Notes on:

Measure, Integration, and Real Analysis

David Duan

Last Updated: May 6, 2023 (Draft)

Contents

Chapter 1

Maps

Contents

MAT137 solely focused on maps from $\mathbb{R} \to \mathbb{R}$,and now we will begin the study of MAT237 with the characteristics of maps from $\mathbb{R}^n \to \mathbb{R}^m$. There are many ways of referring to them but all of these defines a function

$$f : A \to B, \quad A \subseteq \mathbb{R}^n, B \subseteq \mathbb{R}^m$$

This section will refrain from being too rigour as we try to build intuitions and explore examples of multivariable functions.

1.1 Curves

1.1.1 Curves

Definition 1.1.1 The **length** $\ell(I)$ of an open interval $I \subseteq \mathbb{R}$ is defined by

$$\ell(I) = \begin{cases} b - a & I = (a, b) \text{ for some } a, b \in \mathbb{R} \text{ with } a < b \\ 0 & I = \varnothing \\ \infty & I = (-\infty, a) \text{ or } I = (a, \infty) \text{ for some } a \in \mathbb{R} \\ \infty & I = (-\infty, \infty) \end{cases}$$

Let A be an arbitrary subset of $\mathbb{R}$. The *size* of A should be at most the sum of the lengths of a sequence of open intervals whose union contains A. Taking the infimum of all such sums gives a reasonable definition of the size of A, denoted $|A|$ and called the *outer measure* of A.

Definition 1.1.2 The **outer measure** $|A|$ of a set $A \subseteq \mathbb{R}$ is defined by

$$|A| = \inf \left\{ \sum_{k=1}^{\infty} \ell(I_k) \,\middle|\, I_1, I_2, \ldots \text{ are open intervals such that } A \subseteq \bigcup_{k=1}^{\infty} I_k \right\}.$$

The following example shows that finite sets have outer measure 0.

Example 1.1.3 Suppose $A = \{a_1, \ldots, a_n\}$ is a finite set of real numbers. Let $\varepsilon > 0$ and define the sequence $I_1, I_2, \ldots$ of open intervals by

$$I_k = \begin{cases} (a_k - \varepsilon, a_k + \varepsilon) & k \leq n \\ \varnothing & k > n \end{cases}$$

Then $I_1, I_2, \ldots$ is a sequence of open intervals whose union contains A. Clearly,

$\sum_{k=1}^{\infty} \ell(I_k) = 2\varepsilon n$. Since the outer measure takes the infimum of all sequences of open intervals whose union contains A, we have $|A| \leq 2\varepsilon n$. Since ε is arbitrarily small, we conclude that $|A| = 0$. ◊

1.1.2 Good Properties of Outer Measure

We extend Example 1.1.3 to countable subsets of $\mathbb{R}$.

Lemma 1.1.4 Every countable subset of $\mathbb{R}$ has outer measure 0.

Proof. Suppose $A = \{a_1, a_2, \ldots\}$ is a countable subset of $\mathbb{R}$. Let $\varepsilon > 0$. For $k \in \mathbb{Z}^+$, let

$$I_k = \left(a_k - \frac{\varepsilon}{2^k}, a_k + \frac{\epsilon}{2^k}\right).$$

Then $I_1, I_2, \ldots$ is a sequence of open intervals whose union contains A. Since $\sum_{k=1}^{\infty} \ell(I_k) = \varepsilon$, we have $|A| \leq 2\varepsilon$. Since ε is arbitrarily small, we conclude that $|A| = 0$. □

The next result shows that outer measure does the right thing with respect to set inclusion.

Lemma 1.1.5 If $A, B \subseteq \mathbb{R}$ with $A \subseteq B$, then $|A| \leq |B|$.

Proof. Any sequence of open intervals whose union contains B must also contain A. □

We expect that the size of a subset of $\mathbb{R}$ should not change if the set is shifted to the right or to the left. Let $t \in \mathbb{R}$ and $A \subseteq \mathbb{R}$. The **translation** $t + A$ is defined by

$$t + A = \{t + a \mid a \in A\}.$$

The next result shows that the outer measure is **translation invariant**.

Lemma 1.1.6 Let $t \in \mathbb{R}$ and $A \subseteq \mathbb{R}$. Then $|t + A| = |A|$.

Proof. Suppose $I_1, I_2, \ldots$ is a sequence of open intervals whose union contains A. Then $t + I_1, t +$

$I_2, \ldots$ is a sequence of open intervals whose union contains $t + A$. Thus

$$|t + A| \leq \sum_{k=1}^{\infty} \ell\,(t + I_k) = \sum_{k=1}^{\infty} \ell\,(I_k)\,.$$

Taking the infimum of the last term over all sequences $I_1, I_2, \ldots$ of open intervals whose union contains A, we have $|t + A| \leq |A|$.

To get the inequality in the other direction, note that $A = -t + (t + A)$. Thus applying the inequality from the previous paragraph, with A replaced by $t + A$ and t replaced by $-t$, we have $|A| = |-t + (t + A)| \leq |t + A|$. Hence $|t + A| = |A|$. □

The next result proves the **countable subadditivity** of outer measure.

Lemma 1.1.7 Let $A_1, A_2, \ldots \subseteq \mathbb{R}$. Then

$$\left| \bigcup_{k=1}^{\infty} A_k \right| \leq \sum_{k=1}^{\infty} |A_k|.$$

Proof. If $|A_k| = \infty$ for some $k \in \mathbb{Z}^+$, then the inequality above clearly holds. Now assume $|A_k| < \infty$ for all $k \in \mathbb{Z}$. Let $\varepsilon > 0$. For each $k \in \mathbb{Z}^+$, let $I_{1,k}, I_{2,k}, \ldots$ be a sequence of open intervals whose union contains A_k such that

$$\sum_{j=1}^{\infty} \ell(I_{j,k}) \leq \frac{\varepsilon}{2^k} + |A_k|.$$

Adding up such summations for all A_k's and using the fact that $\sum_{i=1}^{\infty}(1/2^i) = 1$, we have

$$\sum_{k=1}^{\infty} \sum_{j=1}^{\infty} \ell\left(I_{j,k}\right) \leq \varepsilon + \sum_{k=1}^{\infty} |A_k|\,. \tag{1.1}$$

Let us now rearrange the doubly indexed collection of open intervals $\{I_{j,k} : j, k \in \mathbb{Z}^+\}$ into a sequence of open intervals whose union contains $\bigcup_{k=1}^{\infty} A_k$ as follows. (See Remark 1.1.8 for intuition.) In step k $(= 2, 3, 4, \ldots)$ we adjoin the $k - 1$ intervals whose indices add up to k :

$$\underbrace{I_{1,1}}_{C_2}, \underbrace{I_{1,2}, I_{2,1}}_{C_3}, \underbrace{I_{1,3}, I_{2,2}, I_{3,1}}_{C_4}, \underbrace{I_{1,4}, I_{2,3}, I_{3,2}, I_{4,1}}_{C_5}, \underbrace{I_{1,5}, I_{2,4}, I_{3,3}, I_{4,2}, I_{5,1}}_{C_6}, \ldots$$

Let's give each interval in this enumerate a new index, say m, i.e.,

$$\bigcup_{m=1}^{\infty} I'_m = \bigcup_{k=1}^{\infty} \bigcup_{j=1}^{\infty} I_{j,k}.$$

By the definition of outer measure, we have

$$\begin{aligned}\left|\bigcup_{k=1}^{\infty} A_k\right| &= \inf\left\{\sum_{m=1}^{\infty} \ell(I_m) \,\middle|\, I_1, I_2, \ldots \text{ are open intervals such that } \bigcup_{m=1}^{\infty} A_k \subseteq \bigcup_{m=1}^{\infty} I_m\right\} \\ &\leq \sum_{m=1}^{\infty} \ell(I'_m) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \ell(I_{j,k}) \leq \varepsilon + \sum_{k=1}^{\infty} |A_k|\,.\end{aligned}$$

Since ε is arbitrarily small, we conclude that $\left|\bigcup_{k=1}^{\infty} A_k\right| \leq \sum_{k=1}^{\infty} |A_k|$ as desired. □

Remark 1.1.8 In the proof above, we rearranged the countable set of intervals

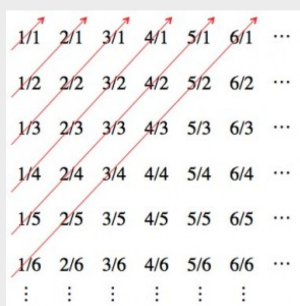
$$\{I_{j,k}\}_{j,k=1}^{\infty}$$

into the form

$$\underbrace{I_{1,1}}_{C_2},\underbrace{I_{1,2}, I_{2,1}}_{C_3}, \underbrace{I_{1,3}, I_{2,2}, I_{3,1}}_{C_4}, \underbrace{I_{1,4}, I_{2,3}, I_{3,2}, I_{4,1}}_{C_5}, \underbrace{I_{1,5}, I_{2,4}, I_{3,3}, I_{4,2}, I_{5,1}}_{C_6}, \ldots$$

This provides a way for us to enumerate this countable number of intervals so that we could perform union or summation on them.

You should connect this back to how we proved the set of rationals is countable:



If at least one of the index variables in $\{I_{j,k}\}_{j,k}$ is finite, say, $j \in \{1, \ldots, n\}$, then if we put the intervals into a table with j columns and k rows, we will have a finite number n of columns. This means that we could first enumerate all intervals $I_{j,k}$ with $j = 1$ (i.e., intervals in the first column), followed by all intervals with $j = 2$ (i.e., intervals in the second column), followed by all intervals with $j = 3$ (i.e., intervals in the third column), etc.:

$$\{\{I_{j,k}\}_{j=1}^{n}\}_{k=1}^{\infty} = (I_{1,1}, I_{1,2}, \ldots, I_{2,1}, I_{2,2}, \ldots, I_{3,1}, I_{3,2}, \ldots, I_{n,1}, I_{n,2}, \ldots).$$

However, both j and k are infinite in our case, so we cannot do this. The rearrangement we had above makes it clear that every interval $I_{j,k}$ will appear somewhere in the list, making it possible to set up a bijection between each interval $I_{j,k}$ and its position in the list, which is an element of $\mathbb{N}$.

Note that countable subadditivity implies finite subadditivity, meaning that

$$|A_1 \cup \cdots \cup A_n| \leq |A_1| + \cdots + |A_n|$$

for all $A_1, \ldots, A_n \subset \mathbb{R}$, because we can take $A_k = \varnothing$ for $k > n$ in Lemma 1.1.7.

1.1.3 Outer Measure of Closed Bounded Interval

One more property nice to have is that for all $a, b \in \mathbb{R}$, $a < b$, the outer measure of the closed interval $[a, b]$ is $b - a$. The $|[a, b]| \leq b - a$ is easy to show as $(a - \infty, b + \infty), \emptyset, \emptyset, \ldots$ is a sequence of open intervals whose union contains $[a, b]$. The other direction, however, is more subtle.

Definition 1.1.9 Let $A \subseteq \mathbb{R}$.

- A collection $\mathcal{C}$ of open subsets of $\mathbb{R}$ is called an **open cover** of A if A is contained in the union of all the sets in $\mathcal{C}$.
- An open cover $\mathcal{C}$ if A is said to have a **finite subcover** if A is contained in the union of some finite list of sets in $\mathcal{C}$.

The following proof uses the *completeness* property of the real numbers (by asserting that the supremum of a certain non-empty bounded set exists, i.e., the *least upper bound property*).

Theorem 1.1.10 [Heine-Borel] Every open cover of a closed bounded subset of $\mathbb{R}$ has a finite subcover.

Proof. Let F be a closed bounded subset of $\mathbb{R}$ and $\mathcal{C}$ an open cover of F.

First consider the case where $F = [a, b]$ for some $a, b \in \mathbb{R}$ with $a < b$. Then $\mathcal{C}$ is an open cover of $[a, b]$. Define the set of points in the interval $[a, b]$ such that the subinterval $[a, d]$ is covered by a finite number of sets from $\mathcal{C}$:

$$D = \{d \in [a, b] : [a, d] \text{ has a finite subcover from } \mathcal{C}\}.$$

Our goal is to show that $D = [a, b]$, i.e., the entire interval $[a, b]$ has a finite subcover from $\mathcal{C}$. Note that $a \in D$ (because $a \in G$ for some $G \in \mathcal{C}$, i.e., the point a itself must be covered by some open set $G \in \mathcal{C}$), so D is non-empty. Let $s = \sup D$, so $s \in [a, b]$. We want to show that $s = b$.

By definition of open cover, there exists an open set $G \in \mathcal{C}$ such that $s \in G$. Since G is open and $s \in G$, we can find $\delta > 0$ such that $(s - \delta, s + \delta) \subseteq G$. Since $s = \sup D$, there exist $d \in (s - \delta, s]$ and $n \in \mathbb{Z}^+$ and $G_1, \ldots, G_n \in \mathcal{C}$ such that $[a, d]$ has a finite subcover from $\mathcal{C}$:

$$[a, d] \subseteq G_1 \cup \cdots \cup G_n.$$

Next, since $(s - \delta, s + \delta) \subseteq G$, i.e., G also covers all points in $[s, s + \delta)$, we see that for all $d' \in [s, s + \delta)$,

$$[a, d'] \subseteq G \cup G_1 \cup \cdots \cup G_n. \tag{1.2}$$

In words, $[a, d']$ has a finite subcover from C for all $d' \in [s, s + \delta)$.

But then $d' \in D$ for all $d' \in [s, s+\delta) \cap [a,b]$, which implies that $s = b$. (Indeed, if $s < b$, then we can find $s' \in (s,b)$ such that $s' \in D$, which means that $s < s'$ is no longer the supremum of D.) Furthermore, substituting $d' = b$ in (1.2) shows that $[a,b]$ has a finite subcover from C, completing the proof in the case where $F = [a,b]$.

Now suppose F is an arbitrary closed bounded subset of $\mathbb{R}$ and that C is an open cover of F. Let $a, b \in \mathbb{R}$ be such that $F \subseteq [a,b]$. Now $C \cup \{\mathbb{R} \setminus F\} \supseteq F \cup \{\mathbb{R} \setminus F\} = \mathbb{R}$ is an open cover of $\mathbb{R}$ and hence is an open cover of $[a,b]$. By our first case (with the open cover of $[a,b]$ being $(C \cup \{\mathbb{R} \setminus F\})$, there exist $G_1, \dots, G_n \in C$ such that

$$[a,b] \subseteq G_1 \cup \cdots \cup G_n \cup (\mathbb{R} \setminus F).$$

Finally $F \subseteq [a,b]$ implies that $F \subseteq G_1 \cup \cdots \cup G_n$, completing the proof. □

We are now ready to prove the other direction.

Proposition 1.1.11 Suppose $a, b \in \mathbb{R}$ with $a < b$. Then $|[a,b]| = b - a$.

Proof. We have argued that $|[a,b]| \leq b - a$. For the other direction, suppose $I_1, I_2, \dots$ is a sequence of open intervals that covers $[a,b]$. By Heine-Borel Theorem, we can find a finite subcover $I_1, \dots, I_n$ for $[a,b]$. We now prove by induction on n that

$$[a,b] \subseteq I_1 \cup \cdots \cup I_n \implies \sum_{k=1}^{n} \ell(I_k) \geq b - a. \tag{1.3}$$

This will then imply that

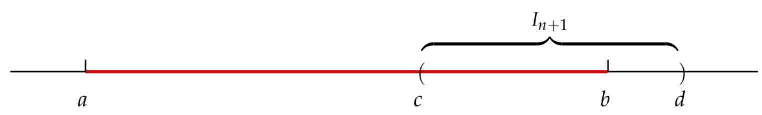
$$\sum_{k=1}^{\infty} \ell(I_k) \geq \sum_{k=1}^{n} \ell(I_k) \geq b - a, \tag{1.4}$$

completing the proof that $|[a,b]| \geq b - a$.

If $n = 1$, then (1.3) clearly implies (1.4) (the equality holds). Now suppose $n > 1$ and (1.3) implies (1.4) for all choices of $a, b \in \mathbb{R}$ with $a < b$. Suppose $I_1, \dots, I_n, I_{n+1}$ are open intervals such that

$$[a,b] \subseteq I_1 \cup \cdots \cup I_n \cup I_{n+1}.$$

Then b is in at least one of the intervals $I_1, \dots, I_n, I_{n+1}$. By relabeling, we can assume that $b \in I_{n+1}$. Suppose $I_{n+1} = (c,d)$. If $c \leq a$, then $\ell(I_{n+1}) \geq b - a$ and we are done. Now assume that $a < c < b < d$, as shown in the graph below.



Hence, $[a, c] \subseteq I_1 \cup \cdots \cup I_n$. By our induction hypothesis, we have $\sum_{k=1}^{n} \ell(I_k) \geq c - a$. Thus

$$\sum_{k=1}^{n+1} \ell(I_k) \geq (c - a) + \ell(I_{n+1}) = (c - a) + (d - c) = d - a \geq b - a,$$

completing the proof. □

The same result holds for open intervals.

Corollary 1.1.12 The outer measure of each open interval equals its length.

Proof. Trivial. □

The next result shows that non-trivial intervals are uncountable. This application of outer measure to prove a result that seems unconnected with outer measure is an indication that outer measure has serious mathematical value.

Proposition 1.1.13 Every interval in $\mathbb{R}$ that contains at least two distinct elements is uncountable.

Proof. Suppose I is an interval that contains $a, b \in \mathbb{R}$ with $a < b$. Then

$$|I| \geq |[a, b]| = b - a > 0,$$

where the first inequality above holds because outer measure preserves order and the equality above comes from Proposition 1.1.11. Since every countable subset of $\mathbb{R}$ has outer measure 0, we conclude that I is uncountable. □

To summarize, we have shown that the outer measure defined above has the following properties:

- Countable sets have outer measure 0.
- Outer measure preserves order, i.e., $A \subseteq B \implies |A| \leq |B|$.
- Outer measure is translation invariant, i.e., $|t + A| = |A|$.
- Outer measure is countably additive, i.e., $|\bigcup_{k=1}^{\infty} A_k| \leq \sum_{k=1}^{\infty} |A_k|$.
- Outer measure of an interval is equal to its length: $|[a, b]| = |(a, b)| = |[a, b)| = |(a, b]| = b - a$.

1.1.4 Outer Measure is Not Additive

So far, we have seen quite nice properties of outer measure. Unfortunately, we now come to an unpleasant property of outer measure. If outer measure were a perfect way to assign a size as an extension of the lengths of intervals, then the outer measure of the union of two disjoint sets would equal the sum of the outer measures of the two sets. Sadly, the next result states that outer measure does not have this property. (In the next section, we begin the process of getting around the next result, which will lead us to measure theory.)

Proposition 1.1.14 There exist disjoint subsets A and B of $\mathbb{R}$ such that

$$|A \cup B| \neq |A| + |B|.$$

Proof. For $x \in [-1, 1]$, let $\tilde{x}$ be the set of numbers in $[-1, 1]$ that differ from x by a rational number. In other words,

$$\tilde{x} = \{c \in [-1, 1] : x - c \in \mathbb{Q}\}.$$

First, we claim that if $a, b \in [-1, 1]$ and $\tilde{a} \cap \tilde{b} \neq \varnothing$, then $\tilde{a} = \tilde{b}$.

- Indeed, suppose exists $d \in \tilde{a} \cap \tilde{b}$. By definition of $\tilde{a}$ and $\tilde{b}$, d differs from both a and b by a rational number, i.e., $a - d$ and $b - d$ are rational numbers. Since $\mathbb{Q}$ is closed under subtraction, we conclude that $a - b$ is a rational number. Again since $\mathbb{Q}$ is closed under addition, the equation $a - c = (a - b) + (b - c)$ implies that if $c \in [-1, 1]$, then $a - c$ is a rational number iff $b - c$ is a rational number, so $c \in \tilde{a} \Leftrightarrow c \in \tilde{b}$ and thus $\tilde{a} = \tilde{b}$. We conclude that the $\sim$ operation partitions $[-1, 1]$ into equivalence classes.

Take $c = 0 \in \mathbb{Q}$, we see that $a \in \tilde{a}$ for each $a \in [-1, 1]$. Thus, $[-1, 1] = \bigcup_{a \in [-1,1]} \tilde{a}$.

Let V be a set that contains exactly one element in each of the distinct sets in $\{\tilde{a} : a \in [-1, 1]\}$. In other words, for every $a \in [-1, 1]$, the set $V \cap \tilde{a}$ has exactly one element. [1]

Let $r_1, r_2, \ldots$ be a sequence of distinct rational number such that $\{r_1, r_2, \ldots\} = [-2, 2] \cap \mathbb{Q}$. Then

$$[-1, 1] \subseteq \bigcup_{k=1}^{\infty} (r_k + V),$$

where the set inclusion above holds because if $a \in [-1, 1]$, then letting v be the unique element of $V \cap \tilde{a}$, we have $a - v \in \mathbb{Q}$, which implies that $a = r_k + v \in r_k + V$ for some $k \in \mathbb{Z}^+$.

[1] This steps involves the Axiom of Choice, as discussed after this proof. The set V arises by choosing one element from each equivalence class.

By the order-preserving property of outer measure (Lemma 1.1.5) and the countable subadditivity of outer measure (Lemma 1.1.7), we have

$$|[-1,1]| \leq \sum_{k=1}^{\infty} |r_k + V|.$$

Since $|[-1,1]| = 2$ (Proposition 1.1.11) and the outer measure is translation invariant (Lemma 1.1.6), we can rewrite the inequality above as $2 \leq \sum_{k=1}^{\infty} |V|$. Thus, $|V| > 0$.

Next, we claim that the sets $r_1 + V, r_2 + V, \ldots$ are disjoint.

- Suppose there exists $t \in (r_j + V) \cap (r_k + V)$. Then $t = r_j + v_1 = r_k + v_2$ for some $v_1, v_2 \in V$, which implies that $v_1 - v_2 = r_k - r_j \in \mathbb{Q}$. By construction of V, we get $v_1 = v_2$, which implies that $r_j = r_k$ and thus $j = k$.

Let $n \in \mathbb{Z}$. Clearly

$$\bigcup_{k=1}^{n} (r_k + V) \subseteq [-3,3]$$

because $V \subseteq [-1,1]$ and each $r_k \in [-2,2]$. The set inclusion above implies that

$$\left| \bigcup_{k=1}^{n} (r_k + V) \right| \leq 6. \tag{1.5}$$

However,

$$\sum_{k=1}^{n} |r_k + V| = \sum_{k=1}^{n} |V| = n|V|. \tag{1.6}$$

Now (1.5) and (1.6) suggest that we choose $n \in \mathbb{Z}^+$ such that $n|V| > 6$. Then,

$$\left| \bigcup_{k=1}^{n} (r_k + V) \right| \leq 6 < n|V| = \sum_{k=1}^{n} |r_k + V|. \tag{1.7}$$

If we had $|A \cup B| = |A| + |B|$ for all disjoint subsets A, B of $\mathbb{R}$, then by induction on n we get

$$\left| \bigcup_{k=1}^{n} A_k \right| = \sum_{k=1}^{n} |A_k|$$

for all disjoint subsets $A_1, \ldots, A_n$ of $\mathbb{R}$. However, (1.7) tells us that no such result holds. Thus, we conclude that there exist disjoint subsets A, B of $\mathbb{R}$ such that $|A \cup B| \neq |A| + |B|$. □

Remark. The **Axiom of Choice**, which belongs to set theory, states that if $\mathcal{E}$ is a set whose elements are disjoint nonempty sets, then there exists a set D that contains exactly one element in each set that is an element of $\mathcal{E}$. We used the Axiom of Choice to construct the set V that was used in the last proof.

1.2 Measurable Spaces and Functions

In the previous section, we see that the outer measure does not have all the desirable properties. In particular, it is not additive. Were we just unlucky? Could we find another function μ that has all the desirable properties? The next result shows that such an ideal μ does not exist.

Proposition 1.2.1 There does not exist a function μ that satisfies all four properties:

1. μ is a function from the set of subsets of $\mathbb{R}$ to $[0, \infty]$.
2. $\mu(I) = \ell(I)$ for every open interval I of $\mathbb{R}$.
3. μ satisfies **countable additivity**, i.e., for every disjoint sequence $A_1, A_2, \ldots$ of subsets of $\mathbb{R}$, we have
$$\mu\left(\bigcup_{k=1}^{\infty} A_k\right) = \sum_{k=1}^{\infty} \mu(A_k).$$
4. μ is *transition invariant*, i.e., $\mu(t + A) = \mu(A)$ for every $A \subseteq \mathbb{R}$ and every $t \in \mathbb{R}$.

Proof. Suppose there is a function μ that satisfies all of the properties listed above. We want to show that μ has all the properties of outer measure that were used in the proof of Proposition 1.1.14.

First, (2) implies that $\mu(\varnothing) = 0$ because the empty set is an open interval with length 0. If $a, b \in \mathbb{R}$ with $a < b$, then $(a, b) \subseteq [a, b] \subseteq (a - \varnothing, b + \varepsilon)$ for every $\varepsilon > 0$. Thus, $b - a \leq \mu([a, b]) \leq b - a + 2\varepsilon$ for every $\varepsilon > 0$. Hence $\mu([a, b]) = b - a$.

Next, let $A \subseteq B \subseteq \mathbb{R}$. Write B as the union of the disjoint sequence $A, B \setminus A, \varnothing, \varnothing, \ldots$. By countable additivity of μ, we have $\mu(B) = \mu(A) + \mu(B \setminus A) + 0 + 0 + \cdots + \mu(A) + \mu(B \setminus A) \geq \mu(A)$.

If $A_1, A_2, \ldots$ is a sequence of subsets of $\mathbb{R}$, then $A_1, A_2 \setminus A_1, A_3 \setminus (A_1 \cup A_2), \ldots$ is a disjoint sequence of subsets of $\mathbb{R}$ whose union is $\bigcup_{k=1}^{\infty} A_k$. Thus,

$$\begin{aligned}\mu\left(\bigcup_{k=1}^{\infty} A_k\right) &= \mu\left(A_1 \cup (A_2 \setminus A_1) \cup (A_3 \setminus (A_1 \cup A_2)) \cup \cdots\right) \\ &= \mu\left(A_1\right) + \mu\left(A_2 \setminus A_1\right) + \mu\left(A_3 \setminus (A_1 \cup A_2)\right) + \cdots \leq \sum_{k=1}^{\infty} \mu\left(A_k\right),\end{aligned}$$

where the second equality follows from the countable additivity of μ.

We have shown that μ has all the properties of outer measure that were used in the proof of Proposition 1.1.14. Repeating the proof of Proposition 1.1.14, we see that there exist disjoint subsets A, B of $\mathbb{R}$ such that $\mu(A \cup B) \neq \mu(A) + \mu(B)$. Thus, the disjoint sequence $A, B, \varnothing, \varnothing, \ldots$ does not satisfy the countable additivity property required by (3). Contradiction. □

1.2.1 σ-Algebras

The last result shows that we need to give up one of the desirable properties in our goal of extending the notion of size from intervals to more general subsets of $\mathbb{R}$. Let us repeat the properties below.

1. μ is a function from the set of subsets of $\mathbb{R}$ to $[0, \infty]$.
2. $\mu(I) = \ell(I)$ for every open interval I of $\mathbb{R}$.
3. μ satisfies countable additivity.
4. $\mu(t + A) = \mu(A)$ for every $A \subseteq \mathbb{R}$ and every $t \in \mathbb{R}$.

We cannot give up (2) because the size of an interval needs to be its length. We cannot give up (3) because countable additivity is needed to prove theorems about limits. We cannot give up (4) because a reasonable notion of size has to be translation invariant. Thus, we are forced to relax the requirement that the size is defined for all subsets of $\mathbb{R}$. We will limit the domain of the size function to the collection of subsets that are closed under *complementation* and *countable unions*.

Definition 1.2.2 Suppose X is a set and $\mathcal{S}$ is a set of subsets of X. Then $\mathcal{S}$ is called a σ**-algebra** on X if the following hold:

- $\varnothing \in \mathcal{S}$;
- (closed under complementation) $E \in \mathcal{S} \implies X \setminus E \in \mathcal{S}$;
- (closed under countable unions) $E_1, E_2, \ldots \in \mathcal{S} \implies \bigcup_{k=1}^{\infty} E_k \in \mathcal{S}$.

Example 1.2.3 Let X be a set.

- $\{\varnothing, X\}$ is a σ-algebra on X.
- $\mathcal{P}(X)$ (the power set of X) is a σ-algebra on X.
- The set of all subsets E of X such that E is countable or $X \setminus E$ is countable is a σ-algebra on X.[a]

[a] To see the set is closed under complementation, let $E \in \mathcal{S}$. WLOG, assume E is countable. Then $X \setminus E \in \mathcal{S}$ because its complement $X \setminus (X \setminus E) = E$ is countable. To see the set is closed under countable union, use the fact that the countable union of countable sets is countable.

Lemma 1.2.4 Suppose $\mathcal{S}$ is a σ-algebra on a set X. Then

- $X \in \mathcal{S}$;

- if $D, E \in \mathcal{S}$, then $D \cup E \in \mathcal{S}$, $D \cap E \in \mathcal{S}$, and $D \setminus E \in \mathcal{S}$;
- (closed under countable intersections) $E_1, E_2, \ldots \in \mathcal{S} \implies \bigcap_{k=1}^{\infty} E_k \in \mathcal{S}$.

Proof. Omitted. □

1.2.2 Borel Subsets of $\mathbb{R}$

There is a smallest σ-algebra on a set X containing a given set $\mathcal{A}$ of subsets of X.

Proposition 1.2.5 Suppose X is a set and $\mathcal{A}$ is a set of subsets of X. Then the intersection of all σ-algebras on X that contain $\mathcal{A}$ is a σ-algebra on X.

Proof. There is at least one σ-algebra on X because the σ-algebra consisting of all subsets of X, $\mathcal{P}(X)$, contains $\mathcal{A}$. Let $\mathcal{S}$ be the intersection of all σ-algebras on X that contain $\mathcal{A}$. Then $\varnothing \in \mathcal{S}$ because $\varnothing$ is an element of each σ-algebra on X that contains $\mathcal{A}$.

Let $E \in \mathcal{S}$. Then E is in every σ-algebra on X that contains $\mathcal{A}$. Then $X \setminus E$ is in every σ-algebra on X that contains $\mathcal{A}$. Hence $X \setminus E \in \mathcal{S}$.

Let $E_1, E_2, \ldots$ be a sequence of elements of $\mathcal{S}$. Then each E_k is in every σ-algebra on X contains $\mathcal{A}$. Then $\bigcup_{k=1}^{\infty} E_k$ is in every σ-algebra on X that contains $\mathcal{A}$. Hence $\bigcup_{k=1}^{\infty} E_k \in \mathcal{S}$, which completes the proof that $\mathcal{S}$ is a σ-algebra on X. □

Definition 1.2.6 The smallest σ-algebra on $\mathbb{R}$ containing all open subsets of $\mathbb{R}$ is called the collection of **Borel subsets** of $\mathbb{R}$.[a] An element of this σ-algebra is called a **Borel set**. This σ-algebra is sometimes called the **Borel algebra** on $\mathbb{R}$.

[a] Equivalently, the collection of **Borel subsets** of $\mathbb{R}$ is the smallest σ-algebra on $\mathbb{R}$ containing all the open intervals.

Example 1.2.7 We give some examples of Borel sets.

- Every closed subset of $\mathbb{R}$ is a Borel set because every closed subset of $\mathbb{R}$ is the complement of an open subset of $\mathbb{R}$.
- Every countable subset of $\mathbb{R}$ is a Borel set because if $B = \{x_1, x_2, \ldots\}$, then $B = \bigcup_{k=1}^{\infty} \{x_k\}$, which is a Borel set because it is a countable union of closed sets which are Borel.
- Every half-open interval $[a, b)$ is a Borel set because $[a, b) = \bigcap_{k=1}^{\infty} (a - 1/k, b)$.
- If $f : \mathbb{R} \to \mathbb{R}$ is a function, then the set of points at which f is continuous is the intersection of a sequence of open sets and thus is a Borel set.

We will see later that there exist subsets of $\mathbb{R}$ that are not Borel sets. However, any subset of $\mathbb{R}$ that you can write down concretely is a Borel set.

1.2.3 Inverse Images

We take a detour to discuss the definition and algebraic properties of *inverse images*.

Definition 1.2.8 If $f : X \to Y$ is a function and $A \subseteq Y$, then the **inverse image** of A under f, denoted $f^{-1}(A)$, is defined by $f^{-1}(A) = \{x \in X : f(x) \in A\}$.

Lemma 1.2.9 Suppose $f : X \to Y$ is a function. Then

1. $f^{-1}(Y \setminus A) = X \setminus f^{-1}(A)$ for every $A \subset Y$;
2. $f^{-1}\left(\bigcup_{A\in\mathcal{A}} A\right) = \bigcup_{A\in\mathcal{A}} f^{-1}(A)$ for every set $\mathcal{A}$ of subsets of Y;
3. $f^{-1}\left(\bigcap_{A\in\mathcal{A}} A\right) = \bigcap_{A\in\mathcal{A}} f^{-1}(A)$ for every set $\mathcal{A}$ of subsets of Y.

Proof. To prove (a), suppose $A \subset Y$. For $x \in X$ we have

$$x \in f^{-1}(Y \setminus A) \iff f(x) \in Y \setminus A \iff f(x) \notin A \iff x \notin f^{-1}(A) \iff x \in X \setminus f^{-1}(A).$$

To prove (b), suppose $\mathcal{A}$ is a set of subsets of Y. Then

$$\begin{aligned} x \in f^{-1}\left(\bigcup_{A\in\mathcal{A}} A\right) &\iff f(x) \in \bigcup_{A\in\mathcal{A}} A \\ &\iff f(x) \in A \text{ for some } A \in \mathcal{A} \\ &\iff x \in f^{-1}(A) \text{ for some } A \in \mathcal{A} \\ &\iff x \in \bigcup_{A\in\mathcal{A}} f^{-1}(A). \end{aligned}$$

Part (c) can be proven in the same fashion as (b). □

Lemma 1.2.10 For $f : X \to Y$ and $g : Y \to W$, we have

$$(g \circ f)^{-1}(A) = f^{-1}(g^{-1}(A))$$

for every $A \subseteq W$.

Proof. Suppose $A \subset W$. For $x \in X$ we have

$$x \in (g \circ f)^{-1}(A) \Leftrightarrow (g \circ f)(x) \in A \Leftrightarrow g(f(x)) \in A \Leftrightarrow f(x) \in g^{-1}(A) \Leftrightarrow x \in f^{-1}\left(g^{-1}(A)\right)$$

Thus $(g \circ f)^{-1}(A) = f^{-1}\left(g^{-1}(A)\right)$. $\square$

1.2.4 Measurable Functions

The next definition tells us which real-valued functions behave reasonably with respect to a σ-algebra on their domain.

Definition 1.2.11 A **measurable space** is an ordered pair $(X, \mathcal{S})$, where X is a set and $\mathcal{S}$ is a σ-algebra on X. An element of $\mathcal{S}$ is called an S**-measurable set**, or just a **measurable set** if S is clear from the context.

A *measurable function* is a function between two measurable spaces such that the pre-image of any measurable set in the codomain is a measurable set in the domain. Intuitively, it is a function that preserves the structure of measurable sets under pre-images.

Definition 1.2.12 Suppose $(X, \mathcal{S})$ is a measurable space. A function $f : X \to \mathbb{R}$ is called $\mathcal{S}$**-measurable** (or just **measurable** if $\mathcal{S}$ is clear from the context) if $f^{-1}(B) \in \mathcal{S}$ for every Borel set $B \subseteq \mathbb{R}$.

Example 1.2.13 If $\mathcal{S} = \{\varnothing, X\}$, then the only $\mathcal{S}$-measurable functions from X to $\mathbb{R}$ are the constant functions.

- It is easy to see that the pre-image of a Borel set $B \subseteq \mathbb{R}$ under a constant function is either $\varnothing$ (if the Borel set does not contain the constant value) or X (if the Borel set contains the constant value).
- If $f : X \to \mathbb{R}$ is not constant, say range$(f) = \{-1, 1\}$. Then $f^{-1}((-2, 0)) \neq \varnothing$ (because there exists $A_1 \in X$ such that $f(x) = -1$) and $f^{-1}((-2, 0)) \neq X$, so $f^{-1}((-2, 0)) \notin \mathcal{S}$ (because there exists $A_2 \in X$ such that $f(x) \neq -1$)

Example 1.2.14 If $\mathcal{S}$ is the set of all subsets of X, then every function from X to $\mathbb{R}$ is $\mathcal{S}$ measurable. Indeed, the pre-image of f for any Borel set $B \subseteq \mathbb{R}$ is always a subset of X, which is contained in $\mathcal{S}$.

Example 1.2.15 If $\mathcal{S} = \{\varnothing, (-\infty, 0), [0, \infty), \mathbb{R}\}$ (which is a σ-algebra on $\mathbb{R}$), then a function $f : \mathbb{R} \to \mathbb{R}$ is $\mathcal{S}$-measurable if and only if f is constant on $(-\infty, 0)$ and f is constant on $[0,\infty)$. This can be viewed as a generalization of the first example.

► ***Sufficient Condition for Measurable Functions***

The definition of an $\mathcal{S}$-measurable function requires the inverse image of every Borel subset of $\mathbb{R}$ to be in $\mathcal{S}$. The next result shows that to verify that a function is $\mathcal{S}$-measurable, we can check the inverse images of a much smaller collection of subsets of $\mathbb{R}$. In words, we just need to check whether the pre-image of all open intervals (a,∞) under f is in $\mathcal{S}$.[2]

Proposition 1.2.16 Suppose $(X,\mathcal{S})$ is a measurable space and $f : X \to \mathbb{R}$ is a function such that

$$f^{-1}((a,\infty)) = \{x \in X : f(x) > a\} \in \mathcal{S}$$

for all $a \in \mathbb{R}$. Then f is an $\mathcal{S}$-measurable function.

Proof. Let $\mathcal{T} = \{A \subseteq \mathbb{R} : f^{-1}(A) \in \mathcal{S}\}$ be the set of all subsets of $\mathbb{R}$ whose pre-image under f is in the σ-algebra $\mathcal{S}$. We want to show that every Borel set of $\mathbb{R}$ is in $\mathcal{T}$. To do this, we will first show that $\mathcal{T}$ is a σ-algebra on $\mathbb{R}$.

- Certainly $\varnothing \in \mathcal{T}$ as $f^{-1}(\varnothing) = \varnothing \in \mathcal{S}$.
- If $A \in \mathcal{T}$, then $f^{-1}(A) \in \mathcal{S}$. By Lemma 1.2.9(1),

  $$f^{-1}(\mathbb{R} \setminus A) = X \setminus f^{-1}(A) \in \mathcal{S}.$$

  Thus, $\mathbb{R} \setminus \mathcal{A} \in \mathcal{T}$. In other words, $\mathcal{T}$ is closed under complementation.
- If $A_1, A_2, \ldots \in \mathcal{T}$, then $f^{-1}(A_1), f^{-1}(A_2), \ldots \in \mathcal{S}$. By Lemma 1.2.9(2)

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

  Thus, $\bigcup_{k=1}^{\infty} A_k \in \mathcal{T}$ and $\mathcal{T}$ is closed under countable unions. Thus, $\mathcal{T}$ is a σ-algebra on $\mathbb{R}$.

By hypothesis, $\mathcal{T}$ contains $\{(a,\infty) : a \in \mathbb{R}\}$. Because $\mathcal{T}$ is closed under complementation and $(-\infty,b] = \mathbb{R} \setminus (b,\infty)$, $\mathcal{T}$ also contains $\{(-\infty,b] : b \in \mathbb{R}\}$. Because the σ-algebra $\mathcal{T}$ is closed under finite intersection and $(a,b] = (-\infty,b] \cap (a,\infty)$, we see that $\mathcal{T}$ contains $\{(a,b] : a,b \in \mathbb{R}\}$. Because

$$(a,b) = \bigcup_{k=1}^{\infty} (a, b-1/k] \qquad \text{and} \qquad (-\infty,b) = \bigcup_{k=1}^{\infty} (-k, b-1/k]$$

[2] Actually, we could replace the collection of sets $\{(a,\infty) : a \in \mathbb{R}\}$ by any collection of subsets of $\mathbb{R}$ such that the smallest σ-algebra containing that collection contains the Borel subsets of $\mathbb{R}$.

and $\mathcal{T}$ is closed under countable unions, we can conclude that $\mathcal{T}$ contains every open subset of $\mathbb{R}$ by the arbitrary choice of a and b. Thus, the σ-algebra $\mathcal{T}$ contains the smallest σ-algebra on $\mathbb{R}$ that contains all open subsets of $\mathbb{R}$. In other words, $\mathcal{T}$ contains every Borel subset of $\mathbb{R}$. Thus, f is an $\mathcal{S}$-measurable function. $\square$

► ***Borel Measurable Functions***

We have been dealing with $\mathcal{S}$-measurable functions from X to $\mathbb{R}$ in the context of an arbitrary set X and a σ-algebra $\mathcal{S}$ on X. An important special case of this setup is when X is a Borel subset of $\mathbb{R}$ and $\mathcal{S}$ is the set of Borel subsets of $\mathbb{R}$ that are contained in X. In this special case, the $\mathcal{S}$-measurable functions are called Borel measurable.

Definition 1.2.17 Suppose $X \subseteq \mathbb{R}$. A function $f : X \to \mathbb{R}$ is called **Borel measurable** if $f^{-1}(B)$ is a Borel set for every Borel set $B \subseteq \mathbb{R}$.

Remark 1.2.18 Let's connect Borel measurable function and Borel sets.

- If $X \subseteq \mathbb{R}$ and there exists a Borel measurable function $f : X \to \mathbb{R}$, then X must be a Borel set (because $X = f^{-1}(\mathbb{R})$).
- If $X \subseteq \mathbb{R}$ and $f : X \to \mathbb{R}$ is a function, then f is a Borel measurable function if and only if $f^{-1}((a, \infty))$ is a Borel set for every $a \in \mathbb{R}$ (by Proposition 1.2.16).

Suppose X is a set and $f : X \to \mathbb{R}$ is a function. The measurability of f depends upon the choice of a σ-algebra on X. If the σ-algebra is called $\mathcal{S}$, then we can discuss whether f is an $\mathcal{S}$-measurable function. If X is a Borel subset of $\mathbb{R}$, then $\mathcal{S}$ might be the set of Borel sets contained in X, in which case the phase *Borel measurable* means the same as $\mathcal{S}$-measurable. However, whether or not $\mathcal{S}$ is a collection of Borel sets, we consider inverse images of Borel subsets of $\mathbb{R}$ when determining whether a function is $\mathcal{S}$-measurable.

► ***More Results on Borel Measurable Functions***

Proposition 1.2.19 Every continuous real-valued function defined on a Borel subset of $\mathbb{R}$ is a Borel measurable function.

Proof. Suppose $X \subseteq \mathbb{R}$ is a Borel set and $f : X \to \mathbb{R}$ is continuous. To prove that f is Borel measurable, fix $a \in \mathbb{R}$. If $x \in X$ and $f(x) > a$, then (by the continuity of f), there exists $\delta_x > 0$ such that $f(y) > a$ for all $y \in (x - \delta_x, x + \delta_x) \cap X$. Thus,

$$f^{-1}((a, \infty)) = \left(\bigcup_{x \in f^{-1}((a,\infty))} (x - \delta_x, x + \delta) \right) \cap X.$$

The union inside the large parentheses above is an open subset of $\mathbb{R}$; hence its intersection with X is a Borel set. Thus, we can conclude that $f^{-1}((a, \infty))$ is a Borel set. By Proposition 1.2.16, f is a

Borel measurable function. □

Proposition 1.2.20 Every increasing function defined on a Borel subset of $\mathbb{R}$ is a Borel measurable function.

Proof. Suppose $X \subset \mathbb{R}$ is a Borel set and $f : X \to \mathbb{R}$ is increasing. To prove that f is Borel measurable, fix $a \in \mathbb{R}$. Let $b = \inf f^{-1}((a, \infty))$. Then it is easy to see that

$$f^{-1}((a, \infty)) = (b, \infty) \cap X \quad \text{or} \quad f^{-1}((a, \infty)) = [b, \infty) \cap X.$$

Either way, we can conclude that $f^{-1}((a, \infty))$ is a Borel set. Now Proposition 1.2.16 implies that f is a Borel measurable function. □

Proposition 1.2.21 Suppose $(X, \mathcal{S})$ is a measurable space and $f : X \to \mathbb{R}$ is an $\mathcal{S}$-measurable function. Suppose g is a real-valued Borel measurable function defined on a subset of $\mathbb{R}$ that includes the range of f. Then $g \circ f : X \to \mathbb{R}$ is an $\mathcal{S}$-measurable function.

Proof. Suppose $B \subset \mathbb{R}$ is a Borel set. Then (Lemma 1.2.9)

$$(g \circ f)^{-1}(B) = f^{-1}\left(g^{-1}(B)\right).$$

Because g is a Borel measurable function, $g^{-1}(B)$ is a Borel subset of $\mathbb{R}$. Because f is an $\mathcal{S}$-measurable function, $f^{-1}\left(g^{-1}(B)\right) \in \mathcal{S}$. Thus the equation above implies that $(g \circ f)^{-1}(B) \in \mathcal{S}$. Thus $g \circ f$ is an $\mathcal{S}$-measurable function. □

Proposition 1.2.22 Suppose $(X, \mathcal{S})$ is a measurable space and $f, g : X \to \mathbb{R}$ are $\mathcal{S}$-measurable. Then

- $f + g, f - g, fg$ are all $\mathcal{S}$-measurable functions.
- If $g(x) \neq 0$ for all $x \in X$, then f/g is an $\mathcal{S}$-measurable function.

Proof. To show $f + g$ is $\mathcal{S}$-measurable, suppose $a \in \mathbb{R}$. We will show that

$$(f + g)^{-1}((a, \infty)) = \bigcup_{r \in \mathbf{Q}} \left(f^{-1}((r, \infty)) \cap g^{-1}((a - r, \infty))\right),$$

which implies that $(f + g)^{-1}((a, \infty)) \in \mathcal{S}$. We will omit the details here. □

Proposition 1.2.23 Suppose $(X, \mathcal{S})$ is a measurable space and $f_1, f_2, \ldots$ is a sequence of $\mathcal{S}$-measurable functions from X to $\mathbb{R}$. Suppose $\lim_{k\to\infty} f_k(x)$ exists for each $x \in X$. Define $f : X \to \mathbb{R}$ by

$$f(x) = \lim_{k\to\infty} f_k(x).$$

Then f is an $\mathcal{S}$-measurable function.

Proof. Suppose $a \in \mathbb{R}$. We will show that

$$f^{-1}((a,\infty)) = \bigcup_{j=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} f_k^{-1}\left(\left(a + \frac{1}{j}, \infty\right)\right) \tag{1.8}$$

which implies that $f^{-1}((a,\infty)) \in \mathcal{S}$.

To prove (1.8), first suppose $x \in f^{-1}((a,\infty))$. Thus there exists $j \in \mathbb{Z}^+$such that $f(x) > a + \frac{1}{j}$. The definition of limit now implies that there exists $m \in \mathbb{Z}^+$such that $f_k(x) > a + \frac{1}{j}$ for all $k \geq m$. Thus x is in the right side of (1.8), proving that the left side of (1.8) is contained in the right side.

To prove the inclusion in the other direction, suppose x is in the right side of (1.8). Thus there exist $j, m \in \mathbb{Z}^+$such that $f_k(x) > a + \frac{1}{j}$ for all $k \geq m$. Taking the limit as $k \to \infty$, we see that $f(x) \geq a + \frac{1}{j} > a$. Thus x is in the left side of (1.8), completing the proof of (1.8). Thus f is an $\mathcal{S}$-measurable function. □

Let's summarize the results we learned here:

- Every continuous function defined on a Borel set is Borel measurable.
- Every increasing function defined on a Borel set is Borel measurable.
- Composition of measurable functions is measurable.
- The sum, difference, product, and quotient of two measurable functions are measurable.
- The point-wise limit of a sequence of $\mathcal{S}$-measurable functions is $\mathcal{S}$-measurable.

The last one is a highly desirable property (recall that the set of Riemann integrable functions on some interval is not closed under taking point-wise limits.)

► ***Generalizing to Extended Reals***

Occasionally we need to consider functions defined on extended reals, i.e., $[-\infty, \infty] = \mathbb{R} \cup \{-\infty, \infty\}$.

Definition 1.2.24 A subset of $[-\infty, \infty]$ is called a **Borel set** if its intersection with $\mathbb{R}$ is a Borel set.

Definition 1.2.25 Let $(X, \mathcal{S})$ be a measurable space. A function $f : X \to [-\infty, \infty]$ is called **$\mathcal{S}$-measurable** if $f^{-1}(B) \in \mathcal{S}$ for every Borel set $B \subseteq [-\infty, \infty]$.

Proposition 1.2.26 Suppose $(X, \mathcal{S})$ is a measurable space and $f : X \to [-\infty, \infty]$ is a function such that

$$f^{-1}((a, \infty]) \in \mathcal{S}$$

for all $a \in \mathbb{R}$. Then f is an $\mathcal{S}$-measurable function.

Proof. Omitted. □

We end this section by showing that the point-wise infimum and point-wise supremum of a sequence of $\mathcal{S}$-measurable functions is $\mathcal{S}$-measurable.

Proposition 1.2.27 Suppose $(X, \mathcal{S})$ is a measurable space and $f_1, f_2, \ldots$ is a sequence of $\mathcal{S}$-measurable functions from X to $[-\infty, \infty]$. Define $g, h : X \to [-\infty, \infty]$ by

$$g(x) = \inf\left\{f_k(x) : k \in \mathbb{Z}^+\right\} \quad \text{and} \quad h(x) = \sup\left\{f_k(x) : k \in \mathbb{Z}^+\right\}.$$

Then g and h are $\mathcal{S}$-measurable functions.

Proof. Let $a \in \mathbb{R}$. The definition of the supremum implies that

$$h^{-1}((a, \infty]) = \bigcup_{k=1}^{\infty} f_k^{-1}((a, \infty]),$$

as you should verify. The equation above, along with Proposition 1.2.26, implies that h is an

$\mathcal{S}$-measurable function. Next, note that

$$g(x) = -\sup\left\{-f_k(x) : k \in \mathbb{Z}^+\right\}$$

for all $x \in X$. Thus the result about the supremum implies that g is an $\mathcal{S}$-measurable function. $\square$

1.3 Measures and Their Properties

1.3.1 Definition and Examples

Definition 1.3.1 Suppose X is a set and $\mathcal{S}$ is a σ-algebra on X. A **measure** on $(X, \mathcal{S})$ is a function $\mu : \mathcal{S} \to [0, \infty]$ such that $\mu(\varnothing) = 0$ and

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

for every disjoint sequence $E_1, E_2, \ldots$ of sets in $\mathcal{S}$.

The concept of a measure as defined here is sometimes called a *positive* measure.

Example 1.3.2 Let X be a set. The **counting measure** is the measure μ defined on the σ-algebra of all subsets of X by setting $\mu(E) = n$ if E is a finite set containing exactly n elements and $\mu(E) = \infty$ if E is not a finite set.

Example 1.3.3 Suppose X is a set, $\mathcal{S}$ a σ-algebra on X, and $c \in X$. Define a **Dirac measure** δ_c on $(X, \mathcal{S})$ by

$$\delta_c(E) = \begin{cases} 1 & c \in E \\ 0 & c \notin E. \end{cases}$$

Example 1.3.4 Suppose $\mathcal{S}$ is the σ-algebra on $\mathbb{R}$ consisting of all subsets of $\mathbb{R}$. Then the function take takes a set $E \subseteq \mathbb{R}$ to $|E|$ (the outer measure of E) is not a measure because it is not finitely additive. (Proposition 1.1.14)

Example 1.3.5 Suppose $\mathcal{B}$ is the σ-algebra on $\mathbb{R}$ consisting of all Borel subsets of $\mathbb{R}$. In the next section, we will show that the outer measure is a measure on $(\mathbb{R}, \mathcal{B})$.

Definition 1.3.6 A **measure space** is an ordered triple $(X, \mathcal{S}, \mu)$, where X is a set, $\mathcal{S}$ is a σ-algebra on X, and μ is a measure on $(X, \mathcal{S})$.

1.3.2 Properties of Measures

Lemma 1.3.7 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $D, E \in \mathcal{S}$ are such that $D \subseteq E$. Then

1. $\mu(D) \leq \mu(E)$;
2. $\mu(E \setminus D) = \mu(E) - \mu(D)$ provided that $\mu(D) < \infty$. [a]

[a] The hypothesis that $\mu(D) < \infty$ is needed here to avoid undefined expressions $\infty - \infty$.

Proof. Because $E = D \cup (E \setminus D)$ and this is a disjoint union, we have

$$\mu(E) = \mu(D) + \mu(E \setminus D) \geq \mu(D),$$

which proves (1). If $\mu(D) < \infty$, then subtracting $\mu(D)$ from both sides proves (2). □

The *countable additivity* property of measure applies only to *disjoint* countable unions. The following *countable subadditivity* property applies to countable unions that may not be disjoint unions.

Lemma 1.3.8 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $E_1, E_2, \ldots \in \mathcal{S}$. Then

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

Proof. Let $D_1 = \varnothing$ and $D_k = E_1 \cup \cdots \cup E_{k-1}$ for $k \geq 2$. Then

$$E_1 \setminus D_1, \quad E_2 \setminus D_2, \quad E_3 \setminus D_3, \quad \ldots$$

is a disjoint sequence of subsets of X whose union equals $\bigcup_{k=1}^{\infty} E_k$. Thus

$$\begin{aligned} \mu\left(\bigcup_{k=1}^{\infty} E_k\right) &= \mu\left(\bigcup_{k=1}^{\infty} (E_k \setminus D_k)\right) \\ &= \sum_{k=1}^{\infty} \mu\left(E_k \setminus D_k\right) \\ &\leq \sum_{k=1}^{\infty} \mu\left(E_k\right), \end{aligned}$$

where the second line above follows from the countable additivity of μ and the last line above follows from Lemma 1.3.7.

The next result shows that measures behave well with increasing unions.

Lemma 1.3.9 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $E_1 \subseteq E_2 \subseteq \cdots$ is an increasing sequence of sets in $\mathcal{S}$. Then

$$\mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \lim_{k\to\infty} \mu(E_k).$$

Proof. If $\mu(E_k) = \infty$ for some $k \in \mathbb{Z}^+$, then the equation above holds because both sides equal ∞. Now assume $\mu(E_k) < \infty$ for all $k \in \mathbb{Z}^+$. For convenience, let $E_0 = \varnothing$. Then

$$\bigcup_{k=1}^{\infty} E_k = \bigcup_{j=1}^{\infty} (E_j \setminus E_{j-1}),$$

where the union on the right side is a disjoint union. Thus

$$\begin{aligned}
\mu\left(\bigcup_{k=1}^{\infty} E_k\right) &= \sum_{j=1}^{\infty} \mu(E_j \setminus E_{j-1}) \\
&= \lim_{k\to\infty} \sum_{j=1}^{k} \mu(E_j \setminus E_{j-1}) \\
&= \lim_{k\to\infty} \sum_{j=1}^{k} \left(\mu(E_j) - \mu(E_{j-1})\right) \\
&= \lim_{k\to\infty} \mu(E_k) && \text{alternating sum}
\end{aligned}$$

as desired. □

The next result shows that measures also behave well with decreasing intersections. However, note that the hypothesis $\mu(E_1) < \infty$ cannot be removed.

Lemma 1.3.10 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $E_1 \supseteq E_2 \supseteq \cdots$ is a decreasing sequence of sets in $\mathcal{S}$, with $\mu(E_1) < \infty$. Then

$$\mu\left(\bigcap_{k=1}^{\infty} E_k\right) = \lim_{k\to\infty} \mu(E_k)$$

Proof. One of De Morgan's Laws tells us that

$$E_1 \setminus \bigcap_{k=1}^{\infty} E_k = \bigcup_{k=1}^{\infty} (E_1 \setminus E_k)$$

Now $E_1 \setminus E_1 \subseteq E_1 \setminus E_2 \subseteq E_1 \setminus E_3 \subseteq \cdots$ is an increasing sequence of sets in $\mathcal{S}$. Applying Lemma 1.3.9 to the equation above gives

$$\mu\left(E_1 \setminus \bigcap_{k=1}^{\infty} E_k\right) = \lim_{k\to\infty} \mu\left(E_1 \setminus E_k\right)$$

Use Lemma 1.3.7 to rewrite the equation above as

$$\mu\left(E_1\right) - \mu\left(\bigcap_{k=1}^{\infty} E_k\right) = \mu\left(E_1\right) - \lim_{k\to\infty} \mu\left(E_k\right),$$

which implies our desired result.

The next result is intuitively plausible-we expect that the measure of the union of two sets equals the measure of the first set plus the measure of the second set minus the measure of the set that has been counted twice.

Lemma 1.3.11 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $D, E \in \mathcal{S}$ with

$$\mu(D \cap E) < \infty.$$

Then

$$\mu(D \cup E) = \mu(D) + \mu(E) - \mu(D \cap E).$$

Proof. We have

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

The right side of the equation above is a disjoint union. Thus

$$\begin{aligned}
\mu(D \cup E) &= \mu(D \setminus (D \cap E)) + \mu(E \setminus (D \cap E)) + \mu(D \cap E) \\
&= (\mu(D) - \mu(D \cap E)) + (\mu(E) - \mu(D \cap E)) + \mu(D \cap E) \\
&= \mu(D) + \mu(E) - \mu(D \cap E),
\end{aligned}$$

as desired. □

1.4 Lebesgue Measure

Recall from Proposition 1.1.14 that there exists disjoint sets $A, B \in\subseteq \mathbb{R}$ such that $|A \cup B| \neq |A| + |B|$. Thus outer measure, despite its name, is not a measure on the σ-algebra of all subsets of $\mathbb{R}$. Our main goal in this section is to prove that outer measure, when restricted to the Borel subsets of $\mathbb{R}$, is a measure. Throughout this section, be careful about trying to simplify proofs by applying properties of measures to outer measure, even if those properties seem intuitively plausible. For example, there are subsets $A \subseteq B \subseteq \mathbb{R}$ with $|A| < \infty$ but $|B \setminus A| \neq |B| - |A|$ (check Lemma 1.3.7).

Lemma 1.4.1 Suppose A and G are disjoint subsets of $\mathbb{R}$ and G is open. Then

$$|A \cup G| = |A| + |G|.$$

Proof. If $|G| = \infty$, then $|A \cup G| = \infty = |A| + |G|$. Now assume $|G| < \infty$. By subadditivity of outer measure (Lemma 1.1.7), $|A \cup G| \leq |A| + |G|$. Let's prove the other direction.

First, consider the case where $G = (a, b)$ for some $a, b \in \mathbb{R}$ with $a < b$. We can assume that $a, b \in A$ (because changing a set by at most two points does not change its outer measure). Let $I_1, I_2, \ldots$ be a sequence of open intervals whose union contains $A \cup G$. For each $n \in \mathbb{Z}^+$, let

$$J_n = I_n \cap (-\infty, a), \quad K_n = I_n \cap (a, b), \quad L_n = I_n \cap (b, \infty).$$

Then $\ell(I_n) = \ell(J_n) + \ell(K_n) + \ell(L_n)$. Now $J_1, L_1, J_2, L_2, \ldots$ is a sequence of open intervals whose union contains A and $K_1, K_2, \ldots$ is a sequence of open intervals whose union contains G. Thus,

$$\sum_{n=1}^{\infty} \ell(I_n) = \sum_{n=1}^{\infty} (\ell(J_n) + \ell(L_n)) + \sum_{n=1}^{\infty} \ell(K_n) \geq |A| + |G|.$$

The inequality above implies that $|A \cup G| \geq |A| + |G|$, completing the proof that $|A \cup G| = |A| + |G|$ in this special case.

Using induction on m, we can now conclude that if $m \in \mathbb{Z}^+$ and G is a union of m disjoint open intervals that are all disjoint from A, then $|A \cup G| = |A| + |G|$.

Now suppose G is an arbitrary open subset of $\mathbb{R}$ that is disjoint from A. Then $G = \bigcup_{n=1}^{\infty} I_n$ for some sequence of disjoint open intervals $I_1, I_2, \ldots$, each of which is disjoint from A. For each $m \in \mathbb{Z}^+$, we have

$$|A \cup G| \geq \left| A \cup \left(\bigcup_{n=1}^{m} I_n \right) \right| = |A| + \sum_{n=1}^{m} \ell(I_n) \geq |A| + |G|,$$

completing the proof that $|A \cup G| = |A| + |G|$. $\square$

Lemma 1.4.2 Suppose A and F are disjoint subsets of $\mathbb{R}$ and F is closed. Then

$$|A \cup G| = |A| + |F|.$$

Proof. Suppose $I_1, I_2, \ldots$ is a sequence of open intervals whose union contains $A \cup F$. Let $G = \bigcup_{k=1}^{\infty} I_k$. Then G is an open set with $A \cup F \subseteq G$. Hence $A \subseteq G \setminus F$, which implies that $|A| \leq |G \setminus F|$.

Because $G \setminus F = G \cap (\mathbb{R} \setminus F)$, we know that $G \setminus F$ is an open set. Apply Lemma 1.4.1 to the disjoint union $G = F \cup (G \setminus F)$ gives $|G| = |F| + |G \setminus F|$. This implies that

$$|A| \leq |G \setminus F|$$
$$|A| + |F| \leq |G \setminus F| + |F| = |G| \leq \sum_{k=1}^{\infty} \ell(I_k).$$

Recall the definition of outer measure:

$$|A \cup F| = \inf \left\{ \sum_{k=1}^{\infty} \ell(I_k) \;\middle|\; I_1, I_2, \ldots \text{ are open intervals such that } A \cup F \subseteq \bigcup_{k=1}^{\infty} I_k \right\}.$$

Now for all covers $\bigcup_{k=1}^{\infty} I_k$ of $A \cup F$, we have

$$|A| + |F| \leq \sum_{k=1}^{\infty} \ell(I_k).$$

Then $|A| + |F|$ must also be less than or equal to the infimum over all $\sum_{k=1}^{\infty} \ell(I_k)$. Hence, we conclude that $|A| + |F| \leq |A \cup F|$, completing the proof for $|A| + |F| = |A \cup F|$. □

The next result shows us that we could approximate a Borel set by a closed set.

Proposition 1.4.3 Suppose $B \subseteq \mathbb{R}$ is a Borel set. Then for every $\varepsilon > 0$, there exists a closed set $F \subseteq B$ such that $|B \setminus F| < \varepsilon$.

Proof. Define

$$\mathcal{L} = \{D \subseteq \mathbb{R} \mid \forall \varepsilon > 0 : \text{there exists a closed set } F \subseteq D \text{ such that } |D \setminus F| < \varepsilon\}.$$

The strategy of the proof is to show that $\mathcal{L}$ is a σ-algebra. Then because $\mathcal{L}$ contains every closed subset of $\mathbb{R}$ (if $D \subseteq \mathbb{R}$ is closed, take $F = D$ in the definition of $\mathcal{L}$), by taking complements we can conclude that $\mathcal{L}$ contains every open subset of $\mathbb{R}$ and thus every Borel subset of $\mathbb{R}$.

To get started with proving that $\mathcal{L}$ is a σ-algebra, we want to prove that $\mathcal{L}$ is closed under countable intersections. Let $D_1, D_2, \ldots$ be a sequence in $\mathcal{L}$. Let $\varepsilon > 0$. For each $k \in \mathbb{Z}^+$, there exists a closed set F_k such that

$$F_k \subseteq D_k \quad \text{and} \quad |D_k \setminus F_k| < \frac{\varepsilon}{2^k}.$$

Thus $\bigcap_{k=1}^{\infty} F_k$ is closed and

$$\bigcap_{k=1}^{\infty} F_k \subseteq \bigcap_{k=1}^{\infty} D_k \quad \text{and} \quad \left(\bigcap_{k=1}^{\infty} D_k\right) \setminus \left(\bigcap_{k=1}^{\infty} F_k\right) \subseteq \bigcup_{k=1}^{\infty} (D_k \setminus F_k).$$

By the countable additivity of outer measure, we get

$$\left|\left(\bigcap_{k=1}^{\infty} D_k\right) \setminus \left(\bigcap_{k=1}^{\infty} F_k\right)\right| \leq \left|\bigcup_{k=1}^{\infty} (D_k \setminus F_k)\right| \leq \sum_{k=1}^{\infty} |D_k \setminus F_k| < \varepsilon.$$

Thus $\bigcap_{k=1}^{\infty} D_k \in \mathcal{L}$, proving that $\mathcal{L}$ is closed under countable intersections.

Now we want to prove that $\mathcal{L}$ is closed under complementation. Suppose $D \in \mathcal{L}$ and $\varepsilon > 0$. We want to show that there is a closed subset of $\mathbb{R} \setminus D$ whose set difference with $\mathbb{R} \setminus D$ has outer measure less than ε, which will allow us to conclude that $\mathbb{R} \setminus D \in \mathcal{L}$.

First we consider the case where $|D| < \infty$. Let $F \subseteq D$ be closed set such that $|D \setminus F| < \varepsilon/2$. The definition of outer measure implies that there exists an open set G such that $D \subseteq G$ and $|G| < |D| + \varepsilon/2$. Now $\mathbb{R} \setminus G$ is a closed set and $\mathbb{R} \setminus G \subseteq \mathbb{R} \setminus D$. Moreover, $(\mathbb{R} \setminus D) \setminus (\mathbb{R} \setminus G) = G \setminus D \subseteq G \setminus F$. Thus,

$$\begin{aligned}
|(\mathbb{R} \setminus D) \setminus (\mathbb{R} \setminus G)| &\subseteq |G \setminus F| \\
&= |G| - |F| && \text{Lemma 1.4.2; } F \text{ is closed} \\
&= (|G| - |D|) + (|D| - |F|) \\
&< \frac{\varepsilon}{2} + |D \setminus F| && \text{subadditivity applied to } D = (D \setminus F) \cup F \\
&< \varepsilon.
\end{aligned}$$

This shows that $\mathbb{R} \setminus D \in \mathcal{L}$ as desired. Now, still assuming that $D \in \mathcal{L}$ and $\varepsilon > 0$, we consider the case where $|D| = \infty$. For $k \in \mathbb{Z}^+$, let $D_k = D \cap [-k, k]$. Because $D_k \in \mathcal{L}$ and $|D_k| < \infty$, the previous case implies that $\mathbb{R} \setminus D_k \in \mathcal{L}$. Clearly, $D = \bigcup_{k=1}^{\infty} D_k$. Thus

$$\mathbb{R} \setminus D = \bigcap_{k=1}^{\infty} (\mathbb{R} \setminus D_k).$$

Since $\mathcal{L}$ is closed under countable intersections, the equation above implies that $\mathbb{R} \setminus D \in \mathcal{L}$, which completes the proof that $\mathcal{L}$ is a σ-algebra. □

Now we can prove that the outer measure of the disjoint union of two sets is what we expect if at least one of the two sets is a Borel set.

Proposition 1.4.4 Suppose A, B are disjoint subsets of $\mathbb{R}$ and B is a Borel set. Then

$$|A \cup B| = |A| + |B|.$$

Proof. Let $\varepsilon > 0$. Let F be a closed set such that $F \subset B$ and $|B \setminus F| < \varepsilon$ (Proposition 1.4.3). Thus

$$\begin{aligned} |A \cup B| &\geq |A \cup F| \\ &= |A| + |F| \\ &= |A| + |B| - |B \setminus F| \\ &\geq |A| + |B| - \varepsilon, \end{aligned}$$

where the second and third lines above follow from Lemma 1.4.2 [use $B = (B \setminus F) \cup F$ for the third line]. Because the inequality above holds for all $\varepsilon > 0$, we have $|A \cup B| \geq |A| + |B|$, which implies that $|A \cup B| = |A| + |B|$. □

Proposition 1.4.5 There is exists $B \subset \mathbb{R}$ such that $|B| < \infty$ and B is not a Borel set.

Proof. Proposition 1.1.14 showed that there exist disjoint sets $A, B \subset \mathbb{R}$ such that $|A \cup B| \neq |A| + |B|$. For any such sets, we must have $|B| < \infty$ because otherwise both $|A \cup B|$ and $|A| + |B|$ equal ∞ (as follows from the inequality $|B| \leq |A \cup B|$). Now 1.4.4 implies that B is not a Borel set.

Theorem 1.4.6 Outer measure is a measure on $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the σ-algebra of Borel subsets of $\mathbb{R}$.

Proof. Suppose $B_1, B_2, \dots$ is a disjoint sequence of Borel subsets of $\mathbb{R}$. For each $n \in \mathbb{Z}^+$, we have

$$\left| \bigcup_{k=1}^{\infty} B_k \right| \geq \left| \bigcup_{k=1}^{n} B_k \right| = \sum_{k=1}^{n} |B_k|.$$

Taking limit as $n \to \infty$, we have $|\bigcup_{k=1}^{\infty} B_k| \geq \sum_{k=1}^{\infty} |B_k|$. The other direction follows from countable subadditivity of outer measure. Hence, $|\bigcup_{k=1}^{\infty} B_k| = \sum_{k=1}^{\infty} |B_k|$ as desired. □

Definition 1.4.7 **Lebesgue measure** is the measure on $(\mathbb{R}, \mathcal{B})$ where $\mathcal{B}$ is the Borel algebra of $\mathbb{R}$, that assigns to each Borel set its outer measure.

1.4.1 Lebesgue Measurable Sets

We have accomplished the major goal of this section, which was to show that outer measure restricted to Borel sets is a measure. As we will see in this subsection, outer measure is actually a measure on a somewhat larger class of sets called the *Lebesgue measurable sets*.

There are many equivalent definitions of Lebesgue measurable sets; the definition in one approach becomes a theorem in another approach. The approach chosen here has the advantage of emphasizing that a Lebesgue measurable set differs from a Borel set by a set with outer measure 0. The intuition here is that sets with outer measure 0 should be considered small and do not matter much.

Definition 1.4.8 A set $A \subseteq \mathbb{R}$ is called **Lebesgue measurable** if there exists a Borel set $B \subseteq A$ such that $|A \setminus B| = 0$.

Every Borel set is a Lebesgue measurable set because if $A \subseteq \mathbb{R}$ is a Borel set, then we can take $B = A$ in the definition above.

The next result gives several equivalent conditions for being Lebesgue measurable. The equivalence of (1) and (4) are out definition and thus is not discussed in the proof.

Although there exist Lebesgue measurable sets that are not Borel sets, you are unlikely to encounter one. The most important application of the result below is that if $A \subseteq \mathbb{R}$ is a Borel set, then A satisfies conditions (2), (3), (5), and (6). Condition (3) implies that every Borel set is almost a countable union of closed sets, and condition (6) implies that every Borel set is almost a countable intersection of open sets.

Theorem 1.4.9 Let $A \subseteq \mathbb{R}$. The following are equivalent.

1. A is Lebesgue measurable.

Approximation with closed subsets:

2. For each $\varepsilon > 0$, there exists a closed set $F \subset A$ with $|A \setminus F| < \varepsilon$.
3. There exist closed sets $F_1, F_2, \ldots$ contained in A such that $\left|A \setminus \bigcup_{k=1}^{\infty} F_k\right| = 0$.
4. There exists a Borel set $B \subset A$ such that $|A \setminus B| = 0$.

Approximation with open supersets:

5. For each $\varepsilon > 0$, there exists an open set $G \supset A$ such that $|G \setminus A| < \varepsilon$.
6. There exist open sets $G_1, G_2, \ldots$ containing A such that $\left|\left(\bigcap_{k=1}^{\infty} G_k\right) \setminus A\right| = 0$.
7. There exists a Borel set $B \supset A$ such that $|B \setminus A| = 0$.

Proof. Let $\mathcal{L}$ denote the collection of sets $A \subseteq \mathbb{R}$ that satisfy (2). We have already proved that every Borel set is in $\mathcal{L}$ (Proposition 1.4.3). We will use the fact that $\mathcal{L}$ is a σ-algebra on $\mathbb{R}$ (see proof of Proposition 1.4.3). In addition to containing the Borel sets, $\mathcal{L}$ also contains every set with outer measure 0 (because if $|A| = 0$, we can take $F = \varnothing$ in (b)).

(2) $\Rightarrow$ (3): Suppose (2) holds. For each $n \in \mathbb{Z}^+$, there exists a closed set $F_n \subseteq A$ with $|A \setminus F_n| < 1/n$. Then for each $n \in \mathbb{Z}^+$, we have

$$A \setminus \bigcup_{k=1}^{\infty} F_k \subseteq A \setminus F_n \implies \left| A \setminus \bigcup_{k=1}^{\infty} F_k \right| \leq |A \setminus F_n| \leq \frac{1}{n} \implies \left| A \setminus \bigcup_{k=1}^{\infty} F_k \right| = 0.$$

(3) $\Rightarrow$ (4): Every countable union of closed sets is a Borel set.

(4) $\Rightarrow$ (2): Suppose (4) holds, so there exists a Borel set $B \subseteq A$ such that $|A \setminus B| = 0$. We know that $B \in \mathcal{L}$ (because B is a Borel set) and $A \setminus B \in \mathcal{L}$ (because $A \setminus B$ has outer measure 0). Since $\mathcal{L}$ is a σ-algebra, $A = B \cup (A \setminus B)$ must also be in $\mathcal{L}$.

So far, we know that (2) $\Leftrightarrow$ (3) $\Leftrightarrow$ (4).

(2) $\Rightarrow$ (5): Suppose (2) holds, so $A \in \mathcal{L}$. Let $\varepsilon > 0$. Since $\mathbb{R} \setminus A \in \mathcal{L}$ (holds because $\mathcal{L}$ is closed under complementation), there exists a closed set $F \subseteq \mathbb{R} \setminus A$ such that $|(\mathbb{R} \setminus A) \setminus F| < \varepsilon$. Now $\mathbb{R} \setminus F$ is an open set with $\mathbb{R} \setminus F \supseteq A$. Because $(\mathbb{R} \setminus F) \setminus A) = (\mathbb{R} \setminus A) \setminus F$, the inequality above implies that $|(\mathbb{R} \setminus F) \setminus A)| < \varepsilon$ as desired.

(5) $\Rightarrow$ (6): Suppose (5) holds. Then for each $n \in \mathbb{Z}^+$, there exists an open set $G_n \supseteq A$ such that $|G_n \setminus A| < 1/n$. Now for each $n \in \mathbb{Z}^+$, we have

$$\left(\bigcap_{k=1}^{\infty} G_k \right) \setminus A \subseteq G_n \setminus A \implies \left| \left(\bigcap_{k=1}^{\infty} G_k \right) \setminus A \right| \leq |G_n \setminus A| \leq \frac{1}{n} \implies \left| \left(\bigcap_{k=1}^{\infty} G_k \right) \setminus A \right| = 0.$$

(6) $\Rightarrow$ (7): Every countable intersection of open sets is a Borel set.

(7) $\Rightarrow$ (2): Suppose (7) holds. Then there exists a Borel set $B \supseteq A$ such that $|B \setminus A| = 0$. Since $B \in \mathcal{L}$ and $\mathbb{R} \setminus (B \setminus A) \in \mathcal{L}$ and $\mathcal{L}$ is a σ-algebra, $A = B \cap (\mathbb{R} \setminus (B \setminus A)) \in \mathcal{L}$.

This completes our chain of implications. $\square$

We give one additional equivalent definition: $A \subseteq \mathbb{R}$ is Lebesgue measure if and only if for every $\varepsilon > 0$, there exists a set G that is the union of finitely many disjoint bounded open intervals such that $|A \setminus G| + |G \setminus A| < \varepsilon$. This tells us that every Borel set with finite measure is almost a finite disjoint union of bounded open intervals. (Proof left as exercise. The condition $|A \setminus G| + |G \setminus A| < \varepsilon$ means that G and A are almost the same.)

Theorem 1.4.10

- The set $\mathcal{L}$ of Lebesgue measurable subsets of $\mathbb{R}$ is a σ-algebra on $\mathbb{R}$.
- Outer measure is a measure on $(\mathbb{R}, \mathcal{L})$.

Proof. We have noted that the set $\mathcal{L}$ of Lebesgue measurable subsets of $\mathbb{R}$ is a σ-algebra on $\mathbb{R}$. For the second bullet point, suppose $A_1, A_2, \ldots$ is a disjoint sequence of Lebesgue measurable sets. By definition, for each $k \in \mathbb{Z}^+$ there exists a Borel set $B_k \subseteq A_k$ such that $|A_k \setminus B_k| = 0$. Now

$$\left|\bigcup_{k=1}^{\infty} A_k\right| \geq \left|\bigcup_{k=1}^{\infty} B_k\right| = \sum_{k=1}^{\infty} |B_k| = \sum_{k=1}^{\infty} |A_k|,$$

where the second line above holds because $B_1, B_2, \ldots$ is a disjoint sequence of Borel sets and outer measure is a measure on the Borel sets; the last line above holds because $B_k \subset A_k$ and by subadditivity of outer measure we have

$$|A_k| = |B_k \cup (A_k \setminus B_k)| \leq |B_k| + |A_k \setminus B_k| = |B_k|.$$

The inequality above, combined with countable subadditivity of outer measure, implies that

$$\left|\bigcup_{k=1}^{\infty} A_k\right| = \sum_{k=1}^{\infty} |A_k|,$$

completing the proof of the second bullet point.

If A is a set with outer measure 0, then A is Lebesgue measurable. Our definition of the Lebesgue measurable sets thus implies that the set of Lebesgue measurable sets is the smallest σ-algebra on $\mathbb{R}$ containing the Borel sets and the sets with outer measure 0. Thus the set of Lebesgue measurable sets is also the smallest σ-algebra on $\mathbb{R}$ containing the open sets and the sets with outer measure 0.

Also, note that because outer measure is not finitely additive, there exist subsets of $\mathbb{R}$ that are not Lebesgue measurable. Let us extend the definition of Lebesgue measure from $\mathcal{B}$ to $\mathcal{L}$.

Definition 1.4.11 **Lebesgue measure** is the measure on $(\mathbb{R}, \mathcal{L})$, where $\mathcal{L}$ is the σ-algebra of Lebesgue measurable subsets of $\mathbb{R}$, that assigns to each Lebesgue measurable set its outer measure.

1.4.2 Cantor Set and Cantor Function

We have seen that every countable set has outer measure 0. Does the converse hold, i.e., is every set with outer measure 0 countable? The *Cantor set*, which is introduced in this subsection, provides the answer to this question.

1.5 Convergence of Measurable Functions

Recall that a measurable space is a pair $(X, \mathcal{S})$, where X is a set and $\mathcal{S}$ is a σ-algebra on X. A function $f : X \to \mathbb{R}$ is $\mathcal{S}$-measurable if $f^{-1}(B) \in \mathcal{S}$ for every Borel set $B \subseteq \mathbb{R}$. In this section, we will study measurable functions, with an emphasis on results that depend upon measures.

1.5.1 Pointwise and Uniform Convergence

Like the difference between *continuity* and *uniform continuity*, the difference between pointwise convergence and uniform convergence lies in the order of the quantifiers. Uniform convergence implies pointwise convergence but converse is not true.

Definition 1.5.1 Suppose X is a set, $f_1, f_2, \ldots$ is a sequence of functions from X to $\mathbb{R}$, and f is a function from X to $\mathbb{R}$.

- The sequence $f_1, f_2, \ldots$ **converges pointwise** on X to f if

$$\forall x \in X : \lim_{k\to\infty} f_k(x) = f(x).$$

  In other words, $f_1, f_2, \ldots$ **converges pointwise** on X to f if

$$\forall x \in X : \forall \varepsilon > 0 : \exists n \in \mathbb{Z}^+ : k \geq n \implies |f_k(x) - f(x)| < \varepsilon.$$

- The sequence $f_1, f_2, \ldots$ **converges uniformly** on X to f if

$$\forall \varepsilon > 0 : \exists n \in \mathbb{Z}^+ : \forall x \in X : k \geq n \implies |f_k(x) - f(x)| < \varepsilon.$$

Pointwise limit of continuous functions need not be continuous. However, the uniform limit of continuous functions is continuous.

Proposition 1.5.2 Suppose $B \subseteq \mathbb{R}$ and $f_1, f_2, \ldots$ is a sequence of functions from B to $\mathbb{R}$ that converges uniformly on B to a function $f : B \to \mathbb{R}$. Suppose $b \in B$ and f_k is continuous at b for each $k \in Z^+$. Then f is continuous at b.

Proof. Suppose $\varepsilon > 0$. Let $n \in \mathbb{Z}^+$ be such that $|f_n(x) - f(x)| < \varepsilon/3$ for $x \in B$. Because f_n is continuous at b, there exists $\delta > 0$ such that $|f_n(x) - f_n(b)| < \varepsilon/3$ for all $x \in (b - \delta, b + \delta) \cap B$. Now suppose $x \in (b - \delta, b + \delta) \cap B$. Then

$$|f(x) - f(b)| \leq |f(x) - f_n(x)| + |f_n(x) - f_n(b)| + |f_n(b) - f(b)| < \varepsilon.$$

Thus f is continuous at b. □

1.5.2 Egorov's Theorem

A sequence of functions that converges pointwise need not converge uniformly. However, the next result says that a pointwise convergent sequence of functions on a measure space with finite total measure (i.e., the measure of the entire set is finite) almost converges uniformly, in the sense that it converges uniformly except on a set that can have arbitrarily small measure.

Theorem 1.5.3 [Egorov] Let $(X, \mathcal{S}, \mu)$ be a measure space with $\mu(X) < \infty$. Suppose $f_1, f_2, \ldots$ is a sequence of $\mathcal{S}$-measurable functions from X to $\mathbb{R}$ that converges pointwise on X to a function $f : X \to \mathbb{R}$. Then for every $\varepsilon > 0$, there exists a set $E \in \mathcal{S}$ such that

- $\mu(X \setminus E) < \varepsilon$, and
- $f_1, f_2, \ldots$ converges uniformly to f on E.

Remark/Intuition. The ultimate goal for this proof is to construct a set $E \in \mathcal{S}$ (and equivalently, $E \subseteq X$) with all the desirable properties. We know that the sequence of functions $f_1, f_2, \ldots$ converges pointwise on X to a function f, i.e., for each $x \in X$ we can find $m \in \mathbb{Z}^+$ such that $k \geq m$ implies $f_k(x)$ is close to $f(x)$. A classic trick when working with multiple limits is to unify the variables to make it work in all cases. For example, Since each m here depends on the particular point $x \in X$, taking the maximum of all m which gives us an $m_{\max}$ that works for all x. Now in our case, instead of finding an m that works for all $x \in X$, we wish to find all $x \in X$ that works with all m. On this subset of X, the sequence of functions $f_1, f_2, \ldots$ must converge pointwise to f. To do this, we will play with union and intersection.

Proof. Let $\varepsilon > 0$. Temporarily fix $n \in \mathbb{Z}^+$. Recall the pointwise convergence says

$$\forall x \in X : \forall \varepsilon > 0 : \exists m \in \mathbb{Z}^+ : k \geq m \implies |f_k(x) - f(x)| < \varepsilon.$$

This implies that (the coloring of variables may help you interpret the expression below)

$$\bigcup_{m=1}^{\infty} \bigcap_{k=m}^{\infty} \left\{ x \in X : |f_k(x) - f(x)| < \frac{1}{n} \right\} = X. \tag{1.9}$$

For $m \in \mathbb{Z}^+$, let

$$A_{m,n} = \bigcap_{k=m}^{\infty} \left\{ x \in X : |f_k(x) - f(x)| < \frac{1}{n} \right\}.$$

Then clearly $A_{1,n} \subset A_{2,n} \subset \cdots$ is an increasing sequence of sets and (1.9) can be rewritten as

$$\bigcup_{m=1}^{\infty} A_{m,n} = X.$$

By Lemma 1.3.9, we have

$$\lim_{m\to\infty} \mu\left(A_{m,n}\right) = \mu\left(\bigcup_{m=1}^{\infty} A_{m,n}\right) = \mu(X).$$

Thus there exists $m_n \in \mathbb{Z}^+$ (m_n depends on n) such that

$$\mu(X) - \mu\left(A_{m_n,n}\right) < \frac{\varepsilon}{2^n}. \tag{1.10}$$

Now take the intersection of all $A_{m_n,n}$:

$$E = \bigcap_{n=1}^{\infty} A_{m_n,n}.$$

Then

$$\begin{aligned}
\mu(X \setminus E) &= \mu\left(X \setminus \bigcap_{n=1}^{\infty} A_{m_n,n}\right) \\
&= \mu\left(\bigcup_{n=1}^{\infty} \left(X \setminus A_{m_n,n}\right)\right) \\
&\leq \sum_{n=1}^{\infty} \mu\left(X \setminus A_{m_n,n}\right) \\
&< \sum_{n=1}^{\infty} \frac{\varepsilon}{2^n} && (1.10) \\
&= \varepsilon,
\end{aligned}$$

proving the first bullet.

To complete the proof, we must verify that $f_1, f_2, \ldots$ converges uniformly to f on E. To do this, suppose $\varepsilon' > 0$. Let $n \in \mathbb{Z}^+$ be such that

$$\frac{1}{n} < \varepsilon'.$$

Then by definition

$$E = \bigcap_{n=1}^{\infty} A_{m_n,n} \implies E \subset A_{m_n,n},$$

which implies that

$$\forall x \in E : \forall k \geq m_n : |f_k(x) - f(x)| < \frac{1}{n} < \varepsilon'$$

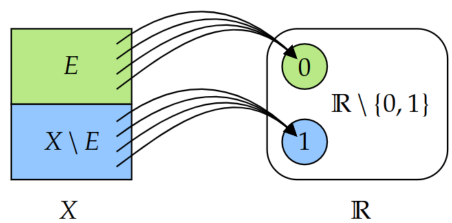
Hence $f_1, f_2, \ldots$ does indeed converge uniformly to f on E. □

1.5.3 Approximation by Simple Functions

Definition 1.5.4 Let $E \subseteq X$. The **characteristic function** of E is defined by

$$\chi_E : X \longrightarrow \mathbb{R}$$
$$x \longmapsto \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$$

The set X that contains E is not explicitly included in the notation χ_E because X will always be clear from the context.



Example 1.5.5 Suppose $(X, \mathcal{S})$ is a measurable space, $E \subseteq X$ and $B \subseteq \mathbb{R}$. Since χ_E maps all elements in E to 1 and maps all elements in $X \setminus E$ to 0, the pre-image of B depends on whether B contains 0 and 1, i.e.,

$$\chi_E^{-1}(B) = \begin{cases} E & \text{if } 0 \notin B \text{ and } 1 \in B, \\ X \setminus E & \text{if } 0 \in B \text{ and } 1 \notin B, \\ X & \text{if } 0 \in B \text{ and } 1 \in B, \\ \varnothing & \text{if } 0 \notin B \text{ and } 1 \notin B. \end{cases}$$

Thus we see that χ_E is an $\mathcal{S}$-measurable function if and only if $E \in \mathcal{S}$.

Definition 1.5.6 Let $(X, \mathcal{S})$ be a measurable space. A **simple function** on X is a measurable function $f : X \longrightarrow \mathbb{R}$ that takes only finitely many distinct values.

In order for f to be called a simple function, f must be measurable, $f(x)$ must be real (cannot be $\pm\infty$) for each $x \in X$, and range(f) must be finite. In other words, a simple function is a measurable function whose range is a subset of $\mathbb{R}$. It is easy to see that the class of simple functions on a measurable space $(X, \mathcal{S})$ is closed with respect to addition, scalar multiplication, and products.

Suppose $(X, \mathcal{S})$ is a measurable space, $f : X \to \mathbb{R}$ is a simple function, and $c_1, \ldots, c_n$ are the distinct nonzero values of f. Define $E_k = f^{-1}(\{c_k\})$ to be the set of values in X that gets mapped to c_k. Then we can write f as a sum of characteristic functions:

$$f = c_1 \chi_{E_1} + \cdots + c_n \chi_{E_n} = \sum_{i=1}^{n} c_i \chi_{E_i}.$$

Using what we learned in the example above, we see that the simple function f is an $\mathcal{S}$-measurable function if and only if $E_1, \ldots, E_n \in \mathcal{S}$.

Much of the power of simple functions lies in the next theorem, which states that every measurable function f can be written as a pointwise limit of a sequence of simple functions f_k. In fact, we will be able to construct the simple functions f_k so that they increase pointwise to f and the convergence is uniform on any set where f is bounded.

Theorem 1.5.7 Suppose $(X, \mathcal{S})$ is a measure space and $f : X \to [-\infty, \infty]$ is $\mathcal{S}$-measurable. Then there exists a sequence of functions $f_1, f_2, \ldots : X \to \mathbb{R}$ such that

1. each f_k is a simple $\mathcal{S}$-measurable function;
2. $\forall k \in \mathbb{Z}^+ : \forall x \in X : |f_k(x)| \leq |f_{k+1}(x)| \leq |f(x)|$;
3. $\forall x \in X : \lim_{k\to\infty} f_k(x) = f(x)$;
4. if f is bounded, $f_1, f_2, \ldots$ converges uniformly on X to f.

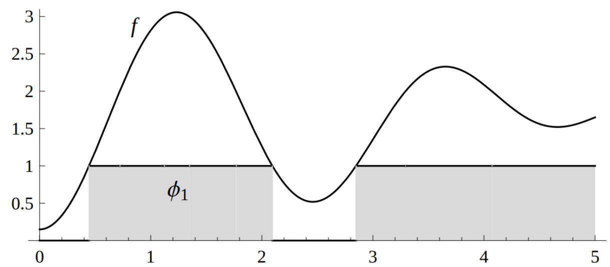
Proof. We will first prove the case where f is non-negative, then generalize to general f. The idea of the proof is that we construct f_k by simply rounding f down to the nearest integer multiple of 2^{-k}. However, if f is unbounded, then this would give f_k infinitely many values, while a simple function can only take finitely many values. Hence we stop the rounding down process at some finite height. Typically choices for this height are k and or 2^k; we will use the former.

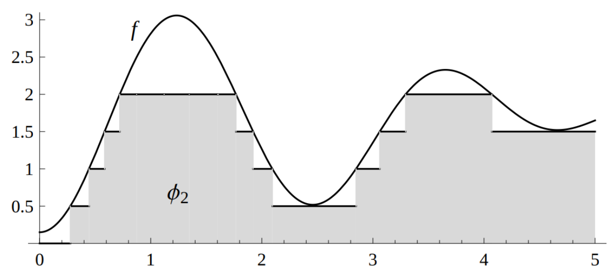
- For $k = 1$, we define f_1 by rounding f down to the nearest integer, with the caveat that we stop at height 1. Specifically,
$$f_1(x) = \begin{cases} 0 & f \leq f(x) < 1 \\ 1 & f(x) \geq 1 \end{cases}$$

- For f_2 we round down to the nearest integer multiple of $\frac{1}{2}$, except we never exceed height 2.

$$f_2(x) = \begin{cases} 0, & 0 \le f(x) < \frac{1}{2} \\ \frac{1}{2}, & \frac{1}{2} \le f(x) < 1 \\ 1, & 1 \le f(x) < \frac{3}{2} \\ \frac{3}{2}, & \frac{3}{2} \le f(x) < 2 \\ 2, & f(x) \ge 2 \end{cases}$$

In the graphs below, f_1, f_2 are labeled ϕ_1, ϕ_2.





In general, given $k \in \mathbb{Z}^+$, we define f_k by

$$f_k(x) = \begin{cases} \frac{j-1}{2^k} & \frac{j-1}{2^k} \le f(x) < \frac{j}{2^k}, j = 1, \dots, k2^k \\ k & f(x) \ge k \end{cases}$$

The sets

$$\left\{ \frac{j-1}{2^k} \le f < \frac{j}{2^k} \right\} \quad \text{and} \quad \{f \ge k\}$$

are measurable because f is measurable, so it follows that f_k is measurable. Further, by construction we have $f_k(x) \le f_{k+1}(x)$ for every x, and

$$f(x) \le k \implies |f(x) - f_k(x)| \le 2^{-k}.$$

If $f(x)$ is finite, then k will eventually exceed $f(x)$, so we have $f_k(x) \to f(x)$ in this case. In fact, if

$f(x) \leq M < \infty$ for all x in some set E, then

$$\sup_{x \in E} |f(x) - f_k(x)| \leq 2^{-k} \quad \text{for } k \geq M.$$

Hence f_k converges uniformly to f on E in this case. On the other hand, if $f(x) = \infty$, then $\phi_k(x) = k$ for every k, so $\phi_k(x) \to f(x)$ as $k \to \infty$, and thus we still have pointwise convergence in this case.

To generalize to general f (that can take both positive and negative values), replace $f(x)$ by $|f(x)|$, so we will be rounding down f when $f(x)$ is positive and rounding up f when $f(x)$ is negative. $\square$

We will also give the proof in Axler's book. It is crucial to notice that we are partitioning the y-axis. This intuition will be important for Lebesgue integration covered in the next chapter.

Proof. The idea of the proof is that for each $k \in \mathbb{Z}^+$ and $n \in \mathbb{Z}$, the interval $[n, n+1)$ is divided into 2^k equally sized half-open subintervals. If $f(x) \in [0, k)$, we define $f_k(x)$ to be the left endpoint of the subinterval into which $f(x)$ fails; if $f(x) \in (-k, 0)$, we define $f_k(x)$ to be the right endpoint of the subinterval into which $f(x)$ fails; and if $|f(x)| \geq k$, we define $f_k(x)$ be the $\pm k$.

Specifically, let

$$f_k(x) = \begin{cases} \frac{m}{2^k} & \text{if } 0 \leq f(x) < k \text{ and } m \in \mathbb{Z} \text{ is such that } f(x) \in \left[\frac{m}{2^k}, \frac{m+1}{2^k}\right) \\ \frac{m+1}{2^k} & \text{if } -k < f(x) < 0 \text{ and } m \in \mathbb{Z} \text{ is such that } f(x) \in \left[\frac{m}{2^k}, \frac{m+1}{2^k}\right) \\ k & \text{if } f(x) \geq k \\ -k & \text{if } f(x) \leq -k. \end{cases}$$

Each $f^{-1}\left(\left[\frac{m}{2^k}, \frac{m+1}{2^k}\right)\right) \in \mathcal{S}$ because f is an $\mathcal{S}$-measurable function. Thus each f_k is an $\mathcal{S}$-measurable simple function; in other words, (1) holds.

Also, (2) holds because of how we have defined f_k.

The definition of f_k implies that

$$|f_k(x) - f(x)| \leq \frac{1}{2^k} \tag{1.11}$$

for all $x \in X$ such that $f(x) \in [-k, k]$. Thus we see that (3) holds.

Finally, (1.11) shows that (4) holds. $\square$

1.5.4 Luzin's Theorem

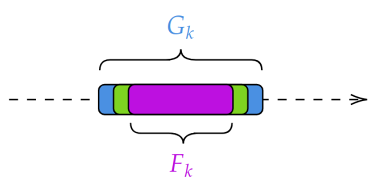
The next result says that an arbitrary Borel measurable function is *almost* continuous, in the case that its restriction to a large closed set is continuous. Be careful about the interpretation of the conclusion of Luzin's theorem that $f|_B$ is a continuous function on B. This is not the same as saying that f (on its original domain) is continuous at each point of B. For example, $\chi_{\mathbf{Q}}$ is discontinuous at every point of $\mathbf{R}$. However, $\chi_{\mathbf{Q}}|_{\mathbf{R}\setminus\mathbf{Q}}$ is a continuous function on $\mathbf{R} \setminus \mathbf{Q}$ (because this function is identically 0 on its domain).

Theorem 1.5.8 [Luzin] Suppose $g : \mathbf{R} \to \mathbf{R}$ is a Borel measurable function. Then for every $\varepsilon > 0$, there exists a closed set $F \subseteq \mathbf{R}$ such that $|\mathbf{R} \setminus F| < \varepsilon$ and $g|_F$ is a continuous function on F.

Proof. First consider the special case where $g = d_1\chi_{D_1} + \cdots + d_n\chi_{D_n}$ for some distinct nonzero $d_1, \ldots, d_n \in \mathbf{R}$ and some disjoint Borel sets $D_1, \ldots, D_n \subset \mathbf{R}$. Suppose $\varepsilon > 0$. For each $k \in \{1, \ldots, n\}$, there exist (by Theorem 1.4.9) a closed set $F_k \subset D_k$ and an open set $G_k \supset D_k$ such that

$$|G_k \setminus D_k| < \frac{\varepsilon}{2n} \quad \text{and} \quad |D_k \setminus F_k| < \frac{\varepsilon}{2n}.$$

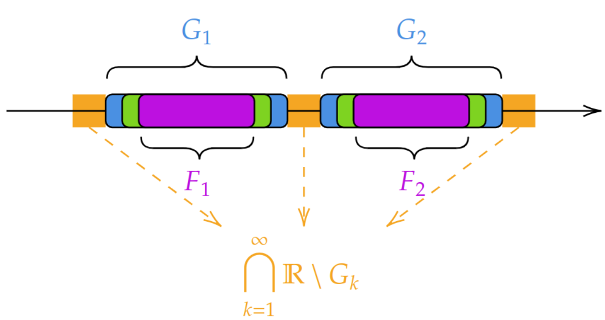
Intuitively, we are approximating the Borel set D_k (green interval, partially blocked by F_k) with a closed subset F_k (purple interval) and an open superset G_k (blue interval, partially blocked by D_k).



Because $G_k \setminus F_k = (G_k \setminus D_k) \cup (D_k \setminus F_k)$, we have $|G_k \setminus F_k| < \varepsilon/n$ for each $k \in \{1, \ldots, n\}$. Let

$$F = \left(\bigcup_{k=1}^{n} F_k\right) \cup \bigcap_{k=1}^{n} (\mathbf{R} \setminus G_k).$$

Then F is a closed subset of $\mathbf{R}$ and $\mathbf{R} \setminus F = \bigcup_{k=1}^{n} (G_k \setminus F_k)$. Thus $|\mathbf{R} \setminus F| < \varepsilon$.



Here, F is the union of all purple and orange blocks; $\mathbb{R} \setminus \mathbb{F}$ is the union of all green and blue blocks *shown*.

Because $F_k \subset D_k$ and $g = d_1\chi_{D_1} + \cdots + d_n\chi_{D_n}$, we see that g is identically d_k on F_k. Thus $g|_{F_k}$ is continuous for each $k \in \{1, \dots, n\}$. Because

$$\bigcap_{k=1}^{n} (\mathbb{R} \setminus G_k) \subset \bigcap_{k=1}^{n} (\mathbb{R} \setminus D_k),$$

we see that g is identically 0 on $\bigcap_{k=1}^{n} (\mathbb{R} \setminus G_k)$. Thus $g|_{\bigcap_{k=1}^{n}(\mathbb{R}\setminus G_k)}$ is continuous. Putting all this together, we conclude that $g|_F$ is continuous, completing the proof in this special case.[3] $\diamond$

Now consider an arbitrary Borel measurable function $g : \mathbb{R} \to \mathbb{R}$. This is where our proof regarding convergence of functions become useful. By Theorem 1.5.7, there exists a sequence $g_1, g_2, \dots$ of functions from $\mathbb{R}$ to $\mathbb{R}$ that converges pointwise on $\mathbb{R}$ to g, where each g_k is a simple Borel measurable function.

Suppose $\varepsilon > 0$. By the special case already proved, for each $k \in \mathbb{Z}^+$, there exists a closed set $C_k \subset \mathbb{R}$ such that $|\mathbb{R} \setminus C_k| < \frac{\varepsilon}{2^{k+1}}$ and $g_k|_{C_k}$ is continuous. (This C_k corresponds to the special set F appeared in the prove for the special case.) Let

$$C = \bigcap_{k=1}^{\infty} C_k$$

Thus C is a closed set and $g_k|_C$ is continuous for every $k \in \mathbb{Z}^+$. Intuitively, g_k behaves well on C_k, so if we take the intersection C of all C_k's, then all g_k's will behave well on C.

Next,

$$\mathbb{R} \setminus C = \bigcup_{k=1}^{\infty} (\mathbb{R} \setminus C_k) \implies |\mathbb{R} \setminus C| = \sum_{k=1}^{\infty} |\mathbb{R} \setminus C_k| < \sum_{k=1}^{\infty} \frac{\varepsilon}{2^{k+1}} = \frac{\varepsilon}{2}.$$

For each $m \in \mathbb{Z}$, the sequence of functions restricted to the interval $(m, m+1)$

$$g_1|_{(m,m+1)}, g_2|_{(m,m+1)}, \dots$$

converges pointwise on $(m, m+1)$ to $g|_{(m,m+1)}$. Thus by Egorov's Theorem (1.5.3), for each $m \in \mathbb{Z}$, there is a Borel set $E_m \subset (m, m+1)$ such that $g_1, g_2, \dots$ converges uniformly to g on E_m and

$$|(m, m+1) \setminus E_m| < \frac{\varepsilon}{2^{|m|+3}}.$$

Thus $g_1, g_2, \dots$ converges uniformly to g on $C \cap E_m$ for each $m \in \mathbb{Z}$. Recall from Proposition 1.5.2 that the uniform limit of continuous functions is continuous. Because each $g_k|_C$ is continuous, we conclude that $g|_{C\cap E_m}$ being the uniform limit of $g_k|_{C\cap E_m}$'s is continuous for each $m \in \mathbb{Z}$.

[3] We used the following fact. Suppose $F_1, \dots, F_n$ are disjoint closed subsets of $\mathbb{R}$. Then if $g : F_1 \cup \cdots \cup F_n \to \mathbb{R}$ is a function such that $g|_{F_k}$ is a continuous function for each k, then g is a continuous function.

Thus $g|_D$ is continuous, where

$$D = \bigcup_{m \in \mathbb{Z}} (C \cap E_m)$$

Because

$$\mathbb{R} \setminus D \subset \mathbb{Z} \cup \left(\bigcup_{m \in \mathbb{Z}} ((m, m+1) \setminus E_m) \right) \cup (\mathbb{R} \setminus C),$$

we have $|\mathbb{R} \setminus D| < \varepsilon$.

Now approximate the Borel set D with a closed subset $F \subseteq D$ such that $|D \setminus F| < \varepsilon - |\mathbb{R} \setminus D|$. Then

$$|\mathbb{R} \setminus F| = |(\mathbb{R} \setminus D) \cup (D \setminus F)| \leq |\mathbb{R} \setminus D| + |D \setminus F| < \varepsilon.$$

Because the restriction of a continuous function to a smaller domain is also continuous, $g|_F$ is continuous, completing the proof. □

We need the following result to get another version of Luzin's Theorem.

Lemma 1.5.9 Every continuous function on a closed subset of $\mathbb{R}$ can be extended to a continuous function on all of $\mathbb{R}$. More precisely, if $F \subseteq \mathbb{R}$ is closed and $g : F \to \mathbb{R}$ is continuous, then there exists a continuous function $h : \mathbb{R} \to \mathbb{R}$ such that $h|_F = g$.

Proof. Suppose $F \subseteq \mathbb{R}$ is closed and $g : F \to \mathbb{R}$ is continuous. Then $\mathbb{R} \setminus F$ is the union of a collection of disjoint open intervals $\{I_k\}$. For each such interval of the form (a, ∞) or of the form $(-\infty, a)$, define $h(x) = g(a)$ for all x in the interval.

For each interval I_k of teh form (b, c) with $b < c$ and $b, c \in \mathbb{R}$, define h on $[b, c]$ to be the linear function such that $h(b) = g(b)$ and $h(c) = g(c)$.

Define $h(x) = g(x)$ for all $x \in \mathbb{R}$ for which $h(x)$ has not been defined by the previous two paragraphs. Then $h : \mathbb{R} \to \mathbb{R}$ is continuous and $h|_F = g$. □

The next result gives a slightly modified way to state Luzin's Theorem. Intuitively, it's saying that the value of a Borel measurable function can be changed on a set with small Lebesgue measure to produce a continuous function.

Theorem 1.5.10 [Luzin] Suppose $E \subseteq \mathbb{R}$ and $g : E \to \mathbb{R}$ is a Borel measurable function. Then for every $\varepsilon > 0$, there exists a closed set $F \subseteq E$ and a continuous function $h : \mathbb{R} \to \mathbb{R}$ such that $|E \setminus F| < \varepsilon$ and $h|_F = g|_F$.

Proof. Suppose $\varepsilon > 0$. Extend g to a function $\tilde{g} : \mathbb{R} \to \mathbb{R}$ by defining

$$\tilde{g}(x) = \begin{cases} g(x) & \text{if } x \in E, \\ 0 & \text{if } x \in \mathbb{R} \setminus E. \end{cases}$$

By the first version of Luzin's Theorem, there is a closed set $C \subset \mathbb{R}$ such that $|\mathbb{R} \setminus C| < \varepsilon$ and $\tilde{g}|_C$ is a continuous function on C. Now approximate the Borel set $C \cap E$ with a closed set $F \subset C \cap E$ such that $|(C \cap E) \setminus F| < \varepsilon - |\mathbb{R} \setminus C|$. Then

$$|E \setminus F| \leq |((C \cap E) \setminus F) \cup (\mathbb{R} \setminus C)| \leq |(C \cap E) \setminus F| + |\mathbb{R} \setminus C| < \varepsilon.$$

Now $\tilde{g}|_F$ is a continuous function on F. Also, $\tilde{g}|_F = g|_F$ (because $F \subset E$). Now use Lemma 1.5.9 to extend $\tilde{g}|_F$ to a continuous function $h : \mathbb{R} \to \mathbb{R}$. □

1.5.5 Lebesgue Measurable Functions

Definition 1.5.11 A function $f : A \to \mathbb{R}$, where $A \subseteq \mathbb{R}$, is called **Lebesgue measurable** if $f^{-1}(B)$ is a Lebesgue measurable set for every Borel set $B \subseteq \mathbb{R}$.

If $f : A \to \mathbb{R}$ is a Lebesgue measurable function, then A is a Lebesgue measurable subset of $\mathbb{R}$ (because $A = f^{-1}(\mathbb{R})$). If A is a Lebesgue measurable subset of $\mathbb{R}$, then the definition above is the standard definition of an $\mathcal{S}$-measurable function, where $\mathcal{S}$ is a σ-algebra of all Lebesgue measurable subsets of A.

Although there exist Lebesgue measurable sets that are not Borel sets, it's unlikely to encounter one. Similarly, a Lebesgue measurable function that is not Borel measurable is unlikely to arise in practice. A great way to simplify the potential confusion about Lebesgue measurable functions being defined by inverse images of Borel sets is to consider only Borel measurable functions.

The next result states that if we adopt the philosophy that what happens on a set of outer measure 0 does not matter much, then we might as well restrict our attention to Borel measurable functions.

Proposition 1.5.12 Suppose $f : \mathbb{R} \to \mathbb{R}$ is a Lebesgue measurable function. Then there exists a Borel measurable function $g : \mathbb{R} \to \mathbb{R}$ such that

$$|\{x \in \mathbb{R} : g(x) \neq f(x)\}| = 0.$$

Proof. Let's use Theorem 1.5.7 to approximate f with a series of Lebesgue measurable simple functions $f_1, f_2, \ldots : \mathbb{R} \to \mathbb{R}$ converging pointwise on $\mathbb{R}$ to f. Suppose $k \in \mathbb{Z}^+$. Then there exists $c_1, \ldots, c_n \in \mathbb{R}$ and disjoint Lebesgue measurable sets $A_1, \ldots, A_n \subseteq \mathbb{R}$ such that

$$f_k = c_1\chi_{A_1} + \cdots + c_n\chi_{A_n}.$$

For each $j \in \{1, \ldots, n\}$, there exists a Borel set $B_j \subseteq A_j$ such that $|A_j \setminus B_j| = 0$. Let

$$g_k = c_1\chi_{B_1} + \cdots + c_n\chi_{B_n}.$$

Then g_k is a Borel measurable function and

$$|\{x \in \mathbb{R} : g_k(x) \neq f_k(x)\}| = 0.$$

If $x \notin \bigcup_{k=1}^{\infty}\{x \in \mathbb{R} : g_k(x) \neq f_k(x)\}$, then $g_k(x) = f_k(x)$ for all $k \in \mathbb{Z}^+$ and $\lim_{k\to\infty} g_k(x) = f(x)$. Let

$$E = \{x \in \mathbb{R} : \lim_{k\to\infty} g_k(x) \text{ exists in } \mathbb{R}\}.$$

Then E is a Borel subset of $\mathbb{R}$. Also,

$$\mathbb{R} \setminus E \subseteq \bigcup_{k=1}^{\infty} \{x \in \mathbb{R} : g_k(x) \neq f_k(x)\}$$

and thus $|\mathbb{R} \setminus E| = 0$. For $x \in \mathbb{R}$, let

$$g(x) = \lim_{k\to\infty} (\chi_E g_k)(x). \tag{1.12}$$

If $x \in E$, then the limit above exists by definition of E; if $x \in \mathbb{R} \setminus E$, then the limit above exists because $(\chi_E g_k)(x) = 0$ for all $k \in \mathbb{Z}^+$.

For each $k \in \mathbb{Z}^+$, the function $\chi_E g_k$ is Borel measurable. Then (1.12) implies that g is a Borel measurable function (by Proposition 1.2.23). Now

$$\{x \in \mathbb{R} : g(x) \neq f(x)\} \subseteq \bigcup_{k=1}^{\infty} \{x \in \mathbb{R} : g_k(x) \neq f_k(x)\},$$

we conclude that $|\{x \in \mathbb{R} : g(x) \neq f(x)\}| = 0$, completing the proof. □

1.6 Chapter Summary

1.6.1 Outer Measure on $\mathbb{R}$

The **length** $\ell(I)$ of an open interval $I \subseteq \mathbb{R}$ matches our intuition:

- the empty interval $\varnothing$ has size 0;
- a bounded interval of the form (a, b) has size $b - a$;
- an unbounded interval of the form (a, ∞), $(-\infty, b)$, or $(-\infty, \infty)$ has size ∞.

The **outer measure** $|A|$ of a set $A \subseteq \mathbb{R}$ as the infimum over the "sizes" of all open covers of A.

$$|A| = \inf \left\{ \sum_{k=1}^{\infty} \ell(I_k) \;\middle|\; I_1, I_2, \dots \text{ are open intervals such that } A \subseteq \bigcup_{k=1}^{\infty} I_k \right\}.$$

- Countable sets have outer measure 0.
- Outer measure preserves order.
- Outer measure is translation invariant.
- Outer measure is countably subadditive.
- Outer measure of an interval is equal to its length.
- Outer measure is NOT countably additive.

1.6.2 Measurable Spaces and Functions

No function μ can satisfy all properties listed in Proposition 1.2.1 and we are forced to relax the requirement that *size* is defined for all subsets of $\mathbb{R}$. More precisely, we will limit the domain of the size function to the collection of subsets that are closed under *complementation* and *countable unions*. A set of subsets satisfying these properties is called a **σ-algebra**. An important example is the **Borel algebra** on $\mathbb{R}$, which contains all open subsets of $\mathbb{R}$ and is the smallest σ-algebra on $\mathbb{R}$.

A **measurable space** is an ordered pair $(X, \mathcal{S})$, where X is a set and $\mathcal{S}$ is a σ-algebra on X. An element of $\mathcal{S}$ is called an **$\mathcal{S}$-measurable set**. A function $f : X \to \mathbb{R}$ is an **$\mathcal{S}$-measurable function** if $f^{-1}(B) \in \mathcal{S}$ for every Borel set $B \subseteq \mathbb{R}$. Intuitively, it preserves the structure of measurable sets under pre-images. To verify whether $f : X \to \mathbb{R}$ is measurable, we only need to check the pre-image of a much smaller collection of subsets of $\mathbb{R}$, e.g., whether $f^{-1}((a, \infty)) \in \mathcal{S}$ for all $a \in \mathbb{R}$.

When X is a Borel subset of $\mathbb{R}$ and $\mathcal{S}$ is the set of Borel subsets of $\mathbb{R}$ that are contained in X, a $\mathcal{S}$-measurable function is called a **Borel measurable** function. We proved that

- Every continuous function defined on a Borel set is Borel measurable.
- Every increasing function defined on a Borel set is Borel measurable.
- Composition of measurable functions is measurable.
- The sum, difference, product, and quotient of two measurable functions are measurable.
- The point-wise limit of a sequence of $\mathcal{S}$-measurable functions is $\mathcal{S}$-measurable.

1.6.3 Measures and Their Properties

A **measure** on $(X, \mathcal{S})$ is a function $\mu : \mathcal{S} \to [0, \infty]$ such that $\mu(\varnothing) = 0$ and has countable additivity for disjoint sets. We proved the following properties of measures:

- For $D \subseteq E$, we have $\mu(D) \leq \mu(E)$ and $\mu(E \setminus D) = \mu(E) - \mu(D)$ (provided that $\mu(D) < \infty$).
- μ has countable subadditivity for arbitrary (not necessarily disjoint) countable union.
- Measures behave well with increasing unions and decreasing intersections.
  - $E_1 \subseteq E_2 \subseteq \cdots \implies \mu\left(\bigcup_{k=1}^{\infty} E_k\right) = \lim_{k\to\infty} \mu(E_k)$.
  - $E_1 \supseteq E_2 \supseteq \cdots \implies \mu\left(\bigcap_{k=1}^{\infty} E_k\right) = \lim_{k\to\infty} \mu(E_k)$.
- For $D, E \in \mathcal{S}$, we have $\mu(D \cap E) = \mu(D) + \mu(E) - \mu(D \cap E)$ (provided that $\mu(D \cap E) < \infty$).

1.6.4 Lebesgue Measure

For disjoint sets $A, B \subseteq \mathbb{R}$, the equality $|A \cup B| = |A| + |B|$ holds if at least one of A, B is open, closed, or Borel. Restricting outer measure to Borel subsets of $\mathbb{R}$ produces a real measure.

A set $A \subseteq \mathbb{R}$ is **Lebesgue measurable** if it can be approximated by a Borel set. (Thus, every Borel set is Lebesgue measurable.) Every Lebesgue measurable set can be approximated with closed subsets or open supersets. The set of Lebesgue measurable sets is a σ-algebra on $\mathbb{R}$, denoted $\mathcal{L}$. This is the smallest σ-algebra on $\mathbb{R}$ containing the open sets and the sets with outer measure 0. Restricting outer measure to Lebesgue measurable sets produces the **Lebesgue measure**.

1.6.5 Convergence of Measurable Functions

Pointwise limit of continuous functions is not always continuous, but the uniform limit of continuous functions is continuous. **Egorov's Theorem** states that a pointwise convergence sequence functions on a measure space with finite total measure almost convergences uniformly.

A **simple function** on X is a measurable function that takes only finitely many distinct values. Every measurable function can be written as a pointwise limit of a sequence of functions. In fact, we can construct the sequence of simple functions so that they increase pointwise to f and the convergence is uniform on any set where f is bounded.

Every continuous function on a closed subset of $\mathbb{R}$ can be extended to a continuous function on the entire $\mathbb{R}$. **Luzin's Theorem** states that any Borel measurable function restricted to a large closed set is continuous. Equivalently, the value of a Borel measurable function can be changed on a set with small Lebesgue measure to produce a continuous function.

A function is **Lebesgue measurable** if $f^{-1}(B)$ is a Lebesgue measurable set for every Borel set $B \subseteq \mathbb{R}$. Although there exist Lebesgue measurable sets that are not Borel sets, it's often easier to treat Borel measurability and Lebesgue measurability equally. For each Lebesgue measurable function, there exists a Borel measurable function that only differ on a measure zero set.

Chapter 2

Integration

Contents

2.1 Integration with Respect to a Measure

We will first define the integral of a non-negative function with respect to a measure. Then by writing a real-valued function as the difference of two non-negative functions, we will define the integral of a real-valued function with respect to a measure.

2.1.1 Integration of Nonnegative Functions

The first definition should remind you of the definition of *partition* in Riemann integration.

Definition 2.1.1 Suppose $\mathcal{S}$ is a σ-algebra on a set X. An **$\mathcal{S}$-partition** of X is a *finite* collection $A_1, \ldots, A_m$ of *disjoint* sets in $\mathcal{S}$ such that $A_1 \cup \cdots \cup A_m = X$.

The next definition introduces clean notation for the "lowest" (infimum) and the "highest" (supremum) values f achieves on set A.

Definition 2.1.2 If f is a real-valued function and $A \subseteq \text{dom}(f)$, then

$$\inf_A f := \inf\{f(x) : x \in A\} \quad \text{and} \quad \sup_A f := \sup\{f(x) : x \in A\}.$$

The next definition should remind you of *lower Riemann sum*, so let's briefly review it. For a bounded function $f : [a,b] \to \mathbb{R}$, its lower Riemann sum over $[a,b]$ with respect to partition $P = \{x_0, x_1, \ldots, x_n\}$ is given by

$$L(f, P, [a,b]) = \sum_{j=1}^{n} (x_j - x_{j-1}) \inf_{[x_{j-1}, x_j]} f.$$

In particular, $(x_j - x_{j-1})$ is the "width" of the subinterval $[x_j, x_{j-1}]$ while $\inf_{[x_{j-1}, x_j]} f$ is an estimation of f over $[x_j, x_{j-1}]$ from below and represents the "height" of f over the given subinterval.

Since we are now working with an arbitrary measure and thus X need not be a subset of $\mathbb{R}$. More importantly, even in the case when X is a closed interval $[a,b]$ in $\mathbb{R}$ and μ is Lebesgue measure on the Borel subsets of $[a,b]$, the sets $A_1, \ldots, A_m$ in the definition below do not need to be subintervals of $[a,b]$ as they do for the lower Riemann sum—they need only be Borel sets.

Given an $\mathcal{S}$-partition $A_1, \ldots, A_m$, we wish to construct a similar *lower Lebesgue sum*. How do we expression the equivalent of "width" and "height" over each A_j? The height is easy: we can just take the infimum of f over A_j, i.e., $\inf_{A_j} f$. For width, recall that we generalized the notion of length (of an interval) to *measure* (of an arbitrary set), so the equivalent of "width" will be $\mu(A_j)$. This motivates the following definition.

Definition 2.1.3 Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $f : X \to [0, \infty]$ is an $\mathcal{S}$-measurable function, and P is an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X. The **lower Lebesgue sum** $\mathcal{L}(f, P)$ is defined by

$$\mathcal{L}(f, P) = \sum_{j=1}^{m} \mu(A_j) \inf_{A_j} f.$$

Suppose $(X, \mathcal{S}, \mu)$ is a measure space. We will denote the integral of an $\mathcal{S}$-measurable function f with respect to μ by $\int f \, \mathrm{d}\mu$. Out basic requirements for an integral are that

1. $\forall E \in \mathcal{S} : \int \chi_E \, \mathrm{d}\mu = \mu(E)$;
2. $\int (f + g) \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$.

As we will see, the following definition satisfies both requirements (although not obvious).

Definition 2.1.4 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f : X \to [0, \infty]$ is an $\mathcal{S}$-measurable function. The **integral** of f with respect to μ, denoted $\int f \, \mathrm{d}\mu$, is defined by

$$\int f \, \mathrm{d}\mu = \sup\{\mathcal{L}(f, P) : P \text{ is an } \mathcal{S}\text{-partition of } X\}.$$

Let's look at why our definition makes sense. Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f : X \to [0, \infty]$ is an $\mathcal{S}$-measurable function. Each $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X leads to an approximation of f from below by the $\mathcal{S}$-measurable simple function

$$\sum_{j=1}^{m} (\inf_{A_j} f) \chi_{A_j}.$$

This suggest that

$$\mu\left(\sum_{j=1}^{m} (\inf_{A_j} f) \chi_{A_j}\right) = \sum_{j=1}^{m} \mu(A_j) \inf_{A_j} f$$

(recall that $\inf f$ is a scalar) should be an approximation from below of our intuitive notion of $\int f \, \mathrm{d}\mu$. Taking the supremum of these approximations leads to our definition of $\int f \, \mathrm{d}\mu$.

Integration with respect to a measure can be called **Lebesgue integration**. We will start by showing that Lebesgue integration behaves as expected on the characteristic function as well as simple functions (represented as linear combinations of characteristic functions of disjoint sets).

▸ ***Integrating the Characteristic Function***

Lemma 2.1.5 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $E \in \mathcal{S}$. Then

$$\int \chi_E \, d\mu = \mu(E).$$

Proof. If P is the $\mathcal{S}$-partition of X consisting of E and its complement $X \setminus E$, then clearly

$$\mathcal{L}(\chi_E, P) = \mu(E) \inf_E f + \mu(X \setminus E) \inf_{X \setminus E} f = \mu(E) \cdot 1 + \mu(X \setminus E) \cdot 0 = \mu(E).$$

Thus,

$$\int \chi_E \, d\mu = \sup\{\mathcal{L}(f, P) : P \text{ is an } \mathcal{S}\text{-partition of } X\} \geq \mathcal{L}(\chi_E, P) = \mu(E).$$

To prove the inequality in the other direction, suppose P is an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X. Then

$$\mu(A_j) \inf_{A_j} \chi_E = \begin{cases} \mu(A_j) & A_j \subseteq E \\ 0 & \text{otherwise} \end{cases}$$

Thus,

$$\mathcal{L}(\chi_E, P) = \sum_{j=1}^{m} \mu(A_j) = \sum_{\{j : A_j \subseteq E\}} \mu(A_j) = \mu\Bigl(\bigcup_{\{j : A_j \subseteq E\}} A_j\Bigr) \leq \mu(E).$$

Thus, $\int \chi_E \, d\mu \leq \mu(E)$, completing the proof. □

Example 2.1.6 Suppose λ is Lebesgue measure on $\mathbb{R}$. As a special case of the result above, we have $\int \chi_{\mathbb{Q}} \, d\lambda = 0$ (because $|\mathbb{Q}| = 0$). Recall that $\chi_{\mathbb{Q}}$ is not Riemann integrable on $[0, 1]$. Thus, even at this early stage in our development of integration with respect to a measure, we have fixed one of the deficiencies of Riemann integration. Note also that Lemma 2.1.5 also implies that

$$\int \chi_{[0,1] \setminus \mathbb{Q}} \, d\lambda = 1$$

because $|[0, 1| \setminus \mathbb{Q}| = 1$, which is what we want. In contrast, the lower Riemann integral of $\chi_{[0,1] \setminus \mathbb{Q}}$ on $[0, 1]$ equals 0, which is not what we want.

► ***Integrating Non-Negative Simple Functions on Disjoint Sets***

Lemma 2.1.7 Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $E_1, \ldots, E_n$ are disjoint sets in $\mathcal{S}$, and $c_1, \ldots, c_n \in [0, \infty]$. Then

$$\int \left(\sum_{k=1}^{n} c_k \chi_{E_k} \right) \, \mathrm{d}\mu = \sum_{k=1}^{n} c_k \mu(E_k).$$

Proof. WLOG, assume that $E_1, \ldots, E_n$ is an $\mathcal{S}$-partition of X (by replacing n by $n+1$ and setting $E_{n+1} = X \setminus (E_1 \cup \cdots \cup E_n)$ and $c_{n+1} = 0$). If P is the $\mathcal{S}$-partition $E_1, \ldots, E_n$ of X, then

$$\mathcal{L}\left(\sum_{k=1}^{n} c_k \chi_{E_k}, P \right) = \sum_{k=1}^{n} c_k \mu(E_k).$$

Thus,

$$\int \left(\sum_{k=1}^{n} c_k \chi_{E_k} \right) \, \mathrm{d}\mu \geq \mathcal{L}\left(\sum_{k=1}^{n} c_k \chi_{E_k}, P \right) = \sum_{k=1}^{n} c_k \mu(E_k).$$

To prove the other direction, suppose that P is an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X. Then

$$\begin{aligned}
\mathcal{L}\left(\sum_{k=1}^{n} c_k \chi_{E_k}, P \right) &= \sum_{j=1}^{m} \mu\left(A_j\right) \min_{\{i : A_j \cap E_i \neq \varnothing\}} c_i && \text{here, inf = min} \\
&= \sum_{j=1}^{m} \sum_{k=1}^{n} \mu\left(A_j \cap E_k\right) \min_{\{i : A_j \cap E_i \neq \varnothing\}} c_i \\
&\leq \sum_{j=1}^{m} \sum_{k=1}^{n} \mu\left(A_j \cap E_k\right) c_k \\
&= \sum_{k=1}^{n} c_k \sum_{j=1}^{m} \mu\left(A_j \cap E_k\right) = \sum_{k=1}^{n} c_k \mu\left(E_k\right).
\end{aligned}$$

The inequality above implies that $\int \left(\sum_{k=1}^{n} c_k \chi_{E_k} \right) \, \mathrm{d}\mu \leq \sum_{k=1}^{n} c_k \mu\left(E_k\right)$, completing the proof.

The next unsurprising result shows that integration is order preserving.

Lemma 2.1.8 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g : X \to [0, \infty]$ are $\mathcal{S}$-measurable functions such that $f(x) \leq g(x)$ for all $x \in X$. Then $\int f \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu$.

Proof. Suppose P is an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X. Then $\inf_{A_j} f \leq \inf_{A_j} g$ for each $j = 1, \ldots, m$.

Thus $\mathcal{L}(f, P) \leq \mathcal{L}(g, P)$. Hence $\int f \, d\mu \leq \int g \, d\mu$. $\square$

2.1.2 Monotone Convergence Theorem

We start with a mild restatement of the definition of the integral of a non-negative function.

Lemma 2.1.9 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f : X \to [0, \infty]$ is $\mathcal{S}$-measurable. Then

$$\begin{aligned} \int f \, \mathrm{d}\mu = \sup \Bigg\{ \sum_{j=1}^{m} c_j \mu(A_j) : & A_1, \ldots, A_m \text{ are disjoint sets in } S \\ & c_1, \ldots, c_m \in [0, \infty), \\ & \forall x \in X : f(x) \geq \sum_{j=1}^{m} c_j \chi_{A_j}(x) \Bigg\}. \end{aligned} \tag{2.1}$$

Intuition. This is just saying that each $\sum_{j=1}^{m} c_j \chi_{A_j}(x)$ is an approximation of f from below and we are taking the supremum over all such approximations.

Proof. First, by Lemmas 2.1.7 and 2.1.8, the LHS is bigger than or equal to the RHS.

To prove that the RHS is bigger than or equal to the LHS, first assume that $\inf_A f < \infty$ for every $A \in \mathcal{S}$ with $\mu(A) > 0$. Then for P an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of non-empty subsets of X, take $c_j = \inf_{A_j} f$, which shows that $\mathcal{L}(f, P)$ is in the set on the right side of (2.1) is bigger than or equal to the LHS.

The only remaining case to consider is when there exists a set $A \in \mathcal{S}$ such that $\mu(A)$ and $\inf_A f = \infty$ (which implies that $f(x) = \infty$ for all $x \in A$). In this case, for arbitrary $t \in (0, \infty)$, we can take $m = 1$, $A_1 = A$, and $c_1 = t$. These choices show that the right side of (2.1) is at least $t\mu(A)$. Since t is an arbitrary positive number, this shows that the RHS of 2.1 equals ∞, which is of course greater or equal to the LHS, completing the proof. $\square$

The next result allows us to interchange limits and integrals in certain circumstances.

Theorem 2.1.10 [Monotone Convergence Theorem] Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $0 \leq f_1 \leq f_2 \leq \cdots$ is an increasing sequence of $\mathcal{S}$-measurable functions. Define $f : X \to [0, \infty]$ by

$$f(x) = \lim_{k \to \infty} f_k(x).$$

Then

$$\lim_{k\to\infty}\int f_k \,\mathrm{d}\mu = \int \lim_{k\to\infty} f_k \,\mathrm{d}\mu = \int f \,\mathrm{d}\mu.$$

Proof. By Proposition 1.2.27, the function f is $\mathcal{S}$-measurable. Since $f_k(x) \le f(x)$ for every $x \in X$, Lemma 2.1.8 tells us that $\int f_k \, d\mu \le \int f \, d\mu$. Thus,

$$\lim_{k\to\infty} \int f_k \, d\mu \le \int f \, d\mu.$$

To prove the inequality in the other direction, suppose $A_1, \ldots, A_m$ are disjoint sets in $\mathcal{S}$ and $c_1, \ldots, c_m \in [0, \infty)$ are such that

$$\forall x \in X : f(x) \ge \sum_{j=1}^{m} c_j \chi_{A_j}(x). \tag{2.2}$$

Let $t \in (0, 1)$. For $k \in \mathbb{Z}^+$, let

$$E_k = \left\{ x \in X : f_k(x) \ge t \sum_{j=1}^{m} c_j \chi_{A_j}(x) \right\}.$$

Then $E_1 \subseteq E_2 \subseteq \cdots$ is an increasing sequence of sets in $\mathcal{S}$ whose union equals X. By Lemma 1.3.9,

$$\forall j \in \{1, \ldots, m\} : \lim_{k\to\infty} \mu(A_j \cap E_k) = \mu(A_j).$$

If $k \in \mathbb{Z}^+$, then

$$f_k(x) \ge \sum_{j=1}^{m} t c_j \chi_{A_j \cap E_k}(x)$$

for every $x \in X$. By 2.1.9,

$$\int f_k \, d\mu \ge t \sum_{j=1}^{m} c_j \mu(A_j \cap E_k).$$

Taking the limit as $k \to \infty$ of both sides of the inequality above gives

$$\lim_{k\to\infty} \int f_k \, d\mu \ge t \sum_{j=1}^{m} c_j \mu(A_j).$$

Now taking the limit as t increases to 1 shows that

$$\lim_{k\to\infty} \int f_k \, d\mu \ge \sum_{j=1}^{m} c_j \mu(A_j).$$

Taking the supremum of the inequality above over all $\mathcal{S}$-partitions $A_1, \ldots, A_m$ of X and all $c_1, \ldots, c_m \in [0, \infty)$ satisfying (2.2) shows (using Lemma 2.1.9) that we have $\lim_{k\to\infty} \int f_k \, d\mu \ge \int f \, d\mu$, completing the proof. $\square$

The proof that the integral for general non-negative $\mathcal{S}$-measurable functions is additive will use the Monotone Convergence Theorem and our next result.

It is easy to see that the representation of a simple function $h : X \to [0, \infty]$ in the form $\sum_{k=1}^{n} c_k \chi_{E_k}$ is not unique. Requiring the numbers $c_1, \ldots, c_n$ to be distinct and $E_1, \ldots, E_n$ to be nonempty and disjoin with $E_1 \cup \cdots \cup E_n = X$ produces what is called the **standard representation** of a simple function [take $E_k = h^{-1}(\{c_k\})$, where $c_1, \ldots, c_n$ are distinct values of h].

The following lemma shows that all representations (including those with sets that are not disjoint) of a simple measurable function give the same sum that we expect from integration.

Lemma 2.1.11 Suppose $(X, \mathcal{S}, \mu)$ is a measure space. Suppose

- $a_1, \ldots, a_m, b_1, \ldots, b_n \in [0, \infty]$ and
- $A_1, \ldots, A_m, B_1, \ldots, B_n \in \mathcal{S}$

are such that $\sum_{j=1}^{m} a_j \chi_{A_j} = \sum_{k=1}^{n} b_k \chi_{B_k}$. Then

$$\sum_{j=1}^{m} a_j \mu\left(A_j\right) = \sum_{k=1}^{n} b_k \mu\left(B_k\right).$$

Proof. WLOG, assume $A_1 \cup \cdots \cup A_m = X$ (otherwise add the term $0\chi_{X \setminus (A_1 \cup \cdots \cup A_m)}$).

Suppose A_1 and A_2 are not disjoint. Then we can write

$$a_1 \chi_{A_1} + a_2 \chi_{A_2} = a_1 \chi_{A_1 \setminus A_2} + a_2 \chi_{A_2 \setminus A_1} + (a_1 + a_2)\, \chi_{A_1 \cap A_2}, \tag{2.3}$$

where the three sets appearing on the right side of the equation above are disjoint. Now

- $A_1 = (A_1 \setminus A_2) \cup (A_1 \cap A_2)$ and
- $A_2 = (A_2 \setminus A_1) \cup (A_1 \cap A_2)$;

each of these unions is a disjoint union. Thus

- $\mu(A_1) = \mu(A_1 \setminus A_2) + \mu(A_1 \cap A_2)$ and
- $\mu(A_2) = \mu(A_2 \setminus A_1) + \mu(A_1 \cap A_2)$.

Hence

$$a_1 \mu(A_1) + a_2 \mu(A_2) = a_1 \mu(A_1 \setminus A_2) + a_2 \mu(A_2 \setminus A_1) + (a_1 + a_2)\, \mu(A_1 \cap A_2).$$

This, in conjunction with (2.3), shows that if we replace the two sets A_1, A_2 by the three disjoint sets $A_1 \setminus A_2$, $A_2 \setminus A_1$, $A_1 \cap A_2$ and make the appropriate adjustments to the coefficients $a_1, \ldots, a_m$, then the value of the sum $\sum_{j=1}^{m} a_j \mu\left(A_j\right)$ is unchanged (although m has increased by 1).

Repeating this process with all pairs of subsets among $A_1, \ldots, A_m$ that are not disjoint after each step, in a finite number of steps we can convert the initial list $A_1, \ldots, A_m$ into a disjoint list of subsets without changing the value of $\sum_{j=1}^{m} a_j \mu\left(A_j\right)$.

The next step is to make the numbers $a_1, \ldots, a_m$ distinct. This is done by replacing the sets corresponding to each a_j by the union of those sets, and using finite additivity of the measure μ to show that the value of the sum $\sum_{j=1}^{m} a_j \mu\left(A_j\right)$ does not change.

Finally, drop any terms for which $A_j = \varnothing$, getting the standard representation for a simple function. We have now shown that the original value of $\sum_{j=1}^{m} a_j \mu\left(A_j\right)$ is equal to the value if we use the standard representation of the simple function $\sum_{j=1}^{m} a_j \chi_{A_j}$. The same procedure can be used with the representation $\sum_{k=1}^{n} b_k \chi_{B_k}$ to show that $\sum_{k=1}^{n} b_k \mu\left(\chi_{B_k}\right)$ equals what we would get with the standard representation. Thus the equality of the functions $\sum_{j=1}^{m} a_j \chi_{A_j}$ and $\sum_{k=1}^{n} b_k \chi_{B_k}$ implies the equality $\sum_{j=1}^{m} a_j \mu\left(A_j\right) = \sum_{k=1}^{n} b_k \mu\left(B_k\right)$. □

Now we can show that our definition of integration does the right thing with simple measurable functions that might not be expressed in the standard representation. The result below differs from Lemma 2.1.7 mainly because the sets $E_1, \ldots, E_n$ in the result below are not required to be disjoint. Like the previous result, the next result would follow immediately from the linearity integration if that property had already been proved.

Lemma 2.1.12 Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $E_1, \ldots, E_n \in \mathcal{S}$, and $c_1, \ldots, c_n \in [0, \infty]$. Then

$$\int \left(\sum_{k=1}^{n} c_k \chi_{E_k} \right) \mathrm{d}\mu = \sum_{k=1}^{n} c_k \mu\left(E_k\right).$$

Proof. Write the simple function $\sum_{k=1}^{n} c_k \chi_{E_k}$ in the standard representation for a simple function and then use Lemma 2.1.7 and Lemma 2.1.11. □

▸ ***Integrating Non-Negative Measurable Functions***

Lemma 2.1.13 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g : X \to [0, \infty]$ are $\mathcal{S}$-measurable functions. Then

$$\int (f + g)\, \mathrm{d}\mu = \int f\, \mathrm{d}\mu + \int g\, \mathrm{d}\mu.$$

Proof. Lemma 2.1.12 tells us that the desired result holds for simple non-negative $\mathcal{S}$-measurable functions. Thus, we approximate by such functions. Specifically, let $f_1, f_2, \ldots$ and $g_1, g_2, \ldots$ be increasing sequences of simple nonnegative $\mathcal{S}$-measurable functions such that

$$\lim_{k\to\infty} f_k(x) = f(x) \quad \text{and} \quad \lim_{k\to\infty} g_k(x) = g(x)$$

for all $x \in X$ (see Theorem 1.5.7 for the existence of such increasing sequences). Then

$$\begin{aligned}\int (f+g)\,\mathrm{d}\mu &= \lim_{k\to\infty}\int (f_k+g_k)\,\mathrm{d}\mu \\ &= \lim_{k\to\infty}\int f_k\,\mathrm{d}\mu + \lim_{k\to\infty}\int g_k\,\mathrm{d}\mu \\ &= \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu,\end{aligned}$$

where the first and third equalities follow from the Monotone Convergence Theorem and the second equality holds by Lemma 2.1.12. □

The lower Riemann integral is not additive, even for bounded non-negative measurable functions. For example, if $f = \chi_{\mathbf{Q}\cap[0,1]}$ and $g = \chi_{[0,1]\setminus\mathbf{Q}}$, then

$$L(f,[0,1]) = 0, \quad L(g,[0,1]) = 0, \quad L(f+g,[0,1]) = 1.$$

In contract, if λ is Lebesgue measure on the Borel subsets of $[0,1]$, then

$$\int f\,\mathrm{d}\lambda = 0, \quad \int g\,\mathrm{d}\lambda = 1, \quad \int (f+g)\,\mathrm{d}\lambda = 1.$$

More generally, we have just proved that

$$\int (f+g)\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu$$

for every measure μ and for all non-negative measurable functions f and g. Recall that integration with respect to a measure is defined via lower Lebesgue sums in a similar fashion to the definition of the lower Riemann integral via lower Riemann sums. However, we have just seen that the integral with respect to a measure has considerably nicer behaviour (additivity!) than the lower Riemann integral.

2.1.3 Integration of Real-Valued Functions

The following definition gives us a standard way to write an arbitrary real-valued function as the difference of two nonnegative functions.

Definition 2.1.14 Suppose $f : X \to [-\infty, \infty]$ is a function. Define functions f^+ and f^- from X to $[0, \infty]$ by $f^+ = \max\{f, 0\}$ and $f^- = \max\{-f, 0\}$, i.e.,

$$f^+(x) = \begin{cases} f(x) & \text{if } f(x) \geq 0, \\ 0 & \text{if } f(x) < 0 \end{cases}, \quad f^-(x) = \begin{cases} 0 & \text{if } f(x) \geq 0 \\ -f(x) & \text{if } f(x) < 0 \end{cases}$$

Both f^+ and f^- are non-negative functions. Moreover, for any $f : X \to [-\infty, \infty]$, we have

$$f = f^+ - f^- \quad \text{and} \quad |f| = f^+ + f^-.$$

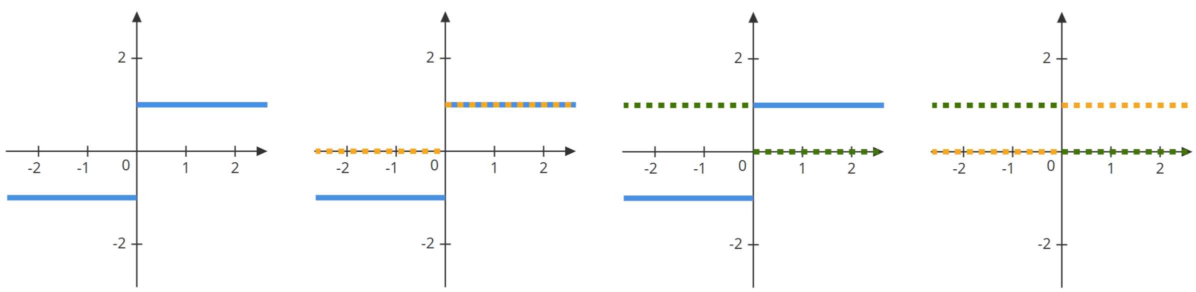


Figure 2.1: From left to right: $\{f\}$, $\{f, f^+\}$, $\{f, f^-\}$, and $\{f^+, f^-\}$.

Example 2.1.15 Consider the function (see plots above)

$$f(x) = \begin{cases} 1 & x > 0 \\ -1 & x \leq 0 \end{cases}$$

Then

$$f^+(x) = \begin{cases} 1 & x > 0 \\ 0 & x \leq 0 \end{cases}, \qquad f^-(x) = \begin{cases} 0 & x > 0 \\ 1 & x \leq 0 \end{cases}$$

You should verify that $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

This decomposition allows us to extend our definition of integration to functions that take on negative as well as positive values.

Definition 2.1.16 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f : X \to [-\infty, \infty]$ is an $\mathcal{S}$-measurable function such that at least one of $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ is finite.[a] The integral of f with respect to μ, denoted $\int f \, d\mu$, is defined by

$$\int f \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu$$

[a]To avoid the undefined expression $\infty - \infty$.

If $f \geq 0$, then $f^+ = f$ and $f^- = 0$; thus this definition is consistent with the previous definition of the integral of a nonnegative function. Below is a function whose integral is not defined.

Example 2.1.17 Suppose λ is Lebesgue measure on $\mathbb{R}$ and $f : \mathbb{R} \to \mathbb{R}$ is defined by

$$f(x) = \begin{cases} 1 & x \geq 0 \\ -1 & x < 0. \end{cases}$$

Then $\int f \, d\lambda$ is not defined because $\int f^+ \, d\lambda = \infty$ and $\int f^- \, d\lambda = \infty$.

Lemma 2.1.18 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g : X \to \mathbf{R}$ are $\mathcal{S}$-measurable functions such that $\int |f| \, d\mu < \infty$ and $\int |g| \, d\mu < \infty$. Then

$$\int (f + g) \, d\mu = \int f \, d\mu + \int g \, d\mu$$

Proof. Recall that $h = h^+ - h^-$. Thus, we can write $(f+g)^+ - (f+g)^- = f + g = f^+ - f^- + g^+ - g^-$. With some algebra, we get $(f+g)^+ + f^- + g^- = (f+g)^- + f^+ + g^+$. Both sides of this equation are sums of nonnegative functions. Thus integrating both sides with respect to μ and using 3.16 gives

$$\int (f+g)^+ \, d\mu + \int f^- \, d\mu + \int g^- \, d\mu = \int (f+g)^- \, d\mu + \int f^+ \, d\mu + \int g^+ \, d\mu$$
$$\int (f+g)^+ \, d\mu - \int (f+g)^- \, d\mu = \int f^+ \, d\mu - \int f^- \, d\mu + \int g^+ \, d\mu - \int g^- \, d\mu,$$

where the left side is not of the form $\infty - \infty$ because $(f+g)^+ \leq f^+ + g^+$and $(f+g)^- \leq f^- + g^-$. The equation above can be rewritten as

$$\int (f+g) \, d\mu = \int f \, d\mu + \int g \, d\mu,$$

completing the proof. □

Note that the condition $\int |f| \,\mathrm{d}\mu < \infty$ (which appears in the statement of Lemma 2.1.18) is equivalent to the condition $\int f^+ \,\mathrm{d}\mu < \infty$ and $\int f^- \,\mathrm{d}\mu < \infty$ (because $|f| = f^+ + f^-$).

Lemma 2.1.19 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f : X \to [-\infty, \infty]$ is a function such that $\int f \,\mathrm{d}\mu$ is defined. If $c \in \mathbf{R}$, then

$$\int cf \,\mathrm{d}\mu = c \int f \,\mathrm{d}\mu$$

Proof. First consider the case where f is a nonnegative function and $c \geq 0$. If P is an $\mathcal{S}$-partition of X, then clearly $\mathcal{L}(cf, P) = c\mathcal{L}(f, P)$. Thus

$$\int cf \,\mathrm{d}\mu = c \int f \,\mathrm{d}\mu.$$

Now consider the general case where f takes values in $[-\infty, \infty]$. Suppose $c \geq 0$. Then

$$\begin{aligned}
\int cf \,\mathrm{d}\mu &= \int (cf)^+ \,\mathrm{d}\mu - \int (cf)^- \,\mathrm{d}\mu \\
&= \int cf^+ \,\mathrm{d}\mu - \int cf^- \,\mathrm{d}\mu \\
&= c\left(\int f^+ \,\mathrm{d}\mu - \int f^- \,\mathrm{d}\mu\right) \\
&= c \int f \,\mathrm{d}\mu,
\end{aligned}$$

where the third line follows from the first paragraph of this proof. Finally, now suppose $c < 0$ (still assuming that f takes values in $[-\infty, \infty]$). Then $-c > 0$ and

$$\begin{aligned}
\int cf \,\mathrm{d}\mu &= \int (cf)^+ \mathrm{d}\mu - \int (cf)^- \mathrm{d}\mu \\
&= \int (-c)f^- \,\mathrm{d}\mu - \int (-c)f^+ \,\mathrm{d}\mu \\
&= (-c)\left(\int f^- \,\mathrm{d}\mu - \int f^+ \,\mathrm{d}\mu\right) \\
&= c \int f \,\mathrm{d}\mu
\end{aligned}$$

completing the proof. □

Together, Lemma 2.1.18 (additivity of integral) and Lemma 2.1.19 (homogeneity of integral) tell us that Lebesgue integration is linear.

The next result generalizes Lemma 2.1.8 to all real-valued functions.

Lemma 2.1.20 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f, g : X \to \mathbf{R}$ are $\mathcal{S}$-measurable functions such that $\int f \, d\mu$ and $\int g \, d\mu$ are defined. Suppose also that $f(x) \le g(x)$ for all $x \in X$. Then $\int f \, d\mu \le \int g \, d\mu$.

Proof. Proof The cases where $\int f \, d\mu = \pm\infty$ or $\int g \, d\mu = \pm\infty$ are left to the reader. Thus we assume that $\int |f| \, d\mu < \infty$ and $\int |g| \, d\mu < \infty$.
The additivity (3.21) and homogeneity (3.20 with $c = -1$) of integration imply that

$$\int g \, d\mu - \int f \, d\mu = \int (g - f) \, d\mu$$

The last integral is nonnegative because $g(x) - f(x) \ge 0$ for all $x \in X$. □

The absolute value of integral is bounded above by the integral of absolute value.

Proposition 2.1.21 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f : X \to [-\infty, \infty]$ is a function such that $\int f \, d\mu$ is defined. Then

$$\left| \int f \, d\mu \right| \le \int |f| \, d\mu$$

Proof. Because $\int f \, d\mu$ is defined, f is an $\mathcal{S}$-measurable function and at least one of $\int f^+ \, d\mu$ and $\int f^- \, d\mu$ is finite. Thus

$$\begin{aligned}
\left| \int f \, d\mu \right| &= \left| \int f^+ \, d\mu - \int f^- \, d\mu \right| \\
&\le \int f^+ \, d\mu + \int f^- \, d\mu \\
&= \int \left(f^+ + f^- \right) \, d\mu \\
&= \int |f| \, d\mu,
\end{aligned}$$

as desired. □

2.2 Limits of Integrals & Integrals of Limits

This section focuses on interchanging limits and integrals. One result in this category is *Monotone Convergence Theorem* we saw in the previous section, which states that if we have a sequence of non-negative measurable functions that are increasing pointwise and the limit of the sequence exists almost everywhere, then the integral of the limit is equal to the limit of the integrals of the functions. We will see more similar results in this section.

2.2.1 Bounded Convergence Theorem

We first introduce the notion of integration on a subset.

Definition 2.2.1 Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $E \in \mathcal{S}$. If $f : X \to [-\infty, \infty]$ is an $\mathcal{S}$-measurable function. Then $\int_E f \, \mathrm{d}\mu$ is defined by

$$\int_E f \, \mathrm{d}\mu = \int \chi_E f \, \mathrm{d}\mu$$

if the RHS of the above equation is defined. Otherwise, $\int_E f \, \mathrm{d}\mu$ is undefined.

Alternatively, you can think of $\int_E f \, \mathrm{d}\mu$ as $\int f|_E \, \mathrm{d}\mu_E$, where $f|_E$ is the function f restricted to E and μ_E is the measure obtained by restricting μ to the elements of $\mathcal{S}$ that are contained in E. Notice that according to the definition above, the notation $\int_X f \, \mathrm{d}\mu$ means the same as $\int f \, \mathrm{d}\mu$.

Lemma 2.2.2 Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $E \in \mathcal{S}$, and $f : X \to [-\infty, \infty]$ is a function such that $\int_E f \, \mathrm{d}\mu$ is defined. Then

$$\left| \int_E f \, \mathrm{d}\mu \right| \leq \mu(E) \sup_E |f|.$$

Proof. Let $c = \sup_E |f|$. We have

$$\begin{aligned}
\left| \int_E f \, \mathrm{d}\mu \right| &= \left| \int \chi_E f \, \mathrm{d}\mu \right| && \text{Definition 2.2.1} \\
&\leq \int \chi_E |f| \, \mathrm{d}\mu && \text{Proposition 2.1.21} \\
&\leq \int c\chi_E \, \mathrm{d}\mu && \text{Lemma 2.1.8} \\
&= c\mu(E) && \text{Lemma 2.1.12}
\end{aligned}$$

$\square$

The next result could be proved as a special case of the *Dominated Convergence Theorem*, which we prove later in this section. Thus you could skip the proof here. However, sometimes you get more insight by seeing an easier proof of an important special case. Thus you may want to read the easy proof of the Bounded Convergence Theorem that is presented next.

Theorem 2.2.3 [Bounded Convergence Theorem] Suppose $(X, \mathcal{S}, \mu)$ is a measure space with $\mu(X) < \infty$. Suppose $f_1, f_2, \ldots$ is a sequence of $\mathcal{S}$-measurable functions from X to $\mathbb{R}$ that converges pointwise on X to a function $f : X \to \mathbb{R}$. If there exists $c \in (0, \infty)$ such that $|f_k(x)| \leq c$ for all $k \in \mathbb{Z}^+$ and all $x \in X$, then

$$\lim_{k \to \infty} \int f_k \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu.$$

Remark/Intuition. Note the key role of *Egorov's Theorem* (Theorem 1.5.3, pointwise convergence is close to uniform convergence on a measure space with finite total masure) in proofs involving interchanging limits and integrals.

Proof. By Proposition 1.2.23, since f is the pointwise limit of $\mathcal{S}$-measurable functions f_k's, f is $\mathcal{S}$-measurable. Suppose c satisfies the hypothesis of this theorem. Let $\varepsilon > 0$. By Egorov's Theorem (Theorem 1.5.3), there exists $E \in \mathcal{S}$ such that

$$\mu(X \setminus E) < \frac{\varepsilon}{4c}$$

and $f_1, f_2, \ldots$ converges uniformly to f on E. Now

$$\begin{aligned}
\left| \int f_k \, \mathrm{d}\mu - \int f \, \mathrm{d}\mu \right| &= \left| \int_{X \setminus E} (f_k - f) \, \mathrm{d}\mu + \int_E (f_k - f) \, \mathrm{d}\mu \right| \\
&= \left| \int_{X \setminus E} f_k \, \mathrm{d}\mu - \int_{X \setminus E} f \, \mathrm{d}\mu + \int_E (f_k - f) \, \mathrm{d}\mu \right| \\
&\leq \int_{X \setminus E} |f_k| \, \mathrm{d}\mu + \int_{X \setminus E} |f| \, \mathrm{d}\mu + \int_E |f_k - f| \, \mathrm{d}\mu \\
&< \frac{\varepsilon}{2} + \mu(E) \sup_E |f_k - f| \, . && \text{Lemma 2.2.2}
\end{aligned}$$

Because $f_1, f_2, \ldots$ converges uniformly to f on E and $\mu(E) < \infty$, the RHS of the inequality above is less than ε for k sufficiently large, which completes the proof. □

2.2.2 Sets of Measure 0 in Integration Theorems

Suppose $(X, \mathcal{S}, \mu)$ is a measure space. If $f, g : X \to [-\infty, \infty]$ are $\mathcal{S}$-measurable functions and two functions only differ on a set of measure zero, i.e.,

$$\mu(\{x \in X : f(x) \neq g(x)\}) = 0,$$

then the definition of an integral implies that $\int f \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu$ (or both integrals are undefined). Because what happens on a set of measure 0 often does not matter, the following definition is useful.

Definition 2.2.4 Suppose $(X, \mathcal{S}, \mu)$ is a measure space. A set $E \in \mathcal{S}$ is said to contain μ-**almost every** element of X if $\mu(X \setminus E) = 0$. If the measure μ is clear from the context, then the phrase **almost every** (or **a. e.**) can be used.

For example, *almost every* real number is irrational (with respect to the usual Lebesgue measure on $\mathbb{R}$) because $|\mathbb{Q}| = 0$. Intuitively, the statement "almost every $\mathcal{T}$" means that statement $\mathcal{T}$ holds on all X except possibly on an "infinitely small" set $X \setminus E$.

Theorems about integrals can almost always be relaxed so that the hypotheses apply only almost everywhere instead of everywhere. For example, consider the *Bounded Convergence Theorem* (Theorem 2.2.3). One of the hypotheses is that

$$\forall x \in X : \lim_{k \to \infty} f_k(x) = f(x)$$

Suppose that the hypotheses of the Bounded Convergence Theorem hold except that the equation above holds only almost everywhere, meaning there is a set $E \in \mathcal{S}$ such that $\mu(X \setminus E) = 0$ and the equation above holds for all $x \in E$. Define new functions $g_1, g_2, \ldots$ and g by

$$g_k(x) = \begin{cases} f_k(x) & x \in E \\ 0 & x \in X \setminus E \end{cases} \qquad g(x) = \begin{cases} f(x) & x \in E \\ 0 & x \in X \setminus E \end{cases}$$

Then $\forall x \in X : \lim_{k \to \infty} g_k(x) = g(x)$. Hence the Bounded Convergence Theorem implies that

$$\lim_{k \to \infty} \int g_k \, \mathrm{d}\mu = \int g \, \mathrm{d}\mu,$$

which immediately implies that

$$\lim_{k \to \infty} \int f_k \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$$

because $\int g_k \, \mathrm{d}\mu = \int f_k \, \mathrm{d}\mu$ and $\int g \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$.

2.2.3 Dominated Convergence Theorem

The next result tells us that if a non-negative function has a finite integral, then its integral over all small sets (in the sense of measure) is small.

Lemma 2.2.5 Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $g : X \to [0, \infty]$ is $\mathcal{S}$-measurable, and $\int g \, d\mu < \infty$. Then for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\int_B g \, d\mu < \varepsilon$$

for every set $B \in \mathcal{S}$ such that $\mu(B) < \delta$.

Proof. Suppose $\varepsilon > 0$. By Lemma 2.1.9, we can approximate g with a simple $\mathcal{S}$-measurable function $h : X \to [0, \infty)$ such that $0 \le h \le g$ and

$$\int g \, d\mu - \int h \, d\mu < \frac{\varepsilon}{2}$$

Let $H = \max\{h(x) : x \in X\}$ and let $\delta > 0$ be such that $H\delta < \varepsilon/2$. Let $B \in \mathcal{S}$ and $\mu(B) < \delta$. Then

$$\begin{aligned} \int_B g \, d\mu &= \int_B (g - h + h) \, d\mu \\ &= \int_B (g - h) \, d\mu + \int_B h \, d\mu \\ &\le \int (g - h) \, d\mu + H\mu(B) \\ &< \frac{\varepsilon}{2} + H\delta \\ &< \varepsilon, \end{aligned}$$

as desired. □

Some theorems, such as Egorov's Theorem, have as a hypothesis that the measure of the entire space is finite, i.e., $\mu(X) < \infty$ when X is the universal set. The next result sometimes allows us to get around this hypothesis by restricting attention to a key set of finite measure.

Lemma 2.2.6 Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $g : X \to [0, \infty]$ is $\mathcal{S}$-measurable, and $\int g \, \mathrm{d}\mu < \infty$. Then for every $\varepsilon > 0$, there exists $E \in \mathcal{S}$ such that $\mu(E) < \infty$ and

$$\int_{X \setminus E} g \, \mathrm{d}\mu < \varepsilon$$

Proof. Suppose $\varepsilon > 0$. Let P be an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X such that

$$\int g \, \mathrm{d}\mu < \varepsilon + \mathcal{L}(g, P). \tag{2.4}$$

Let E be the union of those A_j such that $\inf_{A_j} f > 0$. Then $\mu(E) < \infty$ (because otherwise we would have $\mathcal{L}(g, P) = \infty$, which contradicts the hypothesis that $\int g \, \mathrm{d}\mu < \infty$). Now

$$\begin{aligned} \int_{X \setminus E} g \, \mathrm{d}\mu &= \int g \, \mathrm{d}\mu - \int \chi_E g \, \mathrm{d}\mu \\ &< (\varepsilon + \mathcal{L}(g, P)) - \mathcal{L}(\chi_E g, P) \\ &= \varepsilon, \end{aligned}$$

where the second line follows from (2.4) and the definition of the integral of a nonnegative function, and the last line holds because $\inf_{A_j} f = 0$ for each A_j not contained in E. □

Suppose $(X, \mathcal{S}, \mu)$ is a measure space and $f_1, f_2, \ldots$ is a sequence of $\mathcal{S}$-measurable functions on $\lim_{k\to\infty} f_k(x) = f(x)$ for every (or almost every) $x \in X$. In general, it is not true that $\lim_{k\to\infty} \int f_k \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu$ as pointwise convergence is not strong enough. This naturally motivates the following question: under what conditions would pointwise convergence be strong enough to imply that the integral of limits is equal to the limit of integrals of functions?

We already have two good theorems about interchanging limits and integrals. However, both of these theorems have restrictive hypotheses.

- The *Monotone Convergence Theorem* requires all the functions to be (1) non-negative and that (2) the sequence of functions to be increasing.
- The *Bounded Convergence Theorem* requires (3) the measure of the whole space to be finite and that (4) the sequence of functions to be uniformly bounded by a constant.

The next theorem is the grand result in this area.

1. It does not require the sequence of functions to be nonnegative.
2. It does not require the sequence of functions to be increasing.
3. It does not require the measure of the whole space to be finite.
4. It does not require the sequence of functions to be uniformly bounded.

All these hypotheses are replaced only by a requirement that the *sequence of functions is pointwise bounded by a function with a finite integral*.

(Note that the Bounded Convergence Theorem follows immediately from the result below—take g to be an appropriate constant function and use the hypothesis in the Bounded Convergence Theorem that $\mu(X) < \infty$.)

Theorem 2.2.7 [Dominated Convergence Theorem] Suppose $(X, \mathcal{S}, \mu)$ is a measure space, $f : X \to [-\infty, \infty]$ is $\mathcal{S}$-measurable, and $f_1, f_2, \ldots$ are $\mathcal{S}$-measurable functions from X to $[-\infty, \infty]$ such that

$$\lim_{k\to\infty} f_k(x) = f(x)$$

for almost every $x \in X$. If there exists an $\mathcal{S}$-measurable function $g : X \to [0, \infty]$ such that

$$\int g \, d\mu < \infty \quad \text{and} \quad |f_k(x)| \le g(x)$$

for every $k \in \mathbb{Z}^+$ and almost every $x \in X$, then

$$\lim_{k\to\infty} \int f_k \, d\mu = \int f \, d\mu.$$

2.2.4 Riemann Integrals and Lebesgue Integrals

Theorem 2.2.8 Suppose $a < b$ and $f : [a, b] \to \mathbf{R}$ is a bounded function. Then f is Riemann integrable if and only if

$$| \{x \in [a, b] : f \text{ is not continuous at } x\} |= 0.$$

Furthermore, if f is Riemann integrable and λ denotes Lebesgue measure on $\mathbf{R}$, then f is Lebesgue measurable and

$$\int_a^b f = \int_{[a,b]} f d\lambda$$

Proof. Proof Suppose $n \in \mathbf{Z}^+$. Consider the partition P_n that divides $[a, b]$ into 2^n subintervals of equal size. Let $I_1, \ldots, I_{2^n}$ be the corresponding closed subintervals, each of length $(b - a)/2^n$. Let

$$3.35\ g_n = \sum_{j=1}^{2^n} \left(\inf_{I_j} f\right) \chi_{I_j} \text{ and } \quad h_n = \sum_{j=1}^{2^n} \left(\sup_{I_j} f\right) \chi_{I_j}.$$

The lower and upper Riemann sums of f for the partition P_n are given by integrals. Specifically, 3.36 $L(f, P_n, [a, b]) = \int_{[a,b]} g_n \, \mathrm{d}\lambda \quad$ and $\quad U(f, P_n, [a, b]) = \int_{[a,b]} h_n \, \mathrm{d}\lambda$, where λ is Lebesgue measure on $\mathbf{R}$. The definitions of g_n and h_n given in 3.35 are actually just a first draft of the definitions. A slight problem arises at each point that is in two of the intervals $I_1, \ldots, I_{2^n}$ (in other words, at endpoints of these intervals other than a and b). At each of these points, change the value of g_n to be the infimum of f over the union of the two intervals that contain the point, and change the value of h_n to be the supremum of f over the union of the two intervals that contain the point. This change modifies g_n and h_n on only a finite number of points. Thus the integrals in 3.36 are not affected. This change is needed in order to make 3.38 true (otherwise the two sets in 3.38 might differ by at most countably many points, which would not really change the proof but which would not be as aesthetically pleasing).

Clearly $g_1 \leq g_2 \leq \cdots$ is an increasing sequence of functions and $h_1 \geq h_2 \geq \cdots$ is a decreasing sequence of functions on $[a, b]$. Define functions $f^{\mathrm{L}} : [a, b] \to \mathbf{R}$ and $f^{\mathrm{U}} : [a, b] \to \mathbf{R}$ by

$$f^{\mathrm{L}}(x) = \lim_{n\to\infty} g_n(x) \text{ and } f^{\mathrm{U}}(x) = \lim_{n\to\infty} h_n(x).$$

Taking the limit as $n \to \infty$ of both equations in 3.36 and using the Bounded Convergence Theorem (3.26) along with Exercise 7 in Section 1A, we see that f^L and f^U are Lebesgue measurable functions and 3.37 $L(f, [a, b]) = \int_{[a,b]} f^{\mathrm{L}} \mathrm{d}\lambda$ and $U(f, [a, b]) = \int_{[a,b]} f^{\mathrm{U}} \mathrm{d}\lambda$. Now 3.37 implies that f is Riemann integrable if and only if

$$\int_{[a,b]} \left(f^{\mathrm{U}} - f^{\mathrm{L}}\right) d\lambda = 0$$

Because $f^{\mathrm{L}}(x) \leq f(x) \leq f^{\mathrm{U}}(x)$ for all $x \in [a, b]$, the equation above holds if and only if

$$\left|\left\{x \in [a, b] : f^{\mathrm{U}}(x) \neq f^{\mathrm{L}}(x)\right\}\right| = 0.$$

The remaining details of the proof can be completed by noting that 3.38 $\left\{x \in [a, b] : f^{\mathrm{U}}(x) \neq f^{\mathrm{L}}(x)\right\} = \{x \in [a, b] : f \text{ is not continuous at } x\}$.

We previously defined the notation $\int_a^b f$ to mean the Riemann integral of f. Because the Riemann integral and Lebesgue integral agree for Riemann integrable functions (see 3.34), we now redefine $\int_a^b f$ to denote the Lebesgue integral.

Definition 2.2.9 Suppose $-\infty \leq a < b \leq \infty$ and $f : (a, b) \to \mathbf{R}$ is Lebesgue measurable. Then

- $\int_a^b f$ and $\int_a^b f(x)dx$ mean $\int_{(a,b)} f \, d\lambda$, where λ is Lebesgue measure on $\mathbf{R}$;
- $\int_b^a f$ is defined to be $-\int_a^b f$.

The definition in the second bullet point above is made so that equations such as

$$\int_a^b f = \int_a^c f + \int_c^b f$$

remain valid even if, for example, $a < b < c$.

2.3 Chapter Summary

2.3.1 Integration with Respect to a Measure

Let $\mathcal{S}$ be a σ-algebra on a set X. An $\mathcal{S}$-**partition** is a way of breaking up the set X into a collection of disjoint pieces, where each piece is an element of $\mathcal{S}$. The elements of $\mathcal{S}$ represent the "measurable" subset of X, and an $\mathcal{S}$-partition allows us to break up X into a collection of these measurable subsets.

Let $(X, \mathcal{S}, \mu)$ bee a measure space, $f : X \to [0, \infty]$ an $\mathcal{S}$-measurable function, and P an $\mathcal{S}$-partition $A_1, \ldots, A_m$ of X. The **lower Lebesgue sum** is defined by

$$\mathcal{L}(f, P) = \sum_{j=1}^{m} \mu(A_j) \inf_{A_j} f,$$

where $\mu(A_j)$ denotes the Lebesgue measure and $\inf_A f$ denotes the infimum of f over A. The **Lebesgue integral** of f with respect to μ is defined by

$$\int f \, \mathrm{d}\mu = \sup\{\mathcal{L}(f, P) : P \text{ is an } \mathcal{S}\text{-partition of } X\}.$$

Intuitively, $\sum_{j=1}^{m} \mu(A_j) \inf_{A_j} f$ is an approximation from below of our intuition notion of $\int f \, \mathrm{d}\mu$; taking the supremum of these approximations leads to our definition of $\int f \, \mathrm{d}\mu$.

Define $f^+ = \{f, 0\}$ and $f^- = \max\{-f, 0\}$. The **integral** of a real-valued function f is defined by

$$\int f \, \mathrm{d}\mu = \int f^+ \, \mathrm{d}\mu = \int f^- \, \mathrm{d}\mu.$$

We proved the following good properties of Lebesgue integration.

- The integral of the characteristic function χ_E gives the size of E with respect to measure μ.
- The integral of a simple function on a disjoint union of measurable sets is the sum of the integrals of the function over each of the measurable sets, weighted by the constant value of the function on each set.
- Integration preserves order, i.e., $[\forall x \in X : f(x) \leq g(x)] \implies \int f \, \mathrm{d}\mu \leq \int g \, \mathrm{d}\mu$.
- Lebesgue integration is linear, i.e., $\int c(f + g) \, \mathrm{d}\mu = c \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu$.
- The absolute value of integral is bounded above by the integral of absolute value, i.e.,

$$\left| \int f \, \mathrm{d}\mu \right| \leq \int |f| \, \mathrm{d}\mu.$$

The **Monotone Convergence Theorem** states that if we have a sequence of non-negative measurable functions that are increasing pointwise and the limit of the sequence exists almost everywhere, then the integral of the limit is equal to the limit of the integrals of the functions.