \mu^*(A \cap E_n). \tag{11}$$

But also, by countable subadditivity, we have

$$\lim_{N\to\infty} \mu^*(A \cap \bigcup_{n=1}^{N} E_n) = \sum_{n=1}^{\infty} \mu^*(A \cap E_n) \geq \mu^*(A \cap \bigcup_{n=1}^{\infty} E_n)$$

which proves (9), and thus $\mathscr{M}$ is a σ-algebra. On the other hand, by monotonicity we have for each N

$$\mu^*(A \cap \bigcup_{n=1}^{N} E_n) \leq \mu^*(A \cap \bigcup_{n=1}^{\infty} E_n)$$

so taking $N \to \infty$ and using (11) we obtain

$$\sum_{n=1}^{\infty} \mu^*(A \cap E_n) = \lim_{N\to\infty} \mu^*(A \cap \bigcup_{n=1}^{N} E_n) \leq \mu^*(A \cap \bigcup_{n=1}^{\infty} E_n) \leq \sum_{n=1}^{\infty} \mu^*(A \cap E_n).$$

In particular, choosing $A = \bigcup_{n=1}^{\infty} E_n$ proves that μ^* is countably additive. Since it is immediate that $\mu^*(\varnothing) = 0$, we conclude that $\mu^*|_{\mathscr{M}}$ is a measure.

Finally, that μ^* is a *complete* measure on $\mathscr{M}$ is an immediate consequence of Lemma 3.7. □

4. Construction of Lebesgue measure

In this section, by an *interval* we mean any set $I \subset \mathbb{R}$ of the from (a, b), $[a, b]$, $(a, b]$, $[a, b)$, including $\varnothing$, open and closed half-lines and $\mathbb{R}$ itself. We write $|I| = b - a$ for the length of I, interpreted as $+\infty$ in the line and half-line cases and 0 for $\varnothing$. Recall the definition of Lebesgue outer measure of a set $A \subset \mathbb{R}$ from Example 3.6:

$$m^*(A) = \inf \left\{ \sum_{n=1}^{\infty} |I_n| : A \subset \bigcup_{n=1}^{\infty} I_n \right\}$$

where the I_n are *open* intervals, or empty.

Theorem 4.1. *If $I \subset \mathbb{R}$ is an interval, then $m^*(I) = |I|$.*

Proof. We first consider the case where I is a finite, closed interval $[a, b]$. For any $\epsilon > 0$, the single open interval $(a - \epsilon, b + \epsilon)$ covers I, so $m^*(I) \leq (b - a) + 2\epsilon = |I| + 2\epsilon$, and thus $m^*(I) \leq |I|$. For the reverse inequality, again choose $\epsilon > 0$, and let (I_n) be a cover of I by open intervals such that $\sum_{n=1}^{\infty} |I_n| < m^*(I) + \epsilon$. Since I is compact, there is a finite subcollection $(I_{n_k})_{k=1}^{N}$ of the I_n which covers I. Then

$$\sum_{k=1}^{N} |I_{n_k}| > b - a = |I|. \tag{12}$$

To verify this statement, observe that by passing to a further subcollection, we can assume that none of the intervals I_{n_k} is contained in another one. Then re-index $I_1, \ldots I_N$ so that the left endpoints $a_1, \ldots a_N$ are listed in increasing order. Since these intervals cover I, and there are no containments, it follows that $a_2 < b_1$, $a_3 < b_2$, ...$a_N < b_{N-1}$. (Draw a picture.) Therefore

$$\sum_{k=1}^{N} |I_k| = \sum_{k=1}^{N} (b_k - a_k) = b_N - a_1 + \sum_{k=1}^{N-1} (b_k - a_{k+1}) \geq b_N - a_1 > b - a = |I|.$$

From the inequality (12) we conclude that $|I| < m^*(I) + \epsilon$, and since ϵ was arbitrary we have proved $m^*(I) = |I|$.

Now we consider the cases of bounded, but not closed, intervals $(a, b), (a, b], [a, b)$. If I is such an interval and $\overline{I} = [a, b]$ its closure, then since m^* is an outer measure we have by monotonicity $m^*(I) \leq m^*(\overline{I}) = |I|$. On the other hand, for all $\epsilon > 0$ sufficiently small we have $I_\epsilon := [a + \epsilon, b - \epsilon] \subset I$, so by monotonicity again $m^*(I) \geq m^*(I_\epsilon) = |I| - 2\epsilon$ and letting $\epsilon \to 0$ we get $m^*(I) \geq |I|$.

Finally, the result is immediate in the case of unbounded intervals, since any unbounded interval contains arbitrarily large bounded intervals. □

Theorem 4.2. *Every Borel set $E \in \mathscr{B}_{\mathbb{R}}$ is m^*-measurable.*

Proof. By the Caratheodory extension theorem, the collection of m^*-measurable sets is a σ-algebra, so by Propositions 1.10 and 1.7, it suffices to show that the open rays $(a, +\infty)$ are m^*-measurable. Fix $a \in \mathbb{R}$ and an arbitrary set $A \subset \mathbb{R}$. We must prove

$$m^*(A) \geq m^*(A \cap (a, +\infty)) + m^*(A \cap (-\infty, a]).$$

To simplify the notation put $A_1 = A \cap (a, +\infty)$, $A_2 = A \cap (-\infty, a]$. Let (I_n) be a cover of A by open intervals. For each n let $I'_n = I_n \cap (a, +\infty)$ and $I''_n = I_n \cap (-\infty, a]$. The families $(I'_n), (I''_n)$ are intervals (not necessarily open) that cover A_1, A_2 respectively. Now

$$\sum_{n=1}^{\infty} |I_n| = \sum_{n=1}^{\infty} |I'_n| + \sum_{n=1}^{\infty} |I''_n| \tag{13}$$

$$= \sum_{n=1}^{\infty} m^*(I'_n) + \sum_{n=1}^{\infty} m^*(I''_n) \qquad \text{(by Theorem 4.1)} \tag{14}$$

$$\geq m^*(\bigcup_{n=1}^{\infty} I'_n) + m^*(\bigcup_{n=1}^{\infty} I''_n) \qquad \text{(by subadditivity)} \tag{15}$$

$$\geq m^*(A_1) + m^*(A_2) \qquad \text{(by monotonicity).} \tag{16}$$

Since this inequality holds for all coverings of A by open intervals, taking the infimum we conclude $m^*(A) \geq m^*(A_1) + m^*(A_2)$. □

Definition 4.3. A set $E \subset \mathbb{R}$ is called *Lebesgue measurable* if

$$m^*(A) = m^*(A \cap E) + m^*(A \cap E^c) \tag{17}$$

for all $A \subset \mathbb{R}$. The restriction of m^* to the Lebesgue measurable sets is called *Lebesgue measure*, denoted m. ◁

By Theorem 3.3, m is a measure; by Theorem 4.2, every Borel set is Lebesgue measurable, and by Theorem 4.1 the Lebesgue measure of an interval is its length. It should also be evident by now that m is σ-finite. So, we have arrived at the promised extension of the length function on intervals to a measure. (It turns out that this extension is unique, but we do not prove this here.)

Next we prove that m has the desired invariance properties. Given $E \subset \mathbb{R}$, $x \in \mathbb{R}$, and $t > 0$, let

$$E+x = \{y \in \mathbb{R} : y-x \in E\}, \quad -E = \{y \in \mathbb{R} : -y \in E\}, \quad \text{and } tE = \{y \in \mathbb{R} : y/t \in E\}.$$

It is evident that $\mu^*(E+x) = \mu^*(E)$, $\mu^*(-E) = \mu^*(E)$ and, $\mu^*(tE) = t\mu^*(E)$ since, if I is an interval, then $|I+x| = |I|$, $|-I| = |I|$ and $|tI| = t|I|$. In particular, if both E and $E+x$ are Lebesgue measurable, then $m(E+x) = m(E)$.

Theorem 4.4. *If $E \subset \mathbb{R}$ is Lebesgue measurable, $x \in \mathbb{R}$, and $t > 0$, then the sets $E+x$, $-E$, and tE are Lebesgue measurable. Moreover $m(E+x) = m(E)$, $m(-E) = m(E)$, and $m(tE) = tm(E)$.*

Proof. We give the proof for $E+x$; the others are similar and left as exercises. Accordingly, suppose E is measurable. To prove $E+x$ is measurable, let $A \subset \mathbb{R}$ be given and observe that $A \cap (E+x) = ((A-x) \cap E) + x$ and $A \cap (E+x)^c = ((A-x) \cap E^c) + x$. Thus,

$$\begin{aligned} m^*(A) &= m^*(A-x) && (18)\\ &= m^*((A-x) \cap E) + m^*((A-x) \cap E^c) && (19)\\ &= m^*((A-x) \cap E + x) + m^*((A-x) \cap E^c + x) && (20)\\ &= m^*(A \cap (E+x)) + m^*(A \cap (E+x)^c), && (21)\end{aligned}$$

where measurability of E is used in the second equality. Hence $E+x$ is Lebesgue measurable and $m(E+x) = m(E)$. □

The condition (17) does not make clear which subsets of $\mathbb{R}$ are Lebesgue measurable. We next prove two fundamental approximation theorems, which say that 1) if we are willing to ignore sets of measure zero, then every Lebesgue measurable set is a G_δ or F_σ, and 2) if we are willing to ignore sets of measure ϵ, then every set of finite Lebesgue measure is a union of intervals. (Recall that a set in a topological space is called a G_δ-set if it is a countable intersection of open sets, and an F_σ-set if it is a countable union of closed sets.)

Theorem 4.5. *Let $E \subset \mathbb{R}$. The following are equivalent:*

(a) E is Lebesgue measurable.
(b) For every $\epsilon > 0$, there is an open set $U \supset E$ such that $m^*(U \setminus E) < \epsilon$.
(c) For every $\epsilon > 0$, there is a closed set $F \subset E$ such that $m^*(E \setminus F) < \epsilon$.
(d) There is a G_δ set G such that $E \subset G$ and $m^*(G \setminus E) = 0$.
(e) There is an F_σ set F such that $E \supset F$ and $m^*(E \setminus F) = 0$.

Proof. To prove (a) implies (b) let E a (Lebesgue) measurable set and $\epsilon > 0$ be given. Further, suppose for the moment that $m(E) < \infty$. There is a covering of E by open intervals I_n such that $\sum_{n=1}^{\infty} |I_n| < m(E) + \epsilon$. Put $U = \bigcup_{n=1}^{\infty} I_n$. By subadditivity of m,

$$m(U) \leq \sum_{n=1}^{\infty} m(I_n) = \sum_{n=1}^{\infty} |I_n| < m(E) + \epsilon.$$

Since $U \supset E$ and $m(E) < \infty$ (and both U and E are Lebesgue measurable), we conclude from Theorem 2.4 that $m^*(U \setminus E) = m(U \setminus E) = m(U) - m(E) < \epsilon$.

To remove the finiteness assumption on E, we apply the $\epsilon/2^n$ trick: for each $n \in \mathbb{Z}$ let $E_n = E \cap (n, n+1]$. The E_n are disjoint measurable sets whose union is E, and $m(E_n) < \infty$ for all n. For each n, by the first part of the proof we can pick an open set U_n so that $m(U_n \setminus E_n) < \epsilon/2^{|n|}$. Let U be the union of the U_n; then U is open and $U \setminus E \subset \bigcup_{n=1}^{\infty}(U_n \setminus E_n)$. From the subadditivity of m, we get $m^*(U \setminus E) = m(U \setminus E) < \sum_{n\in\mathbb{Z}} \epsilon 2^{-|n|} = 3\epsilon$.

To prove that (b) implies (d), let $E \subset \mathbb{R}$ be given and for each $n \geq 1$ choose (using (b)) an open set $U_n \supset E$ such that $m^*(U_n \setminus E) < \frac{1}{n}$. Put $G = \bigcap_{n=1}^{\infty} U_n$; then G is a G_δ containing E, and $G \setminus E \subset U_n \setminus E$ for every n. By monotonicity of m^* we see that $m^*(G \setminus E) < \frac{1}{n}$ for every n and thus $m^*(G \setminus E) = 0$. (Note that in this portion of the proof we cannot (and do not!) assume E is measurable.)

To prove (d) implies (a), suppose G is a G_δ set such that $E \subset G$ and $\mu^*(G \setminus E) = 0$. Since G is a G_δ, it is a Borel set and hence Lebesgue measurable by Theorem 4.2. By Lemma 3.7, every m^*-null set is Lebesgue measurable, so $G \setminus E$, and therefore also $E = G \setminus (G \setminus E)$, is Lebesgue measurable.

To prove that (a) implies (c), suppose E is Lebesgue measurable and let $\epsilon > 0$ be given. Thus E^c is Lebesgue measurable and, by the already established implication (a) implies (b), there is an open set U such that $E^c \subset U$ and $m(U \setminus E^c) < \epsilon$. Since $U \setminus E^c = U \cap E = E \setminus U^c$, it follows that $\mu(E \setminus U^c) < \epsilon$. Observing that U^c is closed completes the proof.

Now suppose $E \subset \mathbb{R}$ and (c) holds. Choose a sequence of closed sets (F_n) such that $F_n \subset E$ and $\mu^*(E \setminus F_n) < \frac{1}{n}$. The set $F = \cup_{j=1}^{\infty} F_j$ is an F_σ and, by monotonicity, for each n we have $\mu^*(E \setminus F) \leq \mu^*(E \setminus F_n) < \frac{1}{n}$. Hence $\mu^*(E \setminus F) = 0$. Thus (c) implies (e).

Finally, if (e) holds, then $E = F \cup (E \setminus F)$ for some closed set $F \subset E$ with $\mu^*(E \setminus F) = 0$. Thus, E is the union of a closed (and hence Lebesgue) set and a set of outer measure zero (which is thus Lebesgue). Since the Lebesgue sets are closed under union, E is Lebesgue and the proof is complete. □

Remark 4.6. The conditions in the theorem are *regularity* conditions which link the topology on $\mathbb{R}$ to Lebesgue measure. This theme leads to the notion of a regular Borel measure, discussed in further detail below. ◇

Theorem 4.7. *If E is Lebesgue measurable and $m(E) < \infty$, then for each $\epsilon > 0$ there exists a set A which is a finite union of open intervals such that $m(E \Delta A) < \epsilon$.*

Proof. Let (I_n) be a covering of E by open intervals such that

$$\sum_{n=1}^{\infty} |I_n| < m(E) + \epsilon/2. \tag{22}$$

Since the sum is finite there exists an integer N so that

$$\sum_{n=N+1}^{\infty} |I_n| < \epsilon/2. \tag{23}$$

Let $U = \bigcup_{n=1}^{\infty} I_n$ and $A = \bigcup_{n=1}^{N} I_n$. Then $A \setminus E \subset U \setminus E$, so $m(A \setminus E) \leq m(U) - m(E) < \epsilon/2$ by (22). Similarly $E \setminus A \subset U \setminus A \subset \bigcup_{n=N+1}^{\infty} I_n$, so $m(E \setminus A) < \epsilon/2$ by (23). Therefore $m(E \Delta A) < \epsilon$. □

Thus, while the "typical" measurable set can be quite complicated in the set-theoretic sense (i.e. in terms of the Borel hierarchy), for most questions in analysis this complexity is irrelevant. In fact, Theorem 4.7 is the precise expression of a useful heuristic:

Littlewood's First Principle of Analysis: *Every measurable set $E \subset \mathbb{R}$ with $m(E) < \infty$ is almost a finite union of intervals.*

Definition 4.8. Let X be a topological space. A *neighborhood* U of a point $x \in X$ is an open set such that $x \in U$.

A topological space X is *locally compact* if for each $x \in X$ there is a neighborhood U_x of x and a compact set C_x such that $x \in U_x \subset C_x$.

A topological space is *Hausdorff* if given $x, y \in X$ with $x \neq y$, there exists neighborhoods U and V of x and y respectively such that $U \cap V = \emptyset$. (Distinct points can be separated by open sets.)

A *Borel measure* is a measure on the Borel σ-algebra of a locally compact Hausdorff space.

A Borel measure μ is *outer regular* if, for all $E \in \Sigma$,

$$\mu(E) = \inf\{\mu(U) : U \supset E \text{ and } U \text{ is open}\}$$

and is *inner regular* if

$$\mu(E) = \sup\{\mu(K) : K \subset E \text{ and } K \text{ is compact}\}.$$

Finally μ is *regular* if it is both inner and outer regular. ◁

Theorem 4.9. *If $E \subset \mathbb{R}$ is Lebesgue measurable, then*

$$\begin{aligned} m(E) &= \inf\{m(U) : U \supset E \textit{ and } U \textit{ is open}\} \\ &= \sup\{m(K) : K \subset E \textit{ and } K \textit{ is compact}\} \end{aligned}$$

That is, m is a regular *Borel measure.*

Proof. The first equality follows from monotonicity in the case $m(E) = +\infty$, and from Theorem 4.5(b) (together with the additivity of m) in the case $m(E) < \infty$.

For the second equality, let $\nu(E)$ be the value of the supremum on the right-hand side. By monotonicity $m(E) \geq \nu(E)$. For the reverse inequality, first assume $m(E) < \infty$ and let $\epsilon > 0$. By Theorem 4.5(c), there is a closed subset $F \subset E$ with $m(E \setminus F) < \epsilon/2$. Since $m(E) < \infty$, we have by additivity $m(E) < m(F) + \epsilon/2$, so $m(F) > m(E) - \epsilon/2$. However this F need not be compact. To fix this potential shortcoming, for each $n \geq 1$ let $K_n = F \cap [-n, n]$. Then the K_n are an increasing sequence of compact sets whose union is F. By monotone convergence for sets (Theorem 2.4(c)), there is an n so that $m(K_n) > m(F) - \epsilon/2$. It follows that $m(K_n) > m(E) - \epsilon$, and thus $\nu(E) \geq m(E)$. The case $m(E) = +\infty$ is left as an exercise. □

Remark 4.10. In the preceding theorem, for any set $E \subset \mathbb{R}$ (not necessarily measurable) the infimum of $m(U) = m^*(U)$ over open sets $U \supset E$ defines the Lebesgue outer measure of E. One can also define the *Lebesgue inner measure* of a set $E \subset \mathbb{R}$ as

$$m_*(E) = \sup\{m^*(K) : K \subset E \text{ and } K \text{ is compact}\}.$$

By monotonicity we have $m_*(E) \leq m^*(E)$ for all $E \subset \mathbb{R}$. One can then prove that if $E \subset \mathbb{R}$ and $m^*(E) < \infty$, then E is Lebesgue measurable if and only if $m^*(E) = m_*(E)$, in which case this quantity is equal to $m(E)$.(Some care must be taken in trying to use this definition if $m^*(E) = +\infty$. Why?) ◇

Example 4.11. [The Cantor set] Recall the usual construction of the "middle thirds" Cantor set. Let E_0 denote the unit interval $[0, 1]$. Obtain E_1 from E_0 by deleting the middle third (open) subinterval of E_0, so $E_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Continue inductively; at the n^{th} step delete the middle thirds of all the intervals present at that step. So, E_n is a union of 2^n closed intervals of length 3^{-n}. The *Cantor set* is defined as the intersection $C = \bigcap_{n=0}^{\infty} E_n$. It is well-known (though not obvious) that C is uncountable; we do not prove this fact here. It is clear that C is a closed set (hence Borel) but contains no intervals, since if J is an interval of length ℓ and n is chosen so that $3^{-n} < \ell$, then $J \not\subset E_n$ and thus $J \not\subset C$. The Lebesgue measure of E_n is $(2/3)^n$, which goes to 0 as $n \to \infty$, and thus by monotonicity $m(C) = 0$. So, C is an example of an uncountable, closed set of measure 0. Another way to see that C has measure zero, is to note that at the n^{th} stage ($n \geq 1$) we have deleted a collection of 2^{n-1} disjoint open intervals, each of length 3^{-n}. Thus the Lebesgue measure of $[0, 1] \setminus C$ is

$$\sum_{n=1}^{\infty} 2^{n-1} 3^{-n} = \frac{1}{2} \frac{\frac{2}{3}}{1 - \frac{2}{3}} = 1.$$

Thus $m(C) = 0$. △

Example 4.12. [Fat Cantor sets] Let $\{q_n\}_{n=1}^{\infty}$ be an enumeration of the rationals in $[0, 1]$. Fix a number $0 < \epsilon < 1$. For each integer $n \geq 1$, let U_n be the open interval centered at q_n of length $\epsilon/2$. Let

$$U = \bigcup_{n=1}^{\infty} U_n \quad \text{and} \quad E = [0,1] \setminus U.$$

We estimate $m(U)$ using the union bound:

$$\begin{aligned} m(U) &= m\left(\bigcup_{n=1}^{\infty} U_n\right) && (24)\\ &\leq \sum_{n=1}^{\infty} m(U_n) && (25)\\ &= \sum_{n=1}^{\infty} \frac{\epsilon}{2^n} && (26)\\ &= \epsilon. && (27)\end{aligned}$$

It follows that $m(U \cap [0,1]) \leq m(U) < \epsilon$, and we can then estimate $m(E)$ as

$$m(E) = m([0,1] \setminus U) = 1 - m(U \cap [0,1]) > 1 - \epsilon.$$

We thus have the following facts about the set $E \subset [0,1]$:

- E is closed,
- E contains no intervals (since any interval would contain a rational), and
- $1 - \epsilon < m(E) < 1$.

From this one can prove that E is a Cantor-like set (e.g. with slightly more care one can arrange that E has no isolated points) but unlike the classical Cantor set, E has positive Lebesgue measure. In particular we have constructed a closed set with positive measure but which contains no intervals. △

Example 4.13. [A Lebesgue non-measurable set] Define an equivalence relation on $\mathbb{R}$ by declaring $x \sim y$ if and only if $x - y \in \mathbb{Q}$. This relation partitions $\mathbb{R}$ into disjoint equivalence classes whose union is $\mathbb{R}$. In particular, for each $x \in \mathbb{R}$ its equivalence class is the set $\{x + q : q \in \mathbb{Q}\}$. Since $\mathbb{Q}$ is dense in $\mathbb{R}$, each equivalence class C contains an element of the closed interval $[0,1]$. By the axiom of choice, there is a set $E \subset [0,1]$ that contains exactly one member x_C from each class. We claim the set E is not Lebesgue measurable.

To prove the claim, let $y \in [0,1]$. Then y belongs to some equivalence class C, and hence y differs from x_C by some rational number in the interval $[-1,1]$. Hence

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

On the other hand, since $E \subset [0,1]$ and $|q| \leq 1$ we see that

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

Finally, by the construction of E the sets $E+p$ and $E+q$ are disjoint if p,q are distinct rationals. So if E were measurable, the the sets $E+q$ would be also, and we would have by the countable additivity and monotonicity of m

$$1 \leq \sum_{q\in[-1,1]\cap\mathbb{Q}} m(E+q) \leq 3$$

But by translation invariance, all of the $m(E+q)$ must be equal, which is a contradiction. △

5. Measurable functions

We will state and prove a few "categorical" properties of measurable functions between general measurable spaces, however in these notes we will mostly be interested in functions from a measurable space taking values in the extended positive axis $[0, +\infty]$, the real line $\mathbb{R}$, or the complex numbers $\mathbb{C}$.

Definition 5.1. Let $(X, \mathscr{M})$ and $(Y, \mathscr{N})$ be measurable spaces. A function $f : X \to Y$ is called *measurable* (or $(\mathscr{M}, \mathscr{N})$ measurable) if $f^{-1}(E) \in \mathscr{M}$ for all $E \in \mathscr{N}$. A function $f : X \to \mathbb{R}$ is *measurable* if it is $(\mathscr{M}, \mathscr{B}_{\mathbb{R}})$ measurable unless indicated otherwise. Likewise, a function $f : X \to \mathbb{C}$ is measurable if it is $(\mathscr{M}, \mathscr{B}_{\mathbb{C}})$ measurable (where $\mathbb{C}$ is identified with $\mathbb{R}^2$ topologically). ◁

It is immediate from the definition that if $(X, \mathscr{M}), (Y, \mathscr{N}), (Z, \mathscr{O})$ are measurable spaces and $f : X \to Y, g : Y \to Z$ are measurable functions, then the composition $g \circ f : X \to Z$ is measurable. The following is a routine application of Proposition 1.7.

Proposition 5.2. *Suppose $(X, \mathscr{M})$ and $(Y, \mathscr{N})$ are measurable spaces and the collection of sets $\mathscr{E} \subset 2^Y$ generates $\mathscr{N}$ as a σ-algebra. Then $f : X \to Y$ is measurable if and only if $f^{-1}(E) \in \mathscr{M}$ for all $E \in \mathscr{E}$.* †

Proof. The collection of sets $\{E \subset Y : f^{-1}(E) \in \mathscr{N}\}$ is a σ-algebra which contains $\mathscr{E}$, it therefore contains $\mathscr{N}$. The converse implication is trivial.

□

Corollary 5.3. *Let X, Y be topological spaces equipped with their Borel σ-algebras $\mathscr{B}_X, \mathscr{B}_Y$ respectively. Every continuous function $f : X \to Y$ is $(\mathscr{B}_X, \mathscr{B}_Y)$-measurable (or* Borel measurable *for short). In particular, if $f : X \to \mathbb{F}$ is continuous and X is given its Borel σ-algebra, then f is measurable, where $\mathbb{F}$ is either $\mathbb{R}$ or $\mathbb{C}$,* †

Proof. Since the open sets $U \subset Y$ generate $\mathscr{B}_Y$ and $f^{-1}(U)$ is open (hence in $\mathscr{B}_X$) by hypothesis, this corollary is an immediate consequence of Proposition 5.2. □

Definition 5.4. Let $\mathbb{F} = \mathbb{R}$ or $\mathbb{C}$. A function $f : \mathbb{R} \to \mathbb{F}$ is called *Lebesgue measurable* (resp. *Borel measurable*) if it is $(\mathscr{L}, \mathscr{B}_{\mathbb{F}})$ (resp. $(\mathscr{B}_{\mathbb{R}}, \mathscr{B}_{\mathbb{F}})$) measurable. Here $\mathscr{L}$ is the Lebesgue σ-algebra. ◁

Remark 5.5. Note that since $\mathscr{B}_{\mathbb{R}} \subset \mathscr{L}$, being Lebesgue measurable is a *weaker* condition than being Borel measurable. If f is *Borel* measurable, then $f \circ g$ is Borel or Lebesgue measurable if g is. However if f is only Lebesgue measurable, then $f \circ g$ need not be Lebesgue measurable, even if g is continuous. (The difficulty is that we have no control over $g^{-1}(E)$ when E is a Lebesgue set.) ⋄

It will sometimes be convenient to consider functions that are allowed to take the values $\pm\infty$.

Definition 5.6. [The extended real line] Let $\overline{\mathbb{R}}$ denote the set of real numbers together with the symbols $\pm\infty$. The arithmetic operations $+$ and $\cdot$ can be (partially) extended

to $\overline{\mathbb{R}}$ by declaring
$$\pm\infty + x = x + \pm\infty = \pm\infty$$
for all $x \in \mathbb{R}$,
$$+\infty \cdot x = x \cdot +\infty = +\infty$$
for all nonzero $x \in (0, +\infty)$ (and similar rules for the other choices of signs),
$$0 \cdot \pm\infty = \pm\infty \cdot 0 = 0,$$
The order $<$ is extended to $\overline{\mathbb{R}}$ by declaring
$$-\infty < x < +\infty$$
for all $x \in \mathbb{R}$. ◁

The symbol $+\infty + (-\infty)$ is not defined, so some care must be taken in working out the rules of arithmetic in $\overline{\mathbb{R}}$. Typically we will be performing addition only when all values are finite, or when all values are nonnegative (that is for $x \in [0, +\infty]$). In these cases most of the familiar rules of arithmetic hold (for example the commutative, associative, and distributive laws), and the inequality $\leq$ is preserved by multiplying both sides by the same quantity. However cancellation laws are *not* in general valid when infinite quantities are permitted; in particular from $x \cdot +\infty = y \cdot +\infty$ or $x + +\infty = y + +\infty$ one *cannot* conclude that $x = y$.

The order property allows us to extend the concepts of supremum and infimum, by defining the supremum of a set that is unbounded from above, or set containing $+\infty$, to be $+\infty$; similarly for inf and $-\infty$. This also means every sum $\sum_n x_n$ with $x_n \in [0, +\infty]$ can be meaningfully assigned a value in $[0, +\infty]$, namely the supremum of the finite partial sums $\sum_{n \in F} x_n$.

A set $U \subset \overline{\mathbb{R}}$ will be called *open* if either $U \subset \mathbb{R}$ and U is open in the usual sense, or U is the union of an open subset of $\mathbb{R}$ with a set of the form $(a, +\infty]$ or $[-\infty, b)$ (or both). The collection of these open sets is a topology on the set $\overline{\mathbb{R}}$.

Definition 5.7. [Extended Borel σ-algebra] The *extended Borel σ-algebra over $\overline{\mathbb{R}}$* is the σ-algebra over generated by the Borel sets of $\mathbb{R}$ together with the sets $(a, +\infty]$ for $a > 0$. ◁

Definition 5.8. [Measurable function] Let $(X, \mathscr{M})$ be a measurable space. A function $f : X \to \overline{\mathbb{R}}$ is called *measurable* if it is $(\mathscr{M}, \mathscr{B}_{\overline{\mathbb{R}}})$ measurable; that is, if $f^{-1}(E) \in \mathscr{M}$ for every open set $U \subset \overline{\mathbb{R}}$. ◁

In particular we have the following criteria for measurability which will be used repeatedly:

Corollary 5.9 (Equivalent criteria for measurability)**.** *Let $(X, \mathscr{M})$ be a measurable space.*

(a) A function $f : X \to \overline{\mathbb{R}}$ is measurable if and only if the sets
$$f^{-1}((t, +\infty]) = \{x : f(x) > t\}$$
are measurable for all $t \in \mathbb{R}$; and

(b) A function $f : X \to \mathbb{R}$ or $f : X \to (-\infty, \infty]$ is measurable if and only if $f^{-1}(E) \in \mathscr{M}$ for all $E \in \mathscr{E}$, where $\mathscr{E}$ is any of the collections of sets $\mathscr{E}_j$ appearing in Proposition 1.10.

A function $f : X \to \mathbb{C}$ is measurable if and only if $f^{-1}((a, b) \times (c, d))$ is measurable for every $a, b, c, d \in \mathbb{R}$. (Here $(a, b) \times (c, d)$ is identified with the open box $\{z \in \mathbb{C} : a < Re(z) < b,\ cIm(z) < d\}$.) †

Example 5.10. [Examples of measurable functions]

(a) An indicator function $\mathbf{1}_E$ is measurable if and only if E is measurable. Indeed, the set $\{x : \mathbf{1}_E(x) > t\}$ is either empty, E, or all of X, in the cases $t \geq 1$, $0 \leq t < 1$, or $t < 0$, respectively.
(b) The next series of propositions will show that measurability is preserved by most of the familiar operations of analysis, including sums, products, sups, infs, and limits (provided one is careful about arithmetic of infinities).
(c) Corollary 5.19 below will show that examples (a) and (b) above in fact generate all the examples in the case of $\overline{\mathbb{R}}$ or $\mathbb{C}$ valued functions. That is, every measurable function is a pointwise limit of linear combinations of measurable indicator functions.

△

Proposition 5.11. *Let $(X, \mathscr{M})$ be a measurable space. A function $f : X \to \mathbb{C}$ is measurable if and only if Ref and Imf are measurable.* †

Theorem 5.12. *Let $(X, \mathscr{M})$ be a measurable space.*

(a) If $f, g : X \to \mathbb{C}$ be measurable functions, and $c \in \mathbb{C}$. Then cf, $f + g$, and fg are measurable.
(b) If $f, g : X \to \overline{\mathbb{R}}$ are measurable and, for each x, $\{f(x), g(x)\} \neq \{\pm\infty\}$, then $f + g$ is measurable.
(c) If $f, g : X \to \overline{\mathbb{R}}$ are measurable, then so is fg.

Proof. To prove (b), supppose $f, g : X \to [-\infty, \infty]$ are measurable and $f + g$ is defined. To prove $f + g$ is measurable, observe that, for $t \in \mathbb{R}$ given,

$$\{x \in X : f(x) + g(x) > t\} = \bigcup_{q \in \mathbb{Q}} \{x : f(x) > q\} \cap \{x : g(x) > t - q\},$$

where the fact that if $g(x) < \infty$ and $f(x) > t - g(x)$, then there is a $q \in \mathbb{Q}$ such that $f(x) > q > t - g(x)$ was used to obtain the reverse inclusion in the last equality. (If $g(x) = \infty$, then x is evidently in both sets.) Since all the sets in the last line are measurable, the intersection is finite and the union countable, it follows that $f + g$ is measurable by Corollary 5.9.

Assuming $f, g : X \to [0, \infty]$ are measurable, a proof that fg is measurable can be modeled after the proof for $f + g$. The details are left as an exercise. Likewise, it is an exercise to show that if $f : X \to [-\infty, \infty]$ is measurable, then so are $f^+(x) = \max\{f(x), 0\}$ and $f^-(x) = -\min\{f(x), 0\}$. Of course $f = f^+ - f^-$. With this notation,

$$fg = (f^+g^+ + f^-g^-) + (-f^-g^+ - f^+g^-) = F + G.$$

Because $f^\pm, g^\pm$ take values in $[0,\infty]$ and are measurable all the products $f^\pm g^\pm$ are measurable. Hence, using (b) several times, $F+G$ is measurable (verify that $F+G$ is defined).

The proof of (a) is straightforward using parts (b) and (c) and Proposition 5.11. □

Proposition 5.13. *Let (f_n) be a sequence of $\overline{\mathbb{R}}$-valued measurable functions.*

(a) The functions

$$\sup f_n, \quad \inf f_n, \quad \limsup_{n\to\infty} f_n, \quad \liminf_{n\to\infty} f_n$$

are measurable;

(b) The set on which (f_n) converges is a measurable set; and

(c) If (f_n) converges to f pointwise, then f is measurable.

†

Proof. Since $\inf f_n = -\sup(-f_n)$, it suffices to prove the proposition for sup and lim sup. Let $f(x) = \sup_n f_n(x)$. Then for any real number t, we have $f(x) > t$ if and only if $f_n(x) > t$ for some n. Thus

$$\{x : f(x) > t\} = \bigcup_{n=1}^{\infty} \{x : f_n(x) > t\}.$$

It follows that f is measurable. Likewise $\inf f_n$ is measurable. Consequently, $g_N = \sup_{n\ge N} f_n$ is measurable for each positive integer N and hence $\limsup f_n = \inf g_N$ is also measurable.

Finally, if (f_n) converges to f, then $f = \limsup f_n = \liminf f_n$ is measurable. □

In the previous proposition it is of course essential that the supremum is taken only over a *countable* set of measurable functions; the supremum of an uncountable collection of measurable functions need not be measurable. (Exercise: give a counterexample.)

One simple but important application of this proposition is that if f, g are $\overline{\mathbb{R}}$-valued measurable functions, then $f \wedge g := \min(f,g)$ and $f \vee g := max(f,g)$ are measurable; in particular $f^+ := \max(f,0)$ and $f^- := -\min(f,0)$ are measurable if f is. It also follows that $|f| := f^+ + f^-$ is measurable when f is. Together with Proposition 5.11, this shows that every $\mathbb{C}$ valued measurable function f is a linear combination of four unsigned measurable functions (the positive and negative parts of the real and imaginary parts of f). Thus many statements about general measurable functions can be reduced to the unsigned case.

In discussing convergence of functions on a measure space $(X, \mathscr{M}, \mu)$, we say that $f_n \to f$ *almost everywhere* (or μ-almost everywhere), abbreviated a.e. or μ-a.e., if $f_n \to f$ on a set of full measure. (That is, there is a set $E \in \mathscr{M}$ with $\mu(E^c) = 0$ and $f_n(x) \to f(x)$ for all $x \in E$.)

The next two propositions show that 1) provided that μ is a *complete* measure, a.e. limits of measurable functions are measurable, and 2) one is not likely to make any serious errors in forgetting the completeness requirement.

Proposition 5.14. *Suppose $(X, \mathscr{M}, \mu)$ is a complete measure space and $(Y, \mathscr{N})$ is a measurable space.*

(a) Suppose $f, g : X \to Y$. If f is measurable and $g = f$ a.e., then g is measurable.
(b) If $f_n : X \to \overline{\mathbb{R}}$ are measurable functions and $f_n \to f$ a.e., then f is measurable. The same conclusion holds if $\overline{\mathbb{R}}$ is replaced by $\mathbb{C}$.

†

Proposition 5.15. *Let $(X, \mathscr{M}, \mu)$ be a measure space and $(X, \overline{\mathscr{M}}, \overline{\mu})$ its completion. If $f : X \to \overline{\mathbb{R}}$ is a $\overline{\mathscr{M}}$-measurable function, then there is an $\mathscr{M}$-measurable function g such that $f = g$ $\overline{\mu}$-a.e.*

In particular, if (f_n) is a sequence of $\mathscr{M}$ measurable functions, $f_n : X \to \overline{\mathbb{R}}$, which converges a.e. to a function f, then there is a $\mathscr{M}$ measurable function g such that (f_n) converges a.e. to g. Again, the same conclusion holds with $\overline{\mathbb{R}}$ replaced by $\mathbb{C}$. †

Since our primary interest is in Lebesgue measure, which is already complete, we will not prove these propositions.

Definition 5.16. [Unsigned simple function] A function f on a set X is *unsigned* if its codomain is a subset of $[0, \infty]$. An unsigned function $s : X \to [0, +\infty]$ is called *simple* if it can be expressed as a finite sum

$$s = \sum_{j=0}^{n} c_j \mathbf{1}_{E_j} \tag{28}$$

where each E_j is a subset of X and each $c_j \in [0, +\infty]$. ◁

A *partition* P of the set X is, for some $n \in \mathbb{N}$, a set $P = \{E_0, \dots, E_n\}$ of pairwise disjoint subsets of X whose union is X. If each E_j is measurable, then P is a *measurable partition*.

Proposition 5.17. *A function $s : X \to [0, +\infty]$ is simple if and only if its range is a finite set if and only if there exists a positive integer m, a partition $P = \{F_0, \dots, F_m\}$, and $d_0, \dots, d_m$ in $[0, \infty]$ such that*

$$s = \sum_{k=0}^{m} d_k \mathbf{1}_{F_k}.$$

Thus, a function s is simple if and only if it has a representation as in Equation (28) where $\{E_0, \dots, E_n\}$ forms a partition of X.

Finally, s is a measurable simple function if and only if it has a representation as in Equation (28) where $\{E_0, \dots, E_n\}$ is a measurable partition. †

Theorem 5.18 (The Ziggurat approximation)**.** *Let $(X, \mathscr{M})$ be a measurable space. If $f : X \to [0, +\infty]$ is an unsigned measurable function, then there exists a increasing sequence of unsigned, measurable simple functions $s_n : X \to [0, +\infty)$ such that $s_n \to f$ pointwise increasing on X. If f is bounded, the sequence can be chosen to converge uniformly.*

Proof. For positive integers n and integers $0 \le k < n2^n$, let $E_{n,k} = \{x : \frac{k}{2^n} < f(x) \le \frac{k+1}{2^n}\}$, let $E_{n,n2^n} = \{x : n < f(x)\}$ and define

$$s_n(x) = \sum_{k=0}^{n2^n} \frac{k}{2^n} \mathbf{1}_{E_{n,k}}. \tag{29}$$

Verify that (s_n) is pointwise increasing with limit f and if f is bounded, then the convergence is uniform. □

It will be helpful to record for future use the round-off procedure used in this proof. Let $f : X \to [0, +\infty]$ be an unsigned function. For any $\epsilon > 0$, if $0 < f(x) < +\infty$ there is a unique integer k such that

$$k\epsilon < f(x) \le (k+1)\epsilon.$$

Define the "rounded down" function $f_\epsilon(x)$ to be $k\epsilon$ when $f(x) \in (0, +\infty)$ and equal to 0 or $+\infty$ when $f(x) = 0$ or $+\infty$ respectively. Similarly we can defined the "rounded up" function f^ϵ to be $(k+1)\epsilon$, 0, or $+\infty$ as appropriate. In particular, we have for all $\epsilon > 0$

$$f_\epsilon \le f \le f^\epsilon,$$

and f_ϵ, f^ϵ are measurable if f is. Moreover the same argument used in the above proof shows that $f_\epsilon, f^\epsilon \to f$ pointwise as $\epsilon \to 0$ (and uniformly, if f is bounded).

Finally, by the remarks following Proposition 5.13, the following is immediate (since its proof reduces to the unsigned case):

Corollary 5.19. *Every $\overline{\mathbb{R}}$- or $\mathbb{C}$-valued measurable function is a pointwise limit of measurable simple functions.* †

*******PROOFREADING DONE UP TO HERE TUE 3/12/2019********

6. Integration of simple functions

We will build up the integration theory for measurable functions in three stages. We first define the integral for unsigned simple functions, then extend it to general unsigned functions, and finally to general ($\overline{\mathbb{R}}$ or $\mathbb{C}$-valued) functions. Throughout this section and the next, we fix a measure space $(X, \mathscr{M}, \mu)$; all functions are defined on this measure space.

Suppose $P = \{E_0, \dots, E_n\}$ is a measurable partition of X and

$$s = \sum_{j=0}^{n} c_j \mathbf{1}_{E_j}. \tag{30}$$

If $\{F_0, \dots, F_m\}$ is another measurable partition and

$$s = \sum_{k=0}^{m} d_k \mathbf{1}_{F_k},$$

then it is an exercise to show

$$\sum_{j=0}^{n} c_n\mu(E_n) = \sum_{k=0}^{m} d_m\mu(F_m).$$

It is now possible to make the following definition.

Definition 6.1. Let $(X, \mathscr{M}, \mu)$ be a measure space and $f = \sum_{n=0}^{N} c_n \mathbf{1}_{E_n}$ an unsigned measurable simple function (in its standard representation). The *integral of* f (with respect to the measure μ) is defined to be

$$\int_X f\, d\mu := \sum_{n=0}^{N} c_n\mu(E_n).$$

◁

One thinks of the graph of the function $c\mathbf{1}_E$ as "rectangle" with height c and "base" E; since μ tells us how to measure the length of E the quantity $c \cdot \mu(E)$ is interpreted as the "area" of the rectangle.Note too that the definition explains the convention $0 \cdot \infty = 0$, since the set on which s is 0 should not contribute to the integral.

Let L^+ denote the set of all unsigned measurable functions on $(X, \mathscr{M})$. We begin by collecting some basic properties of the integrals of simple functions, temporarily referred to as *simple integrals*. When X and μ are understood we drop them from the notation and simply write $\int f$ for $\int_X f\, d\mu$.

Theorem 6.2 (Basic properties of simple integrals)**.** *Let $(X, \mathscr{M}, \mu)$ be a measure space and let $f, g \in L^+$ be simple functions.*

(a) (Homogeneity) If $c \geq 0$, then $\int cf = c\int f$.
(b) (Monotonicity) If $f \leq g$, then $\int f \leq \int g$.
(c) (Finite additivity) $\int f + g = \int f + \int g$.
(d) (Almost everywhere equivalence) If $f(x) = g(x)$ for μ-almost every $x \in X$, then $\int f = \int g$.
(e) (Finiteness) $\int f < +\infty$ if and only if is finite almost everywhere and supported on a set of finite measure.
(f) (Vanishing) $\int f = 0$ if and only if $f = 0$ almost everywhere.

Proof. (a) is trivial; we prove (b) and (c) and leave the rest as (simple!) exercises.

To prove (b), write $f = \sum_{j=0}^{n} c_j \mathbf{1}_{E_j}$ and $g = \sum_{k=0}^{m} d_k \mathbf{1}_{F_k}$ for measurable partitions $P = \{E_0, \dots, E_n\}$ and $Q = \{F_0, \dots, F_m\}$ of X. It follows that $R = \{E_j \cap F_k : 0 \leq j \leq n,\ 0 \leq k \leq m\}$ is a measurable partition of X too and

$$f = \sum_{j,k} c_j \mathbf{1}_{E_j \cap F_k}$$

and similarly for g. From the assumption $f \le g$ we deduce that $c_j \le d_k$ whenever $E_j \cap F_k \ne \varnothing$. Thus, in view of the remarks preceding Definition 6.1,

$$\int f = \sum_{j,k} c_j \mu(E_j \cap F_k) \le \sum_{j,k} d_k \mu(E_j \cap F_k) = \int g.$$

For item (c), since $E_j = \bigcup_{k=0}^m E_j \cap F_k$ for each j and $F_k = \bigcup_{j=0}^n F_k \cap E_j$ for each k, it follows from the finite additivity of μ that

$$\int f + \int g = \sum_{j,k} (c_j + d_k) \mu(E_j \cap F_k).$$

Since $f + g = \sum_{j,k} (c_j + d_k) \mathbf{1}_{E_j \cap F_k}$, by essentially the same reasoning the right hand side is equal to $\int (f+g)$. □

If $f : X \to [0, +\infty]$ is a measurable simple function, then so is $\mathbf{1}_E f$ for any measurable set E. We write $\int_E f \, d\mu := \int \mathbf{1}_E f \, d\mu$.

Proposition 6.3. *Let $(X, \mathscr{M}, \mu)$ be a measure space. If f is an unsigned measurable simple function, then the function*

$$\nu(E) := \int_E f \, d\mu$$

is a measure on $(X, \mathscr{M})$. †

Proof. That ν is nonnegative and $\nu(\varnothing) = 0$ are immediate from the definition. Let $(E_n)_{n=1}^\infty$ be a sequence of disjoint measurable sets and let $E = \bigcup_{n=1}^\infty E_n$. Write f as $\sum_{j=1}^m c_j \mathbf{1}_{F_j}$ with respect to a measurable partition $\{F_1, \dots, F_m\}$ and observe

$$\begin{aligned}
\nu(E) &= \int_E f \, d\mu \\
&= \int_X \mathbf{1}_E f \, d\mu \\
&= \sum_{j=1}^m c_j \mu(E \cap F_j) \\
&= \sum_{j=1}^m c_j \mu(F_j \cap \bigcup_{n=1}^\infty E_n) \\
&= \sum_{n=1}^\infty \sum_{j=1}^m c_j \mu(F_j \cap E_n) \\
&= \sum_{n=1}^\infty \nu(E_n).
\end{aligned}$$

□

7. Integration of unsigned functions

We now extend the definition of the integral to all (not necessarily simple) functions in L^+. First note that if $(X, \mathscr{M}, \mu)$ is a measure space and s is a measurable unsigned simple function, then, by Theorem 6.2(b),

$$\int_X s\,d\mu = \sup\{\int_X t\,d\mu : 0 \le s \le t,\ t \text{ is a measurable unsigned simple function}\}.$$

Hence, the following definition is consistent with that of the integral for unsigned simple functions.

Definition 7.1. Let $(X, \mathscr{M}, \mu)$ be a measure space. For an unsigned measurable function $f : X \to [0, +\infty]$, define

$$\int_X f\,d\mu := \sup_{0 \le s \le f; s \text{ simple}} \int_X s\,d\mu. \tag{31}$$

◁

Theorem 7.2 (Basic properties of unsigned integrals)**.** *Let f, g be unsigned measurable functions.*

(a) (Homogeneity) If $c \ge 0$ then $\int cf = c \int f$.
(b) (Monotonicity) If $f \le g$ then $\int f \le \int g$.
(c) (Almost everywhere equivalence) If $f(x) = g(x)$ for μ-almost every $x \in X$, then $\int f = \int g$.
(d) (Finiteness) If $\int f < +\infty$, then $f(x) < +\infty$ for μ-a.e. x.
(e) (Vanishing) $\int f = 0$ if and only if $f = 0$ almost everywhere.
(f) (Bounded) If f is bounded measurable function and $\mu(X) < \infty$, then $\int f\,d\mu < \infty$.

The integral is also additive; however the proof is surprisingly subtle and will have to wait until we have established the Monotone Convergence Theorem.

Proof of Theorem 7.2. (a,b) As in the simple case homogeneity is trivial. Monotonicity is also evident, since any simple function less than f is also less than g.

(c) Let E be a measurable set with $\mu(E^c) = 0$. If s is a simple function, then by Theorem 6.2(e), $\mathbf{1}_E s$ is a simple function and $\int \mathbf{1}_E s = \int s$. Further, if $0 \le s \le f$, then $\mathbf{1}_E s \le \mathbf{1}_E f$. Hence, using monotonicity and taking suprema over simple functions,

$$\int \mathbf{1}_E f \le \int f = \sup_{0 \le s \le f} \int s = \sup_{0 \le s \le f} \int \mathbf{1}_E s \le \sup_{0 \le t \le \mathbf{1}_E f} \int t = \int \mathbf{1}_E f$$

(d) If $f = +\infty$ on a measurable set E, and $\mu(E) > 0$, then $\int f \ge \int n\mathbf{1}_E = n\mu(E)$ for all n, so $\int f = +\infty$. (A direct proof can be obtained from Markov's inequality below.)

(e) If $f = 0$ a.e. and $s \le f$ is a simple function, then by monotonicity $\int s = 0$ so $\int f = 0$. Conversely, suppose there is a set E of positive measure such that $f(x) > 0$ for all $x \in E$. Let $E_n = \{x \in E : f(x) > \frac{1}{n}\}$. Then $E = \bigcup_{n=1}^{\infty} E_n$, so by the pigeonhole principle $\mu(E_N) > 0$ for some N. But then $\frac{1}{N}\mathbf{1}_{E_N} \le f$, so $\int f \ge \frac{1}{N}\mu(E_N) > 0$.

(f) There is a positive real number C so that $0 \le f(x) \le C$ for $x \in X$. Hence, if $0 \le g \le f$ and g is a measurable, then $0 \le g \le C$. Since $\int g\,d\mu \le C\mu(X)$, the constant $C\mu(X)$ is an upper bound for the supremum defining the integral of f and the conclusion follows.

□

Theorem 7.3 (Monotone Convergence Theorem). *If f_n is a sequence of unsigned measurable functions and f_n increases to f pointwise, then $\int f_n \to \int f$.*

Proof. Since (f_n) converges to f and each f_n is measurable, f is measurable. By monotonicity of the integral, the sequence $(\int f_n)$ is increasing and $\int f_n \le \int f$ for all n. Thus the sequence $(\int f_n)$ converges (perhaps to ∞) and $\lim \int f_n \le \int f$. For the reverse inequality, fix $0 < \epsilon < 1$ and let s be a simple function with $0 \le s \le f$. Consider the sets

$$E_n = \{x : f_n(x) \ge (1-\epsilon)s(x)\}.$$

The E_n form an increasing sequence of measurable sets whose union is X. For all n,

$$\int f_n \ge \int_{E_n} f_n \ge (1-\epsilon)\int_{E_n} s.$$

By Monotone convergence for sets (Theorem 2.4(c)) applied to the measure $\nu(E) = \int_E s$ (Proposition 6.3), we see that

$$\lim \int_{E_n} s = \int_X s.$$

Thus $\lim \int f_n \ge (1-\epsilon)\int s$ for every $0 < \epsilon < 1$ and every simple function $0 \le s \le f$. Taking the supremum over s and then over ϵ, we conclude that $\lim \int f_n \ge \int f$. □

Before going on we mention two frequently used applications of the Monotone Convergence Theorem:

Corollary 7.4. *(1) (Vertical truncation) If f is an unsigned measurable function and $f \wedge n := \min(f, n)$, then $\int f \wedge n \to \int f$.*

(2) (Horizontal truncation) If f is an unsigned measurable function and $(E_n)_{n=1}^{\infty}$ is an increasing sequence of measurable sets whose union is X, then $\int_{E_n} f \to \int f$.

†

Proof. Since $f \wedge n$ and $\mathbf{1}_{E_n} f$ are measurable for all n and increase pointwise to f, these follow from the Monotone Convergence Theorem. □

Theorem 7.5 (Additivity of the unisgned integral). *If f, g are unsigned measurable functions, then $\int f + g = \int f + \int g$.*

Proof. By Theorem 5.18, there exist sequences of unsigned, measurable simple functions f_n, g_n which increase pointwise to f, g respectively. Thus $f_n + g_n$ increases to $f + g$, so by Theorem 6.2(c) and the Monotone Convergence Theorem,

$$\int f + g = \lim \int f_n + g_n = \lim \int f_n + \lim \int g_n = \int f + \int g.$$

□

Corollary 7.6 (Tonelli's theorem for sums and integrals)**.** *If (f_n) is a sequence of unsigned measurable functions, then $\int \sum_{n=1}^\infty f_n = \sum_{n=1}^\infty \int f_n$.* †

Proof. Note that $\sum_{n=1}^\infty f_n$ is measurable. By induction on Theorem 7.5, $\int \sum_{n=1}^N f_n = \sum_{n=1}^N \int f_n$ for all $N \geq 1$. Since $\sum_{n=1}^N f_n$ increases pointwise to $\sum_{n=1}^\infty f_n$, we have by the Monotone Convergence Theorem

$$\int \sum_{n=1}^\infty f_n = \lim_{N\to\infty} \int \sum_{n=1}^N f_n = \lim_{N\to\infty} \sum_{n=1}^N \int f_n = \sum_{n=1}^\infty \int f_n.$$

□

If the monotonicity hypothesis is dropped, we can no longer conclude that $\int f_n \to \int f$ if $f_n \to f$ pointwise (see Examples 7.8 below), however the following weaker result holds.

Theorem 7.7 (Fatou's Theorem)**.** *If f_n is a sequence of unsigned measurable functions, then*

$$\int \liminf f_n \leq \liminf \int f_n.$$

Proof. For each n define functions $g_n(x) := \inf_{m\geq n} f_m(x)$. Then the g_n are unsigned measurable functions increasing pointwise to $\liminf f_n$, and $g_n \leq f_n$ for each n. Thus by the Monotone Convergence Theorem and monotonicity

$$\int \liminf f_n = \lim \int g_n = \liminf \int g_n \leq \liminf \int f_n.$$

□

Example 7.8. [Failure of convergence of integrals] This example highlights three modes of failure of the convergence $\int f_n \to \int f$ for sequences of unsigned measurable functions $f_n : \mathbb{R} \to [0, +\infty]$ and Lebesgue measure. In each case $f_n \to 0$ pointwise, but $\int f_n = 1$ for all n:

(1) (Escape to height infinity) $f_n = n\mathbf{1}_{(0,\frac{1}{n})}$
(2) (Escape to width infinity) $f_n = \frac{1}{2n}\mathbf{1}_{(-n,n)}$
(3) (Escape to support infinity) $f_n = \mathbf{1}_{(n,n+1)}$

Note that in the second example the convergence is even uniform. These examples can be though of as "moving bump" functions–in each case we have a rectangle and can vary the height, width, and position. If we think of f_n as describing a density of mass distributed over the real line, then $\int f_n$ gives the total "mass"; Fatou's theorem says mass cannot be created in the limit, but these examples show mass can be destroyed. △

Proposition 7.9 (Markov's inequality)**.** *If f is an unsigned measurable function, then for all $t > 0$*

$$\mu(\{x : f(x) > t\}) \leq \frac{1}{t} \int f$$

†

Proof. Let $E_t = \{x : f(x) > t\}$. Then by definition, $t\mathbf{1}_{E_t} \leq f$, so $t\mu(E_t) = \int t\mathbf{1}_{E_t} \leq \int f$. □

8. Integration of signed and complex functions

Again we work on a fixed measure space $(X, \mathscr{M}, \mu)$. Suppose $f : X \to \overline{\mathbb{R}}$ is measurable. Split f into its positive and negative parts $f = f^+ - f^-$. If at least one of $\int f^+, \int f^-$ is finite, define the *integral of f*

$$\int f = \int f^+ - \int f^-.$$

If both are finite, we say f is *integrable* (or sometimes *absolutely integrable*). Note that f is integrable if and only if $\int |f| < +\infty$; this is immediate since $|f| = f^+ + f^-$. We write

$$\|f\|_1 := \int_X |f| \, d\mu$$

when f is integrable. In the complex case, from the inequalities

$$\max(|\mathrm{Re} f|, |\mathrm{Im} f|) \leq |f| \leq |\mathrm{Re} f| + |\mathrm{Im} f|$$

it is clear that $f : X \to \mathbb{C}$ is absolutely integrable if and only if $\mathrm{Re} f$ and $\mathrm{Im} f$ are. If f is complex-valued and absolutely integrable (that is, f is measurable and $|f|$ is integrable), we define

$$\int f = \int \mathrm{Re} f + i \int \mathrm{Im} f.$$

We also write $\|f\|_1 := \int_X |f| \, d\mu$ in the complex case.

If $f : X \to \overline{\mathbb{R}}$ is absolutely integrable, then necessarily the set $\{x : |f(x)| = +\infty\}$ has measure 0. We may therefore redefine f to be 0, say, on this set, without affecting the integral of f (by Theorem 7.2(c)). Thus when working with absolutely integrable functions, we can (and from now on, will) always assume that f is finite-valued everywhere.

8.1. **Basic properties of the absolutely convergent integral.** The next few propositions collect some basic properties of the absolutely convergent integral. Let $L^1(X, \mathscr{M}, \mu)$ denote the set of all absolutely integrable $\mathbb{C}$-valued functions on X. (If the measure space is understood, as it is in this section, we just write L^1.)

Theorem 8.1 (Basic properties of L^1 functions)**.** *Let $f, g \in L^1$ and $c \in \mathbb{C}$. Then:*

(a) L^1 is a vector space over $\mathbb{C}$ and the map $f \to \int f$ is a linear functional on it.
(b) $\left|\int f\right| \leq \int |f|$.
(c) $\|cf\|_1 = |c|\|f\|_1$.
(d) $\|f+g\|_1 \leq \|f\|_1 + \|g\|_1$.

Proof. To prove L^1 is a vector space, suppose $f, g \in L^1$ and $c \in \mathbb{C}$. Since f, g are measurable, so is $f+g$, thus $|f+g|$ has an integral and since $|f+g| \leq |f|+|g|$ monotonicity and additivity of the unsigned integral, Theorems 7.2 and 7.5, show that $f+g$ is integrable and hence in L^1. Further,

$$\|f+g\|_1 \leq \|f\|_1 + \|g\|_1$$

and item (d) is proved. Next, $\int |cf| = |c| \int |f| < \infty$ (using homogeneity of the unsigned integral in Theorem 7.2) Hence (c) holds. Moreover, it follows that $cf \in L^1$. Thus L^1 is a vector space.

To prove that $f \to \int f$ is linear, first assume f and g are real-valued and $c \in \mathbb{R}$; the complex case then follows essentially by definition. Checking $c \int f = \int cf$ is straightforward. For additivity, let $h = f+g$; then

$$h^+ - h^- = f^+ + g^+ - f^- - g^-$$

and therefore

$$h^+ + f^- + g^- = h^- + f^+ + g^+.$$

Thus

$$\int h^+ + f^- + g^- = \int h^- + f^+ + g^+$$

which rearranges to $\int h = \int f + \int g$ using the additivity of the unsigned integral, and the finiteness of all the integrals involved. Hence f is linear.

If f is real, then

$$\left|\int f\right| = \left|\int f^+ - \int f^-\right| \leq \left|\int f^+ + \int f^-\right| = \int |f|.$$

Hence (b) holds for real-valued functions. When f is complex, assume $\int f \neq 0$ and let $t = \overline{\operatorname{sgn} \int f}$. Then $|t| = 1$ and $|\int f| = \int tf$. It follows that, using in part (b) for the real-valued function $\operatorname{Re} tf$,

$$\left|\int f\right| = \operatorname{Re} \int tf = \int \operatorname{Re} tf \leq \int |tf| = \int |f|.$$

□

Because of cancellation, it is clear that $\int f = 0$ does not imply $f = 0$ a.e. when f is a signed or complex function. However the conclusion $f = 0$ a.e. can be recovered if we assume the vanishing of *all* the integrals $\int_E f$, over all measurable sets E.

Proposition 8.2. *Let $f \in L^1$. The following are equivalent:*

(a) $f = 0$ almost everywhere,
(b) $\int |f| = 0$,
(c) For every measurable set E, $\int_E f = 0$.

†

Proof. Since $f = 0$ a.e. if and only if $|f| = 0$ a.e., (a) and (b) are equivalent by Theorem 7.2(e). Now assuming (b), if E is measurable then by monotonicity and Theorem 8.1(b)

$$\left| \int_E f \right| \leq \int_E |f| \leq \int |f| = 0,$$

so (c) holds.

Now suppose (c) holds and f is real-valued. Let $E = \{f > 0\}$ and note $f^+ = f\mathbf{1}_E$. Hence g is unsigned and, by assumption $\int g = 0$. Thus, by Theorem 7.2, $g = 0$ a.e. Similarly $f^- = 0$ a.e. and thus f is the difference of two functions which are zero a.e. To complete the proof, write f in terms of its real and imaginary parts. □

Corollary 8.3. *If $f, g \in L^1$ and $f = g$ μ-a.e., then $\int f = \int g$.* †

Proof. Apply Proposition 8.2 to $f - g$. □

The preceding proposition and its corollary say that for the purposes of integration, we are free to alter the definition of functions on sets of measure zero. In particular, if $f : X \to \overline{\mathbb{R}}$ is finite valued almost everywhere, then there is another function g which is finite *everywhere* and equal to f a.e. (Simply define g to be 0 (or any other finite value) on the set E where $f = \pm\infty$.)

Another consequence is that we can introduces an equivalence relation on $L^1(X, \mathscr{M}, \mu)$ by declaring $f \sim g$ if and only if $f = g$ a.e. If $[f]$ denotes the equivalence class of f under this relation, we may define the integral on equivalence classes by declaring $\int [f] := \int f$. Corollary 8.3 shows that this is well-defined. It is straightforward to check that $[cf + g] = [cf] + [g]$ for all $f, g \in L^1$ and scalars c (so that $L^1/\sim$ is a vector space), and that the properties of the integral given in Theorem 8.1 all persist if we work with equivalence classes. The advantage is that now $\int [|f|] = 0$ if and only if $[|f|] = 0$. This means that the quantity $\|[f]\|_1$ is a *norm* on $L^1/\sim$. Henceforth will we agree to impose this relation whenever we talk about L^1, but for simplicity we will drop the $[\cdot]$ notation, and also write just L^1 for $L^1/\sim$. So, when we refer to an L^1 function, it is now understood that we refer to the equivalence class of functions equal to f a.e., but in practice this abuse of terminology should cause no confusion.

Just as the Monotone Convergence Theorem is associated to the unsigned integral, there is a convergence theorem for the absolutely convergent integral.

Theorem 8.4 (Dominated Convergence Theorem)**.** *Suppose* $(f_n)_{n=1}^{\infty}$ *is a sequence from* L^1 *which converges pointwise a.e. to a measurable function* f*. If there exists a function* $g \in L^1$ *such that for every* n*, we have* $|f_n| \leq g$ *a.e., then* $f \in L^1$*, and*

$$\lim_{n\to\infty} \int f_n = \int f.$$

Proof. By considering the real and imaginary parts separately, we may assume f and all the f_n are real valued. By hypothesis, $g \pm f_n \geq 0$ a.e. Applying Fatou's theorem to these sequences, we find

$$\int g + \int f \leq \liminf \int (g + f_n) = \int g + \liminf \int f_n$$

and

$$\int g - \int f \leq \liminf \int (g - f_n) = \int g - \limsup \int f_n$$

It follows that $\liminf \int f \geq \int f \geq \limsup \int f$. □

The conclusion $\int f_n \to \int f$ (equivalently, $\left|\int f_n - \int f\right| \to 0$) can be strengthened somewhat:

Corollary 8.5. *If* f_n, f, g *satisfy the hypotheses of the Dominated Convergence theorem, then* $\lim_{n\to\infty} \|f_n - f\|_1 = 0$ *(that is,* $\lim \int |f_n - f| = 0$*).* †

Theorem 8.6 (Density of simple functions in L^1)**.** *If* $f \in L^1$*, then there is a sequence* (f_n) *of simple functions from* L^1 *such that,*

(a) $|f_n| \leq |f|$ *for all* n*,*
(b) $f_n \to f$ *pointwise, and*
(c) $\lim_{n\to\infty} \|f_n - f\|_1 = 0$*.*

Proof. Write $f = u + iv$ with u, v real, and $u = u^+ - u^-, v = v^+ - v^-$. Each of the four functions $u^\pm, v^\pm$ is nonnegative and measurable, so by the ziggurat approximation we can choose four sequences of measurable simple functions $u_n^\pm, v_n^\pm$ increasing pointwise to $u^\pm, v^\pm$ respectively. Now put $u_n = u_n^+ - u_n^-$, $v_n = v_n^+ - v_n^-$, and $f_n = u_n + iv_n$. By construction, each f_n is simple. Moreover

$$|u_n| = u_n^+ + u_n^- \leq u^+ + u^- = |u|,$$

and similarly $|v_n| \leq |v|$, so $|f_n| \leq |f|$. It follows that each f_n is integrable, and $f_n \to f$ pointwise by construction. Finally, since $f \in L^1$, the f_n satisfy the hypothesis of the dominated convergence theorem (with $g = |f|$), so (c) follows from Corollary 8.5. □

9. Additional topics

9.1. **The Lebesgue integral and the Riemann integral.** The treatment in this section follows I. P. Natanson, *Theory of functions of a real variable*, Chapter V.4.

Let $f : [a, b] \to \mathbb{R}$ be a function (for the moment we do not assume f is bounded). Let $x_0 \in [a, b]$ and let $\delta > 0$. Define

$$m_\delta(x_0) = \inf\{f(x)\}, \quad M_\delta(x_0) = \sup\{f(x)\}, \quad x_0 - \delta < x < x_0 + \delta$$

(Strictly speaking, the inf and sup are taken only over those x in $(x_0 - \delta, x_0 + \delta)$ which belong to $[a, b]$.)

From the definitions we have $m_\delta(x_0) \leq f(x_0) \leq M_\delta(x_0)$. If we let δ decrease to 0, then $M_\delta(x_0)$ is nonincreasing, and $m_\delta(x_0)$ is nonincreasing, and hence the limits

$$m(x_0) := \lim_{\delta \to 0^+} m_\delta(x_0), \quad M(x_0) := \lim_{\delta \to 0^+} M_\delta(x_0)$$

exist and satisfy

$$m_\delta(x_0) \leq m(x_0) \leq f(x_0) \leq M(x_0) \leq M_\delta(x_0).$$

The functions $m(x)$ and $M(x)$ are called respectively the *lower* and *upper Baire functions* for f.

Theorem 9.1 (Baire)**.** *A function $f : [a, b] \to \mathbb{R}$ is continuous at $x_0 \in [a, b]$ if and only if*

$$m(x_0) = M(x_0).$$

Proof. Suppose f is continuous at x_0. Given $\epsilon > 0$, there exists $\delta > 0$ such that for $|x - x_0| < \delta$, we have $|f(x) - f(x_0| < \epsilon$. That is, for every $x \in (x_0 - \delta, x_0 + \delta)$, we have

$$f(x_0) - \epsilon < f(x) < f(x_0) + \epsilon$$

and it follows that

$$f(x_0) - \epsilon \leq m_\delta(x_0) \leq M_\delta(x_0) \leq f(x_0) + \epsilon,$$

and thus further that

$$f(x_0) - \epsilon \leq m(x_0) \leq M(x_0) \leq f(x_0) + \epsilon.$$

Since this last inequality holds for all $\epsilon > 0$, we conclude that $m(x_0) = M(x_0) = f(x_0)$.

Conversely, suppose $m(x_0) = M(x_0)$, and note that this common value must equal $f(x_0)$. Let $\epsilon > 0$, from the definition of the Baire functions there exists $\delta > 0$ such that

$$m(x_0) - \epsilon < m_\delta(x_0) \leq m(x_0) \quad \text{and} \quad M(x_0) \leq M_\delta(x_0) < m(x_0) + \epsilon.$$

These inequalities imply

$$f(x_0) - \epsilon < m_\delta(x_0) \quad \text{and} \quad M_\delta(x_0) < f(x_0) + \epsilon.$$

Now let $x \in (x_0 - \delta, x_0 + \delta)$. Then $f(x)$ lies between $m_\delta(x_0)$ and $M_\delta(x_0)$, so that $f(x_0) - \epsilon < f(x) < f(x_0) + \epsilon$. This shows that f is continuous at x_0. □

For a partition P of an interval $[a, b]$, we let $mesh(P)$ denote the maximum length of the subintervals $[x_i, x_{i+1}]$.

Lemma 9.2. *For each $k = 1, 2, \dots$ let $P_k = \{x_0^{(k)} = a, x_1^{(k)}, \dots x_{n_k}^{(k)} = b\}$ be a partition of $[a, b]$, and suppuse that $mesh(P_k) \to 0$ as $k \to \infty$. Let $m_i^{(k)}$ be the infimum of f on the interval $[x_i^{(k)}, x_{i+1}^{(k)}]$. Define an auxiliary function $\varphi_k(x)$ by*

$$\varphi_k(x) = \begin{cases} m_i^{(k)} & \text{for } x \in (x_i^{(k)}, x_{i+1}^{(k)}) \\ 0 & x = x_0^{(k)}, x_1^{(k)}, \dots, x_{n_k}^{(k)} \end{cases}.$$

If x_0 is not equal to any of the points $x_i^{(k)}$ for any k, i, then

$$\lim_{k\to\infty} \varphi_k(x_0) = m(x_0).$$

†

Proof. Fix a point x_0, not equal to any of the $x_i^{(k)}$. Fix k and consider the interval $[x_{i_0}^{(k)}, x_{i_0+1}^{(k)}]$ of the partition P_k which contains x_0. By assumption, x_0 is not an endpoint of this interval, and hence lies in its interior, so there exists $\delta > 0$ for which $(x_0-\delta, x_0+\delta) \subset [x_{i_0}^{(k)}, x_{i_0+1}^{(k)}]$. It follows that $m_{i_0}^{(k)} \le m_\delta(x_0)$ or, what is the same, $\varphi_k(x_0) \le m_\delta(x_0)$.

Letting δ go to 0, we have for all k

$$\varphi_k(x_0) \le m(x_0).$$

Now let $h < m(x_0)$. Then there exists $\delta > 0$ such that $m_\delta(x_0) > h$. Fix this δ; then there exists k_0 such that for $k > k_0$ we have $[x_{i_0}^{(k)}, x_{i_0+1}^{(k)}] \subset (x_0 - \delta, x_0 + \delta)$, where i_0 is chosen (depending on k) so that this closed interval is the one from P_k containing x_0. (The existence of such a k_0 follows from the assumption $mesh(P_k) \to 0$.)

For each k we have $m_{i_0}^{(k)} \ge m_\delta(x_0) > h$, or, what is the same, $\varphi_k(x_0) > h$. Thus, for each $h < m(x_0)$ there is a k_0 such that for all $k > k_0$,

$$h < \varphi_k(x_0) \le m(x_0)$$

which proves that $\varphi_k(x_0) \to m(x_0)$. □

It is evident that an analogous version of the lemma will hold for the upper Baire function $M(x)$. As an immediate application of the lemma we have

Corollary 9.3. *The Baire functions $m(x)$ and $M(x)$ are measurable.* †

Indeed, the lemma shows that m and M are pointwise a.e. limits of sequences of simple functions, and hence (Borel) measurable. (Precisely, the φ_k are (Borel) measurable simple functions, and $\varphi_k(x) \to m(x)$ at every point of $[a, b]$ except possibly at points from the countable set $\{x_i^{(k)}\}$.)

We continue to use the notation from the lemma. To distinguish the two types of integration under discussion, we write $(L)\int$ for the Lebesgue integral and $(R)\int$ for the Riemann integral.

Corollary 9.4. *Let f be a bounded function on $[a, b]$. Then*

$$(L)\int_a^b \varphi_k(x)dx \to (L)\int m(x)dx.$$

†

Indeed, if $|f| \le C$ in $[a, b]$ then $|\varphi_k| \le C, |m| \le C$, so the dominated convergence theorem applies.

Theorem 9.5. *A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if it is continuous almost everywhere on $[a, b]$ (that is, if and only if its set of discontinuities has Lebesgue measure 0.)*

Proof. Let f be bounded. There exists a sequence of partitions P_k with $mesh(P_k) \to 0$ and such that

$$L(P_k, f) \to (R)\underline{\int_a^b} f, \quad U(P_k, f) \to (R)\overline{\int_a^b} f$$

as $k \to \infty$.

Next, we rephrase the last corollary. Note that for each k

$$(L)\int_a^b \varphi_k(x)\,dx = \sum_{i=0}^{n_k-1} \int_{x_i^{(k)}}^{x_{i+1}^{(k)}} \varphi_k(x)\,dx = \sum_{i=0}^{n_k-1} m_i^{(k)} \Delta x_i^{(k)} = L(P_k, f).$$

Thus the corollary says that

$$L(P_k, f) \to (L)\int_a^b m(x)\,dx.$$

Analogously, the upper Darboux sums will approximate the Lebesgue integral of the upper Baire function M:

$$U(P_k, f) \to (L)\int_a^b M(x)\,dx.$$

Thus for any bounded function f we will have

$$U(P_k, f) - L(P_k, f) \to (L)\int_a^b M(x) - m(x)\,dx.$$

We therefore have

$$\left((R)\overline{\int_a^b} f\right) - \left((R)\underline{\int_a^b} f\right) = (L)\int_a^b M(x) - m(x)\,dx.$$

We conclude that the function f is Riemann integrable if and only if

$$(L)\int_a^b [M(x) - m(x)]\,dx = 0.$$

Now, the integrand $M(x) - m(x)$ is nonnegative, so its Lebesgue integral is 0 if and only if $M - m$ is 0 almost everywhere, that is, if and only if $M = m$ almost everywhere.

But we have already shown that $M(x) = m(x)$ if and only if f is continuous at x. This proves the theorem. □

9.2. **Old version a la Rudin.** We temporarily write $\mathscr{R}\int f_a^b$ for the Riemann integral of a bounded function $f : [a,b] \to \mathbb{R}$.

Theorem 9.6. *Suppose $f : [a,b] \to \mathbb{R}$ is a bounded function. Then f is Riemann integrable if and only if f is continuous a.e. on $[a,b]$, in which case f is Lebesgue measurable and $\mathscr{R}\int_a^b f = \int_{[a,b]} f\,dm$.*

Lemma 9.7. *If $f, g : [a,b] \to \mathbb{R}$, g is Lebesgue measurable, and $f = g$ a.e., then f is Lebesgue measurable.* †

Proof of Theorem 9.6. Let $P = \{x_0 = a, x_1, \dots, x_n = b\}$ be a partition of $[a,b]$ and let $E_j = [x_{j-1}, x_j]$ be the j^{th} interval of the partition. As usual let $m_j = \inf_{E_j} f$ and $M_j = \sup_{E_j} f$, and recall the definitions of the upper and lower sums

$$L(P,f) = \sum_{j=1}^{n} m_j |E_j|, \quad U(P,f) = \sum_{j=1}^{n} M_j |E_j|.$$

Define simple functions

$$s^- = \sum_{j=1}^{n} m_j \mathbf{1}_{E_j}, \quad s^+ = \sum_{j=1}^{n} M_j \mathbf{1}_{E_j}$$

and observe that $s^-(x) \le f(x) \le s^+(x)$ for a.e. $x \in [a,b]$. Furthermore, $s^\pm$ are measurable and

$$\int s^- = L(P,f), \quad \int s^+ = U(P,f).$$

Note also that if Q is a refinement of P and $s^\pm_{P,Q}$ are the corresponding simple functions, we have

$$s_P^- \le s_Q^- \le s_Q^+ \le s_P^+$$

almost everywhere. Now take a sequence of partitions P_k so that P_{k+1} is a refinement of P_k, for each k, and such that

$$\lim_k L(P_k, f) = \mathscr{R}\underline{\int_a^b} f \quad \text{and} \quad \lim_k U(P_k, f) = \mathscr{R}\overline{\int_a^b} f.$$

Refining the P_k if necessary, we may assume that each interval of P_k has length at most $\frac{1}{k}$. (So, the "mesh size" of the partitions tends to 0.) Letting $s_k^\pm$ be the corresponding sequences of simple functions, we have that s_k^- is pointwise increasing a.e., and s_k^+ is pointwise decreasing a.e. to (necessarily measurable) functions f^-, f^+ respectively and

$$f^- \le f \le f^+ \quad a.e. \tag{32}$$

By the monotone covergence theorem applied to s_k^- and the dominated convergence theorem applied to s_k^+, we have

$$\int f^- = \lim \int s_k^- = \lim L(P_k, f) = \mathscr{R}\underline{\int_a^b} f \leq \mathscr{R}\overline{\int_a^b} f = \lim U(P_k, f) = \lim \int s_k^+ = \int f^+. \tag{33}$$

It follows that f is Riemann integrable if and only if $\int f^- = \int f^+$, but since $f^- \leq f^+$ a.e., this happens if and only if $f^- = f^+$ a.e., or equivalently $\lim(s_k^+ - s_k^-) = 0$ a.e. But it is straightforward to check that this last condition holds if and only if f is continuous a.e. To see this, note first that if x does not lie in any of the partitions P_k then $\lim(s_k^+(x) - s_k^-(x)) = 0$ if and only if f is continuous at x (this is where the condition on the mesh size is used). Moreover, the set of points that do belong to some partition is countable and hence has measure zero. Finally, if f is Riemann integrable then we have $f = f^- = f^+$ a.e., hence f is measurable (by the lemma) and $\mathscr{R}\int_a^b f = \int f\, dm$ by (33). □

9.3. **Recovering a function from its derivative.** If f is a differentiable function on $[a, b]$, and we know its derivative f', can we recover f from f'? If f' is Riemann integrable, then we know from the Fundamental Theorem of Calculus that

$$f(x) = f(a) + \int_a^x f'(t)dt. \tag{34}$$

However, it is possible that f' is not Riemann integrable, even if it exists for every $x \in [a, b]$ and is bounded. We provide an example of such a function, and then prove that if f' is *bounded*, then it is Lebesgue integrable, and f can be recoverd from f' by the formula (34) if the integral is interpreted as a Lebesgue integral. (We note for this that it crucial that f' exists for *all* $x \in [a, b]$, not just a.e. If f' is only assumed to exist a.e. then there are counterexamples, e.g. the Cantor-Lebesgue function).

Example. Let $E \subset [a, b]$ be closed, nowhere dense subset (i.e., E contains no intervals) with $m(E) > 0$ (for example, we can take E to be a fat Cantor set), and $\inf E = a, \sup E = b$. Its complement $U = [a, b] \setminus E$ in $[a, b]$ is open, and hence is a countable disjoint union of intervals (a_n, b_n), $n = 1, 2, \ldots$. We define a function f on $[a, b]$ as follows: put $f(x) = 0$ for $x \in E$. On each complementary interval (a_n, b_n) we define f by

$$f(x) = (x - a_n)^2 (x - b_n)^2 \sin \frac{1}{(b_n - a_n)(x - a_n)(x - b_n)}.$$

Let us show first that f is differentiable at each point of E, and $f'(x) = 0$ at these points. Indeed, fix $x_0 \in E$ and let $x \in [a, b]$ with $x > x_0$. If $x \in F$, then $f(x) = f(x_0) = 0$. If $x \in (a_n, b_n)$, then $x_0 \leq a_n < x$, so

$$x - x_0 \geq x - a_n$$

and

$$\left| \frac{f(x) - f(x_0)}{x - x_0} \right| \leq (x - a_n)(b - a)^2 \leq (x - x_0)(b - a)^2.$$

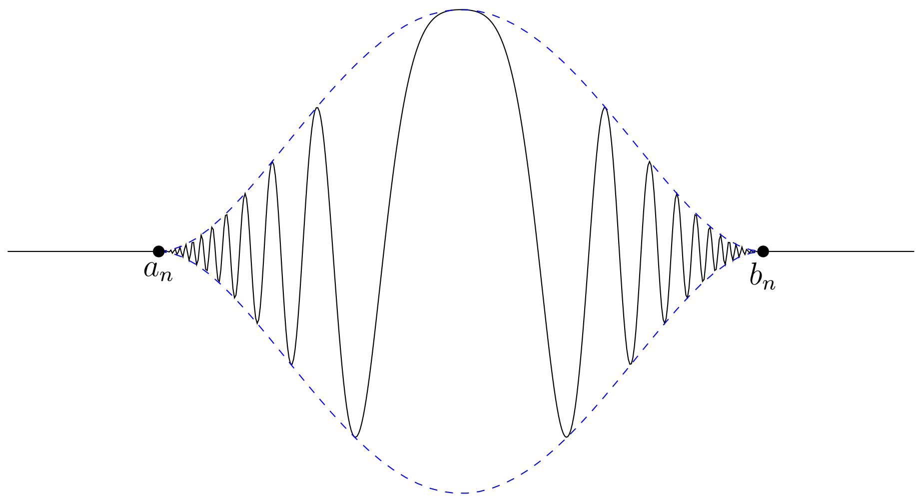

FIGURE 1. The graph of f on one of the intervals (a_n, b_n)

It follows that

$$\lim_{x \to x_0^+} \left| \frac{f(x) - f(x_0)}{x - x_0} \right| = 0.$$

Analogously, the limit as $x \to x_0^-$ is also 0, which proves $f'(x_0) = 0$.

For $x \in (a_n, b_n)$ we evidently have that f is differentiable, and from computing the derivative

$$f'(x) = 2(x - a_n)(x - b_n)(2x - a_n - b_n) \sin \frac{1}{(b_n - a_n)(x - a_n)(x - b_n)} - \\ - \frac{2x - a_n - b_n}{b_n - a_n} \cos \frac{1}{(b_n - a_n)(x - a_n)(x - b_n)}.$$

we see that $|f'(x)| \leq 2(b-a)^3 + 1$, and hence is a bounded function on $[a, b]$. Inspecting the expression for $f'(x)$ for $x \in (a_n, b_n)$, we see that $f'(x)$ does not have a limit as $x \to a_n$ or $x \to b_n$ (the culprit is the cosine term). This shows that f' is not continuous at any of the endpoints a_n, b_n (which belong to E), and using the fact that E is nowhere dense, it is not hard to show that likewise f' is discontinuous at every point of E. But since $m(E) > 0$, we conclude that f' is not Riemann integrable.

Thus, the Riemann integral does not completely solve the problem of recovering a function from its derivative. It turns out the the Lebesgue integral provides a more powerful tool for the solution of this problem.

Theorem 9.8. *Let f be differentiable at* every *point of $[a, b]$. If $f'(x)$ is bounded, then it is Lebesgue integrable, and for all $x \in [a, b]$ we have*

$$f(x) = f(a) + \int_a^x f'(t)\, dt.$$

(Here, of course, the integral is the Lebesgue integral.)

Proof. First of all we note that f must be continuous, since it is differentiable. It will be convenient to extend f to the interval $[a, b+1]$ by putting, for $b < x \leq b+1$

$$f(x) = f(b) + (x-b)f'(b).$$

This extended f is then continuous and differntiable in $[a, b+1]$.

For each $n = 1, 2, 3, \ldots$ put

$$\varphi_n(x) = n\left[f\left(x + \frac{1}{n}\right) - f(x)\right].$$

At each point $x \in [a, b]$ we have $\lim_{n\to\infty} \varphi_n(x) = f'(x)$, and since each of the functions φ_n is continuous, each of them is measurable, and hence f' is measurable. Since f' is also assumed bounded, it is integrable.

Further, by the mean value theorem, for each $x \in [a, b]$ there exists a $\theta \in (0, 1)$ so that

$$\varphi_n(x) = n\left[f\left(x + \frac{1}{n}\right) - f(x)\right] = f'\left(x + \frac{\theta}{n}\right).$$

Since f' is bounded, it follows that the φ_n are uniformly bounded, so by the dominated convergence theorem

$$\int_a^b f'(x)\,dx = \lim_{n\to\infty} \int_a^b \varphi_n(x)\,dx. \tag{35}$$

But

$$\int_a^b \varphi_n(x)\,dx = n\int_a^b f\left(x + \frac{1}{n}\right)\,dx - n\int_a^b f(x)\,dx = \tag{36}$$

$$= n\int_{a+1/n}^{b+1/n} f(x)\,dx - n\int_a^b f(x)\,dx. \tag{37}$$

(The change of variable in the first integral on the right-hand side can be justified as follows: Since f is continuous, it is Riemann integrable, so we can identify the Riemann integral with the Lebesgue integral, and apply the usual change of variables formula valid for the Riemann integral.) We then obtain

$$\int_a^b \varphi_n(x)\,dx = n\int_b^{b+1/n} - n\int_a^{a+1/n} f(x)\,dx.$$

Applying the mean value theorem for integrals to the two integrals on the right-hand side (again using the fact that f is continuous), there exist numbers $\theta_a, \theta_b \in (0, 1)$ so that

$$\int_a^b \varphi_n(x)\,dx = f\left(b + \frac{\theta_b}{n}\right) - f\left(a + \frac{\theta_a}{n}\right)$$

and appealing one last time to the continuity of f we obtain

$$\lim_{n\to\infty} \int_a^b \varphi_n(x)\,dx = f(b) - f(a).$$

Therefore from (35) we find

$$f(b) - f(a) = \int_a^b f'(x)\,dx.$$

Replacing b by arbitrary $x \in [a, b]$, the theorem is proved.

$\square$