Basic Mathematics for Statistics and Economics

Yashaswi Mohanty

Contents

Part 1

Analysis

CHAPTER 1

Measures

1.1. Why is measurement hard?

On the real line $\mathbb{R}$, we may want our measure to satisfy some properties that are consistent with our intuitive notion of "length". Formally, we want a function

$$\lambda : 2^{\mathbb{R}} \longrightarrow [0, \infty]$$

that satisfies

(1) $\lambda(\emptyset) = 0$
(2) $\lambda([a, b]) = b - a$ for $a \leq b \in \mathbb{R}$
(3) Countable additivity: For a countable collection of pairwise-disjoint sets $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathbb{R}$

$$\lambda\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \lambda(A_i) \tag{1}$$

(4) Translation invariance: $\lambda(A + a) = \lambda(A)$ for any $a \in \mathbb{R}$ where $A + a := \{\alpha + a \mid \alpha \in A\}$.

Quite counterintuitively, it turns out that no such function exists! To prove this assertion, we need to construct some special kinds of sets that only exist if we assume the Axiom of Choice.

Example 1.1.1. Define an equivalence relation $\sim$ on $[0, 1]$ such that

$$x \sim y \Leftrightarrow x - y \in \mathbb{Q}.$$

Note that there are uncountably many classes in such a construction as the equivalence class for any given irrational number can contain at most countably many other irrational numbers. For example,

$$\left[\frac{\pi}{4}\right] = \left\{\frac{\pi}{4} + q \mod 1 \mid q \in \mathbb{Q}\right\}.$$

Thus, using the Axiom of Choice, we can construct a set $E \subseteq [0, 1]$ such that E consists of exactly one "representative" from each equivalence class. Next, we can define

$$E_q := \{x + q \mod 1 \mid x \in E\}$$

so that $\{E_q\}_{q\in\mathbb{Q}}$ is a partition of $[0, 1]$. To see that the sets are disjoint, suppose for contradiction that for any distinct $q, \tilde{q} \in \mathbb{Q} \cap [0, 1]$, $E_q \cap E_{\tilde{q}} \neq \emptyset$. If $x \in E_q \cap E_{\tilde{q}}$, then $x - q \in E$ and $x - \tilde{q} \in E$. But they clearly belong to the same equivalence class and this is a contradiction given our construction of E. To see that the union of these sets is $[0, 1]$, consider an arbitrary $y \in [0, 1]$ and observe that since our equivalence relation $\sim$ partitions $[0, 1]$, $y \in [x]$ for some $x \in E$. Then $q^* = y - x \in \mathbb{Q}$ and so $y = x + q^* \mod 1 \in E_{q^*}$. Thus we have that

$$[0, 1] \subseteq \bigcup_{q\in\mathbb{Q}} E_q.$$

Since the reverse inclusion follows by the definition of E_q, we have that $\{E_q\}_{q\in\mathbb{Q}}$ is a partition of $[0, 1]$.

Proposition 1.1.2. *There exists no function $\lambda : 2^{\mathbb{R}} \longrightarrow [0, \infty]$ that satisfies properties (1)-(4) described above*

PROOF. Suppose, for contradiction, that such a function λ exists. We can define the collection of sets $\{E\}_{q\in\mathbb{Q}}$ as in Example 1.1.1 and observe that

$$\begin{aligned} 1 = \lambda([0,1]) &= \lambda\left[\bigcup_{q\in\mathbb{Q}} E_q\right] \\ &= \sum_{q\in\mathbb{Q}} \lambda[E_q] \\ &= \sum_{q\in\mathbb{Q}} c \end{aligned}$$

where the first equality follows from property (2), the second equality follows from the fact that $\{E\}_{q\in\mathbb{Q}}$ is a partition of $[0,1]$, the third equality is due to property (3). The last equality follows as a consequence of translation invariance (property (4)). Since $c \in [0,1]$

$$\sum_{q\in\mathbb{Q}} c = 0 \text{ or } \infty \neq 1$$

which is a contradiction. Thus no such function λ exists. □

This particular example of a *non-measurable* set is called a *Vitali set.* While we used the interval $[0,1]$ to construct such a set, it turns out that this contruction can be extended to any set of positive length in the Lebesgue sense.

1.2. Constructing measures on $\sigma-$algebras

The key issue with our previous definition of a measure on $\mathbb{R}$ is that one cannot have a set-valued function that both has our four desired properties *and* is defined on all subsets of the real line. As a convention, the canonical construction of a measure retains the desired properties in exchange for restricting the class of subsets on which the measure is defined. These subsets are called *measurable* and the standard construction of the Lebesgue measure leads to the class of measurable subsets on the real line to have a special structure of a $\sigma-$*algebra.* Before we define this structure it might be worthwhile looking at various types of structures a class of sets could have

1.2.1. Structures of sets. In the rest of this chapter, we assume that $(\mathcal{X}, \tau)$ is an abstract topological space.

DEFINITION 1.2.1. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call $\mathcal{F}$ a *ring* if

(i) $\emptyset \in \mathcal{F}$
(ii) $A, B \in \mathcal{F} \Rightarrow A \cup B \in \mathcal{F}$
(iii) $A, B \in \mathcal{F} \Rightarrow A \setminus B \in \mathcal{F}$.

Note that the above definition implies that $A \cap B = A \setminus (A \setminus B) \in \mathcal{F}$.

DEFINITION 1.2.2. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call $\mathcal{F}$ an *algebra* if

(i) $\mathcal{F}$ is a ring
(ii) $\mathcal{X} \in \mathcal{F}$.

For example, if we let $\mathcal{X}$ be an arbitrary infinite set, the collection of all finite subsets of $\mathcal{X}$ forms a ring but not an algebra.

DEFINITION 1.2.3. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call $\mathcal{F}$ a σ-ring if

(i) $\mathcal{F}$ is a ring
(ii) $\mathcal{F}$ is closed under countable unions.

DEFINITION 1.2.4. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call $\mathcal{F}$ a σ-algebra if

(i) $\mathcal{F}$ is an algebra.
(ii) $\mathcal{F}$ is closed under countable unions.

Naturally, the power set $2^{\mathcal{X}}$ is a ring, algebra, σ-ring, and σ-algebra all rolled into one.

Remark. Algebras are sometimes referred to as *fields* in the probability literature.

As we said earlier, the notion of a σ-algebra is important because the standard Lebesgue measurable sets form a σ-algebra of subsets of $\mathbb{R}$. However, the other structures we have defined are also important; as we "extend" the notion of the length of an interval on the real line to more complicated sets, we shall first expand our class of measurable sets to a ring of sets.

1.2.2. Lengths of intervals. The mosts intuitive notion of a measure on $\mathbb{R}$ arises from the length of an interval. Thus, in our construction of the Lebesgue measure, we start with the simplest class of sets which consists of intervals in $\mathbb{R}$. Define $\mathcal{L} = \{(a, b] \mid -\infty < a \leq b < \infty\}$ and let $\lambda_1 : \mathcal{L} \longrightarrow [0, \infty]$ be given by $\lambda_1((a, b]) = b - a$. It turns out that our collection of half-open intervals in $\mathbb{R}$ has the structure of a *semi-ring.*

Definition 1.2.5. Let $\mathcal{F} \subseteq 2^{\mathcal{X}}$. We call $\mathcal{F}$ a *semi-ring* if

(i) $\emptyset \in \mathcal{F}$.
(ii) $A, B \in \mathcal{F} \Rightarrow A \cap B \in \mathcal{F}$.
(iii) $A, B \in \mathcal{F} \Rightarrow \exists \{A_i\}_{i=1}^n \in \mathcal{F}$ such that $A_i \cap A_j = \emptyset$ for $i \neq j$ and

$$A \setminus B = \bigcup_{i=1}^{n} A_i$$

Proposition 1.2.6. *$\mathcal{L}$ is a semi-ring.*

Proof. To see (i), note that $\emptyset = (a, a] \in \mathcal{L}$. For (ii), note that for any intervals $A = (a_1, b_1]$ and $B = (a_2, b_2]$, [1]

$$(a_1, b_1] \cap (a_2, b_2] = (\max(a_1, a_2), \min(b_1, b_2)] \in \mathcal{L}.$$

To see (iii), we have to consider two possible cases. First, if A, B are disjoint then $A \setminus B = A \in \mathcal{L}$. If A, B have a non-trivial intersection, then

$$\begin{aligned} A \setminus B = A \cap B^C &= (a_1, b_1] \cap \{(-\infty, a_2] \cup (b_2, \infty)\} \\ &= (a_1, b_1] \cap (-\infty, a_2] \bigcup (a_1, b_1] \cap (b_2, \infty) \\ &= (a_1, \min(b_1, a_2)] \bigcup (\max(a_1, b_2), b_1] \end{aligned}$$

where the components of the union expressed in the last equality are in $\mathcal{L}$. This completes the proof. $\square$

The fact that $\mathcal{L}$ is a semi-ring is important because there's a relatively straightforward way to "expand" a semi-ring into a ring.

Theorem 1.2.7. *Let $\mathcal{F}$ be a semi-ring and let $\mathcal{B}$ be the set of all finite disjoint unions of sets in $\mathcal{F}$. Then $\mathcal{B}$ is a ring.*

Proof. Property (i) in Definition 1.2.1 is trivially satisfied thus we need to prove properties (ii) and (iii). Let $A, B \in \mathcal{B}$. To prove property (iii), we first establish the weaker claim that $A \cap B \in \mathcal{B}$. Observe that

$$\begin{aligned} A &= \bigcup_{i=1}^{n_A} A_i, \ \ A_i \in \mathcal{F}, A_i \cap A_j = \emptyset \text{ for } i \neq j, \\ B &= \bigcup_{i=1}^{n_B} B_i, \ \ B_i \in \mathcal{F}, B_i \cap B_j = \emptyset \text{ for } i \neq j, \end{aligned}$$

[1] If $\max(a_1, a_2) > \min(b_1, b_2)$, then $(\max(a_1, a_2), \min(b_1, b_2)] = \emptyset \in \mathcal{L}$

by the definition of $\mathcal{B}$. Then

$$\begin{aligned} A \cap B &= \left(\bigcup_{i=1}^{n_A} A_i\right) \bigcap \left(\bigcup_{j=1}^{n_B} B_j\right) \\ &= \bigcup_{i=1}^{n_A} \bigcup_{j=1}^{n_B} (A_i \cap B_j) \end{aligned}$$

where $\forall i, j: \ A_i \cap B_j \in \mathcal{F}$ as $\mathcal{F}$ is a semi-ring. Clearly, $A_i \cap B_j$ is disjoint from $A_{i'} \cap B_{j'}$, thus proving the claim. Next, we establish property (iii) by noting that

$$\begin{aligned} A \setminus B &= \left(\bigcup_{i=1}^{n_A} A_i\right) \setminus B \\ &= \left(\bigcup_{i=1}^{n_A} A_i\right) \bigcap B^C \\ &= \bigcup_{i=1}^{n_A} \left(A_i \cap B^C\right) \\ &= \bigcup_{i=1}^{n_A} \left(A_i \cap \left(\bigcap_{j=1}^{n_B} B_j^C\right)\right) \\ &= \bigcup_{i=1}^{n_A} \bigcap_{j=1}^{n_B} \left(A_i \cap B_j^C\right) \\ &= \bigcup_{i=1}^{n_A} \bigcap_{j=1}^{n_B} A_i \setminus B_j \end{aligned}$$

where the $A_i \setminus B_j \in \mathcal{B}$ since $A_i, B_j \in \mathcal{F}$. By the closure under finite intersections property established earlier, $E_i = \bigcap_{j=1}^{n_B} A_i \setminus B_j \in \mathcal{B}$ for any $1 \leq i \leq n_A$. Thus we can rewrite the chain of equalities above as

$$A \setminus B = \bigcup_{i=1}^{n_A} E_i$$

where $E_i \cap E_{i'} = \emptyset$ because $A_i \cap A_{i'} = \emptyset$. Since the finite disjoint union of elements of $\mathcal{B}$ is also a finite disjoint union of elements of $\mathcal{F}$, our claim follows. Finally, to establish property (ii), observe that

$$A \cup B = (A \setminus B) \cup (A \cap B) \cup (B/A)$$

which is a disjoint union of elements in $\mathcal{B}$ and so is also in $\mathcal{B}$ by the same argument as earlier. □

Corollary 1.2.8. *Let $\mathcal{J}$ be the set of all finite disjoint unions of sets in $\mathcal{L}$. Then $\mathcal{J}$ is a ring.*

Proof. By Proposition 1.2.6, $\mathcal{L}$ is a semi-ring. The claim then follows by an application of Theorem theorem 1.2.7. □

Now we can extend our proto-measure λ_1 to a new proto-measure $\lambda_2 : \mathcal{J} \longrightarrow [0, \infty]$ as follows:

$$\lambda_2(A) := \begin{cases} \lambda_1(A), & A \in \mathcal{L} \\ \sum_{i=1}^{n} \lambda_1(B_i), & A = \bigcup_{i=1}^{n} B_i, \{B_i\}_{i=1}^{n} \text{ are disjoint in } \mathcal{L} \end{cases}$$

Proposition 1.2.9. *λ_2 is fiinitely additive on $\mathcal{J}$. That is, for any finite disjoint collection of sets $\{A_i\}_{i=1}^{n} \in \mathcal{J}$*

$$\lambda_2\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \lambda_2(A_i).$$

PROOF. For clarity, we will prove finite additivity for two sets , since the general case follows by induction. Let $A, B \in \mathcal{J}$ such that $A \cap B = \emptyset$. By definition,

$$A = \bigcup_{i=1}^{n_A} A_i, \{A_i\}_{i=1}^{n_A} \text{ are disjoint in } \mathcal{L}$$
$$B = \bigcup_{i=1}^{n_B} B_i, \{B_i\}_{i=1}^{n_B} \text{ are disjoint in } \mathcal{L}$$

and so we have that

$$\begin{aligned}\lambda_2 (A \cup B) &= \lambda_2 \left(\left(\bigcup_{i=1}^{n_A} A_i \right) \cup \left(\bigcup_{i=1}^{n_B} B_i \right) \right) \\ &= \sum_{i=1}^{n_A} \lambda_2 (A_i) + \sum_{i=1}^{n_B} \lambda_2 (B_i) \\ &= \lambda_2 (A) + \lambda_2 (B)\end{aligned}$$

where the second equality follows from associativity of addition along with the fact that $(\bigcup_{i=1}^{n_A} A_i) \cup (\bigcup_{i=1}^{n_B} B_i)$ is a disjoint union of sets in $\mathcal{L}$. □

1.2.3. Structures generated by a class of sets. A key way to "expand" a particular class of sets into a larger structure is to look at the structure *generated* by the class of sets. This idea can be formalized in the following definition, which serves as particular example of this general concept of generation.

DEFINITION 1.2.10. For any $\mathcal{A} \subseteq 2^{\mathcal{X}}$, we refer to the intersection of all rings that contain $\mathcal{A}$ as the ring *generated* by $\mathcal{A}$. Formally, we write

$$\text{ring}(\mathcal{A}) = \bigcap \left\{ \mathcal{R} \subseteq 2^{\mathcal{X}} \text{ is a ring } \mid \mathcal{A} \subseteq \mathcal{R} \right\}.$$

PROPOSITION 1.2.11. *For any* $\mathcal{A} \subseteq 2^{\mathcal{X}}$, $\text{ring}(\mathcal{A})$ *is a ring.*

PROOF. First note that $\text{ring}(\mathcal{A})$ exists since $2^{\mathcal{X}}$ is a ring and so $\left\{ \mathcal{R} \subseteq 2^{\mathcal{X}} \text{ is a ring } \mid \mathcal{A} \subseteq \mathcal{R} \right\}$ is non-empty. Next observe that $\emptyset \in \text{ring}(\mathcal{A})$ vacuously, so property (i) in Definition 1.2.1 is easily satisfied. For property (ii), let $A, B \in \text{ring}(\mathcal{A})$ and observe that $A, B \in \mathcal{R}$ for every $\mathcal{R} \in \left\{ \mathcal{R} \subseteq 2^{\mathcal{X}} \text{ is a ring } \mid \mathcal{A} \subseteq \mathcal{R} \right\}$. Since $\mathcal{R}$ is a ring, $A \cup B \in \mathcal{R}$ for every $\mathcal{R}$ and thus $A \cup B$ is in the intersection i.e. $\text{ring}(\mathcal{A})$. A similar argument establishes property (iii) and thus we can conclude that $\text{ring}(\mathcal{A})$ is a ring (as it should, given its name). □

In our construction of the Lebesgue measure on $\mathbb{R}$, we discovered that $\mathcal{J}$, which is the set of all disjoint unions of half-open intervals in $\mathbb{R}$, is a ring. It turns out that we can make a stronger statement using the language of generators developed here.

PROPOSITION 1.2.12. $\mathcal{J} = \text{ring}(\mathcal{L})$

PROOF. Let $A \in \mathcal{J}$ be arbitrary. Then we can write

$$A = \bigcup_{i=1}^{n_A} A_i$$

where $A_i \in \mathcal{L}$ are pairwise disjoint. Let $\mathcal{R}$ be an arbitrary ring that contains $\mathcal{L}$ and observe that since rings are closed under finite unions, $A \in \mathcal{R}$. Since $\mathcal{R}$ was arbitrary, A is contained by every ring that contains $\mathcal{L}$ and is thus contained in the intersection of all such rings i.e. $\text{ring}(\mathcal{L})$. This proves that $\mathcal{J} \subseteq \text{ring}(\mathcal{L})$.

To see reverse inclusion, recall that $\mathcal{J}$ is a ring that contains $\mathcal{L}$, and so the intersection of all rings that contain $\mathcal{L}$ is certaintly contained in $\mathcal{J}$. This completes the proof. □

In measure theory, the most important structure on sets is the σ-algebra, and the σ-algebra generated by a class of sets $\mathcal{A}$, defined analagously to Definition 1.2.10 about rings and denoted as $\sigma(\mathcal{A})$, plays in an important role in this theory. Using a similar argument as the one shown earlier, one can conclude that $\sigma(\mathcal{A})$ is indeed a σ-algebra. In analysis and probability theory, mathematicians are interested in σ-algebras generated by a special class of sets.

Definition 1.2.13. The σ-algebra generated by the topology τ on set $\mathcal{X}$ is called the *Borel σ-algebra* on $\mathcal{X}$ and is denoted $\mathscr{B}(\mathcal{X})$.

The Borel σ-algebra is interesting because it turns that it is the σ-algebra generated by $\mathcal{L}$ is indeed $\mathscr{B}(\mathbb{R})$, where $\mathbb{R}$ has the usual topology. To prove this fact, we need a little lemma from an introductory course on analysis and topology.

Lemma 1.2.14. *Any open set in the usual topology of $\mathbb{R}$ can be written as a countable disjoint union of open intervals in $\mathbb{R}$.*

Proof. Let O be an open set in $\mathbb{R}$ and let $x \in O$ be arbitrary. Define $I_x \subseteq O$ to be the largest open interval that contains x (that is, I_x is the union of all open intervals in O that contain x). Note that at least one such interval exists because O is open and so there exists some $\varepsilon > 0$ such that $(x-\varepsilon, x+\varepsilon) \subseteq O$. Now for any distinct $x, y \in O$, I_x and I_y are either disjoint or equal since if they were neither, $I_x \cup I_y \subseteq O$ would be a larger interval that contains both x and y. Let $\mathcal{I}$ denote the collection of all disjoint such intervals (that is, we get $\mathcal{I}$ by discarding all the "redundant" intervals in $\{I_x\}_{x\in O}$). We can do this without invoking the Axiom of Choice since there are only countably many intervals in $\mathcal{I}$: every interval $I \in \mathcal{I}$ contains at least one rational number because the rationals are a countably dense subset of $\mathbb{R}$. Thus, since the intervals are disjoint, $\mathcal{I}$ can have at most countably many intervals. Of course

$$O = \bigcup_{I\in\mathcal{I}} I$$

and so our claim follows. $\square$

Proposition 1.2.15. $\sigma(\mathcal{L}) = \mathscr{B}(\mathbb{R})$

Proof. Let O be an open set in $\mathbb{R}$. Then, by Lemma 1.2.14

$$\begin{aligned} O &= \bigcup_{i=1}^{\infty} (a_i, b_i) \\ &= \bigcup_{i=1}^{\infty} \bigcup_{n=1}^{\infty} \left(a_i, b_i - \frac{1}{n}\right] \end{aligned}$$

which is in $\sigma(\mathcal{L})$ by closure under countable unions (property (ii) in Definition 1.2.4). Therefore the topology of $\mathbb{R}$ is in $\sigma(\mathcal{L})$ which implies that $\mathscr{B}(\mathbb{R}) \subseteq \sigma(\mathcal{L})$. The

To see the reverse inclusion, observe that for any $(a, b] \in \mathcal{L}$, we can write

$$(a, b] = (a, b) \cup \{b\} \in \mathscr{B}(\mathbb{R})$$

since $\{b\}$ is closed in $\mathbb{R}$ and closed sets are the complements of open sets and thus contained in $\mathscr{B}(\mathbb{R})$.[2] Therefore $\mathcal{L} \subseteq \mathscr{B}(\mathbb{R})$ and so $\sigma(\mathcal{L}) \subseteq \mathscr{B}(\mathbb{R})$, completing the proof. $\square$

Now we are ready to prove that our proto-measure λ_2 is actually a countably-additive pre-measure on ring$(\mathcal{L})$. But first, we need a lemma about double sums!

Lemma 1.2.16 (Tonelli for series). *Let $\{x_{ij}\}_{i,j\in\mathbb{N}\times\mathbb{N}}$ be a sequence of non-negative (extended) real numbers. Then*

$$\sum_{i,j\in\mathbb{N}^2} x_{ij} = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x_{ij} = \sum_{j=1}^{\infty}\sum_{i=1}^{\infty} x_{ij}.$$

[2] σ-algebras on $\mathcal{X}$ are closed under complements since they are closed under set-differences and contain $\mathcal{X}$.

PROOF. We will prove the first equality since the second then follows by symmetry. Let $F \subset \mathbb{N}^2$ be arbitrary and finite. Then, there exists some $N \in \mathbb{N}$ such that $F \subseteq \{1, 2 \ldots, N\}^2$ and so, by the non-negativity of x_{ij}

$$\sum_{i,j \in F} x_{ij} \leq \sum_{i,j \in \{1,2\ldots,N\}^2} x_{ij} = \sum_{i=1}^{N} \sum_{j=1}^{N} x_{ij} \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_{ij}.$$

This inequality holds for any finite $F \subset \mathbb{N}^2$ and so it holds for the supremum of all such finite sums. That is to say,

$$\sup_{F \subset \mathbb{N}^2 | F \text{ is finite}} \sum_{i,j \in F} x_{ij} \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_{ij}.$$

But recall that for any $\{a_i\}_{i \in \mathcal{I}} \in [0, \infty]$ where $\mathcal{I}$ is any index set

$$\sum_{i \in \mathcal{I}} a_i := \sup_{I \subset \mathcal{I} | I \text{ is finite}} \sum_{i \in I} a_i,$$

and so we have that

$$\sum_{i,j \in \mathbb{N}^2} x_{ij} \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_{ij}.$$

To derive the other inequality, observe that it is sufficient to prove that

$$\sum_{i,j \in \mathbb{N}^2} x_{ij} \geq \sum_{i=1}^{I} \sum_{j=1}^{\infty} x_{ij}$$

for every $I \in \mathbb{N}$. Fix $I = I_0$ and note that

$$\sum_{i=1}^{I_0} \sum_{j=1}^{\infty} x_{ij} = \sum_{i=1}^{I_0} \lim_{J \to \infty} \sum_{j=1}^{J} x_{ij} = \lim_{J \to \infty} \sum_{i=1}^{I_0} \sum_{j=1}^{J} x_{ij}.$$

Thus to prove $\sum_{i,j \in \mathbb{N}^2} x_{ij} \geq \sum_{i=1}^{I_0} \sum_{j=1}^{\infty} x_{ij}$ we need to prove that

$$\sum_{i,j \in \mathbb{N}^2} x_{ij} \geq \sum_{i=1}^{I_0} \sum_{j=1}^{J} x_{ij}$$

for every $J \in \mathbb{N}$. Fix $J = J_0$ and then observe that

$$\sum_{i=1}^{I_0} \sum_{j=1}^{J_0} x_{ij} = \sum_{i,j \in \{1,2,\ldots,I_0\} \times \{1,2,\ldots,J_0\}} x_{ij} \leq \sum_{i,j \in \mathbb{N}^2} x_{ij}$$

where the inequality follows due to non-negativity of x_{ij}. This concludes the proof. $\square$

REMARK. This lemma is a special case of Tonelli's theorem, a fundamental theorem that allows us to construct measures on Cartesian products of measure spaces from the measures on those spaces themselves. This theorem will be motivated and proved in Chapter 7.

PROPOSITION 1.2.17. *λ_2 is a countably additive pre-measure on* ring$(\mathcal{L})$*, that is to say,*

(i) $\lambda_2(\emptyset) = 0$

(ii) For disjoint $\{A_i\}_{i=1}^{\infty} \in \mathcal{J}$ *such that* $\bigcup_{i=1}^{\infty} A_i \in \mathcal{J}$

$$\lambda_2\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \lambda_2(A_i).$$

PROOF. Property (i) is inherited from λ_1. To see property (ii), let $\{A_i\}_{i=1}^{\infty} \in \mathcal{J}$ be disjoint and write $A := \bigcup_{i=1}^{\infty} A_i$ where $A \in \mathcal{J}$ by assumption. First, note that if $\lambda_2(A_i) = \infty$ for any $i \in \mathbb{N}$, then $\infty = \lambda_2(A_i) \leq \lambda_2(\bigcup_{i=1}^{\infty} A_i) = \infty$ where the inequality is due to the monotonicity[3] of λ_2. Thus, in this case, the claim follows vacuously. So, without loss of generality, we can assume that $\lambda_2(A_i) < \infty$ for every $i \in \mathbb{N}$. First, note that for any $n \in \mathbb{N}, \bigcup_{i=1}^{n} A_i \subseteq A$ and so, by the monotonictity and finite additivity of λ_2, we have that

$$\lambda_2(A) \geq \lambda_2\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} \lambda_2(A_i)$$

for every $n \in \mathbb{N}$. Taking limits, we have countable superadditivity:

$$\lambda_2(A) \geq \sum_{i=1}^{\infty} \lambda_2(A_i).$$

In order to deduce the reverse inequality, first suppose that both A and $\{A_i\}$ are in $\mathcal{L}$. Then, we can write

$$A := (a, b]$$

and

$$A_i = (a_i, b_i]$$

for each $i \in \mathbb{N}$. Pick an arbitrary $0 < \epsilon < b - a$ and observe that

$$[a+\epsilon, b] \subseteq \bigcup_{i=1}^{\infty} \left(a_i, b_i + \frac{\epsilon}{2^i}\right)$$

and so by the Heine-Borel theorem, there exists some finite K such that

$$[a+\epsilon, b] \subseteq \bigcup_{k=1}^{K} \left(a_{i_k}, b_{i_k} + \frac{\epsilon}{2^{i_k}}\right).$$

By the finite additivity established in Proposition 1.2.9 and monotonicity, we have that

$$\underbrace{b-a}_{\lambda_2(A)} - \epsilon \leq \sum_{k=1}^{K} b_{i_k} + \frac{\epsilon}{2^{i_k}} - a_{i_k} \leq \underbrace{\sum_{i=1}^{\infty} (b_i - a_i)}_{\sum_{i\in\mathbb{N}} \lambda_2(A_i)} + \epsilon$$

and since ϵ can be arbitrary small the claim follows.

Deducing the general case from the special one outlined above is straightforward. If $A, \{A_i\} \in \mathcal{J}$ then

$$A = \bigcup_{j=1}^{J} B_j$$

where $\{B_j\} \in \mathcal{L}$ are pairwise disjoint. Similarly,

$$A_i = \bigcup_{k=1}^{n_i} C_{ik}$$

where $\{C_{ij}\}_{i\in\mathbb{N}, j\in\mathbb{N}} \in \mathcal{L}$ are pairwise disjoint and $n_i \in \mathbb{N}$. Note that then

$$\lambda_2(A) = \sum_{j=1}^{J} \lambda_2(B_j) \leq \sum_{i=1}^{\infty} \sum_{k=1}^{n_i} \lambda_2(C_{ik}) = \sum_{i=1}^{\infty} \lambda_2(A_i)$$

[3]For any $A, B \in \text{ring}(\mathcal{L})$ such that $A \subseteq B, \lambda_2(B) = \lambda_2(A) + \lambda_2(B \setminus A) \geq \lambda_2(A)$

where the first equality follow from the finite additivity of λ_2 on $\mathcal{J}$, the inequality by the fact that for any $j \in \{1, 2, \ldots, J\}$, there exists a partition of the collection $\{C_{ik}\}$ into subcollections $\left\{C^j_{ik}\right\}_{1 \le j \le J}$ such that

$$B_j = \bigcup_{i,k} C^j_{ik}$$

and so the special case of our result on $\mathcal{L}$ applies (along with an application of Lemma 1.2.16). The final equality again follows by finite additivity. ☐

1.2.4. Outer measures.

Definition 1.2.18. A set valued function

$$\mu^* : 2^{\mathcal{X}} \longrightarrow [0, \infty]$$

is called an outer measure on $\mathcal{X}$ if

(i) $\mu^*(\emptyset) = 0$
(ii) $A \subseteq B \in 2^{\mathcal{X}} \Longrightarrow \mu^*(A) \le \mu^*(B)$
(iii) For $\{A_i\}_{i=1}^{\infty} \in 2^{\mathcal{X}}$

$$\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mu^*(A_i).$$

Example 1.2.19. Given a non-negative extended-real valued function μ on a collection $\mathcal{A} \subseteq 2^{\mathcal{X}}$ such that $\mu(\emptyset) = 0$, define for any $E \subseteq \mathcal{X}$

$$\mu^*(E) := \inf\left\{\sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, E \subseteq \bigcup_{i=1}^{\infty} A_i\right\}$$

Note that this function is defined on $2^{\mathcal{X}}$ since every bounded below subset of the (extended) real numbers has an infimum. Now we prove that the set-function descibed above is indeed an outer measure.

Proposition 1.2.20. *The function $\mu^* : 2^{\mathcal{X}} \longrightarrow [0, \infty]$ defined in Example 1.2.19 is an outer measure*

Proof. For (i), observe that $\emptyset \in \mathcal{A}$ and so $\mu^*(\emptyset) = \mu(\emptyset) = 0$. Next, let $A \subseteq B \subseteq \mathcal{X}$ and observe that

$$\left\{\sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, B \subseteq \bigcup_{i=1}^{\infty} A_i\right\} \subseteq \left\{\sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i\right\}$$

and so

$$\mu^*(B) = \inf\left\{\sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, B \subseteq \bigcup_{i=1}^{\infty} A_i\right\} \ge \inf\left\{\sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i\right\} = \mu^*(A)$$

which gives us (ii). For (iii), let $\{E_i\}_{i=1}^{\infty} \in 2^{\mathcal{X}}$ and assume that $\sum_{i=1}^{\infty} \mu^*(E_i) < \infty$ since otherwise the claim is trivial. Fix $\epsilon > 0$ and choose $A_{ij} \in \mathcal{A}$ such that $E_i \subseteq \bigcup_{j=1}^{\infty} A_{ij}$ and

$$\mu^*(E_i) \le \sum_{j=1}^{\infty} \mu(A_{ij}) < \mu^*(E_i) + \frac{\epsilon}{2^i}$$

for every $i \in \mathbb{N}$[4]. Observe that

$$E := \bigcup_{i=1}^{\infty} E_i \subseteq \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} A_{ij}$$

and so

$$\mu^*(E) \leq \sum_{i,j \in \mathbb{N}^2} \mu(A_{ij}) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \mu(A_{ij}) \leq \sum_{i=1}^{\infty} \mu^*(E_i) + \epsilon$$

where the equality follows by Lemma 1.2.16 and the second inequality is due to properties of the geometric series. Since ϵ was arbitrary, the claim folllows. $\square$

Remark 1.2.21. The outer measure described above is called the *canonical* outer-measure as it as by far the most useful type of outer measure in measure theory. Given a space $\mathcal{X}$, a collection of subsets $\mathcal{A} \subseteq 2^{\mathcal{X}}$, and a countably additive pre-measure μ on $\mathcal{A}$, we can call

$$\mu^*(E) := \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, E \subseteq \bigcup_{i=1}^{\infty} A_i \right\}$$

the canonical outer measure generated by $(\mu, \mathcal{A})$.

Proposition 1.2.22. *Let $\mathcal{A}$, μ, and μ^* be defined as in Example 1.2.19. Then, for any $A \in \mathcal{A}$*

$$\mu^*(A) = \mu(A).$$

Proof. First, observe that A is a cover for itself and that $\emptyset \in \mathcal{A}$ and so

$$\mu^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i \right\} \leq \sum_{i=1}^{\infty} \mu(A_i)$$

where $A_1 = A$ and $A_i = \emptyset$ for $i \neq 1$. Therefore,

$$\mu^*(A) \leq \mu(A).$$

To see the reverse inequality, let $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{A}$ be an arbitrary cover of A. Define,

$$B_i := A \cap \left(A_i \setminus \bigcup_{j=1}^{i-1} A_j \right)$$

and notice that the $\{B_i\}$ is a pairwise disjoint collections whose union is A such that $B_i \subseteq A_i$ for every $i \in \mathbb{N}$. By countable additivity and monotonicity,

$$\mu(A) = \sum_{i=1}^{\infty} \mu(B_i) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

Since $\{A_i\} \subseteq \mathcal{A}$ is an arbitrary cover of A , we have that

$$\mu(A) \leq \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i) \mid A_i \in \mathcal{A}, A \subseteq \bigcup_{i=1}^{\infty} A_i \right\} = \mu^*(A)$$

which completes the proof. $\square$

Now we are (finally!!) ready to extend our pre-measure to a bona-fide measure on a σ-algebra, using the following theorem.

[4]This is possible due to the assumption that $\mu^*(E_i) < \infty$, which implies that the set $\left\{ \sum_{j=1}^{\infty} \mu(A_{ij}) \mid A_{ij} \in \mathcal{A}, E_i \subseteq \bigcup_{j=1}^{\infty} A_{ij} \right\}$ is non-empty. The definition of an infimum then implies that such a cover $\{A_{ij}\}$ exists.

THEOREM 1.2.23 (Caratheodory's Extension Theorem). *Let $\mathcal{X}$ be a set. Given a countably-additive pre-measure μ on ring $\mathcal{A} \subseteq 2^{\mathcal{X}}$ with canonical outer measure μ^* generated by $(\mu, \mathcal{A})$, define the collection*

$$\mathcal{C}(\mu^*) := \left\{A \subseteq \mathcal{X} \text{ such that } \mu^*(E) = \mu^*(A \cap E) + \mu^*\left(A^C \cap E\right) \forall E \in 2^{\mathcal{X}}\right\}.$$

Then

(i) $\mathcal{A} \subseteq \mathcal{C}$.
(ii) $\mathcal{C}(\mu^*)$ is a σ-algebra.
(iii) $\mu^*|_{\mathcal{C}}$ is a countably additive measure on $\mathcal{C}$.

PROOF. First we will show (i). Let $A \in \mathcal{A}$ be arbitrary. By the countable subadditivity of μ^*, we know that

$$\mu^*(E) = \mu^*\left((A \cap E) \bigcup \left(A^C \cap E\right)\right) \leq \mu^*(A \cap E) + \mu^*\left(A^C \cap E\right)$$

for every $E \subseteq \mathcal{X}$. To deduce the reverse inequality, fix E such that $\mu^*(E) < \infty$ because otherwise the claim follows trivially. Pick an $\epsilon > 0$ and find a cover $\{A_i\}_{i=1}^{\infty} \in \mathcal{A}$ of E such that

$$\mu^*(E) \leq \sum_{i=1}^{\infty} \mu(A_i) < \mu^*(E) + \epsilon$$

As in the proof of Proposition 1.2.20, this is possible because $\mu^*(E) < \infty$ and the definition of an infimum. Next, observe that

$$E \cap A \subseteq \bigcup_{i=1}^{\infty}(A_i \cap A),$$
$$E \cap A^C \subseteq \bigcup_{i=1}^{\infty}(A_i \cap A^C)$$

and so

$$\mu^*(E \cap A) \leq \mu^*\left(\bigcup_{i=1}^{\infty}(A_i \cap A)\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i \cap A)$$
$$\mu^*\left(E \cap A^C\right) \leq \mu^*\left(\bigcup_{i=1}^{\infty}(A_i \cap A^C)\right) \leq \sum_{i=1}^{\infty} \mu^*\left(A_i \cap A^C\right)$$

where the first inequality follows due to monotonicity and the second due to subadditivity. Together, these inequalities imply that

$$\begin{aligned}
\mu^*(A \cap E) + \mu^*\left(A^C \cap E\right) &\leq \sum_{i=1}^{\infty} \mu^*(A_i \cap A) + \mu^*\left(A_i \cap A^C\right) \\
&= \sum_{i=1}^{\infty} \mu(A_i \cap A) + \mu\left(A_i \cap A^C\right) \\
&= \sum_{i=1}^{\infty} \mu(A_i) \\
&< \mu^*(E) + \epsilon
\end{aligned}$$

where the first equality is due to the fact that rings are closed under intersections and set-differences along with Proposition 1.2.22 and the second equality is due to the countable additivity of μ. Since ϵ and E are arbitrary, we have that

$$\mu^*(A \cap E) + \mu^*\left(A^C \cap E\right) \leq \mu^*(E)$$

for every $E \subseteq \mathcal{X}$, establishing that $\mathcal{A} \subseteq \mathcal{C}$.

Next we show (ii); that is, we prove $\mathcal{C}$ is a $\sigma-$algebra. Recall Definition 1.2.4 and notice that it is sufficient to prove that (1) $\emptyset, \mathcal{X} \in \mathcal{C}$; (2) if $A \in \mathcal{C}$ then $A^C \in \mathcal{C}$; (3) if $\{A_i\}_{i=1}^{\infty} \in \mathcal{C}$ then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{C}$. Note that $\emptyset, \mathcal{X} \in \mathcal{C}$ because, trivially,

$$\mu^* (E \cap \mathcal{X}) + \mu^* (E \cap \emptyset) = \mu^* (E) .$$

Symmetry between A and A^C in the definition of $\mathcal{C}$ establishes (2). For (3), we first establish closure under finite unions and bootstrap this weaker result to yield the stronger claim. Let $A, B \in \mathcal{C}$ and let $E \subseteq \mathcal{X}$ be arbitrary. Then

$$\begin{aligned}
\mu^* (E) &= \mu^* (E \cap A) + \mu^* \left(E \cap A^C\right) \\
&= \mu^* ((E \cap A) \cap B) + \mu^* \left((E \cap A) \cap B^C\right) + \mu^* \left(E \cap A^C\right) \\
&= \mu^* (E \cap A \cap B) + \mu^* \left(E \cap (A \cap B)^C \cap A\right) + \mu^* \left(E \cap (A \cap B)^C \cap A^C\right) \\
&= \mu^* (E \cap A \cap B) + \mu^* \left(E \cap (A \cap B)^C\right)
\end{aligned}$$

where the second equality is due to the definition of $\mathcal{C}$ and the fact that $B \in \mathcal{C}$, the third equality is due to the identities

$$\begin{aligned}
(A \cap B)^C \cap A &= \left(A^C \cup B^C\right) \cap A = A \cap B^C \\
(A \cap B)^C \cap A^C &= \left(A^C \cup B^C\right) \cap A^C = A^C,
\end{aligned}$$

and the fourth equality follows from the definition of $\mathcal{C}$ and that $A \in \mathcal{C}$. This proves that for any $A, B \in \mathcal{C}$, $A \cap B \in \mathcal{C}$. Property (2) then implies that $A \cup B \in \mathcal{C}$.

To establish closure under countable unions, fix $E \subseteq \mathcal{X}$ and let $\{A_i\}_{i=1}^{\infty} \in \mathcal{C}$ be arbitrary with $B = \bigcup_{i \in \mathbb{N}} A_i$ and define

$$B_n := \bigcup_{i=1}^{n} A_i$$

where $B_n \in \mathcal{C}$ by our result on closure under finite unions. Without loss of generality, we can assume that the $\{A_i\}$ are pairwise disjoint (since we could otherwise replace A_i with $C_i := A_i \setminus \bigcup_{j=1}^{i-1} A_j$ which are disjoint such that $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} C_i$). Then, we have that

$$\begin{aligned}
\mu^* (E) &= \mu^* \left(E \cap B_n^C\right) + \mu^* (E \cap B_n) \\
&= \mu^* \left(E \cap B_n^C\right) + \mu^* (E \cap B_n \cap A_n) + \mu^* \left(E \cap B_n \cap A_n^C\right) \\
&= \mu^* \left(E \cap B_n^C\right) + \mu^* (E \cap A_n) + \mu^* (E \cap B_{n-1})
\end{aligned}$$

where we used the fact that $A_n \in \mathcal{C}$ for the second equality and the disjointness of A_i for the third equality. Observe that the equality $\mu^* (E \cap B_n) = \mu^* (E \cap A_n) + \mu^* (E \cap B_{n-1})$ is a recurrence relation that can be expanded as

$$\mu^* (E \cap B_n) = \sum_{i=1}^{n} \mu^* (E \cap A_i)$$

and so

$$\begin{aligned}
\mu^* (E) &= \mu^* \left(E \cap B_n^C\right) + \sum_{i=1}^{n} \mu^* (E \cap A_i) \\
&\geq \mu^* \left(E \cap B^C\right) + \sum_{i=1}^{n} \mu^* (E \cap A_i)
\end{aligned}$$

for every $n \in \mathbb{N}$ where the inequality is due to the the fact that $B^C \subseteq B_n^C$ and the monotonicity of outer measures. After taking limits, we have that

$$\begin{aligned}\mu^*(E) &\geq \mu^*\left(E \cap B^C\right) + \sum_{i=1}^{\infty} \mu^*(E \cap A_i) \\ &\geq \mu^*\left(E \cap B^C\right) + \mu^*\left(\bigcup_{i \in \mathbb{N}} (E \cap A_i)\right) \\ &= \mu^*\left(E \cap B^C\right) + \mu^*(E \cap B)\end{aligned}$$

where the second inequality follows by countable subadditivity. Another application of countable subadditivity yields

$$\mu^*(E) \leq \mu^*\left(E \cap B^C\right) + \mu^*(E \cap B)$$

and together the two inequalities establish that $B \in \mathcal{C}$, finishing the proof of (ii).

Finally, in order to show that $\mu^*|_{\mathcal{C}}$ is indeed a countably additive measure on $\mathcal{C}$, let $\{A\}_{i=1}^{\infty} \in \mathcal{C}$ be pairwise disjoint, and observe that for $B := \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{C}$ and any $E \subseteq \mathcal{X}$

$$\mu^*(E) \geq \mu^*\left(E \cap B^C\right) + \sum_{i=1}^{\infty} \mu^*(E \cap A_i)$$

due to our previous work. Letting $E = B$, we have

$$\begin{aligned}\mu^*(B) &\geq \sum_{i=1}^{\infty} \mu^*(B \cap A_i) \\ &= \sum_{i=1}^{\infty} \mu^*(A_i).\end{aligned}$$

Since the reverse inequality follows by the subadditivity of the outer measure, our proof is complete. □

Remark 1.2.24. Note that the proof of the facts that $\mathcal{C}(\mu^*)$ is a σ-algebra and $\mu^*|_{\mathcal{C}}$ is countably additive do not depend on the fact $\mathcal{A}$ is a ring; the proofs would hold if $\mathcal{A}$ was any collection of sets and μ^* was any outer measure (as opposed to a *canonical* outer measure).

Remark 1.2.25. Observe that every set $A \subseteq 2^{\mathcal{X}}$ such that $\mu^*(A) = 0$ is in the $\sigma-$algebra $\mathcal{C}(\mu^*)$. To see why, note that $\mu^*(E) \leq \mu^*(A \cap C) + \mu^*\left(A^C \cap E\right)$ for any $E \in 2^{\mathcal{X}}$ by the subadditivity of outer measure. On the other hand, $\mu^*(A \cap E) = 0$ by monotonicity, and so $\mu^*(E) \geq \mu^*\left(A^C \cap E\right) = \mu^*(A \cap E) + \mu^*\left(A^C \cap E\right)$ where the inequality is again monotonicity. This tells us that the measure $\mu^* |_{\mathcal{C}}$ is *complete* in that every subset of a measure zero set is measurable and has measure zero. Complete measures form an important part of measure theory; as we shall see, the Lebesgue measure is complete, and $\mathcal{C}(\lambda^*)$ – called the *Lebesgue $\sigma-$algebra* is the *completion* of the Borel sets $\mathscr{B}(\mathbb{R})$. In other words, you get the Lebesgue sets by adjoining all subsets of measure zero Borel sets to $\mathscr{B}(\mathbb{R})$. This fact is not trivial and shall be proved later in this chapter.

Note that in general such an extension may not be unique and we provide sufficient conditions for uniqueness in Theorem 2.4.7

1.2.5. An extension of the extension theorem. We now have enough machinery to construct the Lebesgue measure on $\mathscr{B}(\mathbb{R})$; in fact, the Caratheodory measurability critierion discussed in the proof of the extension theorem 1.2.23 is strictly larger than the Borel sets, a fact that we hinted at in Remark 1.2.25. To show the Lebesgue measure exists, we observe that λ_2 is a countably-additive pre-measure on $\mathcal{J}$ which is a ring. The canonical outer measure λ^* generated from $(\lambda_2, \mathcal{J})$ then can be restricted to the σ-algebra of measurable sets $\mathcal{C}(\lambda^*)$ as a measure via Caratheodory's extension theorem. That the Borel sets $\mathscr{B}(\mathbb{R}) \subseteq \mathcal{C}(\lambda^*)$ is clear from the fact that $\mathcal{L} \subseteq \mathcal{J} \subseteq \mathcal{C}(\lambda^*)$ and $\sigma(\mathcal{L}) = \mathscr{B}(\mathbb{R})$ (see Proposition1.2.15). The fact that this inclusion is strict is, of course, not obvious; we will return to this point later.

While this strategy to build the Lebesgue meaure works, we can in fact do something more general, which will incidentally also help us establish the properties of the Lebesgue measure. This involves the notion of what is called a *Steiljes* measure, a concept which is particularly useful in probability theory. We will return to this topic at the end of the chapter, after we finish our discussion on general measures on any arbitrary measurable space $(\mathcal{X}, \mathcal{F})$. The main tool wewill use is the following generalization of Caratheodory's extension theorem.

THEOREM 1.2.26 (Extension from semi-rings). *Let $\mathcal{X}$ be a set and $\mathcal{A} \subseteq 2^{\mathcal{X}}$ be a semi-ring. Suppose $\mu : \mathcal{A} \longrightarrow [0, \infty]$ is a set function such that*

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

(ii) For any disjoint $A, B \in \mathcal{A}$ such that $A \cup B \in \mathcal{A}$

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

(iii) For any collection $A_i \in \mathcal{A}$ such that $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$

$$\mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) \leq \sum_{i \in \mathbb{N}} \mu(A_i)$$

then the restriction of the canonical outer measure μ^* generated by $(\mu, \mathcal{A})$ to $\mathcal{C}(\mu^*)$ is a measure.

PROOF. Note that Remark 1.2.24 tells us that $\mathcal{C}(\mu^*)$ is a $\sigma-$algebra and $\mu^*|_{\mathcal{C}}$ is a measure. Thus our two tasks are to show *(i)* that μ^* and μ agree on $\mathcal{A}$ and *(ii)* that $\mathcal{A} \subseteq \mathcal{C}(\mu^*)$. The first result is mostly straightforward; to see that $\mu^*(A) \leq \mu(A)$ for $A \in \mathcal{A}$ we can simply observe that A is a cover for itself. For the reverse inequality, first note that finite additivity and the fact that $\mathcal{A}$ is a semi-ring implies that for any $A, B \in \mathcal{A}$ such that $A \subseteq B, \mu(A) \leq \mu(B)$. Indeed, there exist disjoint $\{C_i\}_{1 \leq i \leq n} \in \mathcal{A}$ where $n \in \mathbb{N}$ such that $B \setminus A = \cup_{1 \leq i \leq n} C_i$ and so $B = A \cup \cup_{1 \leq i \leq n} C_i$ and $\mu(B) = \mu(A) + \sum_{1 \leq i \leq n} \mu(C_i) \geq \mu(A)$. Then for any cover $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{A}$ of A, we have that

$$\mu(A) \leq \sum_{i=1}^{\infty} \mu(A \cap A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

where the first ineqality follows from subadditivity and the fact that $A \cap A_i \in \mathcal{A}$ since semi-rings are closed under intersection along with the fact that $\bigcup_{i \in \mathbb{N}} (A \cap A_i) = A$ and the second inequality follows from monotonicty. Since the cover was arbitrary, we have $\mu^*(A) \geq \mu(A)$.

To prove *(ii),* note that for any $A \in \mathcal{A}$, countable subadditivity of the outer measure implies that

$$\mu^*(E) \leq \mu^*(A \cap E) + \mu^*\left(A^C \cap E\right)$$

for any $E \in 2^{\mathcal{X}}$. To deduce the other inequality, we follow almost exactly the same steps as we did in the proof of Theorem 1.2.23. First, we pick an $E \subseteq 2^{\mathcal{X}}$ such that $\mu^*(E) < \infty$since otherwise the claim follows trivially. Then we use this fact (since only empty subsets of the reals have inifinite infima) to deduce that for any $\epsilon > 0$, there exists a cover $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{A}$of E such that

$$\mu^*(E) \leq \sum_{i=1}^{\infty} \mu(A_i) < \mu^*(E) + \epsilon.$$

Again, we observe that

$$E \cap A \subseteq \bigcup_{i=1}^{\infty} A_i \cap A$$

$$E \cap A^C \subseteq \bigcup_{i=1}^{\infty} A_i \cap A^C$$

Note that $A_i \cap A^C = A_i \setminus A$ and so by the properties of semi-rings there exists, for each i,a disjoint collection of sets $\left\{C_j^i\right\}_{1 \leq j \leq n_i} \in \mathcal{A}$ such that $A_i \cap A^C = \bigcup_{1 \leq j \leq n_i} C_j^i$. Using the monotonicty and countable-subadditivity of outer measures as before, we have that

$$\begin{aligned}
\mu^*(E\cap A)+\mu^*\left(E\cap A^C\right) &\leq \sum_{i=1}^{\infty}\left(\mu^*(A_i\cap A)+\mu^*\left(\bigcup_{1\leq j\leq n_i} C_j^i\right)\right)\\
&\leq \sum_{i=1}^{\infty}\left(\mu^*(A_i\cap A)+\sum_{j=1}^{n_i}\mu^*\left(C_j^i\right)\right)\\
&= \sum_{i=1}^{\infty}\left(\mu(A_i\cap A)+\sum_{j=1}^{n_i}\mu\left(C_j^i\right)\right)\\
&= \sum_{i=1}^{\infty}\left(\mu(A_i)\right)\\
&< \mu^*(E)+\epsilon
\end{aligned}$$

where the first equality follows from part *(i)* and the fact that $A_i\cap A, \left\{C_j^i\right\}_{1\leq j\leq n_i}\in\mathcal{A}$ whereas the second equality follows from finite additivity. Since ϵcan be as small as one wants, our result follows. □

1.3. Abstract measure spaces

Definition 1.3.1. A pair $(\mathcal{X},\mathcal{F})$, where $\mathcal{X}$ is an arbitrary set and $\mathcal{F}$ is a $\sigma-$algebra on $\mathcal{X}$, is called a *measurable space.*

Although we had implicitly defined a measure in the previous section, it's appopriate to write down a formal definition in this section.

Definition 1.3.2. Let $(\mathcal{X},\mathcal{F})$ be a measurable space. A function $\mu:\mathcal{F}\longrightarrow[0,\infty]$ is a *measure* on $\mathcal{X}$ if

(i) $\mu(\emptyset)=0$
(ii) For disjoint $\{A_i\}_{i\in\mathbb{N}}\in\mathcal{F}$

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

The triple $(\mathcal{X},\mathcal{F},\mu)$ is called a *measure space.* If $\mu(\mathcal{X})=1$ then μ is called a *probability measure* and $(\mathcal{X},\mathcal{F},\mu)$ is called a *probability space.*

Definition 1.3.3. Given a measurable space $(\mathcal{X},\mathcal{F})$, any set $A\in\mathcal{F}$ is called a *measurable* set. Conversely, any set $A\subset\mathcal{X}$ such that $A\notin\mathcal{F}$ is referred to as a *non-measurable* set.

While the definition of a measure is simple, it turns out to have some remarkable properties that are useful in the theory of integration and probability that is built on top of measure theory (or, as we shall later see, is equivalent to it).

Proposition 1.3.4. *Let $(\mathcal{X},\mathcal{F})$ be a measurabe space and let*

$$\mu:\mathcal{F}\longrightarrow[0,\infty]$$

be a function. Then μ is a measure if and only if

(i) $\mu(\emptyset)=0$
(ii) For disjoint $A,B\in\mathcal{F}$

$$\mu(A\cup B)=\mu(A)+\mu(B).$$

(iii) For any increasing sequence of sets $A_1\subseteq A_2\ldots$ in $\mathcal{F}$ such that $\bigcup_{i\in\mathbb{N}}A_i=A$

$$\mu(A)=\lim_{i\to\infty}\mu(A_i)$$

PROOF. First we shall establish that Definition 1.3.2 implies properties (i)-(iii) above. Property (i) is inherited straight from the definition; to see (ii), we can let $A_1 = A, A_2 = B$ and $A_j = \emptyset$ for all $j \geq 3$. Then

$$\mu\left(A \cup B\right) = \mu\left(\bigcup_{j \in \mathbb{N}} A_j\right) = \sum_{j=1}^{\infty} \mu\left(A_j\right) = \mu\left(A\right) + \mu\left(B\right)$$

where the second equality is due countably additiivity and the third equality is due to property (i). To see property (iii), let $\{A_i\}_{i\in\mathbb{N}}$ be an increasing sequence of sets such that $A_i \subseteq A_{i+1}$ for every $i \in \mathbb{N}$ and let $A := \bigcup_{i\in N} A_i$. Define

$$B_i := A_i \setminus \bigcup_{j=1}^{i-1} A_j$$

which is the standard "disjointification" of $\{A_i\}_{i\in\mathbb{N}}$ as we have seen earlier. By countable additivity

$$\begin{aligned}
\mu\left(A\right) &= \sum_{i=1}^{\infty} \mu\left(B_i\right) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mu\left(B_i\right) \\
&= \lim_{n\to\infty} \mu\left(\bigcup_{i=1}^{n} B_i\right) \\
&= \lim_{n\to\infty} \mu\left(\bigcup_{i=1}^{n} A_i\right) \\
&= \lim_{n\to\infty} \mu\left(A_n\right)
\end{aligned}$$

where the third equality is due to property (ii). The fourth equality follows from the disjointification and the last equality is due to the increasing nature of the sequence of sets.

Next, we shall establish countable additivity while assuming properties (i)-(iii) in order to complete the equivalence. Let $\{A_i\}_{i\in\mathbb{N}}$ be pairwise disjoint in $\mathcal{F}$. Then, letting $A := \bigcup_{i\in\mathbb{N}} A_i$ we can define

$$B_n := \bigcup_{i=1}^{n} A_i$$

and observe that $\bigcup_{n\in\mathbb{N}} B_n = A$ and $B_n \subseteq B_{n+1}$. Then, by property (iii),

$$\begin{aligned}
\mu\left(A\right) &= \lim_{n\to\infty} \mu\left(B_n\right) \\
&= \lim_{n\to\infty} \sum_{i=1}^{n} \mu\left(A_i\right) \\
&= \sum_{i=1}^{\infty} \mu\left(A_i\right)
\end{aligned}$$

where the second equality is due to finite additivity (property (ii)). This completes the proof. $\square$

REMARK. Property (iii) resembles a continuity condition, and is indeed called *continuity from below* of measures. There is an analagous definition for *continuity from above* which is implied by *continuity from above* for finitely additive measures and pre-measures. If the measures are finite, these two notions of continuity are in fact equivalent.

COROLLARY 1.3.5. *Every measure μ on an arbitrary measurable space $(\mathcal{X}, \mathcal{F})$ is countably subadditive i.e. for any collection $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{F}$*

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

PROOF. We shall first establish *finite* subadditivity and bootstrap this result to countable subadditivity. To see finite subadditivity, let $A, B \in \mathcal{F}$ be arbitrary, and observe that

$$A \cup B = (A \setminus B) \cup B.$$

The two sets on the right hand side are disjoint and so by finite additivity

$$\mu(A \cup B) = \mu(A \setminus B) + \mu(B).$$

Adding $\mu(A \cap B)$ and applying finite additivity again, we deduce that

$$\mu(A \cup B) + \mu(A \cap B) = \mu(A) + \mu(B)$$

which establishes finite subadditivity. To prove the countable analogue, let

$$B_n := \bigcup_{i=1}^{n} A_i$$

and observe that by finite subadditivity

$$\mu(B_n) \leq \sum_{i=1}^{n} \mu(A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$$

where the last inequality follows by the non-negativity of μ. Note that since B_n is an increasing sequence, we can apply Proposition 1.3.4 (iii) to infer that

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{n\to\infty} \mu(B_n) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

□

PROPOSITION 1.3.6. *For a finitely additive measure $\mu : \mathcal{F} \longrightarrow [0, \infty)$, the following statements are equivalent:*

(i) For any increasing sequence of sets $\{A_i\}_{i\in\mathbb{N}}$ such that $A_i \subseteq A_{i+1}$ for all $i \in \mathbb{N}$

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

(ii) For any decreasing sequence of sets $\{A_i\}_{i\in\mathbb{N}}$ such that $A_{i+1} \subseteq A_i$ for all $i \in \mathbb{N}$

$$\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mu(A_i).$$

PROOF. Assuming (i), let $\{A_i\}_{i\in\mathbb{N}}$ be a decreasing sequence of sets and let $A := \bigcap_{i\in\mathbb{N}} A_i$. Then define $B_i = A_1 \setminus A_i$ which is an increasing sequence of sets such that $A_1 \setminus A = \bigcup_{i\in\mathbb{N}} B_i$. By (i),

$$\mu(A_1) - \mu(A) = \mu(A_1 \setminus A) = \lim_{i\to\infty} \mu(B_i) = \mu(A_1) - \lim_{i\to\infty} \mu(A_i)$$

where the first and last equality are due to finite additivity, the finiteness of μ. We can subtract $\mu(A_1)$ from both sides to yield the result.

To establish the converse, assume (ii) and let $\{A_i\}_{i\in\mathbb{N}}$ be an increasing sequence of sets and define $A := \bigcup_{i\in\mathbb{N}} A_i$. Let $B_i := A \setminus A_i$ which is a decreasing sequence of sets such that $\bigcap_{i\in\mathbb{N}} B_i = \emptyset$. By (ii), we have that

$$0 = \mu(\emptyset) = \lim_{i\to\infty} \mu(B_i) = \lim_{i\to\infty} \mu(A \setminus A_i) = \mu(A) - \lim_{i\to\infty} \mu(A_i)$$

where the last equality is again due to finite additivity and the finitenesss of μ. Rearrangement yields the proof. □

Observe how the two results apply without modification to pre-measures as well and so we can establish the countable additivity of λ_2 (see the previous section) using a continuity argument instead of the Heine-Borel argument we previously used (Exercise!).

Proposition 1.3.7. *Let $(\mathcal{X},\mathcal{F})$ be a measurable space and let $\{\mu_i\}_{i\in\mathcal{I}}$ be a collection of measures on $\mathcal{F}$ where $\mathcal{I}$ is at most countable. Then*

$$\mu := \sum_{i\in\mathcal{I}} \mu_i$$

is a measure on $\mathcal{F}$.

Proof. First observe that

$$\mu(\emptyset) = \sum_{i\in\mathcal{I}} \mu_i(\emptyset) = 0.$$

Next, let $\{A_j\}_{j\in\mathbb{N}} \in \mathcal{F}$ be disjoint. Then

$$\begin{aligned}\mu\left(\bigcup_{j\in\mathbb{N}} A_j\right) &= \sum_{i\in\mathcal{I}} \mu_i\left(\bigcup_{j\in\mathbb{N}} A_j\right)\\ &= \sum_{i\in\mathcal{I}}\sum_{j\in\mathbb{N}} \mu_i(A_j)\\ &= \sum_{j\in\mathbb{N}}\sum_{i\in\mathcal{I}} \mu_i(A_j)\\ &= \sum_{j\in\mathbb{N}} \mu(A_j)\end{aligned}$$

where the second equality follows from the countable additivity of μ_i and the third equality follows from the non-negativity of measures and Lemma 1.2.16. This completes the proof. □

1.3.1. $\sigma-$finite measure spaces.

Definition 1.3.8. Let $(\mathcal{X},\mathcal{F},\mu)$ be a measure space. The measure μ is said to be $\sigma-$*finite* if there exists some increasing sequence of sets $\{E_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ such that $\mu(E_i) < \infty$ and

$$\bigcup_{i\in\mathbb{N}} E_i = \mathcal{X}.$$

Proposition 1.3.9. *Let $(\mathcal{X},\mathcal{F},\mu)$ be a measure space. The measure μ being $\sigma-$finite is equivalent to any of the following conditions*

(i) *There exists some* ***pairwise disjoint*** *countable collection of sets $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ such that $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$ and*

$$\bigcup_{i\in\mathbb{N}} A_i = \mathcal{X}$$

(ii) *There exists some countable collection of sets $\{B_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ such that $\mu(B_i) < \infty$ for all $i \in \mathbb{N}$ and*

$$\bigcup_{i\in\mathbb{N}} B_i = \mathcal{X}$$

Proof. First assume that the measure $\mu$is $\sigma-$finite and so there exists some increasing sequence $\{E_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ such that $E_i \subseteq E_{I+1}$, $\mu(E_i) < \infty$ and

$$\bigcup_{i\in\mathbb{N}} E_i = \mathcal{X}.$$

Recall the disjointification

$$A_i := E_i \setminus \bigcup_{j=1}^{i-1} E_j$$

and notice that

$$\begin{aligned}\mu(A_i) &= \mu(E_i) - \mu\left(\bigcup_{j=1}^{i-1} E_j\right)\\ &= \mu(E_i) - \mu(E_{i-1})\\ &< \infty\end{aligned}$$

where the first equality follows from the fact that $\mu(E_i) < \infty$ and $\bigcup_{j=1}^{i-1} E_j = E_{i-1} \subseteq E_i$ along with (finite) additivity. Further,

$$\bigcup_{i\in\mathbb{N}} A_i = \mathcal{X}$$

and so Definition 1.3.8 implies (i).

Next notice that (i) trivially implies (ii) and so all we just need to verify (ii) $\implies$ Definition 1.3.8. To this end, observe that if $\{B_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ is an arbitrary collection that satisfies (ii), then

$$E_n := \bigcup_{i=1}^{n} B_i$$

is an increasing sequence of sets $E_n \subseteq E_{n+1}$ such that

$$\mu(E_n) \leq \sum_{i=1}^{n} \mu(B_i) < \infty$$

and

$$\bigcup_{n\in\mathbb{N}} E_n = \mathcal{X}.$$

□

Proposition 1.3.10. *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let $\{\mu_i\}_{i=1}^{N}$be a finite collection of $\sigma-$finite measure on $\mathcal{F}$. Then the total measure*

$$\mu : \mathcal{F} \longrightarrow \mathbb{R}$$

given by

$$\mu(A) := \sum_{i=1}^{N} \mu_i(A)$$

is also $\sigma-$finite.

Proof. A weaker variant of Proposition 1.3.7 shows that $\mu$is at least a measure on $\mathcal{F}$. We show $\sigma-$finiteness for $N = 2$; the general case follows by induction. Note that if μ_1 and μ_2 are both $\sigma-$finite then by Proposition 1.3.9 there exist $\{E_{1,i}\}_{i\in\mathbb{N}}, \{E_{2,i}\}_{i\in\mathbb{N}} \in \mathcal{F}$ such that $\mu_1(E_{i,1}) < \infty$ and $\mu_2(E_{2,i}) < \infty$ for all $i \in \mathbb{N}$. Further,

$$\bigcup_{i\in\mathbb{N}} E_{1,i} = \bigcup_{i\in\mathbb{N}} E_{2,i} = \mathcal{X}.$$

Then, define

$$C_{i,j} := E_{1,i} \bigcap E_{2,j}$$

and observe that $C_{i,j} \in \mathcal{F}$ and that

$$\begin{aligned}\mu\left(C_{i,j}\right) &= \mu_1\left(E_{1,i} \cap E_{2,j}\right) + \mu_2\left(E_{1,i} \cap E_{2,j}\right) \\ &\leq \mu_1\left(E_{1,i}\right) + \mu_2\left(E_{2,j}\right) \\ &< \infty\end{aligned}$$

for all $(i,j) \in \mathbb{N}^2$. Finally,

$$\bigcup_{i\in\mathbb{N}}\bigcup_{j\in\mathbb{N}} C_{i,j} = \mathcal{X}$$

which by Proposition 1.3.9 establishes the result. □

1.4. The Stieljes measure on $\mathbb{R}$

DEFINITION 1.4.1. A function $F : \mathbb{R} \to \mathbb{R}$ is called a Stieljes function if it is non-decreasing on $\mathbb{R}$ and if it is *right-continuous* i.e. for any $c \in \mathbb{R}$

$$\lim_{x\to c^+} f\left(x\right) = f\left(c\right).$$

PROPOSITION 1.4.2. *Let μ be a measure on $\mathscr{B}(\mathbb{R})$ that is finite on any bounded interval $I \subseteq \mathbb{R}$. Define the function $F : \mathbb{R} \to \mathbb{R}$ such that for $b > a$*

$$F\left(x\right) = \mu\left(\left(-\infty, x\right]\right).$$

Then F is a Stieljes function.

PROOF. Note that F is non-decreasing, since for $x \leq y$ $(-\infty, x] \subseteq (-\infty, y]$ and so $F(x) \leq F(y)$ by the monotonicity of measures. Further, let $x_n \to x$ from the right. Then there exists some decreasing subsequence $x_{n_k} \to x$ as well. Now $(-\infty, x_{n_k}] \subseteq \left(-\infty, x_{n_{k+1}}\right]$ and $\bigcap_k (-\infty, x_{n_k}] = (-\infty, x]$ and so by Propositions 1.3.4 and 1.3.6 (noting that μ is always finite for any such intervals)

$$F\left(x_{n_k}\right) \to F\left(x\right).$$

Next, suppose that there is some $\epsilon > 0$ such that for some subsequence x_{m_k}

$$F\left(x_{m_k}\right) > F\left(x\right) + \epsilon.$$

By convergence, there exists some k_0 such that for all $k \geq k_0$

$$F\left(x_{n_k}\right) < F\left(x\right) + \epsilon.$$

Since $x_{m_k} \to x$, there must be some k such that $x_{m_k} \leq x_{n_{k_0}}$ and so, by the monotonicity of F

$$F\left(x_{m_k}\right) \leq F\left(x_{n_{k_0}}\right) < F\left(x\right) + \epsilon,$$

a contradiction. Therefore $F(x_n) \to F(x)$. □

THEOREM 1.4.3. *Let $F : \mathbb{R} \to \mathbb{R}$ be a Stieljes function. Then the set function on $\mu : \mathcal{L} \to [0, \infty)$ given by*

$$\mu\left(\left(a, b\right]\right) := F\left(b\right) - F\left(a\right)$$

extends to a measure on $\mathscr{B}(\mathbb{R})$.

PROOF. Note that $\mu(\emptyset) = \mu((a,a]) = 0$ for any $a \in \mathbb{R}$. Next, let $A, B \in \mathcal{L}$ such that $\mathrm{A} \cap B = \emptyset$ and $A \cup B \in \mathcal{L}$. Then if $A = (a_1, a_2]$ and $B = (b_1, b_2]$, it must be that $b_1 = a_2$ or $b_2 = a_1$. Suppose without loss of generality that it is the former. Then

$$\begin{aligned}\mu\left(A \cup B\right) &= \mu\left(\left(a_1, b_2\right]\right) \\ &= F\left(b_2\right) - F\left(a_1\right) \\ &= F\left(b_2\right) - F\left(b_1\right) + \left(F\left(b_1\right) - F\left(a_1\right)\right) \\ &= F\left(b_2\right) - F\left(b_1\right) + \left(F\left(a_2\right) - F\left(a_1\right)\right) \\ &= \mu\left(\left(b_1, b_2\right]\right) + \mu\left(\left(a_1, a_2\right]\right) \\ &= \mu\left(A\right) + \mu\left(B\right)\end{aligned}$$

where the fourth equality is due to the fact that $b_1 = a_2$.

Next, let $A, B \in \mathcal{L}$ and such that $A \cup B \in \mathcal{L}$. Since $\mathcal{L}$ is a semi ring, $A \setminus B = \bigcup_{i=1}^{m} C_i$ where $C_i \in \mathcal{L}$ are disjoint and $B \setminus A = \bigcup_{i=1}^{n} D_i$ where $D_i \in \mathcal{L}$ are disjoint. Then $A \cup B = \bigcup_{i=1}^{m} C_i \cup B = \bigcup_{i=1}^{n} D_i \cup A$ and so

$$\mu(A \cup B) = \mu(A) + \mu(B) + \frac{1}{2}\left(\sum_{i=1}^{m} \mu(C_i) + \sum_{j=1}^{n} \mu(D_i)\right)$$

and so by non-negativity

$$\mu(A \cup B) \leq \mu(A) + \mu(B)$$

which is finite subadditivity.

Finally, let $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{L}$ be such that $A := \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{L}$. Let $A_i = (a_i, b_i]$, $A = (a, b]$ and notice that for any $0 < \epsilon < b - a$

$$[a + \epsilon, b] \subseteq \bigcup_{i\in\mathbb{N}} (a_i, b_i + \epsilon).$$

By the Heine Borel theorem, there exists a finite subcover $\{(a_{i_k}, b_{i_k} + \epsilon)\}_{k=1}^{n}$ such that

$$(a + \epsilon, b] \subset [a + \epsilon, b] \subseteq \bigcup_{k=1}^{n} (a_{i_k}, b_{i_k} + \epsilon) \subset \bigcup_{k=1}^{n} (a_{i_k}, b_{i_k} + \epsilon].$$

We can choose this subcover such that its union is in $\mathcal{L}$ without loss of generalityprove thisand so, by the finite subadditivity result and the monotonicity of μ[5]

$$F(b) - F(a + \epsilon) \leq \sum_{k=1}^{n} F(b_{i_k} + \epsilon) - F(a_{i_k}).$$

Letting $\epsilon \to 0$ and applying the right continuity of F, we have that

$$\begin{aligned}\mu(A) = F(b) - F(a) &\leq \sum_{k=1}^{n} F(b_{i_k}) - F(a_{i_k}) \\ &\leq \sum_{i=1}^{\infty} F(b_i) - F(a_i) \\ &= \sum_{i=1}^{\infty} \mu(A_i).\end{aligned}$$

Applying Theorem 1.2.26, our function μ extends to a measure on $\mathscr{B}(\mathbb{R})$. □

Thus Proposition 1.4.2 and Theorem 1.4.3 together show that Stieljes functions and measures that are finite on bounded intervals are in a one-to-one correspondence. For any Stieljes function F , we denote the corresponding measure by μ_F. Stieljes measures enjoy a few important properties.

PROPOSITION 1.4.4. *Let $F : \mathbb{R} \to \mathbb{R}$ be a Stieljes function and let $\mu_F : \mathscr{B}(\mathbb{R}) \to [0, \infty]$ be the corresponding measure. Then for any $x \in \mathbb{R}$, $\mu(\{x\}) = 0$ if and only if F is continuous at x.*

PROOF. Observe that $\{x\} \in \mathscr{B}(\mathbb{R})$ since it is closed. Then,

$$\mu_F(\{x\}) = \lim_{n\to\infty} \mu_F\left(\left(x - \frac{1}{n}, x\right]\right) = \lim_{n\to\infty} F(x) - F\left(x - \frac{1}{n}\right)$$

where the first equality is Proposition 1.3.6. The result then follows. □

We need a couple a lemmas to show that the Lebesgue measure exists with all our requisite properties. In the following, $\lambda^*(A) := \inf\left\{\sum_{j=1}^{\infty} \lambda_1(I_j) \mid A \subseteq \bigcup_{j\in\mathbb{N}} I_j, I_j \in \mathcal{L}\right\}$ is the canonical Lebesgue outer measure.

[5]If $A = (a_1, a_2] \subseteq (b_1, b_2] = B$ then clearly $F(b_2) - F(b_1) \geq F(a_2) - F(a_1)$ by the fact that F is non-decreasing.

LEMMA 1.4.5. *Let $A \subseteq \mathbb{R}$ be an arbitrary set. Then the translated set $A+t := \{a+t \mid a \in A\}$ has Lebesgue outer measure*

$$\lambda^*(A+t) = \lambda^*(A)$$

for all $t \in \mathbb{R}$. Moreover, for any set $A \in \mathscr{B}(\mathbb{R})$, $A+t \in \mathscr{B}(\mathbb{R})$.

PROOF. First, observe that $\lambda_1 : \mathcal{L} \to [0,\infty]$ is translation invariant. Then, if $\{I_j\}_{j\in\mathbb{N}} \in \mathcal{L}$ is a cover for $A \subseteq \mathbb{R}$, then $\{I_j + t\}_{j\in\mathbb{N}}$ is a cover for $A+t$ and so

$$\lambda^*(A+t) \leq \sum_{j\in\mathbb{N}} \lambda_1(I_j+t) = \sum_{j\in\mathbb{N}} \lambda_1(I_j).$$

Taking infimums on the right side yield

$$\lambda^*(A+t) \leq \lambda^*(A).$$

To get the other inequality, we can let $A-t$ play the role of A in the above inequality to yield

$$\lambda^*(A) \leq \lambda^*(A-t)$$

for all $t \in \mathbb{R}$. In particular, the inequality holds if we replace t with $-t$ and the result follows.

Next, for some fixed $t \in \mathbb{R}$ let $\mathcal{B}_t := \{B \in \mathscr{B}(\mathbb{R}) \mid B+t \in \mathscr{B}(\mathbb{R})\} \subseteq \mathscr{B}(\mathbb{R})$. Clearly, $\mathbb{R} \in \mathcal{B}_t$ since $\mathbb{R}+t = \mathbb{R}$. Moreover, let $f : \mathbb{R} \to \mathbb{R}$ be given by $f(x) = x+t$. Then, note that $\mathcal{B}_t = \left\{f^{-1}[B] \in \mathscr{B}(\mathbb{R}) \mid B \in \mathscr{B}(\mathbb{R})\right\}$.[6] Therefore for any $A \in \mathcal{B}_t$ where $A = f^{-1}[B]$ for some $B \in \mathscr{B}(\mathbb{R})$, we have $A^C = \left(f^{-1}[B]\right)^C = f^{-1}\left[B^C\right] \in \mathscr{B}(\mathbb{R})$.Note that $B^C \in \mathscr{B}(\mathbb{R})$ as it is a $\sigma-$algebra and so $A^C \in \mathcal{B}_t$. Next, let $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{B}_t$. Then there exist $B_i \in \mathscr{B}(\mathbb{R})$ such that $A_i = f^{-1}[B_i]$ and $\bigcup_{i\in\mathbb{N}} A_i = \bigcup_{i\in\mathbb{N}} f^{-1}[B_i] = f^{-1}\left[\bigcup_{i\in\mathbb{N}} B_i\right]$ where $\bigcup_{i\in\mathbb{N}} B_i \in \mathscr{B}(\mathbb{R})$. Therefore $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{B}_t$. Therefore $\mathcal{B}_t$ is a $\sigma-$algebra and moreover, since every interval in $\mathcal{L}$ is a translation of another interval in $\mathcal{L}$ so that $\mathcal{L}+t = \mathcal{L}$, we have that $\mathcal{B}_t$ contains $\mathcal{L}$. Therefore, $\mathscr{B}(\mathbb{R}) = \sigma(\mathcal{L}) \subseteq \mathcal{B}_t$. We have shown that $\mathscr{B}(\mathbb{R}) = \mathcal{B}_t$ for any $t \in \mathbb{R}$ and our claim follows. $\square$

REMARK. The type of argument that we used to prove that $\mathscr{B}(\mathbb{R}) = \mathcal{B}_t$ is a type of *generating class* argument. The abstract version of how such arguments go is described later in section 2.4.

A similar result holds for dilations. That is, the Lebesgue measure is scales under dilations.

LEMMA 1.4.6. *Let $A \subseteq \mathbb{R}$ be an arbitrary set. Then for any $\delta \neq 0$, the dilated set $\delta A := \{\delta a \mid a \in A\}$ has Lebesgue outer measure*

$$\lambda^*(\delta A) = |\delta|\lambda^*(A).$$

Moreover, for any set $A \in \mathscr{B}(\mathbb{R})$, $\delta A \in \mathscr{B}(\mathbb{R})$.

PROOF. Let $\delta \neq 0$ be fixed. First, notice that $\lambda_1 : \mathcal{L} \to \mathbb{R}$ scales under dilations in that for any $I \in \mathcal{L}$, $\lambda_1(\delta I) = |\delta|\lambda(I)$. Let $\{I_j\}_{j\in\mathbb{N}}$ be a cover of A. Then clearly, $\{\delta I_j\}_{j\in\mathbb{N}}$ is a cover of δA and so

$$\lambda^*(\delta A) \leq \sum_{j\in\mathbb{N}} \lambda_1(\delta I_j) = |\delta| \sum_{j\in\mathbb{N}} \lambda_1(I_j).$$

Taking infimums,

$$\lambda^*(\delta A) \leq |\delta|\lambda^*(A).$$

Conversely, if $\{\delta I_j\}_{i\in\mathbb{N}}$ is a cover of δA then $\frac{1}{\delta}\{\delta I_j\}_{j\in\mathbb{N}}$ is a cover for A and so

$$\lambda^*(A) \leq \sum_{j\in\mathbb{N}} \lambda_1\left(\frac{1}{\delta}\delta I_j\right) = |\frac{1}{\delta}| \sum_{j\in\mathbb{N}} \lambda_1(\delta I_j).$$

Taking infimums again yields

$$\lambda^*(A) \leq |\frac{1}{\delta}|\lambda^*(\delta A)$$

which completes the proof.

[6]This works because $f^{-1}[f[A]] = f\left[f^{-1}[A]\right] = A$ for any $A \in \mathscr{B}(\mathbb{R})$.

Next, we apply the generating class argument as above, and let $\mathcal{B}_\delta = \{B \in \mathscr{B}(\mathbb{R}) \mid \delta B \in \mathscr{B}(\mathbb{R})\}$. Letting $f(x) = \delta x$ (which is an invertible map since $\delta neq 0$), we can apply the same arguments as earlier to show that $\mathcal{B}_\delta$ is a $\sigma-$algebra and then observe that for any $I \in \mathcal{L} : \delta I \in \mathcal{L}$ so $\mathcal{L} \subseteq \mathcal{B}_t$ and so $\sigma(\mathcal{L}) = \mathscr{B}(\mathbb{R}) \subseteq \mathcal{B}_t$ which completes the proof. ☐

THEOREM 1.4.7 (Existence of the Lebesgue measure). *There exists a complete $\sigma-$algebra $\mathcal{F}$ which contains all the open sets in $\mathbb{R}$ and a $\sigma-$finite measure $\lambda : \mathcal{F} \longrightarrow [0,\infty]$ such that*

(i) For any set $A \in \mathscr{B}(\mathbb{R})$ and any $t \in \mathbb{R}$ we have that

$$\lambda(A+t) = \lambda(A).$$

(ii) For any set $A \in \mathscr{B}(\mathbb{R})$ and any $\delta > 0$

$$\lambda(\delta A) = \delta\lambda(A).$$

(iii) $\lambda((a,b]) = b - a$ for any $a \leq b \in \mathbb{R}$[7]

PROOF. Let $F(x) = x$ and notice that F increasing and continuous and so by Theorem 1.4.3, the function $\lambda : \mathcal{L} \to [0,\infty]$ given by

$$\lambda((a,b]) = b - a$$

extends to a measure on $\mathscr{B}(\mathbb{R})$. To see the $\sigma-$finiteness of λ, note that we can write $\mathbb{R} = \bigcup_{n\in\mathbb{N}}(n, n+1] \cup (-n-1, -n]$ where each component of the union has finite measure. As our discussion in Theorem 1.2.26 shows, the extension is the restriction of the outer measure λ^* to the Caratheodory measurable sets $\mathcal{C}(\lambda^*)$ generated by $(\mathcal{L}, \lambda^*)$, which contain the borel sets $\mathscr{B}(\mathbb{R})$. Let $\mathcal{F} = \mathcal{C}(\lambda^*)$ and notice that our discussion in Remark 1.2.25 tells us that $\mathcal{F}$ is complete. Property (i) follows by Lemma 1.4.5 and property (ii) by Lemma 1.4.6. To see property (iii), we need to look at each case separately. For any compact interval $[a,b] = \{a\} \cup (a,b]$, the result follows by additivity and Proposition 1.4.4. For an open set (a,b) note that $(a,b] = (a,b) \cup \{b\}$ and so again additivity and Proposition 1.4.4applies. Finally, we show the same result for $[a,b) = \{a\} \cup (a,b)$. ☐

1.4.1. The Lebesgue $\sigma-$algebra. Note that we had claimed earlier that $\mathscr{B}(\mathbb{R}) \subsetneq \mathcal{C}(\lambda^*)$. This actually takes a considerable amount of work to prove; we will begin to lay the groundwork for the proof here and give the actual proof in the next chapter. First, we begin with an equivalent definition of the $\sigma-$algebra $\mathcal{C}(\lambda^*)$, which we call the Lebesgue $\sigma-$algebra. This helps us characterize the Lebesgue $\sigma-$algebra as the *completion* of the Borel $\sigma-$algebra.

PROPOSITION 1.4.8. *Let $A \in \mathcal{C}(\lambda^*)$ if and only if for every $\epsilon > 0$ there exists some open set O such that $A \subseteq O$ and*

$$\lambda(O \setminus A) < \epsilon$$

PROOF. Let $A \in \mathcal{C}(\lambda^*)$. First, assume that $\lambda(A) < \infty$. By definition of the Lebesgue outer measure, for every $\epsilon > 0$, there exists some collection of half-open intervals $\{(a_i, b_i]\} \in \mathcal{L}$ such that $A \subseteq \bigcup_{i\in\mathbb{N}}(a_i, b_i]$ and

$$\lambda(A) \leq \sum_{i=1}^{\infty} \lambda((a_i, b_i]) = \sum_{i=1}^{\infty} \lambda((a_i, b_i)) < \lambda(A) + \frac{\epsilon}{2}.$$

where the second equality is property (ii) in Theorem 1.4.7. Next, we consider the open sets $\left(a_i, b_i + \frac{\epsilon}{4^i}\right)$. Clearly, $A \subseteq \bigcup_{i\in\mathbb{N}}\left(a_i, b_i + \frac{\epsilon}{4^i}\right)$. Letting $O = \bigcup_{i\in\mathbb{N}}\left(a_i, b_i + \frac{\epsilon}{4^i}\right)$, we see that O is open since the union of open sets is open. Further, $\lambda(O) = \frac{\epsilon}{2} + \sum_{i=1}^{\infty} b_i - a_i < \lambda(A) + \epsilon$. Therefore, since $O, E \in \mathcal{C}(\lambda^*)$ and $\lambda(O) < \infty$

$$\lambda(O \setminus A) = \lambda(O) - \lambda(A) < \epsilon.$$

[7]The intervals could be open, closed or neither.

Of course, λ is $\sigma-$finite and so for any A with $\lambda(A) = \infty$ we can find a partition $\{A_i\}_{i\in\mathbb{N}}$ of A such that each $\lambda(A_i) < \infty$. For each A_i, there exists some open B_i such that $A_i \subseteq B_i$ and $\lambda(B_i \setminus A_i) < \frac{\epsilon}{2^i}$. Let $B = \bigcup_{i\in\mathbb{N}} B_i$ be open. Then

$$B \setminus A \subseteq \bigcup_{i\in\mathbb{N}} B_i \setminus A_i$$

and so

$$\lambda(B \setminus A) \leq \lambda\left(\bigcup_{i\in\mathbb{N}} (B_i \setminus A_i)\right) \leq \sum_{i=1}^{\infty} \lambda(B_i \setminus A_i) < \epsilon$$

which completes the proof in one direction.

Conversely, now suppose that $A \subseteq \mathbb{R}$ is a set such that for any $\epsilon > 0$ there exists some open set $O \subseteq \mathbb{R}$ such that $A \subseteq O$ and $\lambda(O \setminus A) < \epsilon$. Then, for any

$$\begin{aligned}
\lambda^*(E \cap A) + \lambda^*\left(E \cap A^C\right) &= \lambda^*(E \cap A) + \lambda^*\left(E \cap A^C \cap O\right) + \lambda^*\left(E \cap A^C \cap O^C\right) \\
&\leq \lambda^*(E \cap O) + \lambda^*\left(E \cap A^C \cap O\right) + \lambda^*\left(E \cap O^C\right) \\
&\leq \lambda^*(E \cap O) + \lambda^*\left(E \cap O^C\right) + \epsilon \\
&= \lambda^*(E) + \epsilon
\end{aligned}$$

where the first equality follows by the fact that $O \in \mathcal{C}(\lambda^*)$, the second since $A \subseteq O$, the third by our hypothesis, the fourth again since $O \in \mathcal{C}(\lambda^*)$. Since ϵ was arbitrary, we have that

$$\lambda^*(E) \geq \lambda^*(E \cap A) + \lambda^*\left(E \cap A^C\right)$$

and since the other inequality follows by subadditivity, the result follows. □

Recall from basic topology, that arbitrary unions and finite intersections of open sets are open. Since closed sets are exactly those sets whose complements are open, we have that arbitrary intersections and finite unions of closed sets are closed. Given a topological space $(\mathcal{X}, \tau)$, the collection F_σ are the subsets of $\mathcal{X}$ that are countable unions of closed sets. Similarly, the collection G_δ consists of sets that are countable intersections of open sets. We can use these definitions to formulate more equivalent characterizations of the Lebesgue $\sigma-$algebra.

Proposition 1.4.9. *The following are equivalent for any $A \subseteq \mathbb{R}$*

(i) A is Lebesgue measurable

(ii) There exists a set $G \in G_\delta$ such that $A \subseteq G$ and

$$\lambda(G \setminus A) = 0.$$

(iii) There exists a Borel set B such that $A \subseteq B$ and

$$\lambda(B \setminus A) = 0.$$

(iv) For every $\epsilon > 0$ there exists some closed set $F \subseteq A$ such that

$$\lambda(A \setminus F) < \epsilon$$

(v) There exists a set $F \in F_\sigma$ such that $F \subseteq A$ and

$$\lambda(A \setminus F) = 0$$

(vi) There exists a Borel set B such that $B \subseteq A$ and

$$\lambda(A \setminus B) = 0$$

Proof. $[(i) \implies (ii)]$ First assume (i) and observe that by Proposition 1.4.8, for every $n \in \mathbb{N}$ there exists some open set O_n such that $A \subseteq O_n$ and [8]

$$\lambda(O_n \setminus A) < \frac{1}{n}.$$

[8] Note that $O_n \setminus A = O_n \cap A^C \in \mathcal{C}(\lambda^*)$ since $A, O_n \in \mathcal{C}(\lambda^*)$

Then, notice that $G := \bigcap_{n\in\mathbb{N}} O_n \in G_\delta \subset \mathscr{B}(\mathbb{R})$ and that $A \subseteq G \subseteq O_n$ and so the monotonicity of measures implies that

$$\lambda\left(G \setminus A\right) < \frac{1}{n}$$

for every $n \in \mathbb{N}$. Our result then follows.

$[(ii) \implies (iii)]$ This implication is trivial since G above is a Borel set.

$[(iii) \implies (i)]$ This follows the same argument as in Proposition 1.4.8 to show that $\lambda\left(B \setminus A\right) = 0 \implies \lambda^*\left(E\right) \geq \lambda^*\left(E \cap A\right) + \lambda^*\left(E \cap A^C\right).$

$[(i) \implies (iv)]$ Fix $\epsilon > 0$ and let $A \in \mathcal{C}\left(\lambda^*\right)$. Then $A^C \in \mathcal{C}\left(\lambda^*\right)$ and so by Proposition 1.4.8, there exists some open O such that $A^C \subseteq O$ and

$$\lambda\left(O \setminus A^C\right) < \epsilon.$$

Then notice that $O^C \subseteq A$ is closed and

$$\begin{aligned}
\lambda\left(A \setminus O^C\right) &= \lambda\left(A \cap O\right) \\
&= \lambda\left(O \setminus A^C\right) \\
&< \epsilon
\end{aligned}$$

which completes the argument.

$[(iv) \implies (v)]$ This argument is analagous to $[(i) \implies (ii)]$. Notice that for every $n \in \mathbb{N}$, there exists some closed $F_n \subseteq A$ such that

$$\lambda\left(A \setminus F_n\right) = \lambda\left(A \cap F_n^C\right) < \frac{1}{n}.$$

Let $F := \bigcup_{n\in\mathbb{N}} F_n \subseteq A$ and so

$$\lambda\left(A \setminus F\right) = \lambda\left(A \cap \bigcap_{n\in\mathbb{N}} F_n\right) < \frac{1}{n}$$

for every $n \in \mathbb{N}$ by monotonicity. The result then follows.

$[(v) \implies (vi)]$ This is trivial since F above is a Borel set.

$[(vi) \implies (i)]$ This is analgous to $[(iii) \implies (i)]$. Let $B \subseteq A$ be a Borel set such that $\lambda\left(A \setminus B\right) = 0$. Let $E \subseteq \mathbb{R}$ be arbitrary

$$\begin{aligned}
\lambda^*\left(E \cap A\right) + \lambda^*\left(E \cap A^C\right) &= \lambda^*\left(E \cap A \cap B^C\right) + \lambda^*\left(E \cap A \cap B\right) + \lambda^*\left(E \cap A^C\right) \\
&= \lambda^*\left(E \cap A \cap B\right) + \lambda^*\left(E \cap A^C\right) \\
&\geq \lambda^*\left(E \cap B\right) + \lambda\left(E \cap B^C\right) \\
&= \lambda^*\left(E\right)
\end{aligned}$$

where the first equality is because $B \in \mathcal{C}\left(\lambda^*\right)$, second because $\lambda\left(A \cap B^C\right) = 0$, the inequality because of monotonicity, and the last equality again due to $B \in \mathcal{C}\left(\lambda^*\right)$. This completes the proof. □

Note that Proposition 1.1.2 tells us that the Vitali sets $\{E_q\}_{q\in\mathbb{Q}}$ as described in Example 1.1.1are not measurable since the Lebesgue measure violates countable additivity on such sets. This can be generalized to show that every set of positive Lebesgue measure contains Vitali-like subsets that are not measurable.

Theorem 1.4.10. *For any set $A \in \mathcal{C}\left(\lambda^*\right)$ with $\lambda\left(A\right) > 0$ there exists a subset $B \subset A$ such that $B \notin \mathcal{C}\left(\lambda^*\right)$.*

Proof. First suppose that $\lambda\left(A\right) < \infty$ and notice that $A \subset [-b, b]$ for some $b > 0$ i.e A is bounded. Then $\{bE_q\}_{q\in\mathbb{Q}}$and $\{-bE_q\}_{q\in\mathbb{Q}}$are a disjoint collection such that $\bigcup_{q\in\mathbb{Q}} bE_q \cup -bE_q = [-b, b]$ and so

$\bigcup_{q\in\mathbb{Q}}\underbrace{(bE_q \cup -bE_q)\cap A}_{=:A_q} = A$. Suppose $\{A_q\}_{q\in\mathbb{Q}}$ are measurable, then

$$\begin{aligned}\lambda(A) &= \lambda\left(\bigcup_{q\in\mathbb{Q}} A_q\right) \\ &= \sum_{q\in\mathbb{Q}} \lambda(A_q) \\ &= \sum_{q\in\mathbb{Q}} 2b\lambda(E_q) \\ &= \sum_{q\in\mathbb{Q}} 2bc \\ &= 0 \text{ or } \infty\end{aligned}$$

whre the second equality is countable additivity, the third Lemma 1.4.6 (scaling) and the fourth Lemma 1.4.5(translation invariance), which is a contradiction.

If $\lambda(A)=\infty$ then by $\sigma-$finiteness there exist disjoint $\{B_i\}_{i\in\mathbb{N}} \in \mathcal{C}(\lambda^*)$ such that $\lambda(B_i)<\infty$ and $\bigcup_{i\in\mathbb{N}} B_i = A$. Let B_i^q be akin to A_q above and then our result follows. □

1.4.2. Cantor sets and functions. Our results show that countable sets are measurable and have Lebesgue measure zero. Indeed, every singleton $\{x\}$ is closed and hence measurable and has measure zero since $\{x\} = [x,x]$. Countable additivity then tells us that every countable set has measure zero. In particular, the rational numbers have measure zero. Do there exist uncountable sets that also have measure zero? We can indeed construct such sets. Perhaps more surprisingly, we can use a similar construction to show that there exist sets of *positive* Lebesgue measure, that nevertheless contain no non-trivial intervals within them.

Definition 1.4.11. Let the set $C \subseteq [0,1]$ such that for any $x \in C$

$$x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}$$

where $a_i \in \{0,2\}$. We call C is the standard *Cantor set.*

That is the Cantor set consists of those elements in $[0,1]$ whose ternary (base 3) representation does not contain any 1 in its digits. A hurdle with this definition is that the representation is not unique; as with any decimal or binary representation, we have things like $0.02222222..._3 = 0.1_3$ where the subscripts highlight the fact that we are using ternary representations. A discussion on base-b representations of real numbers can be found in Appendix section D.1.1.2. In this case, we say that an element $x \in [0,1]$ is in the Cantor set C if any one of its representations has coefficients only in $\{0,2\}$. In other words, real numbers like $0.022222..._3 \in C$.

An alternative characterization of the Cantor set is to take the set $C_1 := [0,1]$ and remove the the middle third set $G_1 = \left(\frac{1}{3},\frac{2}{3}\right)$. Then from the remaining $C_2 := [0,1]\setminus G_1 = \left[0,\frac{1}{3}\right] \cup \left[\frac{2}{3},1\right]$ remove the middle third of each interval component; that is, remove $G_2 = \left(\frac{1}{9},\frac{2}{9}\right)\cup\left(\frac{7}{9},\frac{8}{9}\right)$and so on. The recursive description of the sets $\{G_n\}_{n\in\mathbb{N}}$ can be given as follows

$$\begin{aligned}G_1 &= \left(\frac{1}{3},\frac{2}{3}\right) \\ G_n &= \frac{G_{n-1}}{3} \cup \left(\frac{G_{n-1}}{3}+\frac{2}{3}\right).\end{aligned}$$

Symmetrically, we recursively define the sets approximating the Cantor set $\{C_n\}_{n\in\mathbb{N}\cup\{0\}}$

$$\begin{aligned} C_0 &= [0,1] \\ C_n &= \frac{C_{n-1}}{3} \cup \left(\frac{C_{n-1}}{3} + \frac{2}{3}\right) \end{aligned}$$

LEMMA 1.4.12. *$C_{n+1} \subseteq C_n$ for all $n \in \mathbb{N} \cup \{0\}$.*

PROOF. Note that $\left[1, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right] = C_1 \subset C_0 = [0,1]$. Let $f(x) = \frac{x}{3}$ and let $g(x) = \frac{x}{3} + \frac{2}{3}$. Next suppose that $C_{n+1} = f[C_n] \cup g[C_n] \subseteq C_n$. Then $f[C_{n+1}] \subseteq f[C_n]$ and $g[C_{n+1}] \subseteq g[C_n]$. Together, $C_{n+2} = f[C_{n+1}] \cup g[C_{n+1}] \subseteq f[C_n] \cup g[C_n] = C_{n+1}$. This completes the proof. □

PROPOSITION 1.4.13. $C = \bigcap_{n\in\mathbb{N}} C_n$.

PROOF. We will show that for any $x \in [0,1]$ where write $x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}$in its base-3 representation with $a_i \in \{0,1,2\}$, the following claims are equivalent (1) For all $1 \leq i \leq n$ $a_i \neq 1$ (2) $x \in C_n$. We certainly know that $a_1 \neq 1$ if and only if $x \in C_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{1}{3}, 1\right]$. Now assume that that the claims are equivalent for some fixed n and suppose that for $1 \leq i \leq n+1 : a_i \neq 1$. By the induction hypothesis, $x \in C_n$ and so if $a_1 = 0$ then $3x \in [0,1]$ is such that its first n digits after the delimter are 0 or 2 and so $3x \in C_n$ (again by the hypothesis) and so $x \in \frac{C_n}{3}$. If $a_1 = 2$ then $3\left(x - \frac{2}{3}\right) \in C_n$ by the hypothesis and so $x \in \frac{C_n}{3} + \frac{2}{3}$. Therefore, $x \in \frac{C_n}{3} \cup \left(\frac{C_n}{3} + \frac{2}{3}\right) = C_{n+1}$ which completes one implication.

Conversely, suppose that $x \in C_{n+1}$. Since $C_{n+1} \subseteq C_n$, our hypothesis tells us that $a_i \neq 1$ for $1 \leq i \leq n$. Now $x \in \frac{C_n}{3}$ or $x \in \frac{C_n}{3} + \frac{2}{3}$. Clearly, if $a_{n+1} = 1$ then $3x \notin C_n$ and $3\left(x - \frac{2}{3}\right) \notin C_n$ and the proof is complete. □

PROPOSITION 1.4.14. *For every $n \in \mathbb{N}$ $C_n = [0,1] \setminus \bigcup_{i=1}^{n} G_i$.*

PROOF. For $n = 1$ the claim is obvious. Suppose the claim holds for n i.e. $C_n = [0,1] \setminus \bigcup_{i=1}^{n} G_i$ and note that

$$\begin{aligned} C_{n+1} &= \frac{C_n}{3} \cup \left(\frac{C_n}{3} + \frac{2}{3}\right) \\ &= \frac{1}{3}\left([0,1] \setminus \bigcup_{i=1}^{n} G_i\right) \cup \left(\frac{1}{3}\left([0,1] \setminus \bigcup_{i=1}^{n} G_i\right) + \frac{2}{3}\right) \\ &= \left(\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]\right) \setminus \bigcup_{i=1}^{n}\left(\frac{G_i}{3} \cup \left(\frac{G_i}{3} \cup \frac{2}{3}\right)\right) \\ &= C_1 \setminus \bigcup_{i=1}^{n}\left(\frac{G_i}{3} \cup \left(\frac{G_i}{3} \cup \frac{2}{3}\right)\right) \\ &= [0,1] \setminus G_1 \setminus \bigcup_{i=2}^{n+1} G_i \\ &= [0,1] \setminus \bigcup_{i=1}^{n+1} G_i \end{aligned}$$

which completes the proof. □

PROPOSITION 1.4.15. *C is closed under the usual topology of $\mathbb{R}$.*

PROOF. Note that $C_0 = [0,1]$ is closed and $C_n = \frac{C_{n-1}}{3} \cup \left(\frac{C_{n-1}}{3} + \frac{2}{3}\right)$ is closed if C_{n-1} is closed. Therefore, C_n is closed for all $n \in \mathbb{N}$ and so the intersection is closed as well. □

PROPOSITION 1.4.16. *$C \in \mathscr{B}(\mathbb{R})$ and $\lambda(C) = 0$.*

PROOF. Since C is closed its in $\mathscr{B}(\mathbb{R})$. We first prove that $\lambda(C_n) \leq \left(\frac{2}{3}\right)^n$ for any $n \in \mathbb{N}$. Note that for $n = 0$ this is obvious, for $n = 1$ we have that

$$\begin{aligned}\lambda(C_1) &= \lambda\left(\left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]\right) \\ &= \lambda\left(\left[0, \frac{1}{3}\right]\right) + \lambda\left(\left[\frac{2}{3}, 1\right]\right) \\ &= \frac{2}{3}.\end{aligned}$$

Now suppose that $\lambda(C_n) \leq \left(\frac{2}{3}\right)^n$ and observe that

$$\begin{aligned}\lambda(C_{n+1}) &\leq \lambda\left(\frac{C_{n+1}}{3}\right) + \lambda\left(\frac{C_n}{3} + \frac{2}{3}\right) \\ &= \frac{1}{3}\left(\frac{2}{3}\right)^n + \frac{1}{3}\left(\frac{2}{3}\right)^n \\ &= \left(\frac{2}{3}\right)^{n+1}\end{aligned}$$

where the inequality is by subadditivity, and the first equality by the properties (translation invariance and scaling) outlined in Theorem 1.4.7. Then, by Proposition1.3.6

$$\begin{aligned}\lambda(C) &= \lambda\left(\bigcap_{n \in \mathbb{N}} C_n\right) \\ &= \lim_{n\to\infty} \lambda(C_n) \\ &\leq \lim_{n\to\infty} \left(\frac{2}{3}\right)^n \\ &= 0.\end{aligned}$$

□

PROPOSITION 1.4.17. *The Cantor set C is uncountable.*

PROOF. Because the Cantor set can be described as the collection of all ternary strings with only 0s and 2s, we know by Cantor's diagonal argument that such a collection is not countable. □

PROPOSITION 1.4.18. *The Cantor set contains no non-trivial (i.e. non-singleton) intervals.*

PROOF. $I \subset C$ is a non-trivial interval then $\lambda(I) > 0$ which would imply, by monotonicity, that $\lambda(C) > 0$ which contradicts Proposition1.4.16. □

Another way to establish the uncountability of C is to examine the behavior of the following function.

DEFINITION 1.4.19. Let $\psi_C : [0,1] \to [0,1]$ be given as

$$\psi_C(x) = \begin{cases} \sum_{i=1}^{\infty} \frac{a_i}{2^i}, & x = \sum_{i=1}^{\infty} \frac{2a_i}{3^i}, a_i \in \{0,1\} \\ \sup_{y \leq x, y \in C} \psi_C(y) & x \in [0,1] \setminus C \end{cases}.$$

ψ_C is called the *Cantor function.*

Note that the Cantor function f is well defined since every member of the Cantor set has a *unique* base 3 representation where coefficients are only 0 and 2. Here's a practical way to think about the Cantor function. First, for any $x \in C$, $\psi_C(x)$ is computed from the unique base-3 representation (i.e. the one with only 0s and 2s) by turning all the 2s into 1s and then interpreting the resulting string as a binary number. For $x \notin C$, $\psi_C(x)$ is computed by taking the base-3representation of x, truncating after the first 1 and then replacing each 2 before the first 1 with 1s and then interpreting the result as a binary number.

Proposition 1.4.20. $\psi_C[C] = [0,1]$.

Proof. Let $y \in [0,1]$ be arbitrary. Then y has a binary representation $y = \sum_{i=1}^{\infty} \frac{a_i}{2^i}$ where $a_i \in \{0,1\}$. We take the terminating expansion to stave off questions about uniqueness (i.e. the expansion that doesn't end in all 1s). Then,

$$x = \sum_{i=1}^{\infty} \frac{2a_i}{3^i}$$

is the unique element in the Cantor set such that $\psi_C(x) = y$ which completes the proof. □

This yields yet another proof that C is uncountable, since $[0,1]$ is uncountable. Expand

Proposition 1.4.21. *The Cantor function is non-decreasing i.e for any $x, y \in [0,1]$ such that $x \geq y$*

$$\psi_C(x) \geq \psi_C(y).$$

Proof. First suppose both $x, y \in C$ and $x \geq y$. Write the unique expansions $x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}$ and $y = \sum_{i=1}^{\infty} \frac{b_i}{3^i}$ and notice that $a_i \geq b_i$ for all $i \in \mathbb{N}$ and so clearly $\psi_C(x) \geq \psi_C(y)$. The case when both $x, y \notin C$ is trivial by the first part and the definition of the supremum. Now suppose that $x \in C$ and $y \notin C$. If $x \geq y$ then then clearly $\psi_C(x) \geq \psi_C(y)$ by the definition and the first part of the proof. A similar argument holds if $y \geq x$. This completes the proof. □

Proposition 1.4.22. *The Cantor function ψ_C is continuous on $[0,1]$.*

Proof. Suppose that ψ_C is discontinuous at $c \in [0,1]$. Since ψ_C is increasing, we have that the limits $f(c^-)$ and $f(c^+)$ exist and are in $[0,1]$ since it is closed. Moreover, $f(c^-) < f(c^+)$ which implies that there is some real number $y \in [0,1]$such that $f(c^-) < y < f(c^+)$. But then there's no $x \in [0,1]$ such that $f(x) = y$ which is a contradiction. □

Note that that since ψ_C is a non-decreasing and continuous function on $[0,1]$, it can be extended to a Stieljes function in an obvious way, define

$$F_{\psi_C}(x) = \begin{cases} 1, & x > 1 \\ \psi_c(x), & x \in [0,1] \\ 0, & x < 0 \end{cases}$$

The function F_{ψ_C} is still non-decreasing and continuous and so there it induces a Stieljes measure on $\mathscr{B}(\mathbb{R})$ by Theorem 1.4.3. This measure, is in some sense *orthogonal* to the Lebesgue measure, a notion that we will make precise in Chapter 6. This idea is intimately connected to the fact that while ψ_C is differentiable almost everywhere (the meaning of this phrase will be made precise later), it is not the integral of its derivative, a fact which we shall also establish in that chapter.

The Cantor function is symmetric, a fact that is important in establishing moments of the Cantor distribution in probability theory. For now, we use this fact to compute the integral of this function.

Proposition 1.4.23. *For any $x \in [0,1]$, $\psi_C(1-x) = 1 - \psi_C(x)$.*

Proof. First suppose that $x \in C$ in which case $1 - x \in C$. To see this, write $x = \sum_{i=1}^{\infty} \frac{2a_i}{3^i}$ where $a_i \in \{0,1\}$ and then $1 - x = \sum_{i=1}^{\infty} \frac{(2-2a_i)}{3^i}$ where $2 - 2a_i$ is 0 if $a_i = 1$ and 2 if $a_i = 0$. Then, $\psi_C(1-x) = \sum_{i=1}^{\infty} \frac{1-a_i}{2^i} = \sum_{i=1}^{\infty} \frac{1}{2^i} - \sum_{i=1}^{\infty} \frac{a_i}{2^i} = 1 - \psi_C(x)$. Similarly, let $x \in [0,1] \setminus C$ and observe that we can write $x = \sum_{i=1}^{\infty} \frac{a_i}{3^i}$ where $a_i \in \{0,1,2\}$. Moreover, $n = \inf\{i \mid a_i = 1\} < \infty$ and so

$$\begin{aligned}
\psi_C(x) &= \psi_C\left(\sum_{i=1}^{\infty} \frac{a_i}{3^i}\right) \\
&= \psi_C\left(\sum_{i=1}^{n-1} \frac{a_i}{3^i} + \sum_{i=n+1}^{\infty} \frac{2}{3^i}\right) \\
&= \sum_{i=1}^{n-1} \frac{b_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i} \\
&= \sum_{i=1}^{n-1} \frac{b_i}{2^i} + \frac{1}{2^n}
\end{aligned}$$

where $b_i = \frac{a_i}{2} \in \{0,1\}$. Note that the third equality follows by fill. Now similarly, $1 - x \in [0,1] \setminus C$ with $1 - x = \sum_{i=1}^{\infty} \frac{(2-a_i)}{3^i}$ and $2 - a_n = 1$ is the first occurrence of 1 in its ternary expansion. Then

$$\begin{aligned}
\psi_C(1-x) &= \psi_C\left(\sum_{i=1}^{\infty} \frac{(2-a_i)}{3^i}\right) \\
&= \psi_C\left(\sum_{i=1}^{n-1} \frac{(2-a_i)}{3^i} + \sum_{i=n+1}^{\infty} \frac{2}{3^i}\right) \\
&= \sum_{i=1}^{n-1} \frac{1-b_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i} \\
&= \sum_{i=1}^{n-1} \frac{1}{2^i} + 2\sum_{i=n+1}^{\infty} \frac{1}{2^i} - \left(\sum_{i=1}^{n-1} \frac{b_i}{2^i} + \sum_{i=n+1}^{\infty} \frac{1}{2^i}\right) \\
&= \sum_{i=1}^{\infty} \frac{1}{2^i} - \left(\sum_{i=1}^{n-1} \frac{b_i}{2^i} + \frac{1}{2^n}\right) \\
&= 1 - \psi_C(x)
\end{aligned}$$

which completes the proof. □

A corollary of this result is that the integral of the Cantor function over the unit interval is $\frac{1}{2}$.

Corollary 1.4.24. *The Cantor function ψ_C is Riemann integrable on $[0,1]$ with integral*

$$\int_0^1 \psi_C(x)\,dx = \frac{1}{2}.$$

Proof. The function is Riemann integrable since it is continuous (we provide an exact characterization of Riemann integrable functions in Theorem 3.5.12, when we review the Riemann integral). Note that

$$\begin{aligned}
\int_0^1 \psi_C(1-x)\,dx &= -\int_1^0 \psi_C(u)\,du \\
&= \int_0^1 \psi_C(u)\,du
\end{aligned}$$

where we have used the $u-$substitution method from elementary calculus. add reference to later chapters. But since $\psi_C(1-x) = 1 - \psi_C(x)$ and the integral is linear, we have

$$1 = \int_0^1 1dx = 2\int_0^1 \psi_C(x)\,dx$$

and the result follows. □

The Cantor function also has the remarkable property that it is actually constant on each interval component of the complement of the Cantor set in $[0,1]$.This is particularly striking since then it's only strictly increasing inside the Cantor set, which is a set of measure zero, and yet manages to hit every point in $[0,1]$.

PROPOSITION 1.4.25. *The Cantor function is constant on every interval component of* $[0,1]\setminus C$.

PROOF. First note that since $\bigcup_{i=1}^{\infty}G_i=[0,1]\setminus C$ is open since each set G_i is open. Moreover, by Lemma 1.2.14, there are countably many disjoint open intervals $\{O_j\}_{j\in\mathbb{N}}\subseteq[0,1]$such that $\bigcup_{i=1}^{\infty}G_i=\bigcup_{j=1}^{\infty}O_j$. Let $x\in\bigcup G_i$ and notice that then $x\in O_j$ for some j. Define $y_0=\sup\{y\in C\mid y\leq x\}$. Note that $y_0\in C$ since the sets C and $\{y\in[0,1]\mid y\leq x\}$ are both closed. For any $z\in O_j$, $y_0<z$ since otherwise $\min\{z,x\}\leq y_0\leq\max\{z,x\}$ would mean that O_j is not contained in $[0,1]\setminus C$. Next, suppose that there's some $y_1\in C$ such that $y_0<y_1<z$. Clearly $y_1>x$ (since otherwise $y_1\in\{y\in C\mid y\leq x\}$ and so $y_1\leq y_0$) which again means that $z>y_1>x$ and so $O_j\cap C\neq\emptyset$, which is a contradiction. Therefore, it must be that $y_0=\sup\{y\in C\mid y\leq z\}$ and so for any $z\in O_j$, $\psi_C(z)=\psi_C(y_0)$, yielding the result. □

1.4.2.1. *Fat cantor sets.* Folland: Exercise 32, page 40

CHAPTER 2

Measurable functions

2.1. Limits of sets and their indicator functions

Before we embark on a general description of measurable functions, it's useful to look at a special kind of function: the indicator function of a set. These functions are special, because they are essentially the building blocks of all important functions in measure theory. In fact, indicator functions of sets are the key to linking abstract measure theory on one hand, to the theory of integration on the other. As we shall later see, this link is actually an equivalence: measures and integrals are equivalent objects, and so, in the context of this theory, sets and their indicator functions are also in some sense equivalent. While the full scope of this equivalence will only become salient when we discuss integration, this section will shed some light on why we perhaps should expect this *ex-ante*.

DEFINITION 2.1.1. Let $\mathcal{X}$ be a set and let $A \subseteq \mathcal{X}$ be an arbitrary subset. The function

$$\mathbb{1}_A : \mathcal{X} \longrightarrow \{0,1\}$$

defined by

$$\mathbb{1}_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

is called the *indicator function* of set A.

The algebra of sets implies a corresponding Boolean algebra for indicator functions.

FACT 2.1.2. *Let $A, B \subseteq \mathcal{X}$ be arbitrary and let $\mathbb{1}_A, \mathbb{1}_B$ be their respective indicator functions. Then the indicator function of the set $C := A \cup B$ is given by*

$$\mathbb{1}_C = \max\{\mathbb{1}_A, \mathbb{1}_B\} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \mathbb{1}_B$$

where the maximum is taken pointwise. Similarly, the indicator function for set $D := A \cap B$ is given by

$$\mathbb{1}_D = \min\{\mathbb{1}_A, \mathbb{1}_B\} = \mathbb{1}_A \mathbb{1}_B.$$

The indicator function for A^C is given by

$$\mathbb{1}_{A^C} = 1 - \mathbb{1}_A.$$

Note that if A, B are disjoint, then the indicator function of their union is simply the sum of their individual indicators, i.e.

$$\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B.$$

We can extend these facts to describe indicator functions of arbitrary unions and intersections of sets in the obvious way

PROPOSITION 2.1.3. *Let $\mathcal{I}$ be an arbitrary index set and let $\{A_i\}_{i \in \mathcal{I}} \subseteq \mathcal{X}$ be subsets with indicator functions $\{\mathbb{1}_{A_i}\}_{i \in \mathcal{I}}$. Then, the indicator function for $B := \bigcup_{i \in \mathcal{I}} A_i$ is given by*

$$\mathbb{1}_B = \sup_{i \in \mathcal{I}} \mathbb{1}_{A_i}$$

where the supremum is taken pointwise. Similarly, the indicator function for $C := \bigcap_{i \in \mathcal{I}} A_i$ is given by

$$\mathbb{1}_C = \inf_{i \in \mathcal{I}} \mathbb{1}_{A_i}.$$

PROOF. We provide the argument for B; the argument for C is analagous. Observe that

$$\begin{aligned}\mathbb{1}_B(x) = 1 &\iff x \in \bigcup_{i\in\mathcal{I}} A_i \\ &\iff x \in A_{i_0} \text{ for some } i_0 \in \mathcal{I} \\ &\iff \mathbb{1}_{A_{i_0}} = 1 \text{ for some } i_0 \in \mathcal{I} \\ &\iff \sup_{i\in\mathcal{I}} \mathbb{1}_{A_i} = 1\end{aligned}$$

which completes the argument. □

These arguments appear to be rather pedantic, but they are key to defining limiting operations on sets. With a background in undergraduate calculus, it can be quite cumbersome to think of a sequence of sets converging to another set. However, it is quite straightforward to imagine the pointwise convergence of a sequence of *indicator functions* of sets. For example, we have what appears to be a fairly daunting definition for the limit of a sequence of sets.

DEFINITION 2.1.4. Let $\{A_i\}_{i\in\mathbb{N}}$ be a sequence of sets in $2^{\mathcal{X}}$. Then the limit superior of the sequence is given by

$$\limsup_{n\to\infty} A_n := \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i.$$

Similarly, the limit inferior of the sequence is given by

$$\liminf_{n\to\infty} A_n := \bigcup_{n=1}^{\infty}\bigcap_{i=n}^{\infty} A_i.$$

If $\limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n$ then the limit of the sequence is defined and is equal to the limit superior and inferior i.e.

$$\lim_{n\to\infty} A_n := \limsup_{n\to\infty} A_n = \liminf_{n\to\infty} A_n.$$

While these definitions appear arbitrary, they demarcate important concepts in both analysis and probability. To unpack the intuition, let's try to understand what it means for an element $x \in \mathcal{X}$ to be in $\limsup_{n\to\infty} A_n$. If $x \in \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i$, then $x \in \bigcup_{i=n}^{\infty} A_i$ for every $n \in \mathbb{N}$. That it is to say, for any $n \in \mathbb{N}$, there exists an $i \geq n$ such that $x \in A_i$. This essentially says that x is in infinitely many of the sets $\{A_i\}_{i\in\mathbb{N}}$. In the language of probability, the event $\limsup_{n\to\infty} A_n$ is the event of outcomes that occur infinitely often in the collection of events $\{A_i\}_{i\in\mathbb{N}}$.

On the other hand, if $x \in \liminf_{n\to\infty} A_n$,then there exists some $n_0 \in \mathbb{N}$ such that $x \in A_i$ for every $i \geq n_0$. Clearly then, $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$ which mirrors the domination condition for limit superiors and inferiors of sequences of real numbers or real functions. So when does equality hold? Note that if $\{A_i\}_{i\in\mathbb{N}}$ is an increasing sequence of sets

$$\begin{aligned}\limsup_{n\to\infty} A_n &= \bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty} A_i \\ &= \bigcap_{n=1}^{\infty}\bigcup_{i=1}^{\infty} A_i \\ &= \bigcup_{i=1}^{\infty} A_i \\ &= \bigcup_{i=1}^{\infty}\bigcap_{n=i}^{\infty} A_n \\ &= \liminf_{i\to\infty} A_i\end{aligned}$$

where the second and fourth equalitiies follow from the increasing nature of the sets A_i. This shows that the continuity from below condition described in Proposition 1.3.4 is in fact bona-fide continuity.

After developing the theory of integration, we will (seemingly) generalize this continuity result to measurable functions in the form of the famous *monotone convergence theorem* . Of course, once we know that measures and integrals are essentially the same objects, it will be clear that continuity from below and monotone convergence are two sides of the same coin. While this result is better known in it's integral formulation, there's another result that is perhaps better known in its measure-theoretic formulation: the Borel-Cantelli lemma.

THEOREM 2.1.5 (First Borel-Cantelli lemma). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be an arbitrary measure space and let $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ be a sequence of sets. If*

$$\sum_{i=1}^{\infty} \mu(A_i) < \infty$$

then

$$\mu\left(\limsup_{i\to\infty} A_i\right) = 0.$$

PROOF. Define by $B_n := \bigcup_{i=n}^{\infty} A_i$. It's clear that $\{B_n\}_{n\in\mathbb{N}}$ is a decreasing sequence of sets. More over $\mu(B_1) = \mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) \leq \sum_{i=1}^{\infty} \mu(A_i) < \infty$ by subadditivity and so, since μ is finite on $\{B_n\}_{n\in\mathbb{N}}$, we can apply Propositions 1.3.4 and 1.3.6 to establish

$$\begin{aligned}
\mu\left(\bigcap_{n=1}^{\infty} B_n\right) &= \lim_{n\to\infty} \mu(B_n) \\
&= \lim_{n\to\infty} \mu\left(\bigcup_{i=n}^{\infty} A_i\right) \\
&\leq \lim_{n\to\infty} \sum_{i=n}^{\infty} \mu(A_i) \\
&= \lim_{n\to\infty} \left[\sum_{i=1}^{\infty} \mu(A_i) - \sum_{i=1}^{n-1} \mu(A_i)\right] \\
&= 0
\end{aligned}$$

where the inequality follows from subadditivity and the last equality is due to the the assumption that $\sum_{i\in\mathbb{N}} \mu(A_i) < \infty$ along with the fact that a sequence and its tail have the same limit. ☐

REMARK. This version of the Borel-Cantelli lemma is sometimes called the *first* Borel-Cantelli lemma since its converse, which is true under certain conditions, is also called the Borel-Cantelli lemma in the literature. To prevent ambiguity, we refer to the conversee result as the *second* Borel-Cantelli lemma. The second Borel-Cantelli lemma uses the probabilistic concept of independence and is covered in Theorem9.2.7 in the chapter on independence and as such, we will relegate the discussion of the second Borel-Cantelli lemma to when we formally delve into probabiility theory in the second part of these notes.

By now you should be sufficiently convinced that our definitions of the limiting behavior sets indeed make sense. However, if you any doubts, our treatment of indicator functions should help resolve them completely

PROPOSITION 2.1.6. *Let $\{A_i\}_{i\in\mathbb{N}}$ be a collection of subsets of $\mathcal{X}$ and let $\{\mathbb{1}_{A_i}\}_{i\in\mathbb{N}}$ be their corresponding indicator functions. Then*

$$\limsup_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\limsup_{n\to\infty} A_n}$$

and

$$\liminf_{n\to\infty} \mathbb{1}_{A_n} = \mathbb{1}_{\liminf_{n\to\infty} A_n}.$$

PROOF. We prove the first assertion as the second one follows by the same argument. Note that

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{1}_{A_n} &= \inf_{n\in\mathbb{N}} \left\{\sup_{i\geq n} \mathbb{1}_{A_i}\right\} \\ &= \inf_{n\in\mathbb{N}} \left\{\mathbb{1}_{\bigcup_{i\geq n} A_i}\right\} \\ &= \mathbb{1}_{\bigcap_{n\in\mathbb{N}}\bigcup_{i\geq n} A_i} \\ &= \mathbb{1}_{\limsup_{n\to\infty} A_n}\end{aligned}$$

where the second and third equalities follow by Proposition 2.1.3. □

By now, you should have begun to appreciate that indicator functions are essentially functional equivalents of the sets they indicate: sets and their indicator functioins are just two representations of the same object. This is not particularly suprising, given that sets are defined by their membership and indicator functions describe membership. In this context, it should then not be suprising that indicator functions of *measurable sets* are *measurable functions*, even though we have not yet described the latter concept yet. This is indeed true, and moreover, any non-negative measurable function can be built by taking a limit of a linear combination of indicators of sets. But before we can show this, we should first define what a measurable function is!

DEFINITION 2.1.7. Let $(\mathcal{X},\mathcal{F})$ and $(\mathcal{Y},\mathcal{G})$ be two measurable spaces. A function

$$f : \mathcal{X} \longrightarrow \mathcal{Y}$$

is called $\mathcal{F}/\mathcal{G}-$*measurable* if for any $G \in \mathcal{G}$

$$f^{-1}[G] \in \mathcal{F}.$$

REMARK. This definition also resembles a continuity condition; indeed, if $\mathcal{F}$ and $\mathcal{G}$ were topologies rather than $\sigma-$algebras, this would be the definition of a continuous function. It turns out that if $\mathcal{F}$ and $\mathcal{G}$ are Borel $\sigma-$algebras, then continuity implies measurability: this is a fact that we establish in the next section. Moreover, all measurable functions are in some sense *almost* continuous: we make this notion precise when we discuss the deep connections between topology and measure theory in Chapter 8.

Later in these notes, we will stop writing $\mathcal{F}/\mathcal{G}$ explicitly and let the reader infer the $\sigma-$algebras in play from the context.

2.2. Properties of measurable functions

Armed with our definition of measurable functions, we are ready to discuss interesting examples of such functions along with their properties. First, we establish that the measurability of a sets and its indicator function is indeed equivalent, as we had guessed earlier

PROPOSITION 2.2.1. *Let $(\mathcal{X},\mathcal{F})$ be a measurable space. Then for any $A \subseteq \mathcal{X}$, A is measurable (i.e. $A \in \mathcal{F}$) if and only if*

$$\mathbb{1}_A : \mathcal{X} \longrightarrow \{0,1\}$$

is $\mathcal{F}/2^{\{0,1\}}-$measurable.

PROOF. First assume that $A \in \mathcal{F}$ and observe that if $B = \{0,1\}$ then $\mathbb{1}_A^{-1}[B] = \mathcal{X} \in \mathcal{F}$, if $B = \{1\}$ then $\mathbb{1}_A^{-1}[B] = A \in \mathcal{F}$, if $B = \{0\}$ then $\mathbb{1}_A^{-1}[B] = A^C \in \mathcal{F}$, and if $B = \emptyset$ then $\mathbb{1}_A^{-1}[B] = \emptyset \in \mathcal{F}$.
Conversely, assume that $\mathbb{1}_A$ is measurable and notice how $A = f^{-1}[\{1\}] \in \mathcal{F}$ which completes the proof. □

COROLLARY 2.2.2. *Let $(\mathcal{X},\mathcal{F})$ be a measurable space. For any $A \subseteq \mathcal{X}$, $A \in \mathcal{F}$ if and only if*

$$\mathbb{1}_A : \mathcal{X} \longrightarrow \mathbb{R}$$

is $\mathcal{F}/\mathscr{B}(\mathbb{R})-$measurable.

PROOF. First assume that $A \in \mathcal{F}$ and observe that for any set $B \in \mathscr{B}(\mathbb{R})$

$$\begin{aligned}\mathbb{1}_A^{-1}[B] &= \mathbb{1}_A^{-1}\left[(B \setminus \{0,1\}) \bigcup (B \cap \{0,1\})\right] \\ &= \mathbb{1}_A^{-1}[B \setminus \{0,1\}] \bigcup \mathbb{1}_A^{-1}[B \cap \{0,1\}] \\ &= \mathbb{1}_A^{-1}[B \cap \{0,1\}] \in \mathcal{F}\end{aligned}$$

where the second equality follows by the property of preimages and the last equality follows by the fact that $\mathbb{1}_A^{-1}[B \setminus \{0,1\}] = \emptyset$. The inclusion on the last line then is a consequence of Proposition 2.2.1.

Conversely, assume that $\mathbb{1}_A$ is measurable and observe that since $\{1\} \in \mathscr{B}(\mathbb{R})$, the results follows trivially. $\square$

In this case, measurability of our function was easy to establish because the $\sigma-$algebra $2^{\{0,1\}}$ could be explicitly enumerated. Generally, this is not possible as $\sigma-$algebras can be extremely large. Nevertheless, it is possible establish measurability using a smaller class of sets in the target $\sigma-$algebra; this is the crux of generating class arguments which we discusss more abstractly in the next section. Even the simplest of such arguments can be quite powerful, as we shall see with the following result.

THEOREM 2.2.3 (Generic generating class argument). *Let $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ be measurable spaces and let $\mathcal{E} \subseteq \mathcal{G}$ be a collection of sets such that $\sigma(\mathcal{E}) = \mathcal{G}$. A function*

$$f : \mathcal{X} \longrightarrow \mathcal{Y}$$

is $\mathcal{F}/\mathcal{G}-$measurable if and only if

$$f^{-1}[E] \in \mathcal{F}$$

for every $E \in \mathcal{E}$.

PROOF. If f is measurable, then by definition, $f^{-1}[E] \in \mathcal{F}$ for every $E \in \mathcal{E}$ since $\mathcal{E} \subseteq \mathcal{G}$. Conversely, suppose that $f^{-1}[E] \in \mathcal{F}$ for every $E \in \mathcal{E}$ and define

$$\mathcal{D} = \left\{G \in \mathcal{G} \mid f^{-1}[G] \in \mathcal{F}\right\}.$$

By assumption, $\mathcal{E} \subseteq \mathcal{D}$ and with a little effort we can show that $\mathcal{D}$ is in fact a $\sigma-$algebra, which then shows that $\mathcal{G} = \sigma(\mathcal{E}) \subseteq \sigma(\mathcal{D}) = \mathcal{D}$. First, it is clear that $\emptyset \in \mathcal{D}$ as $f^{-1}[\emptyset] = \emptyset \in \mathcal{F}$. Next, for any $A \in \mathcal{D}$, observe that $f^{-1}[A^c] = \left(f^{-1}[A]\right)^C \in \mathcal{F}$ since $\mathcal{F}$ is a $\sigma-$algebra. Finally, for any collection $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{D}$, $f^{-1}\left[\bigcup_{i\in\mathbb{N}} A_i\right] = \bigcup_{i\in\mathbb{N}} f^{-1}[A_i] \in \mathcal{F}$ again because $\mathcal{F}$ is a $\sigma-$algebra. This completes the proof. $\square$

Now we can show that continuous functions between two spaces equipped with Borel $\sigma-$algebras are indeed measurable.

COROLLARY 2.2.4. *Let $(\mathcal{X}, \mathscr{B}(\mathcal{X}))$ and $(\mathcal{Y}, \mathscr{B}(\mathcal{Y}))$ be measurable spaces and let*

$$f : \mathcal{X} \longrightarrow \mathcal{Y}$$

be continuous. Then f is $\mathscr{B}(\mathcal{X})/\mathscr{B}(\mathcal{Y})-$measurable.

PROOF. Let $\mathcal{O}_{\mathcal{Y}}$ be the topology on $\mathcal{Y}$. By definition

$$\sigma(\mathcal{O}_{\mathcal{Y}}) = \mathscr{B}(\mathcal{Y})$$

and by continuity, for any open set $O \in \mathcal{O}_{\mathcal{Y}}$

$$f^{-1}[O] \in \mathcal{O}_{\mathcal{X}} \subseteq \mathscr{B}(\mathcal{X})$$

where $\mathcal{O}_{\mathcal{X}}$ is the topology on $\mathcal{X}$. Thus by Theorem 2.2.3, f is measurable. $\square$

Next we can use the generic generating class argument to establish routine properties of measurable functions. First, we prove the following useful lemma, which is an important property on its own.

LEMMA 2.2.5. *Let $(\mathcal{X},\mathcal{F}),(\mathcal{Y},\mathcal{G})$, and $(\mathcal{Z},\mathcal{H})$ be measurable spaces. If the functions*

$$\begin{aligned} f &: \mathcal{X} \longrightarrow \mathcal{Y} \\ g &: \mathcal{Y} \longrightarrow \mathcal{Z} \end{aligned}$$

are $\mathcal{F}/\mathcal{G}$ and $\mathcal{G}/\mathcal{H}-$measurable respectively, then the composition function $\phi := f \circ g$ is $\mathcal{F}/\mathcal{H}-$measurable.

PROOF. Let $H \in \mathcal{H}$ be arbitrary and note that

$$\begin{aligned} \phi^{-1}[H] &= f^{-1}\left[g^{-1}[H]\right] \\ &= f^{-1}[G] \\ &= F \in \mathcal{F} \end{aligned}$$

where $G := g^{-1}[H] \in \mathcal{G}$ since g is measurable. □

PROPOSITION 2.2.6. *Let $(\mathcal{X},\mathcal{F})$ be a measurable space and let f,g be real-valued Borel-measurable functions on $\mathcal{X}$; that is $f,g : \mathcal{X} \longrightarrow \mathbb{R}$ and are $\mathcal{F}/\mathscr{B}(\mathbb{R})-$measurable. Define $T : \mathcal{X} \longrightarrow \mathbb{R}^2$ as*

$$T(x) := \begin{pmatrix} f(x) \\ g(x) \end{pmatrix}$$

where $\mathbb{R}^2$ is equipped with its Borel $\sigma-$algebra $\mathscr{B}\left(\mathbb{R}^2\right)$. Finally, let the function $\psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$ be continuous with respect to the standard topologies. Then the function

$$h := \psi \circ T : \mathcal{X} \longrightarrow \mathbb{R}$$

is Borel measurable.

PROOF. First, note that by Corollary 2.2.4 and Lemma 2.2.5, we only need to prove the measurability of T in order to deduce the measurability of h. Let R be an open rectangle in $\mathbb{R}^2$ i.e. $R = I_1 \times I_2$ where

$$I_j = (a_j, b_j)$$

for $a_j > b_j \in \mathbb{R}$. Then consider

$$\begin{aligned} T^{-1}[R] &= \{x \in \mathcal{X} \mid T(x) \in R\} \\ &= \{x \in \mathcal{X} \mid (f(x), g(x)) \in I_1 \times I_2\} \\ &= \{x \in \mathcal{X} \mid f(x) \in I_1 \text{ and } g(x) \in I_2\} \\ &= \{x \in \mathcal{X} \mid f(x) \in I_1\} \bigcap \{x \in \mathcal{X} \mid g(x) \in I_2\} \\ &= f^{-1}[I_1] \bigcap g^{-1}[I_2] \in \mathcal{F} \end{aligned}$$

due to the measurability of f,g and the fact that $\sigma-$algebras are closed under intersection.

Let $\mathcal{R}$ denote the collection of all open rectanges in $\mathbb{R}^2$. Since $\mathcal{R}$ is a subset of all open sets in $\mathbb{R}^2$, we know that $\sigma(\mathcal{R}) \subseteq \mathscr{B}\left(\mathbb{R}^2\right)$. To deduce the converse inclusion, recall that a small modification of Lemma 1.2.14 will show that any open set in $\mathbb{R}^2$ can be written as a countable union of disjoint sets in $\mathcal{R}$ i.e. for any open set $O \subseteq \mathbb{R}^2$

$$O = \bigcup_{i \in \mathbb{N}} R_i$$

where $R_i \in \mathcal{R}$. This implies that $O \in \sigma(\mathcal{R})$ and so $\mathscr{B}\left(\mathbb{R}^2\right) \subseteq \sigma(\mathcal{R})$. Applying a generating class argument with $\mathcal{R}$ then shows that T is measurable. □

COROLLARY 2.2.7. *For any real-valued Borel-measurable functions f,g on $(\mathcal{X},\mathcal{F})$, the following functions are also measurable*

(i) $h(x) := f(x) + g(x)$
(ii) $h(x) := f(x)g(x)$
(iii) $h(x) := f(x)/g(x)$ where $g(x) \neq 0$

PROOF. For (i), let $\psi(x,y) := x+y$ (which is a continuous function) and apply Proposition 2.2.6. The other cases are simiilar. $\square$

So we know that measurabiltiy is preserved under addition and multiplication, but the principal concept in analysis is the limit, and we would like measurability of functions to be preserved under limiting operations. The following results help us establish that measurability of real-valued functions is indeed preserved under pointwise limits, when one exists.

LEMMA 2.2.8. *The following sets are generating classes for* $\mathscr{B}(\mathbb{R})$:

$$\{(a,\infty) \mid \forall a \in \mathbb{R}\},\ \{[a,\infty) \mid \forall a \in \mathbb{R}\},\ \{(-\infty,b) \mid \forall b \in \mathbb{R}\},\ \{(-\infty,b] \mid \forall a \in \mathbb{R}\}.$$

PROOF. Recall that the collection of half-open intervals $\mathcal{L} = \{(a,b] \mid -\infty < a \leq b < \infty\}$ is a generating class for $\mathscr{B}(\mathbb{R})$. Observe that for any $a \leq b \in \mathbb{R}$,

$$(a,b] = \bigcap_{n\in\mathbb{N}} \left((a,\infty) \cap \left(b+\frac{1}{n},\infty\right)\right)$$

and so $\mathcal{L} \subseteq \sigma(\{(a,\infty) \mid \forall a \in \mathbb{R}\})$ which implies that $\mathscr{B}(\mathbb{R}) \subseteq \sigma(\{(a,\infty) \mid \forall a \in \mathbb{R}\})$. To see the reverse inclusion, note that for any $a \in \mathbb{R}$

$$(a,\infty) = \bigcup_{b\in\mathbb{N}} (a,b]$$

which shows that $\sigma(\{(a,\infty) \mid \forall a \in \mathbb{R}\}) \subseteq \mathscr{B}(\mathbb{R})$. The other sets can be shown to be generators in a similar fashion. $\square$

PROPOSITION 2.2.9. *For a sequence of real-valued Borel-measurable functions* $\{f_n\}_{n\in\mathbb{N}}$ *on* $(\mathcal{X},\mathcal{F})$, *the functions*

$$g := \sup_{n\in\mathbb{N}} f_n$$

and

$$h := \inf_{n\in\mathbb{N}} f_n$$

are $\mathcal{F}/\mathscr{B}\left(\overline{\mathbb{R}}\right)$ *measurable where* $\overline{\mathbb{R}}$ *represents the extended real line.*

PROOF. Let $a \in \mathbb{R}$ be arbitrary. Note that the set

$$g^{-1}[(a,\infty)] = \{x \in \mathcal{X} \mid g(x) > a\} = \bigcup_{n\in\mathbb{N}} \{x \in \mathcal{X} \mid f_n(x) > a\} = \bigcup_{n\in\mathbb{N}} f_n^{-1}[(a,\infty)]$$

since if $g(x) > a$ then there's at least one $n \in \mathbb{N}$ such that $f_n(x) > a$. Note that by the measurability of f_n, $f_n^{-1}[(a,\infty)] \in \mathcal{F}$ for every $n \in \mathbb{N}$. Since $\mathcal{F}$ is a $\sigma-$algebra, $g^{-1}[(a,\infty)] \in \mathcal{F}$ by closure under countable unions. This establishes the measurability of g by Lemma 2.2.8. A similar argument establishes the measurability of h. $\square$

COROLLARY 2.2.10. *For a sequence of real-valued Borel-measurable functions* $\{f_n\}_{n\in\mathbb{N}}$ *on* $(\mathcal{X},\mathcal{F})$, *the functions*

$$g := \limsup_{n\to\infty} f_n$$

and

$$h := \liminf_{n\to\infty} f_n$$

are measurable (provided they exist) and if $h = g$ *then*

$$\lim_{n\to\infty} f_n = h = g$$

is measurable.

PROOF. Recall that

$$g = \limsup_{n\to\infty} f_n = \inf_{n\in\mathbb{N}} \left\{\sup_{k\geq n} f_k\right\}$$

and apply Proposition 2.2.9. The other results follow in the same fashion. $\square$

Of course, we are often more interested in establishing the measurability of vector-valued function i.e functions whose range is some subset of the Euclidean space $\mathbb{R}^n$. This limit theorems for real-valued functions proved here extend naturally to arbitrary finite dimensional spaces.

PROPOSITION 2.2.11. *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let the sequence of functions $\{f_m\}_{m\in\mathbb{N}}$*

$$f_m : \mathcal{X} \longrightarrow \mathbb{R}^n$$

be $\mathcal{F}/\mathscr{B}(\mathbb{R}^n)-$measurable. The pointwise limit function $f := \lim_{m\to\infty} f_m$ (if it exists) is $\mathcal{F}/\mathscr{B}(\mathbb{R}^n)-$measurable if and only if the projection functions $f_1, \ldots, f_n$ are $\mathcal{F}/\mathscr{B}(\mathbb{R})-$measurable.

PROOF. First assume that the projections $f_1, \ldots, f_n$ are measurable. Then, applying the same proof we used to prove the measurability of the function T in Proposition 2.2.6, we have that

$$f = \begin{pmatrix} f_1 \\ \vdots \\ f_n \end{pmatrix}$$

is measurable.

Conversely, suppose that f is measurable, then recall that the projection functions $\pi_k : \mathbb{R}^n \longrightarrow \mathbb{R}$ given by

$$\pi_k \begin{pmatrix} x_1 \\ \vdots \\ x_k \\ \vdots \\ x_n \end{pmatrix} = x_k$$

is a continuous function and hence measurable for any $1 \le k \le n$. Applying Lemma 2.2.5, we have that $f_k = \pi_k \circ f$ is measurable for any $1 \le k \le n$. $\square$

In general, it appears that a measurability is preserved under a wide range of operations on functions. Additionally, the richness of commonly seen $\sigma-$algebras makes it the case that almost all functions we encounter in everyday mathematics are measurable. This heuristic is a fine one, but it begs the question: precisely how rich does the $\sigma-$algebra of the domain have to be for a given function to be measurable? This has a relatively straightforward answer.

PROPOSITION 2.2.12. *Let $\mathcal{X}$ be a set and let $(\mathcal{Y}, \mathcal{G})$ be a measurable space. Let*

$$T : \mathcal{X} \longrightarrow \mathcal{Y}$$

be a function. Then the "smallest" $\sigma-$algebra on $\mathcal{X}$ that makes T measurable is

$$\sigma(T) := \left\{T^{-1}[G] \mid G \in \mathcal{G}\right\}.$$

That is, $\sigma(T)$ is the intersecion of all $\sigma-$algebras on $\mathcal{X}$ that makes T measurable.

PROOF. Let $\mathcal{R}$ be the collection of all $\sigma-$algebras on $\mathcal{X}$ that makes T measurable. Clearly, $\sigma(T) \subseteq R$ for every $R \in \mathcal{R}$ by the definition of measurability. Thus

$$\sigma(T) \subseteq \bigcap_{R\in\mathcal{R}} R$$

and so all we have to do is show that $\sigma(T) \in \mathcal{R}$ to show the reverse inclusion. To do this, we simply show that $\sigma(T)$ is a $\sigma-$algebra.

Note that $\emptyset = T^{-1}[\emptyset] \in \sigma(T)$. Next, let $A \in \sigma(T)$ be arbitrary and observe that $A = T^{-1}[G]$ for some $G \in \mathcal{G}$. Then $A^C = \left(T^{-1}[G]\right)^C = T^{-1}\left[G^C\right] \in \sigma(T)$ as $G^C \in \mathcal{G}$ since $\mathcal{G}$ is a $\sigma-$algebra. Finally, suppose $\{A_i\}_{i\in\mathbb{N}} \in \sigma(T)$ and note that $A_i = T^{-1}[G_i]$ for $G_i \in \mathcal{G}$ and so

$$\bigcup_{i\in\mathbb{N}} A_i = \bigcup_{i\in\mathbb{N}} T^{-1}[G_i] = T^{-1}\left[\bigcup_{i\in\mathbb{N}} G_i\right] \in \sigma(T)$$

since $\mathcal{G}$ is closed under countable unions. This completes the proof. $\square$

COROLLARY 2.2.13. *Let $\mathcal{X}$ be a set and let $(\mathcal{Y}, \mathcal{G})$ be a measurable space. Let $\mathcal{H} \subset \mathcal{G}$ be a generating class such that $\sigma(\mathcal{H}) = \mathcal{G}$. Then for a function $T : \mathcal{X} \to \mathcal{Y}$*

$$\sigma\left(T^{-1}(\mathcal{H})\right) = \sigma(T).$$

PROOF. Note that $T^{-1}(\mathcal{H}) \subseteq \sigma(T)$ and so $\sigma\left(T^{-1}(\mathcal{H})\right) \subseteq \sigma(T)$. Conversely, note that T is $\sigma\left(T^{-1}(\mathcal{H})\right)/\mathcal{G}$ measurable by the generating class argument in Theorem 2.2.3. But since by Proposition 2.2.12,$\sigma(T)$ is the smallest $\sigma-$algebra on which T is measurable, we have that $\sigma(T) \subseteq \sigma\left(T^{-1}(\mathcal{H})\right)$. $\square$

PROPOSITION 2.2.14. *Let $(\mathcal{X}, \mathcal{F})$, $(\mathcal{Y}, \mathcal{G})$, and $(\mathcal{Z}, \mathcal{H})$ be measurable spaces and let $f : \mathcal{X} \to \mathcal{Y}$ be $\mathcal{F}/\mathcal{G}$ measurable. Then for any $\mathcal{G}/\mathcal{H}$ measurable function $g : \mathcal{Y} \to \mathbb{R}$*

$$\sigma(g \circ f) \subseteq \sigma(f).$$

PROOF. Note that for any $H \in \mathcal{H}$, $(g \circ f)^{-1}[H] = f^{-1}\left[g^{-1}[H]\right] \in \sigma(f)$ which implies the claim. $\square$

Note that $\sigma-$algebras generated by measurable functions constitute an important subclass of all $\sigma-$algebras.

PROPOSITION 2.2.15. *Let $\mathcal{X}$ be a set and let $\{B_i\}_{i\in\mathbb{N}} \subseteq \mathcal{X}$ be a countable collection of sets. Then there exists a Borel-measurable function $f : \mathcal{X} \to \mathbb{R}$ such that*

$$\sigma\left(\{B_i\}_{i\in\mathbb{N}}\right) = \sigma(f).$$

2.3. Constructing measurable functions

2.3.1. Simple functions. In the beginning of this chapter, we foreshadowed how the indicator functions of measurable sets were in fact the building blocks of all real-valued non-negative measurable functions; now we have developed just enough of the theory to show that this is true. First, we should introduce some useful notation to save space in the future. We denote by $\mathcal{M}(\mathcal{X}, \mathcal{F})$ the set of (extended) real-valued Borel-measurable functions on $(\mathcal{X}, \mathcal{F})$. We write $\mathcal{M}^+(\mathcal{X}, \mathcal{F}) \subset \mathcal{M}(\mathcal{X}, \mathcal{F})$ to denote the set of all non-negative (extended) real-valued Borel-measurable functions on $(\mathcal{X}, \mathcal{F})$. Note that the properties we described for the set of real-valued measurable functions on $(\mathcal{X}, \mathcal{F})$ in the previous section carry over to the set of extended real valued measurable functions, as long as no weird $\infty - \infty$ or $\infty \cdot 0$ situations arise.

DEFINITION 2.3.1. A function $s : \mathcal{X} \longrightarrow [0, \infty)$ is called a *simple function* if its range is a finite subset of $[0, \infty)$ It can be represented as an aggregation of indicator functions as

$$s = \sum_{i=1}^{k} \alpha_i \mathbb{1}_{A_i}$$

where $\{\alpha_i\}_{i=1}^k \in [0, \infty)$ is the range of s and $A_i := \{x \in \mathcal{X} \mid s(x) = \alpha_i\}$ partitions the domain $\mathcal{X}$ into preimages of the singletons in the range. Such a representation is called the *standard representation* of s

REMARK. Note that in general, a simple function could be written as a finite linear combination of indicator functions in more than one way. For example, the function $s = 3\mathbb{1}_A + 7\mathbb{1}_B$ can also be expressed as $3\mathbb{1}_{A\setminus B} + 10\mathbb{1}_{A\cap B} + 7\mathbb{1}_{B\setminus A}$, where the latter is the *standard representation* since the sets $A \cap B, A \setminus B, B \setminus A$ (and the implicitly included $\mathcal{X} \setminus (A \cup B)$) form a partition of the domain given by preimages of the singletons in the range (namely $\{0, 3, 7, 10\}$). Standard representations are, of course, unique.

It should be clear that a non-negative simple function $s \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ if and only if $\{A_i\}_{i=1}^k \in \mathcal{F}$; the "if" part follows directly from Corollary 2.2.7. To see the "only if" part, note that if s is measurable then $s^{-1}[\{\alpha_i\}] = A_i \in \mathcal{F}$ since singletons in $\mathbb{R}$ are Borel sets (they are closed). Denote the collection of these measurable simple functions on $(\mathcal{X}, \mathcal{F})$ as $M_{\text{sim}}(\mathcal{X}, \mathcal{F})$. Then, we have shown that $M_{\text{sim}}(\mathcal{X}, \mathcal{F}) \subset \mathcal{M}^+(\mathcal{X}, \mathcal{F})$. It turns out that we can make a stronger claim: the measurable simple functions are in some sense "dense" in the space of non-negative measurable functions.

PROPOSITION 2.3.2. *Let $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ be arbitrary. Then there exists a sequence of simple measurable functions $\{f_n\}_{n\in\mathbb{N}} \in M_{\text{sim}}(\mathcal{X}, \mathcal{F})$ such that*

$$f_n \leq f_{n+1}$$

pointwise for every $n \in \mathbb{N}$ and

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

where the limit is taken pointwise.

PROOF. It turns out that we can establish the existence of such functions f_n constructively. Define

$$f_n(x) := \sum_{k=0}^{4^n-1} \frac{k}{2^n} \mathbb{1}_{\left\{\frac{k}{2^n} \leq f(x) < \frac{k+1}{2^n}\right\}} + 2^n \mathbb{1}_{\{f(x) \geq 2^n\}}$$

and observe that f_n are Borel-measurable simple functions (given the measurability of f) and that $f_n \leq f$ pointwise by definition. Next, note that $\left\{\frac{k}{2^n}\right\}_{k=0}^{4^n} \subset \left\{\frac{k}{2^{n+1}}\right\}_{k=0}^{4^{n+1}}$ and fix $x_0 \in \mathcal{X}$ to be arbitrary. If $\frac{k_0}{2^n} \leq f(x_0) < \frac{k_0+1}{2^n}$ for some $k_0 \in \{0, 1, \ldots, 4^n - 1\}$ then either $\frac{k_0}{2^n} = \frac{2k_0}{2^{n+1}} \leq f(x_0) < \frac{2k_0+1}{2^{n+1}}$ or $\frac{2k_0+1}{2^{n+1}} \leq f(x_0) < \frac{2k_0+2}{2^{n+1}} = \frac{k_0+1}{2^n}$. In either case,

$$f_n(x_0) = \frac{k_0}{2^n} \leq f_{n+1}(x_0) \leq \frac{2k_0+1}{2^{n+1}}$$

by the definitions of our simple functions f_n. Conversely, if $f(x_0) \geq 2^n$ then either $f(x_0) \geq 2^{n+1}$ or there exists some $k_0 \in \left\{2^{2n+1}, \ldots, 4^{n+1} - 1\right\}$ such that

$$\frac{k_0}{2^{n+1}} \leq f(x_0) < \frac{k_0+1}{2^{n+1}}.$$

Again, in either case,

$$f_n(x_0) = 2^n \leq \frac{k_0}{2^{n+1}} \leq f_{n+1}(x_0) \leq 2^{n+1}$$

which shows that $f_n \leq f_{n+1}$ pointwise.

To show convergence, again pick an arbitrary $x_0 \in \mathcal{X}$ and observe that if $f(x_0) = \infty$ then $f_n(x_0) = 2^n \nearrow \infty = f(x_0)$. Conversely, suppose that $f(x_0) < \infty$. Then, since the natural numbers are not bounded above in $\mathbb{R}$, we know that there exists some $n_{x_0} \in \mathbb{N}$ such that for every $n \geq n_{x_0}$: $f(x_0) < 2^n$. Thus for each such n, there exists some $k_n \in \{0, 1, \ldots, 4^n - 1\}$ such that

$$\frac{k_n}{2^n} \leq f(x_0) < \frac{k_n+1}{2^n} \tag{2}$$

which would imply that $f_n(x_0) = \frac{k_n}{2^n}$ for all $n \geq n_{x_0}$. Then

$$0 \leq f(x_0) - f_n(x_0) = f(x_0) - \frac{k_n}{2^n} \leq \frac{k_n+1}{2^n} - \frac{k_n}{2_n} = \frac{1}{2^n}$$

where the first inequality is due to the fact that $f_n \leq f$ pointwise and the second inequality is due to (2). Taking limits then yields the result. □

PROPOSITION 2.3.3. *For any function $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$, the derived functions*[1]

$$f^+ := \max\{f, 0\}$$
$$f^- := \max\{-f, 0\}$$

[1]Again, maxima are taken pointwise.

are contained in $\mathcal{M}^+(\mathcal{X},\mathcal{F})$

PROOF. Define for any $a \in \mathbb{R}$, $F_a := \{x \in \mathcal{X} \mid f^+(x) \in (a,\infty)\}$ and notice that if $a > 0$, then

$$F_a = f^{-1}[(a,\infty)] \in \mathcal{F}$$

by the measurability of f. Conversely, if $a \leq 0$ then

$$F_a = \mathcal{X} \in \mathcal{F}$$

and so, by Lemma 2.2.8 and a generating class argument, $f^+ \in \mathcal{M}^+(\mathcal{X},\mathcal{F})$.

To see that $f^{-1} \in \mathcal{M}^+(\mathcal{X},\mathcal{F})$, notice that by Corollary 2.2.7, $-f \in \mathcal{M}(\mathcal{X},\mathcal{F})$ and apply the same argument. □

REMARK. Another way to prove the above proposition would be to observe that

$$\max\{x,y\} = \frac{x+y+|x-y|}{2}$$

and apply Proposition 2.2.6.

Proposition 2.3.3 is important because any function $f \in \mathcal{M}(\mathcal{X},\mathcal{F})$ can be decomposed as

$$f = f^+ - f^-$$

and since both $f^+, f^- \in \mathcal{M}^+(\mathcal{X},\mathcal{F})$, by Proposition 2.3.2, there are non-negative simple functions $\{s_n^+\}_{n\in\mathbb{N}}, \{s_n^-\}_{n\in\mathbb{N}}$ such that $s_n^+ \nearrow f^+$and $s_n^- \nearrow f^-$. By the linearity of limits, the sequence of functions $h_n := s_n^+ - s_n^-$ converges (although not monotonically) to f. This fact will prove important in the chapter on integration.

PROPOSITION 2.3.4. *Let* $s,t \in \mathcal{M}_{\text{sim}}(\mathcal{X},\mathcal{F})$. *Then*

$$h := s + t \in \mathcal{M}_{\text{sim}}(\mathcal{X},\mathcal{F})$$

and

$$g := st \in \mathcal{M}_{\text{sim}}(\mathcal{X},\mathcal{F}).$$

Moreover, if s *and* t *be given by the standard representations*

$$s = \sum_{i=1}^{I} \alpha_i \mathbb{1}_{A_i}$$
$$t = \sum_{j=1}^{J} \beta_j \mathbb{1}_{B_j}$$

then

$$h = \sum_{i=1}^{I}\sum_{j=1}^{J} (\alpha_i + \beta_j)\mathbb{1}_{A_i\cap B_j}$$
$$g = \sum_{i=1}^{I}\sum_{j=1}^{J} (\alpha_i \beta_j)\mathbb{1}_{A_i\cap B_j}$$

PROOF. First we prove that

$$\sum_{i=1}^{I} \alpha_i \mathbb{1}_{A_i} + \sum_{j=1}^{J} \beta_j \mathbb{1}_{B_j} = \sum_{i=1}^{I}\sum_{j=1}^{J} (\alpha_i + \beta_j)\mathbb{1}_{A_i\cap B_j}.$$

To see this, let $x_0 \in \mathcal{X}$ be arbitrary and observe that since $\{A_i\}_{i=1}^{I}$ and $\{B_j\}_{j=1}^{J}$ are partitions of $\mathcal{X}$, there exists exactly one $1 \leq i_0 \leq I$ and one $1 \leq j_0 \leq J$ such that $x_0 \in A_{i_0}$ and $x \in B_{j_0}$. Then,

$x_0 \in A_{i_0} \cap B_{j_0}$ and $x \notin A_i \cap B_j$ if either $i \neq i_0$ or $j \neq j_0$.[2]Then

$$\begin{aligned}\left(\sum_{i=1}^{I} \alpha_i \mathbb{1}_{A_i} + \sum_{j=1}^{J} \beta_j \mathbb{1}_{B_j}\right)(x_0) &= \sum_{i=1}^{I} \alpha_i \mathbb{1}_{A_i}(x_0) + \sum_{j=1}^{J} \beta_j \mathbb{1}_{B_j}(x_0) \\ &= \alpha_{i_0} + \beta_{j_0} \\ &= \sum_{i=1}^{I} \sum_{j=1}^{J} (\alpha_i + \beta_j) \mathbb{1}_{A_i \cap B_j}(x_0)\end{aligned}$$

which establishes our claim. Note that the representation above need not be standard as the partition $\{A_i \cap B_j\}_{(i,j)\in\{1,\ldots,I\}\times\{1,\ldots,J\}}$ need not be the collection of preimages of singletons in $\mathbf{Ran}(h)$; in fact, it is a *refinement* of the collection of preimages, and so the preimage of every singleton in $\mathbf{Ran}(h)$ can be written as a union of sets in $\{A_i \cap B_j\}_{(i,j)\in\{1,\ldots,I\}\times\{1,\ldots,J\}}$.

Next, for closure under multiplication, observe that

$$\begin{aligned}st &= \left(\sum_{i=1}^{I} \alpha_i \mathbb{1}_{A_i}\right)\left(\sum_{j=1}^{J} \beta_j \mathbb{1}_{B_j}\right) \\ &= \sum_{i=1}^{I} \sum_{j=1}^{J} (\alpha_i \beta_j) \mathbb{1}_{A_i} \mathbb{1}_{B_j} \\ &= \sum_{i=1}^{I} \sum_{j=1}^{J} (\alpha_i \beta_j) \mathbb{1}_{A_i \cap B_j}\end{aligned}$$

where the last equality follows from Fact 2.1.2. For the same reason as before, this representation need not be standard. □

PROPOSITION 2.3.5. *Let $\mathcal{X}$ be an arbitrary set and let $f : \mathcal{X} \to \mathbb{R}$ be a function, then for any $\sigma(f)/\mathscr{B}(\mathbb{R})-$measurable function $g : \mathcal{X} \longrightarrow \mathbb{R}$, there exists a Borel-measurable map $h : \mathbb{R} \to \mathbb{R}$ such that*

$$g = h \circ f.$$

PROOF. First suppose that $g = \mathbb{1}_A$ for some $A \in \sigma(f)$. Then

$$g = \mathbb{1}_A(x) = \mathbb{1}_{f^{-1}[B]}(x) = \mathbb{1}_B(f(x))$$

for some $B \in \mathcal{G}$. Thus in this case, $h = \mathbb{1}_B$.

Now consider the case of a simple function $g = \sum_{i=1}^{n} \alpha_i \mathbb{1}_{A_i}$ for $\alpha_i \in \mathbb{R}, A_i \in \sigma(F)$. Note that by the above example for indicator functions, we can find $h_i : \mathbb{R} \to \mathbb{R}$ such that

$$g = \sum_{i=1}^{n} \alpha_i h_i(f(x))$$

and so the function $h = \sum_{i=1}^{n} \alpha_i h_i$ does the trick.

Next, for $g \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$, by Proposition 2.3.2we can find a sequence of simple functions s_n such that $s_n \nearrow g$ pointwise. Note from our work above that we can write $s_n = h_n \circ f$ where h_n is Borel-measurable as above. Let $h := \lim_{n\to\infty} h_n$ which exists pointwise since it is a monotone limit $(s_n \leq s_{n+1} \implies h_{n+1} \leq h_n)$. In particular, h is Borel-measurable by Corollary 2.2.10 and so for any fixed $x \in \mathcal{X}$

$$\lim_{n\to\infty} s_n(x) = \lim_{n\to\infty} h_n(f(x)) = h(f(x)) = g(x).$$

[2]In other words, $\{A_i \cap B_j\}_{(i,j)\in\{1,\ldots,I\}\times\{1,\ldots,J\}}$forms a partition of $\mathcal{X}$.

Finally, for general measurable $g \in \mathcal{M}(\mathcal{X}, \mathcal{F})$, we can write $g = g^{+} - g^{-}$ and by the previous results, we know there exist Borel-measurable real-valued maps h_1, h_2 such that $g^{+} = h_1 \circ f$, $g^{-} = h_2 \circ f$. Then

$$\begin{aligned} g &= h_1 \circ f - h_2 \circ f \\ &= (h_1 - h_2) \circ f \end{aligned}$$

and $h := h_1 - h_2$ is Borel-measurable by Corollary 2.2.7. This completes the proof. $\square$

2.3.2. Images of measurable sets under measurable functions. We have the shown (or rather, defined) that for measurable functions, preimages of measurable sets are measurable. Does this hold for images under measurable functions as well? That is, for any pair of measurable spaces $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ and a measurable function $f : \mathcal{X} \to \mathcal{G}$, is it true that $f[A] \in \mathcal{G}$ for any $A \in \mathcal{F}$? The answer to this question is "no" and to show this, we continue on a thread we began in Subsection 1.4.2 (read that section again). We shall first construct a strictly increasing analogue of the Cantor function ψ_C from that section.

Proposition 2.3.6. *Let $\phi(x) := \psi_C(x) + x$. Then ϕ is a strictly increasing, continuous function that maps $[0,1]$ onto $[0,2]$. Moreover, $\phi[C] \in \mathscr{B}(\mathbb{R})$ is a set of positive Lebesgue measure.*

Proof. Note that the sum of a non-decreasing and strictly increasing function is strictly increasing and the sum of two continuous functions is continuous. That gives us the first two properties. Now since $\phi(0) = 0$ and $\phi(1) = 2$, by the intermediate value theorem $\phi[[0,1]] = [0,2]$. Next, note that $\phi[C] = \left(\phi^{-1}\right)^{-1}[C]$ which is Borel measurable since C is closed (and so Borel measurable) and ϕ^{-1} is a continuous (and hence Borel measurable) function by Proposition D.2.14. Finally, consider the image $\phi[[0,1] \setminus C]$. We know from Proposition 1.4.25that there exists a countable collection of disjoint open intervals $\{O_j\}_{j \in \mathbb{N}} \subset [0,1]$ such that $\bigcup_{j \in \mathbb{N}} O_j = [0,1] \setminus C$ such that $\psi_C(x) = c_j$ for any $x \in O_j$. Then,

$$\begin{aligned} \phi[[0,1] \setminus C] &= \phi\left[\bigcup_{j \in \mathbb{N}} O_j\right] \\ &= \bigcup_{j \in \mathbb{N}} \phi[O_j] \\ &= \bigcup_{j \in \mathbb{N}} O_j + c_j \end{aligned}$$

where each $O_j + c_j$ is disjoint since ϕ is a bijection and so its image preserves intersections. Then,

$$\begin{aligned} \lambda(\phi[[0,1] \setminus C]) &= \lambda\left(\bigcup_{j \in \mathbb{N}} O_j + c_j\right) \\ &= \sum_{i=1}^{\infty} \lambda(O_j) \\ &= \lambda\left(\bigcup_{j \in \mathbb{N}} O_j\right) \\ &= \lambda([0,1] \setminus C) \\ &= 1 \end{aligned}$$

where in the inequality we used countable additivity and translation invariance. Then, by additivity, finiteness of λ on bounded intervasls, and the fact that the image of a bijective function preserves set differences-

$$\begin{aligned}1=\lambda\left(\phi\left[[0,1]\setminus C\right]\right)&=\lambda\left(\phi\left[[0,1]\right]\setminus\phi\left[C\right]\right)\\&=\lambda\left(\phi\left[[0,1]\right]\right)-\lambda\left(\phi\left[C\right]\right)\\&=\lambda\left([0,2]\right)-\lambda\left(\phi\left[C\right]\right)\\&=2-\lambda\left(\phi\left[C\right]\right).\end{aligned}$$

Therefore

$$\lambda\left(\phi\left[C\right]\right)=1$$

which completes the proof. □

THEOREM 2.3.7. *There exists a measurable set $A\in\mathcal{C}\left(\lambda^*\right)$ such that $\phi\left[A\right]\notin\mathcal{C}\left(\lambda^*\right)$.*

PROOF. Note that by Theorem 1.4.10 and Proposition 2.3.6, $\phi\left[C\right]$ contains a set E such that $E\notin\mathcal{C}\left(\lambda^*\right)$. Let $A=C\cap\phi^{-1}\left[E\right]$ and notice that A is measurable as it is a subset of a measure-zero set and since $\mathcal{C}\left(\lambda^*\right)$ is complete. Then, $\phi\left[A\right]=E$ which is not measurable. □

COROLLARY 2.3.8. *There exists a set $A\in\mathcal{C}\left(\lambda^*\right)$ such that $A\notin\mathscr{B}\left(\mathbb{R}\right)$.*

PROOF. Let $A=C\cap\phi^{-1}\left[E\right]$ as in Theorem 2.3.7. Then $A\in\mathcal{C}\left(\lambda^*\right)$ but $\phi\left[A\right]\notin\mathcal{C}\left(\lambda^*\right)$. But recall that ϕ^{-1} is measurable since it is continuous and $\phi\left[A\right]=\left(\phi^{-1}\right)^{-1}\left[A\right]$ and so if $A\in\mathscr{B}\left(\mathbb{R}\right)$ then $\phi\left[A\right]\in\mathscr{B}\left(\mathbb{R}\right)$ which would be a contradiction. □

2.3.3. An iterative construction of the Cantor function.

2.4. Types of generating class arguments

So far we have seen how a generating class argument, such as the one formalized in Theorem 2.2.3, can help establish the measurability of functions. In general, measurability can be replaced by any property, opening up other avenues to use such arguments. The abstract analogue of Theorem 2.2.3 goes as follows: Let $\mathcal{F}$ be a $\sigma-$algebra on $\mathcal{X}$. We want to show that all sets in $\mathcal{F}$ enjoy some property $(*)$. In order to prove that this is indeed the case we

(1) Find a subclass of sets $\mathcal{E}\subseteq\mathcal{F}$ that enjoys property $(*)$ such that $\sigma\left(\mathcal{E}\right)=\mathcal{F}$.
(2) Define $\mathcal{D}=\{F\in\mathcal{F}\mid F$ enjoys property $(*)\}$.
(3) Observe that if $\mathcal{D}$ is a $\sigma-$algebra then

$$\mathcal{E}\subseteq\mathcal{D}\subseteq\mathcal{F}\Longrightarrow\mathcal{F}=\sigma\left(\mathcal{E}\right)\subseteq\sigma\left(\mathcal{D}\right)=\mathcal{D}\subseteq\mathcal{F}\Longrightarrow\mathcal{F}=\mathcal{D}.$$

In this argument, we do not assume any structure on the special class $\mathcal{E}$. However, imposing structure can add power to generating class arguments, since then we can drop the requirement that subclass $\mathcal{D}$ of "desirable" sets be a $\sigma-$algebra. The fact that we can do this is not *a-priori* obvious, and the point of this section is to develop the theory to show that this can be done with different types of structural assumptions. The results of this section form the backbone of standard techniques in measure theory, and have important applications, one of which we show here.

2.4.1. Dynkin's $\pi-\lambda$ Theorem.

DEFINITION 2.4.1. A collection of sets $\mathcal{D}\subseteq2^{\mathcal{X}}$ is called a *Dynkin system* or a $\lambda-$system if

(i) $\mathcal{X}\in\mathcal{D}$
(ii) $A_1,A_2\in\mathcal{D}$ s.t. $A_2\subseteq A_1\Longrightarrow A_2\setminus A_1\in\mathcal{D}$
(iii) $\{A_i\}_{i\in\mathbb{N}}\in\mathcal{D}$ s.t. $A_i\subseteq A_{i+1}\Longrightarrow\bigcup_{i\in\mathbb{N}}A_i\in\mathcal{D}$

PROPOSITION 2.4.2. *A collection of sets $\mathcal{D}\subseteq2^{\mathcal{X}}$ is a $\lambda-$system if and only if*

(i') $\mathcal{X}\in\mathcal{D}$
(ii') $A\in\mathcal{D}\Longrightarrow A^C\in\mathcal{D}$
(iii') $\{A_i\}_{i\in\mathbb{N}}\in\mathcal{D}$ *s.t* $A_i\cap A_j=\emptyset\Longrightarrow\bigcup_{i\in\mathbb{N}}A_i\in\mathcal{D}$

PROOF. First suppose that $\mathcal{D}$ is a $\lambda-$system. Then, property (i') above is automatically satisfied. For property (ii'), observe that (i),(ii) together imply (ii') since $A^C = \mathcal{X} \setminus A$. Next, let $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{D}$ be pairwise disjoint and define

$$B_n := \bigcup_{i=1}^{n} A_i,$$

observing that $B_n \subseteq B_{n+1}$ and that $\bigcup_{n\in\mathbb{N}} B_n = \bigcup_{i\in\mathbb{N}} A_i$. Then (iii') follows from (iii).

Conversely, assume that $\mathcal{D}$ satisfies $(i') - (iii')$ above. Again, (i) follows from (i'). Next, pick $A_1, A_2 \in \mathcal{D}$ such that $A_2 \subseteq A_1$. By (ii') , $A_1^C \in \mathcal{D}$ and we know that$A_1^C \cap A_2 = \emptyset$ and so by $(ii')\,,(iii')$, we have that

$$A_1 \setminus A_2 = \left(A_1^C \cup A_2\right)^C \in \mathcal{D}$$

which proves (ii). Finally, let $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{D}$ be an increasing sequence of sets (that is, $A_i \subseteq A_{i+1}$ for every $i \in \mathbb{N}$). By (ii),

$$B_i := A_i \setminus A_{i-1} \in \mathcal{D}$$

and are pairwise disjoint. Applying (iii') we have that

$$\bigcup_{i\in\mathbb{N}} A_i = \bigcup_{i\in\mathbb{N}} B_i \in \mathcal{D}$$

completing the proof. $\square$

It is easy to see that a $\sigma-$algebra is a $\lambda-$system, and arguments akin to those in Proposition 1.2.11 will prove that

$$\lambda(\mathcal{E}) := \bigcap \left\{\mathcal{A} \subseteq 2^{\mathcal{X}} \text{ is a } \lambda-\text{system} \mid \mathcal{E} \subseteq \mathcal{A}\right\}$$

is itself a $\lambda-$system, and these two facts together show that

$$\mathcal{E} \subseteq \lambda(\mathcal{E}) \subseteq \sigma(\mathcal{E}).$$

DEFINITION 2.4.3. A collection of sets $\mathcal{E} \subseteq 2^{\mathcal{X}}$ is called a $\pi-$system if it is closed under finite intersections i.e. if $A, B \in \mathcal{E}$ then $A \cap B \in \mathcal{E}$.

Properties of $\pi-$systems and $\lambda-$systems can be combined to yield a $\sigma-$algebra.

LEMMA 2.4.4. *A collection of sets $\mathcal{F} \subseteq 2^{\mathcal{X}}$ is a $\sigma-$algebra if and only if it is both a $\pi-$system and a $\lambda-$system.*

PROOF. Clearly, if $\mathcal{F}$ is a $\sigma-$algebra, then it is both a $\pi-$system and a $\lambda-$system. To see the converse, note that Proposition 2.4.2 gives us (i) containing $\mathcal{X}$ and (ii) closure under complements for free. Thus we need to prove that for any countable (not necessarily disjoint) collection of sets $\{E_i\}_{i\in\mathbb{N}} \in \mathcal{E}$

$$\bigcup_{i\in\mathbb{N}} E_i \in \mathcal{F}.$$

As we usually do, we shall look at the "disjointification"

$$\begin{aligned} F_i &:= E_i \setminus \bigcup_{j=1}^{i-1} E_j \\ &= E_i \cap \left(\bigcap_{j=1}^{i-1} E_j^C\right) \end{aligned}$$

and observe that closure under finite intersections (from being a $\pi-$system) and complements implies that $F_i \in \mathcal{F}$. Moreover, the F_i are pairwise disjoint, and so by closure under countable *disjoint* unions, we have that

$$\bigcup_{i\in\mathbb{N}} E_i = \bigcup_{i\in\mathbb{N}} F_i \in \mathcal{F}$$

which finishes the proof. $\square$

THEOREM 2.4.5 (Dynkin's $\pi-\lambda$ Theorem). *Let $\mathcal{E}\subseteq 2^{\mathcal{X}}$ be a $\pi-$system. Then*

$$\lambda(\mathcal{E})=\sigma(\mathcal{E}).$$

PROOF. We had shown earlier that $\lambda(\mathcal{E})\subseteq\sigma(\mathcal{E})$ and so we now we use the additional hypothesis of closure under finite intersections to show the reverse inclusion. By Lemma 2.4.4, we only need to show that $\lambda(\mathcal{E})$ inherits closure under finite intersections from $\mathcal{E}$, since then $\lambda(\mathcal{E})$ would be a $\sigma-$algebra that contains $\mathcal{E}$ and so $\sigma(\mathcal{E})\subseteq\lambda(\mathcal{E})$.

In order to show that $\lambda(\mathcal{E})$ is indeed closed under finite intersections, define for every $B\in\lambda(\mathcal{E})$

$$\mathcal{D}_B:=\{A\subseteq\mathcal{X}\mid A\cap B\in\lambda(\mathcal{E})\}.$$

It turns out that $\mathcal{D}_B$ is a $\lambda-$system for every $B\in\lambda(\mathcal{E})$. To see this, note that $\mathcal{X}\in\mathcal{D}_B$ since $\mathcal{X}\cap B=B\in\lambda(\mathcal{E})$. Moreover, for any sets $A_1,A_2\in\mathcal{D}_B$ such that $A_2\subseteq A_1$

$$(A_1\setminus A_2)\cap B=A_1\cap A_2^C\cap B=A_1\cap B\setminus A_2=A_1\cap B\setminus A_2\cap B\in\lambda(\mathcal{E})$$

since $A_1\cap B, A_2\cap B\in\lambda(\mathcal{E})$ and $A_2\cap B\subseteq A_1\cap B$ (see property (ii) in Definition 2.4.1). This shows that $A_1\setminus A_2\in\mathcal{D}_B$. Finally, let $\{A_i\}_{i\in\mathbb{N}}\in\mathcal{D}_B$ be an increasing sequence of sets, then

$$A_i\cap B\subseteq A_{i+1}\cap B$$

and so using property (iii) of $\lambda-$systems

$$\left(\bigcup_{i\in\mathbb{N}}A_i\right)\cap B=\bigcup_{i\in\mathbb{N}}(A_i\cap B)\in\lambda(\mathcal{E})$$

which shows that $\bigcup_{i\in\mathbb{N}}A_i\in\mathcal{D}_B$, proving our claim that $\mathcal{D}_B$ is a $\lambda-$system for any $B\in\lambda(\mathcal{E})$. In particular, for any $E\in\mathcal{E}$, $\mathcal{D}_E$ is a $\lambda-$system such that $\mathcal{E}\subseteq\mathcal{D}_E$ because of $\mathcal{E}$ is closed under finite intersections. Since $\mathcal{D}_E$ is a $\lambda-$system,

$$\lambda(\mathcal{E})\subseteq\mathcal{D}_E$$

for every $E\in\mathcal{E}$. Now we have proved that for any $B\in\lambda(\mathcal{E})$ and any $E\in\mathcal{E}$, their intersection $A\cap E\in\lambda(\mathcal{E})$, which seems like it is just a little bit short of what we need. But notice that this means $\mathcal{E}\subseteq\mathcal{D}_B$ for every $B\in\lambda(\mathcal{E})$, which implies that

$$\lambda(\mathcal{E})\subseteq\mathcal{D}_B$$

for every $B\in\lambda(\mathcal{E})$. Thus we have proved that for any $A,B\in\lambda(\mathcal{E}), A\cap B\in\lambda(\mathcal{E})$ which completes the proof. □

COROLLARY 2.4.6. *Let $\mathcal{D}\subseteq 2^{\mathcal{X}}$ be a $\lambda-$system and let $\mathcal{E}\subseteq\mathcal{D}$ be a $\pi-$system. Then*

$$\sigma(\mathcal{E})\subseteq\mathcal{D}.$$

PROOF. By the $\pi-\lambda$ theorem

$$\sigma(\mathcal{E})=\lambda(\mathcal{E})\subseteq\mathcal{D}$$

since $\mathcal{D}$ is a $\lambda-$system that contains $\mathcal{E}$. □

This corollary is useful because it adds structural assumptions on the generating class (by requiring that it be a $\pi-$system), which allows us to loosen the restriction that $\mathcal{D}$ be a $\sigma-$algebra like we used to assume in standard generating class arguments. This technique is powerful, and has important applications such as finding sufficient conditions to show that two measures on a $\sigma-$algebra are equal without having to explicitly check that equality holds on the entire $\sigma-$algebra.

THEOREM 2.4.7 (Uniqueness of Measures). *Let $(\mathcal{X},\mathcal{F})$ be a measurable space and let μ,ν be measures on $\mathcal{F}$. If there exists a $\pi-$system $\mathcal{E}\subseteq 2^{\mathcal{X}}$ such that $\sigma(\mathcal{E})=\mathcal{F}$ and*

$$\mu(E)=\nu(E)$$

for every $E\in\mathcal{E}$, and there exists increasing sequence $\{E_i\}_{i\in\mathbb{N}}\in\mathcal{E}$ such that $\mathcal{X}=\bigcup_{i\in\mathbb{N}}E_i$ and

$$\mu(E_i)=\nu(E_i)<\infty$$

for all $i \in \mathbb{N}$ *then*

$$\mu(F) = \nu(F)$$

for every $F \in \mathcal{F}$.

PROOF. Define

$$\mathcal{D}_i := \{F \in \mathcal{F} \mid \mu(F \cap E_i) = \nu(F \cap E_i)\}$$

and note that $\mathcal{X} \in \mathcal{D}_i$ for every $i \in \mathbb{N}$ since $\mu(E_i) = \nu(E_i)$. Next, let $F \in \mathcal{D}_i$ and observe that

$$\begin{aligned}
\mu\left(F^C \cap E_i\right) &= \mu(E_i \setminus (E_i \cap F)) \\
&= \mu(E_i) - \mu(E_i \cap F) \\
&= \nu(E_i) - \nu(E_i \cap F) \\
&= \nu(E_i \setminus (E_i \cap F)) \\
&= \nu\left(F^C \cap E_i\right)
\end{aligned}$$

where the second and fourth equalities follow from (finite) additivity of measures and the fact that $\mu(E_i \cap F) \leq \mu(E_i) < \infty$. This proves that for any $F \in \mathcal{D}_i$, $F^C \in \mathcal{D}_i$ for every $i \in \mathbb{N}$. Finally, let $\{F_j\}_{j \in \mathbb{N}}$be a pairwise disjoint collection of sets in $\mathcal{D}_i$. Then

$$\begin{aligned}
\mu\left(\left(\bigcup_{k \in \mathbb{N}} F_j\right) \cap E_i\right) &= \mu\left(\bigcup_{j \in \mathbb{N}} (F_j \cap E_i)\right) \\
&= \sum_{j=1}^{\infty} \mu(F_j \cap E_i) \\
&= \sum_{j=1}^{\infty} \nu(F_j \cap E_i) \\
&= \nu\left(\bigcup_{j \in \mathbb{N}} (F_j \cap E_i)\right) \\
&= \nu\left(\left(\bigcup_{k \in \mathbb{N}} F_j\right) \cap E_i\right)
\end{aligned}$$

where the second and fourth equality follow from countable additivity and the third equality follows from the uniqueness of limits. This proves that $\bigcup_{j \in \mathbb{N}} F_j \in \mathcal{D}_i$ for every $i \in \mathbb{N}$, which shows (through Proposition 2.4.2) that every $\mathcal{D}_i$ is a $\lambda-$system.

Now note that since $\mathcal{E}$ is a $\pi-$system (and so closed under finite intersections), we have that

$$\mathcal{E} \subseteq \mathcal{D}_i$$

for every $i \in \mathbb{N}$. By Corollary 2.4.6, we have that

$$\mathcal{F} = \sigma(\mathcal{E}) \subseteq \mathcal{D}_i$$

for every $i \in \mathbb{N}$. This means that for any $F \in \mathcal{F}$ and any $i \in \mathbb{N}$

$$\mu(F \cap E_i) = \nu(F \cap E_i).$$

Let $F \in \mathcal{F}$ be arbitrary and observe that since $\{E_i\}_{i \in \mathbb{N}} \nearrow \mathcal{X}$, $\{E_i \cap F\}_{i \in \mathbb{N}} \nearrow \mathcal{X} \cap F = F$. Then, by the continuity of measures,

$$\mu(F) = \lim_{i \to \infty} \mu(F \cap E_i) = \lim_{i \to \infty} \nu(F \cap E_i) = \nu(F),$$

again by the uniqueness of limits. $\square$

2.4.2. Monotone class theorem.

DEFINITION 2.4.8. A collection of sets $\mathcal{M} \subseteq 2^{\mathcal{X}}$ is called a *monotone class* if

(i) For an increasing sequence of sets $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{M}$: $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{M}$
(ii) For a decreasing sequence of sets $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{M}$: $\bigcap_{i\in\mathbb{N}} A_i \in \mathcal{M}$

Note that every $\sigma-$algebra is already a monotone class. However, we can make an even stronger claim: every $\lambda-$system is monotone class.

PROPOSITION 2.4.9. *Let $\mathcal{D} \subseteq 2^{\mathcal{X}}$ be a $\lambda-$system. Then $\mathcal{D}$ is a monotone class*

PROOF. Property (i) above follows directly from Definition 2.4.1. To see property (ii), let $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{D}$ be a decreasing sequence of sets and observe that

$$B_i = \mathcal{X} \setminus A_i \in \mathcal{D}$$

is an inccreasing sequence of sets such that

$$\bigcup_{i\in\mathbb{N}} B_i \in \mathcal{D}.$$

However,

$$\begin{aligned}\bigcap_{i\in\mathbb{N}} A_i &= \left(\bigcup_{i\in\mathbb{N}} A_i^C\right)^C \\ &= \left(\bigcup_{i\in\mathbb{N}} B_i\right)^C \\ &= \mathcal{X} \setminus \left(\bigcup_{i\in\mathbb{N}} B_i\right)\end{aligned}$$

which is in $\mathcal{D}$. This completes the proof. □

As usual, we define the monotone class generated by class of sets $\mathcal{A}$ as

$$\mathscr{M}(\mathcal{A}) := \bigcap \left\{\mathcal{M} \subseteq 2^{\mathcal{X}} \text{ is a monotone class} \mid \mathcal{A} \subseteq \mathcal{M}\right\}$$

which is itself a monotone class by the usual arguments. Again, since $\sigma-$algebras and $\lambda-$systems are monotone classes, we have that

$$\mathcal{A} \subseteq \mathscr{M}(\mathcal{A}) \subseteq \lambda(\mathcal{A}) \subseteq \sigma(\mathcal{A})$$

where we also use the fact that a $\sigma-$algebra is a $\lambda-$system. Finally, note that an *algebra* of sets $\mathcal{A}$ is also a $\pi-$system and so, by the $\pi - \lambda$ theorem,

$$\sigma(\mathcal{A}) = \lambda(\mathcal{A}).$$

THEOREM 2.4.10 (Monotone class theorem). *Let $\mathcal{A} \subseteq 2^{\mathcal{X}}$ be an algebra of sets. Then*

$$\mathscr{M}(\mathcal{A}) = \sigma(\mathcal{A}).$$

PROOF. We have shown that $\mathscr{M}(\mathcal{A}) \subseteq \sigma(\mathcal{A})$ and so we need to show the reverse inclusion. Since $\mathcal{A}$ is also a $\pi-$system, this is equivalent to showing that $\lambda(\mathcal{A}) \subseteq \mathscr{M}(\mathcal{A})$. To this end, it is sufficient to show that $\mathscr{M}(\mathcal{A})$ is a $\lambda-$system. Note that $\mathcal{X} \in \mathscr{M}(\mathcal{A})$ since $\mathcal{X} \in \mathcal{A}$. Further, $\mathscr{M}(\mathcal{A})$ is closed under limits of increasing sequences of sets since it's a monotone class. Thus we need to prove that for any $A_1, A_2 \in \mathscr{M}(\mathcal{A})$ such that $A_2 \subseteq A_1$ we have that $A_1 \setminus A_2 \in \mathscr{M}(\mathcal{A})$.

For any set $E \in \mathscr{M}(\mathcal{A})$, define

$$M_E := \{F \subseteq \mathscr{M}(\mathcal{A}) \mid E \setminus F, F \setminus E \in \mathscr{M}(\mathcal{A})\}$$

and note that for an increasing collection of sets $\{F_i\}_{i\in\mathbb{N}} \in M_E$

$$\begin{aligned} E \setminus \bigcup_{i\in\mathbb{N}} F_i &= E \cap \left(\bigcap_{i\in\mathbb{N}} F_i^C \right) \\ &= \bigcap_{i\in\mathbb{N}} \left(E \cap F_i^C \right) \\ &= \bigcap_{i\in\mathbb{N}} E \setminus F_i \in M_E \end{aligned}$$

since $E \setminus F_i \supseteq E \setminus F_{i+1}$ and $E \setminus F_i \in \mathscr{M}(\mathcal{A})$ for all $i \in \mathbb{N}$. A similar aragument shows that $\bigcup_{i\in\mathbb{N}} F_i \setminus E \in \mathscr{M}(\mathcal{A})$ which shows that $\bigcup_{i\in\mathbb{N}} F_i \in M_E$. We can apply the same argument to show that for a decreasing sequence of sets $\{F_i\}_{i\in\mathbb{N}} \in M_E$, the intersection $\bigcap_{i\in\mathbb{N}} F_i \in M_E$. This proves that M_E is in fact a monotone class, and since $\mathcal{A}$ is an algebra (and so it's closed under set differences), $\mathcal{A} \subseteq M_E$ which implies that $\mathscr{M}(\mathcal{A}) \subseteq M_E$ for any $E \in \mathscr{M}(\mathcal{A})$. In other words, this proves that for any $E, F \in \mathscr{M}(\mathcal{A})$:

$$E \setminus F, F \setminus E \in \mathscr{M}(\mathcal{A})$$

which completes the proof. □

Note how we used our previous work with the $\pi - \lambda$ theorem to establish the monotone class theorem; it turns out that the monotone class also implies theorem implies the $\pi - \lambda$ theorem, and so these two theorems are in fact equivalent. This should not be particularly surprising to you at this juncture since they are both generating class arguments which make slightly different structural assumptions to yield the same result. We can present these assumptions in a concise and organized fashion, using the notation defined in the abstract generating class argument described at the beginning of this section. In Table 1 $\mathcal{E}$ is the generating class which satisfies the desired property $(*)$ and $\mathcal{D}$ is all the sets in the $\sigma-$algebra $\mathcal{F}$ which satisfies property $(*)$. In the generic argument, we put very strong structural assumptions on $\mathcal{D}$ and none on $\mathcal{E}$. In the $\pi - \lambda$ case, we weaken the assumptions on $\mathcal{D}$ in exchange for imposing (weak) assumptions on $\mathcal{E}$. Finally, in the monotone class approach, we impose a strong structural assumption on $\mathcal{E}$, whereas $\mathcal{D}$ has the minimal structure of a monotone class. The type of argument one uses in practice depends on context, and in many situations one can pick to use either Dynkin's theorem or the monotone class theorem.

Table 1. Types of generating class arguments

Name	Structure of $\mathcal{E}$	Structure of $\mathcal{D}$
Generic	No structure	$\sigma-$algebra
Dynkin's theorem	$\pi-$system	$\lambda-$system
Monotone class theorem	Algebra	Monotone class

2.4.3. Generating classes of functions.

Definition 2.4.11. A set $\mathcal{H}$ of bounded real functions on $\mathcal{X}$ is called a $\lambda-$*space of functions* if

(i) The constant function $\mathbb{1}_{\mathcal{X}} \in \mathcal{H}$
(ii) $\mathcal{H}$ is a vector space over $\mathbb{R}$
(iii) If $h_n \in \mathcal{H}$ such that $h_n \leq h_{n+1}$ pointwise and

$$\lim_{n\to\infty} h_n = h$$

is well defined and bounded then $h \in \mathcal{H}$

$\lambda-$spaces of functions are named so because they generalize $\lambda-$systems of sets, as we show in the following result.

Proposition 2.4.12. *If $\mathcal{H}$ is a $\lambda-$space of functions on $\mathcal{X}$ then*

$$\mathcal{D} := \{A \subseteq \mathcal{X} \mid \mathbb{1}_A \in \mathcal{H}\}$$

is a $\lambda-$system.

Proof. Note that since $\mathbb{1}_{\mathcal{X}} \in \mathcal{H}$ we have that $\mathcal{X} \in \mathcal{D}$. Next, assume that $A_1, A_2 \in \mathcal{D}$ such that $A_2 \subseteq A_1$ and observe that

$$\begin{aligned} \mathbb{1}_{A_1 \setminus A_2} &= \mathbb{1}_{A_1 \cap A_2^C} \\ &= \mathbb{1}_{A_1}(1 - \mathbb{1}_{A_2}) \\ &= \mathbb{1}_{A_1} - \mathbb{1}_{A_2} \in \mathcal{H} \end{aligned}$$

where the second and third equalities follows from Fact 2.1.2, and the inclusion follows from the fact that $\mathcal{H}$ is closed under linear combinations. This prove that $A_1 \setminus A_2 \in \mathcal{H}$. Fiinally, let $\{A_i\}_{i\in\mathbb{N}}$ be an increasing sequence of sets in $\mathcal{D}$ and observe that

$$\begin{aligned} \mathbb{1}_{\bigcup_{i\in\mathbb{N}} A_i} &= \sup_{i\in\mathbb{N}} \mathbb{1}_{A_i} \\ &= \lim_{n\to\infty} \mathbb{1}_{A_i} \in \mathcal{H} \end{aligned}$$

where the first equality follows from 2.1.3 and the second equality is due to the fact that $A \subseteq B \Longrightarrow \mathbb{1}_A \leq \mathbb{1}_B$, and so $\mathbb{1}_{A_i}$ converges pointwise to its supremum. This proves that $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{D}$ which implies that $\mathcal{D}$ is a $\lambda-$system. $\square$

In the next theorem and elsewhere, let $\mathcal{M}_{\text{bdd}}(\mathcal{X}, \mathcal{F})$ and $\mathcal{M}^+_{\text{bdd}}(\mathcal{X}, \mathcal{F})$ denote the set of all bounded real-valued Borel-measurable functions on $(\mathcal{X}, \mathcal{F})$,and the set of all non-negative and bounded real-valued Borel-measurable functions on $(\mathcal{X}, \mathcal{F})$ respectively.

Theorem 2.4.13 ($\pi - \lambda$ theorem for functions). *Let $\mathcal{G}$ be a $\pi-$system of sets and let $\mathcal{H}$ be a $\lambda-$space of functions on $\mathcal{X}$. If*

$$\{\mathbb{1}_A \mid A \in \mathcal{G}\} \subseteq \mathcal{H}$$

then $\mathcal{M}_{\text{bdd}}(\mathcal{X}, \sigma(\mathcal{G})) \subseteq \mathcal{H}$.

Proof. Define $\mathcal{D} := \{D \subseteq \mathcal{X} \mid \mathbb{1}_D \in \mathcal{H}\}$. Then, by Proposition 2.4.12, $\mathcal{D}$ is a $\lambda-$system such that $\mathcal{G} \subseteq \mathcal{D}$. By the $\pi - \lambda$ theorem for sets, $\sigma(\mathcal{G}) \subseteq \mathcal{D}$. This means that the indicator functions of sets in $\sigma(\mathcal{G})$ is contained in $\mathcal{H}$.

Next, let $f \in \mathcal{M}^+_{\text{bdd}}(\mathcal{X}, \sigma(\mathcal{G}))$ be arbitrary. Since $f \in \mathcal{M}^+(\mathcal{X}, \sigma(\mathcal{G}))$, by Proposition 2.3.2 there exists an increasing sequence of simple functions $\{s_n\}_{n\in\mathbb{N}} \in \mathcal{M}^+(\mathcal{X}, \sigma(\mathcal{G}))$ such that

$$f = \lim_{n\to\infty} s_n.$$

Since, f is bounded, $s_n \leq f$ are all bounded. Moreover, since every s_n is a measurable simple function, it's a finite linear combination of indicator functions of sets in $\sigma(\mathcal{G})$. Therefore, $\{s_n\}_{n\in\mathbb{N}} \in \mathcal{H}$ and since f is the monotone limit of s_n, $f \in \mathcal{H}$.

Finally, assume that $f \in \mathcal{M}_{\text{bdd}}(\mathcal{X}, \sigma(\mathcal{G}))$ and note that $f^+, f^- \in \mathcal{H}$ by Proposition 2.3.3 and the last step, and so $f = f^+ - f^- \in \mathcal{H}$. $\square$

CHAPTER 3

Integration

3.1. Constructing the Lebesgue integral

THEOREM 3.1.1. *Let $(\mathcal{X}, \mathcal{F})$ be measurable space. For any measure μ on $\mathcal{F}$, there exists a unique linear functional*

$$\bar{\mu} : \mathcal{M}^+(\mathcal{X}, \mathcal{F}) \longrightarrow [0, \infty]$$

that satisfies the following properties

(i) $\bar{\mu}(\mathbb{1}_A) = \mu(A)$ *for any* $A \in \mathcal{F}$

(ii) (Linearity) *For any* $f, g \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ *and any* $\alpha, \beta \geq 0$

$$\bar{\mu}(\alpha f + \beta g) = \alpha \bar{\mu}(f) + \beta \bar{\mu}(g)$$

(iii) (Monotone convergence) *For a sequence of increasing functions* $\{f_n\}_{n \in \mathbb{N}} \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$

$$\bar{\mu}\left(\lim_{n \to \infty} f_n\right) = \lim_{n \to \infty} \bar{\mu}(f_n)$$

We can prove Theorem 3.1.1 constructively, in a similar fashion to how we proved the existence of the Lebesgue measure in Chapter 1. In this spirit, we shall define a functional on the non-negative measurable simple functions and extend the domain of this functional to more complicated function spaces.

DEFINITION 3.1.2. For any measurable simple function $s \in \mathcal{M}_{\text{sim}}(\mathcal{X}, \mathcal{F})$ with the standard representation

$$s = \sum_{i=1}^{I} \alpha_i \mathbb{1}_{A_i}$$

define, for any measure μ on $\mathcal{F}$, the functional $\bar{\mu}_0$ as

$$\bar{\mu}_0(s) := \sum_{i=1}^{I} \alpha_i \mu(A_i).$$

Immediately, we can see that our proto-integral $\bar{\mu}_0$ behaves quite nicely: it is always non-negative and since indicator functions are special cases of simple functions, we have that

$$\bar{\mu}_0(\mathbb{1}_A) = \mu(A)$$

for any $A \in \mathcal{F}$. Moreover, our functional satisfies the linearity property on the space of simple measurable functions. To see this we shall need the following lemma which solves a minor technical issue that arises due to our definition of $\bar{\mu}$ relying on the standard representation of simple functions.

LEMMA 3.1.3. *For any non-negative measurable simple function* $s \in \mathcal{M}_{\text{sim}}(\mathcal{X}, \mathcal{F})$ *with a standard representation*

$$s = \sum_{i=1}^{I} \alpha_i \mathbb{1}_{A_i}$$

and another representation

$$s=\sum_{j=1}^{J}\beta_i\mathbb{1}_{B_j}$$

where $\{B_j\}_{j=1}^{J}$ *is a partition of* $\mathcal{X}$ *we have*

$$\bar{\mu}_0(s):=\sum_{i=1}^{I}\alpha_i\mu(A_i)=\sum_{j=1}^{J}\beta_j\mu(B_j).$$

where $\bar{\mu}_0$ *is the functional derived from a measure* μ *as in Definition 3.1.2.*

Proof. Note that both A_i and B_j partition $\mathcal{X}$ and so observe that

$$\begin{aligned}\sum_{i=1}^{I}\alpha_i\mu(A_i)&=\sum_{i=1}^{I}\alpha_i\mu\left(\bigcup_{j=1}^{J}(A_i\cap B_j)\right)\\&=\sum_{i=1}^{I}\sum_{j=1}^{J}\alpha_i\mu(A_i\cap B_j)\end{aligned}$$

where the last equality follows by finite addivity. Similarly,

$$\sum_{j=1}^{J}\beta_j\mu(B_j)=\sum_{i=1}^{I}\sum_{j=1}^{J}\beta_j\mu(A_i\cap B_j).$$

Now observe that since $\sum_{i=1}^{I}\alpha_i\mathbb{1}_{A_i}=\sum_{j=1}^{J}\beta_i\mathbb{1}_{B_j}$ we know that $\alpha_i=\beta_j$ if $A_i\cap B_j\neq\emptyset$; conversely, if $A_i\cap B_j=\emptyset$ then $\mu(A_i\cap B_j)=0$ and so we have that

$$\sum_{i=1}^{I}\alpha_i\mu(A_i)=\sum_{i=1}^{I}\sum_{j=1}^{J}\alpha_i\mu(A_i\cap B_j)=\sum_{i=1}^{I}\sum_{j=1}^{J}\beta_j\mu(A_i\cap B_j)=\sum_{j=1}^{J}\beta_j\mu(B_j).$$

□

Proposition 3.1.4. *Let* $s,t\in\mathcal{M}_{\text{sim}}(\mathcal{X},\mathcal{F})$ *and let* μ*be a measure on* $\mathcal{F}$*. Then, for any* $\alpha,\beta\geq 0$*, we have that*

$$\bar{\mu}_0(\alpha s+\beta t)=\alpha\bar{\mu}_0(s)+\beta\bar{\mu}_0(t).$$

Proof. It is sufficient to prove that $\bar{\mu}_0(\alpha s)=\alpha\bar{\mu}_0(s)$ and $\bar{\mu}_0(s+t)=\bar{\mu}_0(s)+\bar{\mu}_0(t)$. To show the first equality, simply notice that if s is given by the standard representation

$$s=\sum_{i=1}^{I}a_i\mathbb{1}_{A_i}$$

then the standard representation of αs when $\alpha>0$ is simply

$$\alpha s=\sum_{i=1}^{I}\alpha a_i\mathbb{1}_{A_i}$$

and so

$$\begin{aligned}\bar{\mu}_0(\alpha s)&=\sum_{i=1}^{I}\alpha a_i\mu(A_i)\\&=\alpha\sum_{i=1}^{I}a_i\mu(A_i)\\&=\alpha\bar{\mu}_0(s).\end{aligned}$$

When $\alpha=0$, $\alpha s=\mathbb{1}_{\emptyset}$ and so $\bar{\mu}_0(\alpha s)=\mu(\emptyset)=0=\alpha\bar{\mu}_0(s)$.

In order to prove the second equality, observe that if the standard representations of s and t are given by

$$s = \sum_{i=1}^{I} a_i \mathbb{1}_{A_i}$$
$$t = \sum_{j=1}^{J} b_j \mathbb{1}_{B_j}$$

then

$$\begin{aligned}
\bar{\mu}_0 (s + t) &= \bar{\mu}_0 \left(\sum_{i=1}^{I} \sum_{j=1}^{J} (a_i + b_j) \mathbb{1}_{A_i \cap B_j} \right) \\
&= \sum_{i=1}^{I} \sum_{j=1}^{J} (a_i + b_j) \mu (A_i \cap B_j) \\
&= \sum_{i=1}^{I} a_i \sum_{j=1}^{J} \mu (A_i \cap B_j) + \sum_{j=1}^{J} b_j \sum_{i=1}^{I} \mu (A_i \cap B_j) \\
&= \sum_{i=1}^{I} a_i \mu (A_i) + \sum_{j=1}^{J} b_j \mu (B_j) \\
&= \bar{\mu}_0 (s) + \bar{\mu}_0 (t)
\end{aligned}$$

where the first equality follows from Proposition 2.3.4, the second equality due to Lemma 3.1.3, and the fourth equality due to finite additivity of μ. This completes the proof. □

Linearity and non-negativity of $\bar{\mu}_0$ tells us that for any simple measurable functions $f \leq g$, we have that

$$\bar{\mu}_0 (f) \leq \bar{\mu}_0 (g) .$$

Indeed, we can decompose $g = (g - f) + f$ where $g - f$ and f are both non-negative simple functions and so

$$\bar{\mu}_0 (g) = \bar{\mu}_0 (g - f) + \bar{\mu}_0 (f)$$

which by non-negativity proves our claim. In other words, our functional $\bar{\mu}_0$ is an *increasing* or *monotone* functional. With this final fact, we are now ready to extend $\bar{\mu}_0$ to the space of non-negative measurable functions.

Proof of Theorem 3.1.1. For any $f \in \mathcal{M}^+ (\mathcal{X}, \mathcal{F})$, define

$$\bar{\mu} (f) := \sup \{ \bar{\mu}_0 (s) \mid s \in \mathcal{M}_{\text{sim}} (\mathcal{X}, \mathcal{F}) \text{ such that } s \leq f \} .$$

We claim that $\bar{\mu}$ is the unique functional (for a given measure μ) that satisfies the properties described in the statement of the theorem. To see this, we first have to show that the functional defined above indeed satisfes the three requisite properties and then show that it is the only such functional. Before we do this, note that the set $\{ \bar{\mu}_0 (s) \mid s \in \mathcal{M}_{\text{sim}} (\mathcal{X}, \mathcal{F}) \text{s.t } s \leq f \}$ is always non-empty, thanks to Proposition 2.3.2. Therefore the supremum is at least 0 since $\bar{\mu}_0 (s) \geq 0$. Moreover, since for any non-negative measurable functions $f \leq g$, $\{ \bar{\mu}_0 (s) \mid s \in \mathcal{M}_{\text{sim}} (\mathcal{X}, \mathcal{F}) \text{ such that } s \leq f \} \subseteq \{ \bar{\mu}_0 (s) \mid s \in \mathcal{M}_{\text{sim}} (\mathcal{X}, \mathcal{F}) \text{ such that } s \leq g \}$,we can conclude that $\bar{\mu} (f) \leq \bar{\mu} (g)$. Thus we already know that our functional $\bar{\mu}$ is both non-negative and monotone. Next, to show property (i), observe that for any simple function $t \in \mathcal{M}_{\text{sim}} (\mathcal{X}, \mathcal{F})$

$$\begin{aligned}
\bar{\mu} (t) &= \sup \{ \bar{\mu}_0 (s) \mid s \in \mathcal{M}_{\text{sim}} (\mathcal{X}, \mathcal{F}) \text{ such that } s \leq t \} \\
&= \bar{\mu}_0 (t)
\end{aligned}$$

where the second equality follows from the monotonicty of $\bar{\mu}_0$. Property (i) then follows from letting $t = \mathbb{1}_A$ for any set $A \in \mathcal{F}$. Next, we prove property (iii) which is monotone convergence.

Let $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}^+(\mathcal{X},\mathcal{F})$ an increasing sequence of measurable functions. Since we are working with the extended non-negative real numbers, we know that

$$f := \lim_{n\to\infty} f_n = \sup_{n\in\mathbb{N}} f_n \in \mathcal{M}^+(\mathcal{X},\mathcal{F})$$

by Proposition 2.2.9. Note that by the monotonicity of $\bar{\mu}$, $\bar{\mu}(f_n) \leq \bar{\mu}(f)$ for all $n \in \mathbb{N}$ and so

$$\lim_{n\to\infty} \bar{\mu}(f_n) \leq \bar{\mu}(f).$$

To deduce the reverse inequality, let $s \leq f$ be a non-negative measurable simple function, fix some $t \in (0,1)$ and define $A_n = \{x \in \mathcal{X} \mid f_n(x) \geq ts(x)\}$. Since $f_n \leq f_{n+1}$, we have that $A_n \subseteq A_{n+1}$. Moreover, for any $x \in \mathcal{X}$, if $f(x) > 0$ then there is some $\epsilon_x > 0$ such that $f(x) - ts(x) = 2\epsilon_x$ and so, by pointwise convergence, there exists some $n_{\epsilon_x} \in \mathbb{N}$ such that

$$f_n(x) - ts(x) = \underbrace{(f(x) - ts(x))}_{2\epsilon_x} - \underbrace{(f(x) - f_n(x))}_{\leq \epsilon_x} \geq \epsilon_x > 0$$

for all $n \geq n_{\epsilon_x}$, proving that $x \in A_n$ for such n. Conversely, if $f(x) = 0$ then $f(x) = f_n(x) = ts(x) = 0$ and so $x \in A_n$ for every $n \in \mathbb{N}$. Together, these two cases show that $\bigcup_{n\in\mathbb{N}} A_n = \mathcal{X}$. Now, by the montonicity of $\bar{\mu}$, observe that

$$\begin{aligned}
\bar{\mu}(f_n) &\geq \bar{\mu}(f_n \mathbb{1}_{A_n}) \\
&\geq \bar{\mu}(ts\mathbb{1}_{A_n}) \\
&= \bar{\mu}_0(ts\mathbb{1}_{A_n}) \\
&= t\bar{\mu}_0(s\mathbb{1}_{A_n}) \\
&= t\bar{\mu}(s\mathbb{1}_{A_n}) \qquad (3)
\end{aligned}$$

where the third and last equalities follow from the result above that $\bar{\mu}$ extends $\bar{\mu}_0$ and the fourth equality is due to the linearity of $\bar{\mu}_0$. Let $s = \sum_{i=1}^{J} b_i \mathbb{1}_{B_i}$ then by Lemma 3.1.3

$$\bar{\mu}(s\mathbb{1}_{A_n}) = \sum_{i=1}^{I} b_i \mu(B_i \cap A_n).$$

Since $A_n \subseteq A_{n+1} \implies B_i \cap A_n \subseteq B_i \cap A_{n+1}$, by continuity from below of measures, we have that

$$\begin{aligned}
\lim_{n\to\infty} \bar{\mu}(s\mathbb{1}_{A_n}) &= \lim_{n\to\infty} \sum_{i=1}^{I} b_i \mu(B_i \cap A_n) \\
&= \sum_{i=1}^{I} b_i \mu\left(\bigcup_{n\in\mathbb{N}} (B_i \cap A_n)\right) \\
&= \sum_{i=1}^{I} b_i \mu(B_i) \\
&= \mu(s). \qquad (4)
\end{aligned}$$

where the third equality follows from the fact that $\bigcup_{n\in\mathbb{N}} A_n = \mathcal{X}$. Together, (3) and (4) imply that

$$\lim_{n\to\infty} \bar{\mu}(f_n) \geq t\bar{\mu}(s).$$

Since this is true for any $t \in (0,1)$, we have that

$$\lim_{n\to\infty} \bar{\mu}(f_n) \geq \bar{\mu}(s).$$

Finally, since $s \leq f$ was an arbitrary simple function, we have that

$$\lim_{n\to\infty} \bar{\mu}(f_n) \geq \sup\{\bar{\mu}_0(s) \mid s \in \mathcal{M}_{\text{sim}}(\mathcal{X},\mathcal{F}) \text{ such that } s \leq f\} = \bar{\mu}(f).$$

Next, we show that $\bar{\mu}$ is a linear functional. As before, it is sufficient to show that for any $f, g \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ and any $\alpha \geq 0$,

$$\begin{aligned}\bar{\mu}(\alpha f) &= \alpha \bar{\mu}(f) \\ \bar{\mu}(f+g) &= \bar{\mu}(f) + \bar{\mu}(g).\end{aligned}$$

First we shall show the homogenous scaling property. If $\alpha = 0$ then the proof is trivial; if $\alpha > 0$ then by the monotone convergence property

$$\bar{\mu}(\alpha f) = \lim_{n \to \infty} \bar{\mu}(s_n)$$

where $\{s_n\}_{n \in \mathbb{N}}$ is an increasing sequence of simple functions such that $s_n \nearrow \alpha f$. Then, $h_n = \frac{s_n}{\alpha}$ is an increasing sequence of simple functions such that $h_n \nearrow f$ and so

$$\begin{aligned}\bar{\mu}(f) &= \lim_{n \to \infty} \bar{\mu}(h_n) \\ &= \lim_{n \to \infty} \frac{1}{\alpha} \bar{\mu}_0(s_n) \\ &= \frac{1}{\alpha} \bar{\mu}(\alpha f).\end{aligned}$$

Rearranging yields the proof. Now let $\{s_n\}_{n \in \mathbb{N}}, \{t_n\}_{n \in \mathbb{N}}$ be increasing sequences of measurable simple functions such that $s_n \nearrow f, t_n \nearrow g$. By the linearity of limits, we have that $s_n + t_n \nearrow f + g$ and so

$$\begin{aligned}\bar{\mu}(f+g) &= \lim_{n \to \infty} \bar{\mu}(s_n + t_n) \\ &= \lim_{n \to \infty} \bar{\mu}_0(s_n + t_n) \\ &= \lim_{n \to \infty} \bar{\mu}_0(s_n) + \lim_{n \to \infty} \bar{\mu}_0(t_n) \\ &= \bar{\mu}(f) + \bar{\mu}(g)\end{aligned}$$

where the first equality follows from the monotone convergence property, the second equality follows from the fact that measurable simple functions are closed under addition and the fact that $\bar{\mu}$ extends $\bar{\mu}_0$, the third equality due to the linearity of $\bar{\mu}_0$, and finally the fourth equality due to a second application of monotone convergence.

To show that our functional $\bar{\mu}$ is unique, suppose that there were two functionals $\bar{\mu}_1, \bar{\mu}_2$ that satisfied properties (i)-(iii) with respect to some measure μ. Then, by property (i), $\bar{\mu}_1(\mathbb{1}_A) = \mu(A) = \bar{\mu}_2(\mathbb{1}_A)$ for any measurable set $A \in \mathcal{F}$. Next, by the linearity property along with equality on indicator functions, $\bar{\mu}_1(s) = \bar{\mu}_2(s)$ for any $s \in \mathcal{M}_{\text{sim}}(\mathcal{X}, \mathcal{F})$. Finally, for any arbitrary $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ and $s_n \in \mathcal{M}_{\text{sim}}(\mathcal{X}, \mathcal{F})$ such that $s_n \nearrow f$, observe that

$$\begin{aligned}\bar{\mu}_1(f) &= \lim_{n \to \infty} \bar{\mu}_1(s_n) \\ &= \lim_{n \to \infty} \bar{\mu}_2(s_n) \\ &= \bar{\mu}_2(f)\end{aligned}$$

where the first equality uses the monotone convergence propert of $\bar{\mu}_1$, the second equality uses the fact that our two functionals are equal on simple functions, and the last equality uses the monotone convergence property of $\bar{\mu}_2$. This completes the proof. $\square$

Definition 3.1.5. A function $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ is called *integrable* with respect to a measure μ(or $\mu-$*integrable* in short) if

$$\bar{\mu}(|f|) < \infty.$$

The collection of all such functions is denoted $\mathcal{L}^1(\mathcal{X}, \mathcal{F}, \mu)$.

When the underlying space and $\sigma-$algebra are clear, we shall simply write $\mathcal{L}^1(\mu)$. The significance of the exponent 1 will become clear in the next chapter.

Proposition 3.1.6. *A function $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ is in $\mathcal{L}^1(\mu)$ if and only if*

$$\bar{\mu}(f^+), \bar{\mu}(f^-) < \infty.$$

PROOF. Note that if $\bar{\mu}(|f|) < \infty$ then

$$\begin{aligned}\bar{\mu}(|f|) &= \bar{\mu}\left(f^{+} + f^{-}\right) \\ &= \bar{\mu}\left(f^{+}\right) + \bar{\mu}\left(f^{-}\right) \\ &< \infty\end{aligned}$$

which would imply that $\bar{\mu}(f^{+}), \bar{\mu}(f^{-}) < \infty$. The converse follows similarly. □

PROPOSITION 3.1.7. *For any measure space $(\mathcal{X}, \mathcal{F}, \mu)$, the space of $\mu-$integrable functions $\mathcal{L}^1(\mu)$ is a vector space over $\mathbb{R}$.*

PROOF. First observe that the zero function $\mathbb{1}_{\emptyset} \in \mathcal{L}^1(\mu)$since $\bar{\mu}(|\mathbb{1}_{\emptyset}|) = \mu(\emptyset) = 0 < \infty$ and so we have an additive identity. Next, notice that $f \in \mathcal{L}^1(\mu) \iff -f \in \mathcal{L}^1(\mu)$ and so we have additive inverses for each function. Commutativity and associativity of addition follow from the definition of addition on spaces of functions, as does distributivity of scalar multiplication over addition. Finally, notice that for any $\alpha \in \mathbb{R}$ and any functions $f, g \in \mathcal{L}^1(\mu)$

$$\begin{aligned}\bar{\mu}(|\alpha f + g|) &\leq \bar{\mu}(|\alpha f| + |g|) \\ &= |\alpha|\,\bar{\mu}(|f|) + \bar{\mu}(|g|) \\ &< \infty\end{aligned}$$

where the first inequality follows from the triangle inequality of $|\cdot|$ and monotonicty of $\bar{\mu}$, and the equality follows from the linearity of $\bar{\mu}$. This proves that $\mathcal{L}^1(\mu)$ is closed under finite linear combinations and so is a vector space. □

DEFINITION 3.1.8. For any function $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ and a measure μ on $\mathcal{F}$, we can define the *Lebesgue integral*

$$\tilde{\mu}(f) := \bar{\mu}\left(f^{+}\right) - \bar{\mu}\left(f^{-}\right)$$

wherever the difference is defined (i.e. at least one of $\bar{\mu}(f^{+})$ and $\bar{\mu}(f^{-})$ is finite).

REMARK. Note that for any non-negative measurable function f, $\tilde{\mu}(f) = \bar{\mu}(f)$ and so $\tilde{\mu}$ extends $\bar{\mu}$ just like $\bar{\mu}$ extended $\bar{\mu}_0$

3.2. Properties of the integral

PROPOSITION 3.2.1. *For any functions $f, g \in \mathcal{L}^1(\mathcal{X}, \mathcal{F}, \mu)$ and scalar $\alpha \in \mathbb{R}$, we have that*

$$\tilde{\mu}(\alpha f + g) = \alpha\tilde{\mu}(f) + \tilde{\mu}(g).$$

PROOF. First we shall prove that $\tilde{\mu}(\alpha f) = \alpha\tilde{\mu}(f)$. Note that if $\alpha = 0$ then the equality follows trivially. If $\alpha > 0$ then

$$\begin{aligned}\tilde{\mu}(\alpha f) &= \bar{\mu}\left((\alpha f)^{+}\right) - \bar{\mu}\left((\alpha f)^{-}\right) \\ &= \bar{\mu}\left(\alpha f \mathbb{1}_{\{\alpha f > 0\}}\right) - \bar{\mu}\left(-\alpha f \mathbb{1}_{\{\alpha f < 0\}}\right) \\ &= \bar{\mu}\left(\alpha f \mathbb{1}_{\{f > 0\}}\right) - \bar{\mu}\left(-\alpha f \mathbb{1}_{\{f < 0\}}\right) \\ &= \alpha\bar{\mu}\left(f^{+}\right) - \alpha\bar{\mu}\left(f^{-}\right) \\ &= \alpha\tilde{\mu}(f)\end{aligned}$$

where the third equality follows from the fact that $\alpha > 0$ and the fourth equality due to the linearity of $\bar{\mu}$. Finally, if $\alpha < 0$ then

$$\begin{aligned}\tilde{\mu}(\alpha f) &= \bar{\mu}\left((\alpha f)^{+}\right) - \bar{\mu}\left((\alpha f)^{-}\right)\\ &= \bar{\mu}\left(\alpha f \mathbb{1}_{\{\alpha f>0\}}\right) - \bar{\mu}\left(-\alpha f \mathbb{1}_{\{\alpha f<0\}}\right)\\ &= \bar{\mu}\left(\alpha f \mathbb{1}_{\{f<0\}}\right) - \bar{\mu}\left(-\alpha f \mathbb{1}_{\{f>0\}}\right)\\ &= \bar{\mu}\left(-\alpha \times -f \mathbb{1}_{\{f<0\}}\right) - \bar{\mu}\left(-\alpha f \mathbb{1}_{\{f>0\}}\right)\\ &= \alpha\bar{\mu}\left(f^{+}\right) - \alpha\bar{\mu}\left(f^{-}\right)\\ &= \alpha\tilde{\mu}(f).\end{aligned}$$

Next, let $h = f + g$ and observe that

$$\begin{aligned}h &= h^{+} - h^{-}\\ &= f + g\\ &= \left(f^{+} - f^{-}\right) + \left(g^{+} - g^{-}\right).\end{aligned}$$

Rearranging, we have that

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

where the functions on each side are non-negative measurable functions and so

$$\bar{\mu}\left(h^{+} + f^{-} + g^{-}\right) = \bar{\mu}\left(h^{+}\right) + \bar{\mu}\left(f^{-}\right) + \bar{\mu}\left(g^{-}\right) = \bar{\mu}\left(h^{-}\right) + \bar{\mu}\left(f^{+}\right) + \bar{\mu}\left(g^{+}\right) = \bar{\mu}\left(h^{-} + f^{+} + g^{-}\right)$$

by linearity of $\bar{\mu}$. As $h, g, f \in \mathcal{L}^1(\mu)$, the integrals of the individual components are not infinite and so we can rearrange the second equality above as

$$\underbrace{\bar{\mu}\left(h^{+}\right) - \bar{\mu}\left(h^{-}\right)}_{\tilde{\mu}(h)} = \underbrace{\bar{\mu}\left(f^{+}\right) - \bar{\mu}\left(f^{-}\right)}_{\tilde{\mu}(f)} + \underbrace{\bar{\mu}\left(g^{+}\right) - \bar{\mu}\left(g^{-}\right)}_{\tilde{\mu}(g)}$$

which completes the proof. □

Corollary 3.2.2. *For any functions $f, g \in \mathcal{L}^1(\mathcal{X}, \mathcal{F}, \mu)$ such that $f \leq g$ pointwise,*

$$\tilde{\mu}(f) \leq \tilde{\mu}(g).$$

Proof. Note that $h = g - f \geq 0$ and so

$$\tilde{\mu}(g - f) = \bar{\mu}(g - f) \geq 0.$$

Then, by Proposition 3.2.1,

$$\tilde{\mu}(g - f) = \tilde{\mu}(g) - \tilde{\mu}(f) \geq 0$$

which completes the proof. □

Corollary 3.2.3. *For any function $f \in \mathcal{L}^1(\mathcal{X}, \mathcal{F}, \mu)$*

$$|\tilde{\mu}(f)| \leq \tilde{\mu}(|f|).$$

Proof. By Corollary 3.2.2 and the triangle inequality for $|\cdot|$

$$\tilde{\mu}(f) \leq \tilde{\mu}(|f|).$$

Similarly, we have that

$$-f \leq |f| \implies \tilde{\mu}(-f) \leq \tilde{\mu}(|f|) \implies \tilde{\mu}(f) \geq -\tilde{\mu}(|f|)$$

where the second implication follows due to linearity. Together, the two inequalities imply that

$$|\tilde{\mu}(f)| \leq \tilde{\mu}(|f|)$$

which is the result. □

3.2.1. Interchanging limits and integrals. At this point we have defined the functionals $\bar{\mu}_0, \bar{\mu}$,and $\tilde{\mu}$ to operate on simple functions, non-negative measurable functions, and all measurable functions respectively. Since $\bar{\mu}$ extends $\bar{\mu}_0$ and $\tilde{\mu}$ extends $\bar{\mu}$, we can dispense with the unnecessary amounts of new notation and simply denote the integral as $\bar{\mu}(f)$ for any function $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$, provided that the integral is defined. As discussed earlier, this is analagous to our construction of the Lebesgue measure in Chapter 1, where we "extended" the measure from simple sets to more complicated sets over the course of the chapter.

Next we show the power of this integration theory by establishing two results which allow us to interchange pointwise limits of functions and with their integrals. We have already proved the following result in Theorem 3.1.1.

THEOREM 3.2.4 (Monotone convergence theorem). *Let $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ be such that $f_n \leq f_{n+1}$ pointwise. Then, for any meaasure μon $\mathcal{F}$,*

$$\lim_{n\to\infty} \bar{\mu}(f_n) = \bar{\mu}\left(\lim_{n\to\infty} f_n\right).$$

The other theorem drops the requirement that the functions f_n be non-negative or that the sequence be monotone, in exchange for asking the sequence f_n to be uniformly bounded by an integrable function.

THEOREM 3.2.5 (Dominated convergence theorem). *Let $\{f_n\}_{n\in\mathbb{N}}$ be a sequence of functions in $\mathcal{L}^1(\mathcal{X}, \mathcal{F}, \mu)$ and suppose there exists some $g \in \mathcal{L}^1(\mu)$ such that*

$$|f_n| \leq g$$

pointwise for every $n \in \mathbb{N}$. Then, if $f := \lim_{n\to\infty} f_n$ is defined, we have that

$$\lim_{n\to\infty} \bar{\mu}(|f_n - f|) = 0$$

which also shows that

$$\lim_{n\to\infty} \bar{\mu}(f_n) = \bar{\mu}(f).$$

PROOF. Note that by the continuity of the absolute value function, $|f_n| \leq g \Longrightarrow |f| \leq g$ pointwise which shows $f \in \mathcal{L}^1(\mu)$. Then, by the triangle inequality

$$|f_n(x) - f(x)| \leq |f_n(x)| + |f(x)| \leq 2g(x)$$

for every $n \in \mathbb{N}$ and every $x \in \mathcal{X}$. Note that

$$\phi_i(x) := \sup_{n\geq i} |f_n(x) - f(x)| \leq 2g(x)$$

for every $i \in \mathbb{N}$ and every $x \in \mathcal{X}$, which proves that $\phi_i \in \mathcal{L}^1(\mu)$ for every $i \in \mathbb{N}$. Next, observe that

$$\begin{aligned}\lim_{i\to\infty} \phi_i &= \limsup_{n\to\infty} |f_n - f| \\ &= \lim_{n\to\infty} |f_n - f| \\ &= 0\end{aligned}$$

where the second and third equalities follows from the assumption that $f_n \longrightarrow f$ pointwise. Further, define

$$\psi_i(x) := 2g(x) - \phi_i(x) \geq 0$$

and observe that since $\phi_i \geq \phi_{i+1}$ pointwise for every $i \in \mathbb{N}$, $\psi_i \leq \psi_{i+1}$ and $\lim_{i\to\infty} \psi_i = 2g$. Then, by the monotone convergence theorem

$$\lim_{i\to\infty} \bar{\mu}(\psi_i) = \bar{\mu}(2g).$$

However, note that

$$\begin{aligned}\bar{\mu}(\psi_i) &= \bar{\mu}(2g - \phi_i) \\ &= \bar{\mu}(2g) - \bar{\mu}(\phi_i)\end{aligned}$$

and so, since $\bar{\mu}(\psi_i), \bar{\mu}(2g) < \infty$,

$$\lim_{i\to\infty} \bar{\mu}(\phi_i) = \bar{\mu}(2g) - \lim_{i\to\infty} \bar{\mu}(\psi_i) = 0.$$

Finally, note that $0 \leq \bar{\mu}(|f_n - f|) \leq \bar{\mu}(\phi_n)$ for every $n \in \mathbb{N}$ by the monotonicty of $\bar{\mu}$ and so

$$\lim_{n\to\infty} \bar{\mu}(|f_n - f|) = 0.$$

By Corollary 3.2.3,

$$0 \leq \lim_{n\to\infty} |\bar{\mu}(f_n - f)| \leq \lim_{n\to\infty} \bar{\mu}(|f_n - f|) = 0$$

which shows that

$$\lim_{n\to\infty} \bar{\mu}(f_n) = \bar{\mu}(f).$$

□

We have shown results that establish interchanging limits and integrals when such limits exist. The next result makes a more general statement about interchanging limit inferiors of non-negative measurable functions (which always exist in the extended real numbers) with their integrals, although we cannot get strict equality.

THEOREM 3.2.6 (Fatou's lemma). *For any functions* $f_n \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$

$$\bar{\mu}\left(\liminf_{n\to\infty} f_n\right) \leq \liminf_{n\to\infty} \bar{\mu}(f_n).$$

PROOF. Define $g_n(x) = \inf_{i\geq n} f_i(x)$. Clearly, $g_n \leq f_n$ pointwise and

$$\liminf_{n\to\infty} f_n = \lim_{n\to\infty} g_n \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$$

by Corollary 2.2.10. Further, $g_n \leq g_{n+1}$ pointwise and so by the monotone convergence theorem

$$\bar{\mu}\left(\liminf_{n\to\infty} f_n\right) = \bar{\mu}\left(\lim_{n\to\infty} g_n\right) = \lim_{n\to\infty} \bar{\mu}(g_n) = \liminf_{n\to\infty} \bar{\mu}(g_n) \leq \liminf_{n\to\infty} \bar{\mu}(f_n)$$

where the fourth equality follows from the fact the when the limit exists, limit superiors and inferiors are both equal to the limit. □

COROLLARY 3.2.7. *For any functions* $f_n \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ *such that* $f_n \leq g$ *for some* $g \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ *such that* $\bar{\mu}(g) < \infty$

$$\bar{\mu}\left(\limsup_{n\to\infty} f_n\right) \geq \limsup_{n\to\infty} \bar{\mu}(f_n).$$

PROOF. Define $h_n := g - f_n \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ and observe that by Fatou's lemma

$$\bar{\mu}\left(\liminf_{n\to\infty} h_n\right) \leq \liminf_{n\to\infty} \bar{\mu}(h_n).$$

Notice that we can rewrite the left hand side as

$$\begin{aligned}\bar{\mu}\left(\liminf_{n\to\infty} h_n\right) &= \bar{\mu}\left(g + \liminf_{n\to\infty} -f_n\right) \\ &= \bar{\mu}\left(g - \limsup_{n\to\infty} f_n\right) \\ &= \bar{\mu}(g) - \bar{\mu}\left(\limsup_{n\to\infty} f_n\right).\end{aligned}$$

Similarly, we can rewrite the right hand side as

$$\liminf_{n\to\infty} \bar{\mu}(h_n) = \bar{\mu}(g) - \limsup_{n\to\infty} \bar{\mu}(f_n).$$

Since $\bar{\mu}(g) < \infty$, we can subtract it from both sides to yield the result. □

Note that Fatou's lemma is essentially a simple corollarly of the monotone convergence theorem; it turns out that we can deduce the monotone convergence theorem from Fatou's lemma as well, meaning that the two theorems are in fact equivalent. Thus we could characterisze Lebesgue integrals as linear functionals that satisfy Fatou's lemma instead of montone convergence.

PROPOSITION 3.2.8. *Let* $(\mathcal{X},\mathcal{F},\mu)$ *be a measure space and let* $\mu' : \mathcal{M}^+(\mathcal{X},\mathcal{F}) \to [0,\infty]$ *be a linear functional satisfying* $\mu(A) = \mu'(\mathbb{1}_A)$*, linearity and Fatou's lemma. Then for any increasing sequence*$f_n \in \mathcal{M}^+(\mathcal{X},\mathcal{F})$

$$\mu'\left(\lim_{n\to\infty} f_n\right) = \lim_{n\to\infty} \mu'(f_n).$$

One important application of the dominated convergence theorem is differentiating under the integral sign.

THEOREM 3.2.9 (Differentiating under the integral sign). *For any* $\theta \in [-\delta,\delta]$, *define* $g(x,\theta) \in \mathcal{L}^1(\mathcal{X},\mathcal{F},\mu)$ *to be differentiable with respect to* θ*, with derivative*

$$g_\theta(x,\theta) = \lim_{\epsilon\to 0} \frac{g(x,\theta+\epsilon) - g(x,\theta)}{\epsilon}.$$

Define the function

$$m(\theta) := \bar{\mu_x}(g(x,\theta))$$

where the subscript x *clarifies the variable of integration. If there exists a function* $G \in \mathcal{L}^1(\mu)$ *such that* $|g_\theta| \le G$ *for every* $\theta \in (-\delta,\delta)$ *then*

$$\frac{dm(\theta)}{d\theta} = \frac{d\bar{\mu_x}(g(x,\theta))}{d\theta} = \bar{\mu_x}(g_\theta(x,\theta)).$$

PROOF. Define the function

$$f_n(x,\theta) := \frac{g\left(x,\theta+\frac{1}{n}\right) - g(x,\theta)}{\frac{1}{n}}$$

and observe by the linearity of the Lebesgue integral that

$$\frac{m\left(\theta+\frac{1}{n}\right) - m(\theta)}{\frac{1}{n}} = \bar{\mu_x}(f_n(x,\theta))$$

and that

$$\lim_{n\to\infty} f_n(x,\theta) = g_\theta(x,\theta).$$

Next, observe that by the mean value theorem

$$f_n(x,\theta) = g_\theta\left(x,\tilde{\theta}_n\right)$$

for some $\theta < \tilde{\theta}_n < \theta + \frac{1}{n}$ for every $n \in \mathbb{N}$, which implies that for large enough n

$$|f_n(x,\theta)| = \left|g_\theta\left(x,\tilde{\theta}_n\right)\right| \le G(x).$$

Applying the dominated convergence theorem, we have that

$$\begin{aligned}
\bar{\mu_x}(g_\theta(x,\theta)) &= \bar{\mu_x}\left(\lim_{n\to\infty} f_n(x,\theta)\right) \\
&= \lim_{n\to\infty} \bar{\mu_x}(f_n(x,\theta)) \\
&= \lim_{n\to\infty} \frac{m\left(\theta+\frac{1}{n}\right) - m(\theta)}{\frac{1}{n}} \\
&= \frac{dm(\theta)}{d\theta}
\end{aligned}$$

which completes the proof. □

3.2.2. New measures from old and their integrals.

PROPOSITION 3.2.10. *Let $(\mathcal{X},\mathcal{F},\mu)$ be a measure space. For any function $f\in\mathcal{M}^{+}(\mathcal{X},\mathcal{F})$, the set valued function $\nu:\mathcal{F}\longrightarrow[0,\infty]$ given by*

$$\nu(A):=\bar{\mu}(f\mathbb{1}_{A})$$

is a measure on $\mathcal{F}$. Moreover, for any $A\in\mathcal{F}$, if $\mu(A)=0$ then $\nu(A)=0$.

PROOF. Note that $\nu(\emptyset)=\bar{\mu}(f\mathbb{1}_{\emptyset})=\bar{\mu}(\mathbb{1}_{\emptyset})=\mu(\emptyset)=0$. Next, let $\{A_i\}_{i\in\mathbb{N}}\in\mathcal{F}$ be disjoint and define

$$B_n=\bigcup_{i=1}^{n}A_i$$

which is an increasing sequence of sets such that $\bigcup_{n\in\mathbb{N}}B_n=\bigcup_{i\in\mathbb{N}}A_i$. Then

$$\begin{aligned}
\nu\left(\bigcup_{i\in\mathbb{N}}A_i\right)&=\bar{\mu}\left(f\mathbb{1}_{\bigcup_{i\in\mathbb{N}}A_i}\right)\\
&=\bar{\mu}\left(f\mathbb{1}_{\lim_{n\to\infty}B_n}\right)\\
&=\bar{\mu}\left(f\lim_{n\to\infty}\mathbb{1}_{B_n}\right)\\
&=\bar{\mu}\left(f\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{1}_{A_i}\right)\\
&=\lim_{n\to\infty}\bar{\mu}\left(\sum_{i=1}^{n}f\mathbb{1}_{A_i}\right)\\
&=\lim_{n\to\infty}\sum_{i=1}^{n}\bar{\mu}(f\mathbb{1}_{A_i})\\
&=\sum_{i=1}^{\infty}\bar{\mu}(f\mathbb{1}_{A_i})\\
&=\sum_{i=1}^{\infty}\nu(A_i)
\end{aligned}$$

where the second equality follows from the discussion on convergence of sets in Section 2.1, the third equality due to Proposition 2.1.6, the fourth equality from induction on Fact 2.1.2, the fifth equality from the monotone convergence theorem, and the sixth equality due to the linearity of $\bar{\mu}$. Finally, let $A\in\mathcal{F}$ be a $\mu-$measure zero set and define $\{s_n\}_{n\in\mathbb{N}}\in\mathcal{M}_{\text{sim}}(\mathcal{X},\mathcal{F})$ to be an increasing sequence of simple functions which converges to f with standard representation

$$s_n=\sum_{i=1}^{I_n}a_{i.n}\mathbb{1}_{A_{i,n}}.$$

Then the $s_n\mathbb{1}_A\nearrow f\mathbb{1}_A$and

$$\begin{aligned}
\bar{\mu}(s_n\mathbb{1}_A)&=\bar{\mu}\left(\sum_{i=1}^{I_n}a_{i,n}\mathbb{1}_{A_{i,n}\cap A}\right)\\
&=\sum_{i=1}^{I_n}a_{i,n}\mu(A_{i,n}\cap A)\\
&=0
\end{aligned}$$

where the last equality follows from the monotonicty of measures. Applying the monotone convergence theorem, we have that

$$\begin{aligned}\nu(A) &= \bar{\mu}(f\mathbb{1}_A) \\ &= \lim_{n\to\infty}\bar{\mu}(s_n\mathbb{1}_A) \\ &= 0\end{aligned}$$

which completes the proof. □

REMARK 3.2.11. A measure ν on $\mathcal{F}$ with the relation $\mu(A) = 0 \Longrightarrow \nu(A) = 0$ for every $A \in \mathcal{F}$ and a refence measure μ is called *absolutely continuous* with respect to μ. This relation is denoted as $\nu << \mu$ symbolically. We discuss absolute continuity in Chapter 6, where we prove the converse of this theorem under a minor restriction. The non-negative measurable function f which generates the new measure ν is called the *density* of ν with respect to μ. If ν is a probability measure then f is called the *probability density function* of ν with respect to μ.

COROLLARY 3.2.12. *Let μ, ν be measures on $(\mathcal{X}, \mathcal{F})$ and let f be a density of ν with respect to μ. The unique Lebesgue integral associated with the measure ν is given by*

$$\bar{\nu}(g) = \bar{\mu}(fg),$$

for any function $g \in \mathcal{M}(\mathcal{X}, \mathcal{F})$, provided the right-hand side is defined.

PROOF. Note that for any set $A \in \mathcal{F}$, $\bar{\nu}(\mathbb{1}_A) = \bar{\mu}(f\mathbb{1}_A) = \nu(A)$. Further, $\bar{\nu}$ inherits linearity and monotone convergence on $\mathcal{M}^+(\mathcal{X}, \mathcal{F})$ from $\bar{\mu}$ and so by Theorem 3.1.1, $\bar{\nu}$ is the unique integral (with respect to measure ν) on $\mathcal{M}^+(\mathcal{X}, \mathcal{F})$, which of course extends uniquely to $\mathcal{M}(\mathcal{X}, \mathcal{F})$ through Definition 3.1.8. □

PROPOSITION 3.2.13. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $(\mathcal{Y}, \mathcal{G})$ be a measurable space . For any $\mathcal{F}/\mathcal{G}$ measurable function $f : \mathcal{X} \to \mathcal{Y}$, the function $f\mu : \mathcal{G} \longrightarrow [0, \infty]$ given by*

$$f\mu(B) := \mu\left(f^{-1}[B]\right)$$

is a measure on $\mathscr{B}(\mathbb{R})$.

PROOF. Note that $f\mu(\emptyset) = \mu\left(f^{-1}[\emptyset]\right) = \mu(\emptyset)$. Next, let $\{B_i\}_{i\in\mathbb{N}} \in \mathcal{G}$ be disjoint and observe that

$$\begin{aligned}f\mu\left(\bigcup_{i\in\mathbb{N}} B_i\right) &= \mu\left(f^{-1}\left[\bigcup_{i\in\mathbb{N}} B_i\right]\right) \\ &= \mu\left(\bigcup_{i\in\mathbb{N}} f^{-1}[B_i]\right) \\ &= \sum_{i=1}^{\infty}\mu\left(f^{-1}[B_i]\right) \\ &= \sum_{i=1}^{\infty} f\mu(B_i)\end{aligned}$$

where the second equality is a property of inverse maps and the third equality is due to the countable additivity of μ. □

REMARK. Here the measure $f\mu$ is called the *image measure* generated by f via μ. If μ is a probability measure, then $f\mu$ is called the *probability distribution* of f.

COROLLARY 3.2.14. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $(\mathcal{Y}, \mathcal{G})$ be a measurable space. Define $f\mu$ to be the image measure of a measurable function $f : \mathcal{X} \to \mathcal{Y}$ with respect to measure μ. Then, for any function $g \in \mathcal{M}(\mathcal{Y}, \mathcal{G})$, the Lebesgue integral associated with measaure $f\mu$ is given by*

$$\bar{f\mu}(g) := \bar{\mu}(g \circ f)$$

provided the right hand side is defined.

PROOF. First note that for any $B \in \mathcal{G}$

$$\begin{aligned}\bar{f\mu}\left(\mathbb{1}_B\right) &= \bar{\mu}\left(\mathbb{1}_B \circ f\right) \\ &= \bar{\mu}\left(\mathbb{1}_{f^{-1}[B]}\right) \\ &= \mu\left(f^{-1}[B]\right) \\ &= f\mu\left(B\right)\end{aligned}$$

which satisfies the first required property of Theorem 3.1.1. Linearity and monotone convergence follow from $\bar{\mu}$ and so the uniqueness criterion of the theorem tells us that $\bar{f\mu}$ is indeed the unique integral associated with the image measure $f\mu$. □

Recall from Proposition(1.3.7) that the countable sum of measures on $(\mathcal{X}, \mathcal{F})$ is a measure on $(\mathcal{X}, \mathcal{F})$. The integral resulting from this compound measure can be decomposed into the integrals from constituent summand measures in certain situations.

PROPOSITION 3.2.15. *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let $\{\mu_i\}_{i\in\mathbb{N}}$ be a countable collection of measures on $\mathcal{F}$ with their respective integrals $\{\bar{\mu}_i(\cdot)\}_{i\in\mathbb{N}}$. Then the integral associated with the sum measure*

$$\mu := \sum_{i=1}^{\infty} \mu_i$$

is given by

$$\bar{\mu}(f) := \sum_{i=1}^{\infty} \bar{\mu}(f)$$

for any $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$.

PROOF. We first show this holds for $f = \mathbb{1}_A$for any $A \in \mathcal{F}$; indeed, an application of Proposition 1.3.7 is sufficient for this purpose. Next, for any $s \in M_{\text{sim}}(\mathcal{X}, \mathcal{F})$, we have that

$$s = \sum_{i=1}^{I} \alpha_i \mathbb{1}_{A_i}$$

where $\alpha_i > 0$, $A_i := \{x \in \mathcal{X} : s(x) = \alpha_i\}$, and $I \in \mathbb{N}$. Thus,

$$\begin{aligned}\bar{\mu}(s) &= \sum_{i=1}^{I} \alpha_i \mu(A_i) \\ &= \sum_{i=1}^{I} \alpha_i \sum_{j=1}^{\infty} \mu_j(A_i) \\ &= \sum_{j=1}^{\infty} \sum_{i=1}^{I} \mu_j(A_i) \\ &= \sum_{j=1}^{\infty} \bar{\mu_j}(s)\end{aligned}$$

where the first equality is by the definition of integrals on simple functions and the third is by the linearity of limits of sequences. This establishes the result for simple functions. Finally, let $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ be arbitrary and observe by Proposition 2.3.2 that there exists some increasing sequence

$\{s_n\} \in M_{\text{sim}}(\mathcal{X},\mathcal{F})$ such that $s_n \nearrow f$ and

$$\begin{aligned}\bar{\mu}(f) &= \bar{\mu}\left(\lim_{n\to\infty} s_n\right)\\ &= \lim_{n\to\infty} \bar{\mu}(s_n)\\ &= \lim_{n\to\infty} \sum_{i=1}^{\infty} \bar{\mu}_i(s_n)\\ &= \sum_{i=1}^{\infty} \bar{\mu}_i\left(\lim_{n\to\infty} s_n\right)\\ &= \sum_{i=1}^{\infty} \bar{\mu}_i(f)\end{aligned}$$

where the second equality follows by monotone convergence and the fourth equality follows by two applications of monotone convergence (you can think of the countable sum as integration with respect to the counting measure). This completes the proof. □

3.2.3. Equivalence of integrals and measures. Note that while in Theorem 3.1.1 we constructed the integral from a seemingly more primitive concept of a measure, it turns out that a measure can be constructed out a linear functional on the space of measurable functions such that the functional is the integral with respect to the measure we have constructed. In this sense, measures and integrals are really equivalent. At this point, this may seem to be a trivial observation (and the proof of this result is indeed trivial); this shift of perspective, however, offers powerful simplifications to questions concerning the existence and uniqueness of measures. Indeed, our approach to product measures in Chapter 7 would rely on this insight, allowing us to prove both the existence of product measures, and the representation of integrals with respect to product measures as iterated integrals, in a single stroke.

THEOREM 3.2.16. *Let $(\mathcal{X},\mathcal{F})$ be a measurable space and let*

$$\Lambda : \mathcal{M}^+(\mathcal{X},\mathcal{F}) \longrightarrow [0,\infty]$$

be a linear functional that satisfies monotone convergence i.e for any sequence $f_n \in \mathcal{M}(\mathcal{X},\mathcal{F})$ such that $f_n \leq f_{n+1}$ and $f := \lim_{n\to\infty} f_n \in \mathcal{F}$ we have

$$\Lambda(f) = \lim_{n\to\infty} \Lambda(f_n).$$

Further, suppose there exists a function $g \in \mathcal{M}^+(\mathcal{X},\mathcal{F})$ such that $\Lambda(g) < \infty$. Then the function

$$\lambda : \mathcal{F} \longrightarrow [0,\infty]$$

given by

$$\lambda(A) := \Lambda(\mathbb{1}_A)$$

for any $A \in \mathcal{F}$ is a measure and Λ is the integral with respect to λ.

PROOF. First, note that

$$\begin{aligned}\Lambda(g) &= \Lambda(g + \mathbb{1}_\emptyset)\\ &= \Lambda(g) + \Lambda(\mathbb{1}_\emptyset)\\ &= \Lambda(g) + \lambda(\emptyset)\end{aligned}$$

by linearity and since $\Lambda(g) < \infty$, we can subtract it from both sides to deduce $\lambda(\emptyset) = 0$. To establish countable additivity, note that for a disjoint collection $A_n \in \mathcal{F}$,

$$\begin{aligned} \lambda\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \Lambda\left(\sum_{n=1}^{\infty} \mathbb{1}_{A_n}\right) \\ &= \sum_{n=1}^{\infty} \Lambda\left(\mathbb{1}_{A_n}\right) \\ &= \sum_{n=1}^{\infty} \lambda\left(A_n\right) \end{aligned}$$

where the first equality is by (an extension of) Proposition 2.1.3 and the second by linearity and monotone convergence. Of course, since Λ satisfies the properties of Theorem 3.1.1, it is the unique integral induced by the measure λ; this completes the proof. ◻

3.3. Null sets

Definition 3.3.1. Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space. A set $A \in \mathcal{F}$ is called a $\mu-$null set if it's a measure zero set with respect to μ i.e.

$$\mu(A) = 0.$$

The collection of all $\mu-$null sets in $\mathcal{F}$ is denoted N_μ.

Note we will often omit the "μ" when describing a $\mu-$null set and simply say "null set" if the measure is clear from context.

Proposition 3.3.2. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space. The set of all null sets N_μ is closed under countable unions.*

Proof. Let $\{A_i\}_{i\in\mathbb{N}} \in N_\mu$ be arbitrary. Then, by countable subadditivity

$$\begin{aligned} \mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) &\le \sum_{i=1}^{\infty} \mu\left(A_i\right) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mu\left(A_i\right) \\ &= 0. \end{aligned}$$

◻

Proposition 3.3.3. *For any function $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ and a measure μ on $\mathcal{F}$, if the integral $\bar{\mu}(f) = 0$ then the set $\{x \in \mathcal{X} \mid f(x) > 0\} \in N_\mu$.*

Proof. Observe that for any $n \in \mathbb{N}$ we have the pointwise inequality

$$h_n := \mathbb{1}_{\left\{f \ge \frac{1}{n}\right\}} \le nf$$

where both sides are measurable since f is Borel-measurable. By the monotonicity and linearity of the integral, we have

$$\mu\left(\left\{x \in \mathcal{X} \mid f(x) \ge \frac{1}{n}\right\}\right) = \bar{\mu}\left(h_n\right) \le n\bar{\mu}(f) = 0.$$

Since $f(x) \ge \frac{1}{n} \implies f(x) \ge \frac{1}{n+1}$ and $\bigcup_{n\in\mathbb{N}} \left\{x \in \mathcal{X} \mid f(x) \ge \frac{1}{n}\right\} = \{x \in \mathcal{X} \mid f(x) > 0\}$, by the continuity from below of measures (or, equivalently, the monotone convergence theorem)

$$\mu(\{x \in \mathcal{X} \mid f(x) > 0\}) = \lim_{n\to\infty} \mu\left(\left\{x \in \mathcal{X} \mid f(x) \ge \frac{1}{n}\right\}\right) = 0$$

which completes the proof. ◻

REMARK. When two function $f, g \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ are such that $f = g$ on $\mathcal{X} \setminus A$ for some $A \in N_\mu$, we say that the functions are equal $\mu-$almost everywhere. In the literature, this is often shortened to writing $\mu-$a.e or simply a.e if the measure is clear from context. If μ is a probability measure, then we say $f = g$ almost surely, which is often shortened to a.s in the literature.[1] From this it's clear that the set $\{x \in \mathcal{X} \mid f(x) \neq g(x)\} \subseteq A$. Of course, if we can show that $\{x \in \mathcal{X} \mid f(x) \neq g(x)\} \in \mathcal{F}$ then $\{x \in \mathcal{X} \mid f(x) \neq g(x)\} \in N_\mu$ by the monotonicity of measures.

Corresponding to the notion of almost-everywhere equality of measurable functions, there's a notion of almost-everywhere equality of sets. Recall from set theory that the symmetric difference of two sets A, B is given

$$A\Delta B := (A \setminus B) \bigcup (B \setminus A)$$

Given a measure μon some measurable space $(\mathcal{X}, \mathcal{F})$ and sets $A, B \in \mathcal{F}$, we say $A \overset{\mu-\text{a.e}}{=} B$ if $\mu(A\Delta B) = 0$. Of course, in the spirit of Theorem 3.2.16, this definition must coincide with the one for measurable functions in that we must have that $\mathbb{1}_A \overset{\mu-\text{a.e}}{=} \mathbb{1}_B$. This is easily shown in the following result.

PROPOSITION 3.3.4. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space. For any $A, B \in \mathcal{F}$, $A \overset{\mu-a.e}{=} B$ if and only if $\mathbb{1}_A \overset{\mu-a.e}{=} \mathbb{1}_B$*

PROOF. Note that $\mu(A\Delta B) = 0 \iff \mu\left(A \cap B^C\right) = \mu\left(B \cap A^C\right) = 0$. Further, observe that

$$\{x \in \mathcal{X} \mid \mathbb{1}_A(x) \neq \mathbb{1}_B(x)\} = \underbrace{\{x \in \mathcal{X} \mid \mathbb{1}_A(x) > \mathbb{1}_B(x)\}}_{=A\cap B^C} \bigcup \underbrace{\{x \in \mathcal{X} \mid \mathbb{1}_A(x) < \mathbb{1}_B(x)\}}_{=B\cap A^C}$$

which completes the proof. $\square$

LEMMA 3.3.5. *Let $f, g \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ be arbitrary. Then, the set*

$$A := \{x \in \mathcal{X} \mid f(x) \neq g(x)\}$$

is measurable i.e. $A \in \mathcal{F}$.

PROOF. Let $h := (f - g)\,\mathbb{1}_{\{f=g=\infty \text{or} f=g=-\infty\}^C}$and observe that h is measurable by Corollary 2.2.7. Further, $A \subseteq \{x \in \mathcal{X} \mid f(x) = g(x) = \infty \text{ or } f(x) = g(x) = -\infty\}^C$ and so

$$A = \{x \in \mathcal{X} \mid h(x) \neq 0\} = \{x \in \mathcal{X} \mid h(x) > 0\} \bigcup \{x \in \mathcal{X} \mid h(x) < 0\}$$

is in $\mathcal{F}$ by the measurability of h. $\square$

PROPOSITION 3.3.6. *For any function $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ and a measure μ on $\mathcal{F}$, if the integral $\bar{\mu}(f) < \infty$ then $f < \infty$ $\mu-$almost everywhere.*

PROOF. Observe that the pointwise equality

$$\mathbb{1}_{\{f=\infty\}} \leq \frac{f}{n}$$

where both sides are measurable since f is Borel-measurable. By the monotonicity and linearity of the integral and our assumptions,

$$0 \leq \mu(\{x \in \mathcal{X} \mid f(x) = \infty\}) \leq \frac{1}{n}\bar{\mu}(f) < \infty.$$

Since weak inequalities are preserved under limits, we have that

$$0 \leq \mu(\{x \in \mathcal{X} \mid f(x) = \infty\}) \leq \lim_{n\to\infty} \frac{1}{n}\bar{\mu}(f) = 0$$

which completes the proof. $\square$

PROPOSITION 3.3.7. *Let $f, g \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ be arbitrary and let μbe a measure on $\mathcal{F}$. If $f = g$ $\mu-$almost everywhere, then*

$$\bar{\mu}(f) = \bar{\mu}(g).$$

---

[1] Sometimes, this is denoted even more compactly as $f \overset{\text{a.e}}{=} g$.

Proof. Define the set

$$A := \{x \in \mathcal{X} \mid f(x) \neq g(x)\}$$

which has measure zero by Lemma 3.3.5 and our assumption. Then, I claim that the pointwise inequality

$$\min\{g, n\} \leq n\mathbb{1}_A + f$$

holds for every $n \in \mathbb{N}$. To see this, we can look at the following cases:

(1) $g(x) > n$ and $g(x) \neq f(x)$: In this case, we see that the inequality resolves to

$$n \leq n + f(x)$$

which is true by the non-negativity of f.

(2) $g(x) \leq n$ and $g(x) \neq f(x)$: In this case, the inequality resolves to

$$g(x) \leq n + f(x)$$

which is again true by the non-negativity of f.

(3) $g(x) > n$ and $g(x) = f(x)$: In this case, the inequality resolves to

$$n \leq f(x)$$

which is true since $f(x) = g(x)$.

(4) $g(x) \leq n$ and $g(x) = f(x)$: In this case, the inequality resolves to

$$g(x) \leq f(x)$$

which is true by assumption.

Note that both the left and right hand side are measurable by Proposition 2.2.6and so, integrating both sides, we have

$$\begin{aligned}\bar{\mu}(\min\{g, n\}) &\leq n\mu(A) + \bar{\mu}(f) \\ &= \bar{\mu}(f)\end{aligned}$$

since $\mu(A) = 0$. Finally, observe that $\min\{g, n\} \leq \min\{g, n+1\}$ and that $\lim_{n\to\infty} \min\{g, n\} = g$. Then, applying the monotone convergence theorem, we have

$$\bar{\mu}(g) = \lim_{n\to\infty} \bar{\mu}(\min\{g, n\}) \leq \bar{\mu}(f).$$

We can deduce the reverse inequality with the analagous pointwise inequality

$$\min\{f, n\} \leq n\mathbb{1}_A + g$$

which proves the result. □

Corollary 3.3.8. *Let $f, g \in L^1(\mathcal{X}, \mathcal{F}, \mu)$ be such that they are equal almost everywhere. Then*

$$\bar{\mu}(f) = \bar{\mu}(g).$$

Proof. If $f = g$ a.e then

$$f^+ - f^- \overset{\text{a.e}}{=} g^+ - g^- \iff f^+ + g^- \overset{\text{a.e}}{=} g^+ + f^-$$

and so, integrating both sides and applying Proposition 3.3.7 we have

$$\bar{\mu}\left(f^+ + g^-\right) = \bar{\mu}\left(g^+ + f^-\right).$$

Applying linearity and observing that the integral of each component is finite by 3.1.6, we get

$$\bar{\mu}(f) = \bar{\mu}\left(f^+\right) - \bar{\mu}\left(f^-\right) = \bar{\mu}\left(g^+\right) - \bar{\mu}\left(g^-\right) = \bar{\mu}(g)$$

as desired. □

PROPOSITION 3.3.9. *Let $f, g \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ (or let $f, g \in \mathcal{L}^1(\mu)$) where μ is a measure on $\mathcal{F}$ such that*

$$\bar{\mu}(f\mathbb{1}_F) = \bar{\mu}(g\mathbb{1}_F)$$

for every $F \in \mathcal{F}$. Then

$$f = g$$

$\mu-$almost everywhere.

PROOF. Assume the hypothesis is true and define

$$A := \{x \in \mathcal{X} \mid f(x) \neq g(x)\} = \underbrace{\{x \in \mathcal{X} \mid f(x) > g(x)\}}_{A_1} \bigcup \underbrace{\{x \in \mathcal{X} \mid f(x) < g(x)\}}_{A_2}$$

and suppose for contradiction that $\mu(A) > 0$. Then one of A_1 or A_2 has positive measure. Assume, without loss of generality, that $\mu(A_1) > 0$ and further define

$$A_{1,n} := \left\{x \in \mathcal{X} \mid f(x) - \frac{1}{n} \geq g(x)\right\} = \left\{x \in \mathcal{X} \mid z(x) \geq \frac{1}{n}\right\}$$

where $z = f - g$ and so is measurable, which in turn implies that $A_{1,n} \in \mathcal{F}$ for all $i \in \mathbb{N}$. Further, $A_{1,n} \subseteq A_{1,n+1}$ and $\bigcup_{n\in\mathbb{N}} A_{1,n} = A$, and so by continuity from below

$$\lim_{n\to\infty} \mu(A_{1,n}) = \mu(A_1) > 0.$$

By the definition of limts, there exists some $n_0 \in \mathbb{N}$ such that

$$\mu(A_{1,n}) > 0 \; \forall n \geq n_0$$

and for such n we also have, by the monotonicty of integration

$$\bar{\mu}(z\mathbb{1}_{A_{1,n}}) \geq \bar{\mu}\left(\frac{1}{n}\mathbb{1}_{A_{1,n}}\right) = \frac{1}{n}\mu(A_{1,n}) > 0$$

which implies, by the linearity of integration, that

$$\bar{\mu}(f\mathbb{1}_{A_{1,n}}) > \bar{\mu}(g\mathbb{1}_{A_{1,n}})$$

which contradicts our hypotehsis. This completes the proof. □

DEFINITION 3.3.10 (Convergence almost everywhere). Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and define $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ to be a sequence of functions. The sequence is f_n is said to *converge almost everywhere* to a function $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ if

$$\mu\left(\left\{x \in \mathcal{X} \mid \lim_{n\to\infty} f_n(x) \neq f(x)\right\}\right) = 0.$$

In this case, we write

$$\lim_{n\to\infty} f_n \stackrel{\text{a.e}}{=} f$$

or say $f_n \longrightarrow f$ $\mu-$a.e (or $f_n \xrightarrow{\text{a.e}} f$).

THEOREM 3.3.11 (Generalized monotone convergence theorem). *Let $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ be such that $f_n \leq f_{n+1}$ $\mu-$almost everyhwere; that is, there is some $A \in N_\mu$ such that $f_n(x) \leq f_{n+1}(x)$ for all $x \in \mathcal{X} \setminus A$. Then if*

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

on $\mathcal{X} \setminus A$, we have that

$$\lim_{n\to\infty} \bar{\mu}(f_n) = \bar{\mu}(f).$$

PROOF. Define $g_n = f_n \mathbb{1}_{\mathcal{X}\setminus A}$ and observe that

$$g_n \leq g_{n+1}$$

pointwise for all $n \in \mathbb{N}$and

$$\lim_{n\to\infty} g_n = f\mathbb{1}_{\mathcal{X}\setminus A}.$$

By the standard monotone convergence theorem,

$$\lim_{n\to\infty} \bar{\mu}(g_n) = \bar{\mu}\left(f\mathbb{1}_{\mathcal{X}\setminus A}\right).$$

But note that since $g_n \overset{\text{a.e}}{=} f_n$ and $f \overset{\text{a.e}}{=} f\mathbb{1}_{\mathcal{X}\setminus A}$, by Proposition 3.3.7,

$$\lim_{n\to\infty} \bar{\mu}(f_n) = \lim_{n\to\infty} \bar{\mu}(g_n) = \bar{\mu}\left(f\mathbb{1}_{\mathcal{X}\setminus A}\right) = \bar{\mu}(f)$$

which completes the proof. □

We can similarly strengthen the dominated convergence theorem.

THEOREM 3.3.12 (Generalized dominated convergence theorem). *Let $\{f_n\}_n \in \mathcal{L}^1(\mathcal{X}, \mathcal{F}, \mu)$ and suppose there exists some $g \in \mathcal{L}^1(\mu)$ be such that*

$$|f_n| \leq g$$

$\mu-$almost everywhere. Then, if there exists some $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ such that

$$\lim_{n\to\infty} f_n \overset{\text{a.e}}{=} f$$

we have that $f \in \mathcal{L}^1(\mu)$ and

$$\lim_{n\to\infty} \bar{\mu}(|f_n - f|) = 0$$

and

$$\lim_{n\to\infty} \bar{\mu}(f_n) = \bar{\mu}(f).$$

PROOF. Without loss of generality[2], assume that there exists some set $A \in N_\mu$ such that

$$|f_n(x)| \leq g(x)$$

for all $x \in \mathcal{X} \setminus A$ and every $n \in \mathbb{N}$, and that

$$\lim_{n\to\infty} f_n(x) = f(x)$$

for all $x \in \mathcal{X} \setminus A$. Then, consider the functions $h_n := f_n \mathbb{1}_{\mathcal{X}\setminus A}$and observe that $g^* := g\mathbb{1}_{\mathcal{X}\setminus A} \in \mathcal{L}^1(\mu)$ since $g^* \overset{\text{a.e}}{=} g \implies |g^*| \overset{\text{a.e}}{=} |g| \implies \bar{\mu}(|g^*|) = \bar{\mu}(|g|) < \infty$. Next, note that

$$|h_n| \leq g^*$$

everywhere for each $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} h_n = f\mathbb{1}_{\mathcal{X}\setminus A}$$

pointwise and so, applying the usual dominated convergence theorem

$$\lim_{n\to\infty} \bar{\mu}(h_n) = \bar{\mu}\left(f\mathbb{1}_{\mathcal{X}\setminus A}\right).$$

As, before, by Corollary 3.3.8

$$\lim_{n\to\infty} \bar{\mu}(f_n) = \lim_{n\to\infty} \bar{\mu}(h_n) = \bar{\mu}\left(f\mathbb{1}_{\mathcal{X}\setminus A}\right) = \bar{\mu}(f).$$

□

In the context of our discussion on the equivalence between integrals and measures, we had foreshadowed how the monotone convergence theorem and the continuity from below of measures were basically the same concept. We had also said that the Borel-Cantelli lemma could be understood through from both an integration and measure-theoretic perspective. We can make this precise with the following result.

[2]This can be justfied by Proposition 3.3.2

THEOREM 3.3.13 (Generalized Borel-Cantelli lemma). *Let $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}^+(\mathcal{X},\mathcal{F})$ and let μ be measure on $\mathcal{F}$. If*

$$\sum_{n=1}^{\infty} \bar{\mu}(f_n) < \infty$$

then

$$\sum_{n=1}^{\infty} f_n < \infty$$

$\mu-$almost everywhere.

PROOF. Note that

$$\begin{aligned}\bar{\mu}\left(\sum_{n=1}^{\infty} f_n\right) &= \bar{\mu}\left(\lim_{N\to\infty}\sum_{n=1}^{N} f_n\right)\\ &= \lim_{N\to\infty}\bar{\mu}\left(\sum_{n=1}^{N} f_n\right)\\ &= \sum_{n=1}^{\infty}\bar{\mu}(f_n) < \infty\end{aligned}$$

where the second equality follows from the monotone convergence theorem and the third equality due to the linearity of the integral. By Proposition 3.3.6

$$\sum_{n=1}^{\infty} f_n < \infty$$

$\mu-$almost everywhere. □

It should be clear that we can recover our original Borel-Cantelli lemma by letting $f_n = \mathbb{1}_{A_n}$ for sets $\{A_n\}_{n\in\mathbb{N}} \in \mathcal{F}$.

3.4. Convergence of measurable functions

So far, we have discussed two modes of convergence for measurable functions explicitly: pointwise convergence and almost-everywhere convergence, the latter of which is implied by the former. We have also implicitly defined another type of convergence through the dominated convergence theorem. We can make this explicit with the following definition

DEFINITION 3.4.1 (Convergence in L^1). A sequence of functions $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{L}^1(\mathcal{X},\mathcal{F},\mu)$ is said to converge to a function $f \in \mathcal{L}^1(\mathcal{X},\mathcal{F},\mu)$ in $\mathcal{L}^1$ if

$$\lim_{n\to\infty}\bar{\mu}(|f_n - f|) = 0.$$

In this case, we write

$$f_n \xrightarrow{\mathcal{L}^1} f.$$

DEFINITION 3.4.2 (Convergence in measure). A sequence of functions $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X},\mathcal{F})$ is said to *converge in measure* with respect to measure μ on $\mathcal{F}$ if for every $\epsilon > 0$

$$\lim_{n\to\infty}\mu(\{x\in\mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}) = 0.$$

In this case, we write

$$f_n \xrightarrow{\mu} f.$$

Immediately, we would like to know if these "limits" are well behaved in some sense. That is, we woud like to ensure that they satisfy some basic properties that we expect limits to satisfy.

PROPOSITION 3.4.3. *Let $(\mathcal{X},\mathcal{F},\mu)$ be a measure space and let $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X},\mathcal{F})$ converge to $f \in \mathcal{M}(\mathcal{X},\mathcal{F})$ in any of the modes described earlier. Then $|f_n - f| \to 0$ in the same mode.*

Proof. We first show this for almost everywhere convergence. Suppose $f_n \xrightarrow{\text{a.e}} f$ and consider the function $g_n := |f_n - f|$. Note that if for any $x \in \mathcal{X}$

$$\lim_{n\to\infty} g_n(x) = 0 \Longleftrightarrow \lim_{n\to\infty} f_n(x) = f(x)$$

by the definition of convergence of sequences. The result then follows by observing that the points where this does not occur are identical (and thus so are their measures). The result for the other two modes are trivial. □

Proposition 3.4.4 (Linearity of convergence). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $\{f_n\}_{n\in\mathbb{N}}, \{g_n\} \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ converge to $f, g \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ respectively in any of the modes described earlier. For any $c \in \mathbb{R}$*

$$cf_n + g_n \to cf + g$$

in the same mode of convergence.

Proof. In the case of almost sure convergence, let the null sets where $\lim_{n\to\infty} f_n(x) \neq f(x)$ and $\lim_{n\to\infty} g_n(x) \neq g(x)$ be N_f and N_g respectively. Then $N = N_f \cup N_g$ is a null set and so

$$\lim_{n\to\infty} cf_n + g_n = cf + g$$

on N^C which establishes linearity.

In the case of convergence in measure, fix $\epsilon > 0$ and observe that

$$\begin{aligned}\{x \in \mathcal{X} \mid |cf_n(x) + g(x) - cf(x) - g(x)| > \epsilon\} &\subseteq \{x \in \mathcal{X} \mid |c||f_n(x) - f(x)| + |g_n(x) - g(x)| > \epsilon\} \\ &\subseteq \left\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \frac{\epsilon}{2c}\right\} \cup \left\{x \in \mathcal{X} \mid |g_n(x) - g(x)| > \frac{\epsilon}{2}\right\}.\end{aligned}$$

Subadditvity and monotonicty implies

$$\begin{aligned}\mu\left(\{x \in \mathcal{X} \mid |cf_n(x) + g(x) - cf(x) - g(x)| > \epsilon\}\right) &\leq \mu\left(\left\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \frac{\epsilon}{2c}\right\}\right) \\ &\quad + \mu\left(\left\{x \in \mathcal{X} \mid |g_n(x) - g(x)| > \frac{\epsilon}{2}\right\}\right).\end{aligned}$$

Taking limits yields the result.

The case for L^1 convergence will be established in chapter 4 where we show that L^1 convergence corresponds to convergence in a (semi) norm, which automatically implies the result; for now we take it as given. □

Proposition 3.4.5 (Squeeze theorem). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $f_n \leq g_n \leq h_n$ be in $\mathcal{M}(\mathcal{X}, \mathcal{F})$ such that $f_n, h_n \to \psi$ in one of the three modes of convergence. Then $g_n \to \psi$ in the same mode of convergence.*

Proof. Suppose that $f_n, h_n \xrightarrow{\text{a.s}} \psi$. Let N_f and N_h be the null sets where pointwise convergence fails for the sequences f_n and h_n respectively. On the complement of their union, we have pointwise convergence. Take any point $x \in N_f^C \cap N_h^C$;for any $\epsilon > 0$ there's some $n_{x,\epsilon} \in \mathbb{N}$ such that for all $n \geq n_{x,\epsilon}$

$$|f_n(x) - \psi(x)| < \epsilon$$

and

$$|h_n(x) - \psi(x)| < \epsilon$$

and so

$$-\epsilon < f_n(x) - \psi(x) \leq g_n(x) - g(x) \leq h_n(x) < \epsilon \Longleftrightarrow |g_n(x) - \psi(x)| < \epsilon$$

which completes the proof.

Now for the convergence in measure, observe that for any $\epsilon > 0$,

$$\{x \in \mathcal{X} \mid |h_n(x) - \psi(x)| \leq \epsilon\} \bigcap \{x \in \mathcal{X} \mid |f_n(x) - \psi(x)| \leq \epsilon\} \subseteq \{x \in \mathcal{X} \mid |g_n(x) - \psi(x)| \leq \epsilon\}$$

and so

$$\{x \in \mathcal{X} \mid |g_n(x) - \psi(x)| > \epsilon\} \subseteq \{x \in \mathcal{X} \mid |h_n(x) - \psi(x)| > \epsilon\} \bigcup \{x \in \mathcal{X} \mid |f_n(x) - \psi(x)| > \epsilon\}$$

and so

$$\mu(\{x \in \mathcal{X} \mid |g_n(x) - \psi(x)| > \epsilon\}) \leq \mu(\{x \in \mathcal{X} \mid |h_n(x) - \psi(x)| > \epsilon\}) + \mu(\{x \in \mathcal{X} \mid |f_n(x) - \psi(x)| > \epsilon\})$$

by monotonicty and sub-additivity. Taking limits yields the result.

For L^1 convergence, note that we have $f_n - \psi \leq g_n - \psi \leq h_n - \psi$ and so[3]

$$|g_n - \psi| \leq \max\{|f_n - \psi|, |h_n - \psi|\} \leq |f_n - \psi| + |h_n - \psi|.$$

The linearity of integrals and the squeeze theorem for real sequences then implies the result. □

Finally, for our limiting operations to make sense, limits should be unique. Unfortunately, this is not strictly true; however all limits are almost everywhere unique. For the case of convergence in measure and L^1 convergence, we shall show that this follows from the fact that the topologies for these modes of convergence are given by (pseudo) metrics (in the special case of L^1, by a semi-norm); the case of almost-everywhere convergence, the result follows by the fact that the union of null sets is null. Therefore we have the following.

Proposition 3.4.6 (Uniqueness of limits). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $\{f_n\}_{n\in\mathbb{N}}, f, g \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ be such that $f_n \longrightarrow f$ and $f_n \longrightarrow g$ in any of the modes of convergence. Then $f \overset{a.s}{=} g$.*

It would be good to know the conditions under which one type of convergence implies another, one of which we have already explored in the form of the (generalized) dominated convergence theorem, which gives us sufficient conditions for almost everywhere convergence implies $\mathcal{L}^1$ convergence. The next <few> results link the other types of convergence.

Proposition 3.4.7. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space such that $\mu(\mathcal{X}) < \infty$ and let $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ such that*

$$f_n \xrightarrow{a.e} f.$$

Then

$$f_n \xrightarrow{\mu} f.$$

Proof. Define

$$A := \left\{x \in \mathcal{X} \mid \lim_{n\to\infty} f_n(x) \neq f(x)\right\}$$

and observe that by our hypothesis $\mu(A) = 0$. Let's unpack what this means carefully. Fix any $x_0 \in A$ and observe that there exists some $\epsilon > 0$ such that $|f_n(x_0) - f(x_0)| > \epsilon$ for infinitely many $n \in \mathbb{N}$. Recall the discussion on convergence of sets in Section 2.1 and notice that we can formalize our intuition with the equalities

$$\begin{aligned} A &= \bigcup_{\epsilon>0} \bigcap_{n_0\in\mathbb{N}} \bigcup_{n\geq n_0} \{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\} \\ &= \bigcup_{\epsilon>0} \limsup_{n\to\infty} \{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\} \end{aligned}$$

where the set $\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\} \in \mathcal{F}$ by Lemma 2.2.5 and so $\limsup_{n\to\infty} \{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\} \in \mathcal{F}$ by closure under countable unions and intersections. Then,

$$0 \leq \mu\left(\limsup_{n\to\infty} \{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}\right) \leq \mu(A) = 0$$

for any $\epsilon > 0$ by the monotonicity of measures. Finally, recall that by Proposition 2.1.6

$$\mathbb{1}_{\limsup_{n\to\infty}\{x\in\mathcal{X} \mid |f_n(x)-f(x)|>\epsilon\}} = \limsup_{n\to\infty} \mathbb{1}_{\{x\in\mathcal{X} \mid |f_n(x)-f(x)|>\epsilon\}} \leq 1$$

where $\bar{\mu}(1) = \mu(\mathcal{X}) < \infty$. Then, by the reverse Fatou Lemma

$$0 \leq \limsup_{n\to\infty} \mu(\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}) \leq \mu\left(\limsup_{n\to\infty} \{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}\right) = 0.$$

[3] If $a \leq b \leq c$ then, $|b| \leq \max\{b, -b\} \leq \max\{c, -a\} \leq \max\{|c|, |a|\}$.

By the non-negativity of measures

$$\begin{aligned} 0 &\leq \liminf_{n\to\infty} \mu\left(\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}\right) \\ &\leq \lim_{n\to\infty} \mu\left(\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}\right) \\ &\leq \limsup_{n\to\infty} \mu\left(\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}\right) \\ &= 0 \end{aligned}$$

which completes the proof. $\square$

Note that if we drop the finiteness assumption the result does not hold: take $\mathbb{1}_{[n,n+1]} \longrightarrow 0$ pointwise and the Lebesgue measure λ on $\mathbb{R}$ and note that for any $\epsilon \in (0,1)$

$$\lim_{n\to\infty} \lambda\left(\left\{x \in \mathbb{R} \mid \mathbb{1}_{[n,n+1]} > \epsilon\right\}\right) = \lim_{n\to\infty} \lambda\left([n, n+1]\right) = 1.$$

Note that crucial failure here is that we cannot dominate $\mathbb{1}_{[n,n+1]}$ uniformly (i.e. for all n) with any integrable function g and so a violation of dominated convergence leads to a failure of convergence in measure.

The converse to this result is not true either, that is, convergence in measure does not imply almost everywhere convergence, even when the measure in question in is finite.

Example 3.4.8. Observe that for any $n \in \mathbb{N}$, there exists some $k \in \mathbb{N} \cup \{0\}$ such that $2^k \leq n < 2^{k+1}$. By construction, there's only one such k and so we can denote it $k(n)$. Consider the following collection of sets in $\mathscr{B}(\mathbb{R})$

$$E_n := \left[\frac{n - 2^{k(n)}}{2^{k(n)}}, \frac{n + 1 - 2^{k(n)}}{2^{k(n)}}\right]$$

and observe that $E_n \subseteq [0, 1]$ for all $n \in \mathbb{N}$. To see this, note that the lower bound is smallest when $n = 2^{k(n)}$ and the upper bound is largest when $n = 2^{k(n)+1} - 1$. which correspond to 0 and 1 respectively. Now for $\epsilon \in (0, 1)$

$$\lim_{n\to\infty} \lambda\left(\{x \in \mathbb{R} \mid \mathbb{1}_{E_n} > \epsilon\}\right) = \lim_{n\to\infty} \lambda\left(E_n\right) = \lim_{n\to\infty} \frac{1}{2^{k(n)}} = 0$$

as $k(n) \to \infty$. Further, for $\epsilon \geq 1$

$$\lambda\left(\{x \in \mathbb{R} \mid \mathbb{1}_{E_n} > \epsilon\}\right) = 0$$

for every $n \in \mathbb{N}$ and so

$$\mathbb{1}_{E_n} \xrightarrow{\mu} 0.$$

Now for $x \in [0, 1]$, we have that $x \in E_n$ for infinitely many $n \in \mathbb{N}$. We can establish this by noting that

$$\bigcup_{2^{k(n)} \leq n \leq 2^{k(n)} - 1} E_n = [0, 1]$$

and so for each $k \in \mathbb{N} \cup \{0\}$ there's some n_k such that $x \in E_{n_k}$. Then the subsequence $\mathbb{1}_{E_{n_k}} = 1$ and so the our sequence $\mathbb{1}_{E_n}$doesnt converge pointwise to zero anywhere in $[0, 1]$. In particular, it doesn't converge almost everywhere.

However, convergence in measure implies almost everywhere subsequential convergence. To show this, we will first establish a useful lemma.

Lemma 3.4.9. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ be such that*

$$\sum_{n=1}^{\infty} \mu\left(\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}\right) < \infty$$

for every $\epsilon > 0$ and some $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$. Then

$$f_n \xrightarrow{a.e} f.$$

Proof. Applying the Borel-Cantelli lemma, we know that for each $\epsilon > 0$

$$\mu\left(\limsup_{n\to\infty}\{x\in\mathcal{X}\mid |f_n(x)-f(x)|>\epsilon\}\right)=0.$$

In particular, this is true for $\epsilon = \frac{1}{p}$ for every $p\in\mathbb{N}$. By the Archimedean property of natural numbers and our discussion in the proof of Proposition 3.4.7, the set

$$\left\{x\in\mathcal{X}\mid \lim_{n\to\infty}f_n(x)\neq f(x)\right\}=\bigcup_{p\in\mathbb{N}}\limsup_{n\to\infty}\left\{x\in\mathcal{X}\mid |f_n(x)-f(x)|>\frac{1}{p}\right\}$$

and so by Proposition 3.3.2

$$\mu\left(\left\{x\in\mathcal{X}\mid \lim_{n\to\infty}f_n(x)\neq f(x)\right\}\right)=0$$

which completes the proof. □

Proposition 3.4.10. *Let $(\mathcal{X},\mathcal{F},\mu)$ be a measure space and let $\{f_n\}_{n\in\mathbb{N}}\in\mathcal{M}(\mathcal{X},\mathcal{F})$ be a sequence of functions such that*

$$f_n\xrightarrow{\mu}f$$

for some $f\in\mathcal{M}(\mathcal{X},\mathcal{F})$. Then there exists a sequence of natural numbers $\{n_k\}_{k\in\mathbb{N}}$ such that

$$f_{n_k}\xrightarrow{a.e}f.$$

Proof. Note that by convergence in measure, for any $\epsilon,\delta>0$ there exists some $n^*_{\epsilon,\delta}\in\mathbb{N}$ such that for all $n\geq n^*_{\epsilon,\delta}$

$$\mu\left(\{x\in\mathcal{X}\mid |f_n(x)-f(x)|>\epsilon\}\right)<\delta.$$

In particular, this is true for $\epsilon=\delta=2^{-k}$ for all $k\in\mathbb{N}$. Define the sequence $\{n_k\}_{k\in\mathbb{N}}$ by

$$n_k:=\begin{cases}n^*_{1,1}, & k=1\\ n^*_{2^{-k},2^{-k}}, & n_{k-1}<n^*_{2^{-k},2^{-k}}\\ n^*_{2^{-k},2^{-k}}+1, & n_{k-1}=n^*_{2^{-k},2^{-k}}\end{cases}$$

and observe that n_k is a strictly increasing sequence of natural numbers and that

$$\sum_{k=1}^{\infty}\mu\left(\left\{x\in\mathcal{X}\mid |f_{n_k}(x)-f(x)|>2^{-k}\right\}\right)<\sum_{k=1}^{\infty}2^{-k}=1<\infty.$$

Applying Lemma 3.4.9 furnishes the result. □

Proposition 3.4.11 (Markov's inequality). *Let $(\mathcal{X},\mathcal{F},\mu)$ be a measure space. For any function $f\in\mathcal{L}^1(\mu)$ and any $a>0$*

$$\mu\left(\{x\in\mathcal{X}\mid |f(x)|>a\}\right)\leq\frac{\bar{\mu}(|f|)}{a}.$$

Proof. Note the pointwise inequality

$$a\mathbb{1}_{\{x\in\mathcal{X}||f(x)|>a\}}\leq|f(x)|$$

and observe that by the monotonicity of the integral

$$a\bar{\mu}\left(\mathbb{1}_{\{x\in\mathcal{X}||f(x)|>a\}}\right)\leq\bar{\mu}(|f|)$$

which finishes the proof. □

Proposition 3.4.12. *For a sequence of functions $\{f_n\}_{n\in\mathbb{N}}\in\mathcal{L}^1(\mathcal{X},\mathcal{F},\mu)$ such that*

$$f_n\xrightarrow{\mathcal{L}^1}f$$

for some $f\in\mathcal{L}^1(\mu)$, we have

$$f_n\xrightarrow{\mu}f.$$

PROOF. Note that for any fixed $\epsilon > 0$

$$0 \leq \mu\left(\{x \in \mathcal{X} \mid |f_n(x) - f(x)| > \epsilon\}\right) \leq \frac{1}{\epsilon}\bar{\mu}\left(|f_n - f|\right)$$

by the Markov inequality. Taking the limit proves the result. □

Again, the converse of this result is false.

EXAMPLE 3.4.13. The sequence of functions $f_n := n\mathbb{1}_{\left(0,\frac{1}{n}\right]}$ on the measure space $\left([0,1], \mathscr{B}([0,1]), \lambda\right)$ converges pointwise to zero: for any $x \in \mathbb{R}$, eventually $\mathbb{1}(x) = 0$ for large $n$and so $f_n \to 0$ which means that $f_n \xrightarrow{\lambda} 0$ as well by Proposition 3.4.7. But $\bar{\lambda}(|f_n|) = 1$ which doesn't converge to 0.

THEOREM 3.4.14 (Egorov's theorem). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space such that $\mu(\mathcal{X}) < \infty$ and fix some $\epsilon > 0$. If $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ is a sequence functions such that*

$$f_n \xrightarrow{a.e} f$$

where $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ then there exists some $A \in \mathcal{F}$ such that $\mu(A) < \epsilon$ and

$$\lim_{n\to\infty} \sup_{x\in\mathcal{X}\setminus A} |f_n(x) - f(x)| = 0$$

PROOF. Define the set

$$A_{n,k} := \bigcup_{i=n}^{\infty} \left\{x \in \mathcal{X} \mid |f_i(x) - f(x)| > \frac{1}{k}\right\}$$

and recall from our earlier discussion that

$$\mu\left(\bigcap_{n\in\mathbb{N}} A_{n,k}\right) = 0$$

for every $k \in \mathbb{N}$ by almost everywhere convergence. Then, since $\mu(\mathcal{X}) < \infty$, we can apply the continuity from above of measures to deduce that

$$\lim_{n\to\infty} \mu(A_{n,k}) = 0.$$

This means there exists some $n_k \in \mathbb{N}$ such that for all $n \geq n_k$

$$\mu(A_{n,k}) < \frac{\epsilon}{2^k}$$

and so

$$\mu\left(\bigcup_{k\in\mathbb{N}} A_{n_k,k}\right) \leq \sum_{k=1}^{\infty} \mu(A_{n_k,k}) \leq \epsilon.$$

If we let $A := \bigcup_{k\in\mathbb{N}} A_{n_k,k}$ then

$$\mathcal{X} \setminus A = \bigcap_{k\in\mathbb{N}} \bigcap_{i=n_k}^{\infty} \left\{x \in \mathcal{X} \mid |f_i(x) - f(x)| \leq \frac{1}{k}\right\}.$$

Fix any $k \in \mathbb{N}$, and observe that for any $n \geq n_k$ and any $x \in \mathcal{X} \setminus A$

$$|f_n(x) - f(x)| \leq \frac{1}{k}$$

which means

$$\sup_{x\in\mathcal{X}\setminus A} |f_n(x) - f(x)| \leq \frac{1}{k}.$$

Since k was unspecified, we have uniform convergence. □

3.4.1. Topologies of various convergence notions.

Definition 3.4.15. Let X be a set and let

$$d : X \times X \to \mathbb{R}$$

be a function that satisfies

(i) $d(x,y) = d(y,x) \forall x,y \in X$ (Symmetry)

(ii) $d(x,z) \leq d(x,y) + d(y,z) \forall x,y,z \in X$ (Triangle inequality)

Then (X,d) is called a *pseudometric space* and d is called a *pseudo-metric.*

Proposition 3.4.16. *Let $(\mathcal{X},\mathcal{F},\mu)$ be a finite measure space i.e. $\mu(\mathcal{X}) < \infty$. Then the function*

$$d(f,g) := \bar{\mu}\left(\frac{|f-g|}{1+|f-g|}\right)$$

defines a pseudo-metric. Moreover, for any $\{f_n\}_{n\in\mathbb{N}}, f \in \mathcal{M}(\mathcal{X},\mathcal{F})$

$$f_n \xrightarrow{\mu} f$$

if and only if

$$d(f_n,f) \longrightarrow 0.$$

Proof. Note that

$$0 \leq \frac{|f-g|}{1+|f-g|} \leq 1$$

and so the function d is well defined as an integral on a finite measure space of bounded a function. Note that the symmetry of d is obvious given the symmetry of the absolute value function. To see the triangle inequality, first observe that the function $g(x) = \frac{x}{1+x}$ is increasing on $x > 0$ (it has a derivative $\frac{1}{(1+x)^2}$) and so, by the triangle inequality for absolute value

$$\begin{aligned}
\frac{|f-h|}{1+|f-h|} &\leq \frac{|f-g|+|g-h|}{1+|f-g|+|g-h|} \\
&= \frac{|f-g|}{1+|f-g|+|g-h|} + \frac{|g-h|}{1+|f-g|+|g-h|} \\
&\leq \frac{|f-g|}{1+|f-g|} + \frac{|g-h|}{1+|g-h|}.
\end{aligned}$$

Linearity and monotonicity of integration yields the triangle inequality for d. Now suppose $f_n \xrightarrow{\mu} f$, let $\epsilon > 0$ and observe that

$$\begin{aligned}
d(f_n,f) &= \bar{\mu}\left(\frac{|f_n-f|}{1+|f_n-f|}\right) \\
&= \bar{\mu}\left(\frac{|f_n-f|}{1+|f_n-f|}\mathbb{1}_{\{|f_n-f|>\epsilon\}}\right) + \bar{\mu}\left(\frac{|f_n-f|}{1+|f_n-f|}\mathbb{1}_{\{|f_n-f|\leq\epsilon\}}\right) \\
&\leq \mu\left(|f_n-f| > \epsilon\right) + \frac{\epsilon}{1+\epsilon}\mu\left(|f_n-f| \leq \epsilon\right) \\
&\leq \mu\left(|f_n-f| > \epsilon\right) + \frac{\epsilon}{1+\epsilon}\mu\left(\mathcal{X}\right).
\end{aligned}$$

Therefore by convergence in measure

$$\lim_{n\to\infty} d(f_n,f) \leq \frac{\epsilon}{1+\epsilon}\mu\left(\mathcal{X}\right).$$

Of course, ϵ can be arbitrarily small and so we have convergence in the pseudo-metric.

Conversely, suppose $\lim_{n\to\infty} d(f_n,f) = 0$ and fix some $\epsilon > 0$. Then

$$0 \leq \frac{\epsilon}{1+\epsilon}\mu\left(|f_n-f| > \epsilon\right) \leq \bar{\mu}\left(\frac{|f_n-f|}{1+|f_n-f|}\mathbb{1}_{\{|f_n-f|>\epsilon\}}\right) \leq \bar{\mu}\left(\frac{|f_n-f|}{1+|f_n-f|}\right)$$

and in the limit
$$\frac{\epsilon}{1+\epsilon}\mu\left(|f_n - f| > \epsilon\right) \longrightarrow 0 \implies \mu\left(|f_n - f| > \epsilon\right) \longrightarrow 0$$
which completes the proof. □

Note that $d(f,g) = 0 \iff f \overset{\text{a.e}}{=} g$. This is a simple corrolary of Proposition 3.3.3. Propositions 3.4.10 and 3.4.7 gives us a sort of equivalence between almost everywhere equivalence as a corollary of a basic fact about sequences in topological spaces.

LEMMA 3.4.17. *Let $(\mathcal{X},\tau)$ be a topological space. The sequence$\{x_n\}_{n\in\mathbb{N}} \in \mathcal{X}$ converges to $x \in \mathcal{X}$ if and only if every subsequence of x_n has a further subsequence that converges to x.*

PROOF. Convergence implies subsequential convergence for every subseqeuence (and in particulary, subsubsequences). Conversely, suppose that $x_n \nrightarrow x$. Then there's an open set $\mathcal{U} \in \tau$ that contains x such that there are infinitely many elements of x_n outside $\mathcal{U}$. In other words, there exists some subsequence $x_{n_k} \notin \mathcal{U}$ for all $k \in \mathbb{N}$. Then no subsequence of x_{n_k} converges to x. □

PROPOSITION 3.4.18. *Let $(\mathcal{X},\mathcal{F},\mu)$ be a finite measure space i.e $\mu(\mathcal{X}) < \infty$. Then for any sequence $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X},\mathcal{F})$ and $f \in \mathcal{M}(\mathcal{X},\mathcal{F})$*
$$f_n \xrightarrow{\mu} f$$
if and only if every subsequence f_{n_k} has a further subsequence $f_{n_{k_j}}$ such that
$$f_{n_{k_j}} \xrightarrow{a.e} f.$$

PROOF. Suppose $f_n \xrightarrow{\mu} f$ and let f_{n_k} be an arbitrary subsequence. This subsequence also converges to f in measure and so by Proposition 3.4.10 there's some sub-subsequence $f_{n_{k_j}} \xrightarrow{\text{a.e}} f$. Conversely, suppose every subsequence f_{n_k} has a further subsequence $f_{n_{k_j}} \xrightarrow{\text{a.e}} f$. Then by Proposition 3.4.7 $f_{n_{k_j}} \xrightarrow{\mu} f$. By Proposition 3.4.18, convergence in measure is (pseudo)-metrizable and thus topological and so by Lemma 3.4.17 $f_n \xrightarrow{\mu} f$. □

COROLLARY 3.4.19. *There is no topology for almost everywhere convergence.*

PROOF. Note that if almost everywhere convergence was topological, then by Lemma 3.4.17 and Proposition 3.4.18, $f_n \xrightarrow{\mu} f \implies f_n \xrightarrow{\text{a.e}} f$. But this has been ruled out by Example 3.4.8. □

Thus we have established that the topology of convergence in measure comes from a (pseudo)-metric, whereas almost everywhere convergence is not topologizable at all! Convergence in L^1 , on the other hand, corresponds to a (semi)-normed convergence as shown in Chapter 5. Proposition 3.4.18 implies a form of dominated convergence theorem for convergence in measure.

PROPOSITION 3.4.20. *Let $(\mathcal{X},\mathcal{F},\mu)$ be a finite measure space i.e $\mu(\mathcal{X}) < \infty$. Suppose $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X},\mathcal{F})$ such that $|f_n| \le g$ where $g \in \mathcal{L}^1(\mathcal{X},\mathcal{F},\mu)$. Then*
$$f_n \xrightarrow{\mu} f \implies f_n \xrightarrow{\mathcal{L}^1} f.$$

PROOF. Note that if $f_n \xrightarrow{\mu} f$, then for any subsequence f_{n_j} of f_n there exists a further subseqeunce $f_{n_{j_k}}$ such that
$$f_{n_{j_k}} \xrightarrow{\text{a.e}} f.$$
Since $|f_{n_{j_k}}| \le g \in \mathcal{L}^1(\mu)$, the dominated convergence theorem implies that
$$f_{n_{j_k}} \xrightarrow{\mathcal{L}^1} f.$$
But L^1 convergence is topological and so by Lemma 3.4.17,
$$f_n \xrightarrow{\mathcal{L}^1} f.$$
□

3.4.2. Uniform integrability and uniform absolute continuity. The dominated convergence theorem says that for sequence of random variables that are dominated by an integrable function, pointwise convergence implies convergence in L^1. We would like to extend this notion more broadly.

DEFINITION 3.4.21. Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $\mathcal{C} \subseteq \mathcal{M}(\mathcal{X}, \mathcal{F})$ be a collection of measurable functions. We say $\mathcal{C}$ is *uniformly integrable* if for every $\epsilon > 0$, there exists some $M \in \mathbb{N}$ such that

$$\bar{\mu}\left(|f| \mathbb{1}_{\{|f|>M\}}\right) < \epsilon$$

for every $f \in \mathcal{C}$.

DEFINITION 3.4.22. Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $\mathcal{C} \subseteq \mathcal{M}(\mathcal{X}, \mathcal{F})$ be a collection of measurable functions. We say $\mathcal{C}$ is *uniformly absolutely continuous* if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for any $f \in \mathcal{C}$

$$F \in \mathcal{F}, \mu(F) < \delta \implies |\bar{\mu}(f \mathbb{1}_F)| < \epsilon.$$

REMARK 3.4.23. Note that if $\mathcal{C}$ is a trivial class i.e. $\mathcal{C} = \{f\}$, then the absolute continuity property always holds as observed in Proposition 3.2.10 and Remark 3.2.11. Of course, it is not clear that this $\epsilon - \delta$ characterization of absolute continuity is the same as the one discussed in that remark; later on in Chapter 6 (see Proposition 6.2.3) we show they are indeed equiavelent for finite measures.

PROPOSITION 3.4.24. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a finite measure space i.e. $\mu(\mathcal{X}) < \infty$. Then a class of functions $\mathcal{C} \subseteq \mathcal{M}(\mathcal{X}, \mathcal{F})$ is uniformly integrable if and only if it is uniformly absolutely continuous and*

$$\sup_{f \in \mathcal{C}} \bar{\mu}(|f|) < \infty.$$

PROOF. Suppose $\mathcal{C} \subseteq \mathcal{M}(\mathcal{X}, \mathcal{F})$ is uniformly integrable. Fix some $\epsilon > 0$ and observe that by uniform integrability there is some $M \in \mathbb{N}$ such that

$$\bar{\mu}\left(|f| \mathbb{1}_{\{|f|>M\}}\right) < \frac{\epsilon}{2}.$$

Then for any $F \in \mathcal{F}$,

$$\begin{aligned}
|\bar{\mu}(f \mathbb{1}_F)| &\leq \bar{\mu}(|f| \mathbb{1}_F) \\
&= \bar{\mu}\left(|f|\left(\mathbb{1}_{\{|f|>M\}} + \mathbb{1}_{\{|f| \leq M\}}\right) \mathbb{1}_F\right) \\
&= \bar{\mu}\left(|f| \mathbb{1}_{\{|f|>M\}} \mathbb{1}_F\right) + \bar{\mu}\left(|f| \mathbb{1}_{\{|f| \leq M\}} \mathbb{1}_F\right) \\
&\leq \bar{\mu}\left(|f| \mathbb{1}_{\{|f|>M\}}\right) + \bar{\mu}\left(|f| \mathbb{1}_{\{|f| \leq M\}} \mathbb{1}_F\right) \\
&\leq \bar{\mu}\left(|f| \mathbb{1}_{\{|f|>M\}}\right) + \bar{\mu}\left(M \mathbb{1}_{\{|f| \leq M\}} \mathbb{1}_F\right) \\
&\leq \frac{\epsilon}{2} + M\mu(\{|f| \leq M\} \cap F) \\
&\leq \frac{\epsilon}{2} + M\mu(F)
\end{aligned}$$

where all the inequalities follow by the monotonicity of the integral. Letting $\delta = \frac{\epsilon}{2M}$ then shows uniform absolute continuity. Letting $F = \mathcal{X}$ then yields

$$\bar{\mu}(|f|) \leq \frac{\epsilon}{2} + M\mu(\mathcal{X})$$

for any $f \in \mathcal{C}$ which gives us the boundedness in L^1.

Conversely, assume that $\mathcal{C}$ is uniformly absolutely continuous and $\sup_{f \in \mathcal{C}} \bar{\mu}(|f|) < \infty$. Let $\epsilon > 0$ and note that there exists some $\delta > 0$ such that $F \in \mathcal{F}, \mu(F) < \delta \implies |\bar{\mu}(f \mathbb{1}_F)| < \epsilon$. Note that by Markov's inequality

$$\mu(|f| > M) \leq \frac{\bar{\mu}(|f|)}{M}.$$

for any $M \in \mathbb{N}$. Since $\bar{\mu}(|f|) < \infty$,we can take M large enough to so that $\mu(|f| > M) < \delta$ which by uniform absolute continuity implies $\left|\bar{\mu}\left(|f| \mathbb{1}_{\{|f|>M\}}\right)\right| = \bar{\mu}\left(|f| \mathbb{1}_{\{|f|>M\}}\right) < \epsilon$. This shows uniform integrability, completing the proof. ☐

Note that uniform integrability generalizes the notion of the domination by an integrable function, which is an essential ingredient for the dominated convergence theorem.

PROPOSITION 3.4.25. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $\mathcal{C} \subseteq \mathcal{M}(\mathcal{X}, \mathcal{F})$ be a collection of measurable functions. If for every $f \in \mathcal{C}$, we have that $|f| \leq g$ such that $g \in \mathcal{L}^1(\mu)$ then $\mathcal{C}$ is uniformly integrable.*

PROOF. Let $\epsilon > 0$ be fixed. Since g is integrable, Proposition 3.3.6 tells us that $|g| < \infty$ almost everywhere. Then

$$\mathbb{1}\{|g| > M\} \leq \frac{|g|}{M} \overset{\text{a.e}}{\to} 0$$

and so

$$|g|\mathbb{1}\{|g| > M\} \overset{\text{a.e}}{\to} 0.$$

Since $|g| \geq |g|\mathbb{1}\{|g| > M\}$, dominated convergence implies that

$$\bar{\mu}(|g|\mathbb{1}\{|g| > M\}) < \epsilon$$

for some large enough M. Then, since $|f| \leq |g|$,

$$\bar{\mu}(|f|\mathbb{1}\{|f| > M\}) \leq \bar{\mu}(|g|\mathbb{1}\{|g| > M\}) < \epsilon.$$

□

Classes of uniformly integrable functions are preserved under linear combinations.

PROPOSITION 3.4.26. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $c \in \mathbb{R}$ be a scalar. Further, let $\mathcal{C}, \mathcal{D} \subseteq \mathcal{M}(\mathcal{X}, \mathcal{F})$ be uniformly integrable collections of functions. Then*

$$\mathcal{C} + c\mathcal{D} := \{f + cg \mid f \in \mathcal{C}, g \in \mathcal{D}\}$$

is a uniformly integrable collection.

PROOF. Note that when $c = 0$ then the result is triviall true so let $c \neq 0$. Fix $\epsilon > 0$ and note that there exists some M such that for any $g \in \mathcal{D}$

$$\bar{\mu}\left(|g|\mathbb{1}_{\{|g|>M\}}\right) < \frac{\epsilon}{|c|}.$$

Then, for $N = M|c|$

$$\bar{\mu}\left(|cg|\mathbb{1}_{\{|cg|>N\}}\right) = |c|\bar{\mu}\left(|g|\mathbb{1}_{\{|g|>M\}}\right) < \epsilon.$$

Next, observe that COMPLETE LATER □

In the main result for this section, we generalize the dominated convergence theorem.

THEOREM 3.4.27 (Vitali's Convergence Theorem). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a finite measure space i.e. $\mu(\mathcal{X}) < \infty$. If $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ are uniformly integrable and $f_n \xrightarrow{\mu} f$ then $f \in \mathcal{L}^1(\mu)$ and*

$$f_n \xrightarrow{\mathcal{L}^1} f.$$

PROOF. Note that if $f_n \xrightarrow{\mu} f$ then by Proposition 3.4.10 there exists some subsequence $f_{n_j} \xrightarrow{\text{a.e}} f$. By Fatou's lemma and the fact that absolute value is continuous, we have

$$\bar{\mu}(|f|) = \bar{\mu}\left(\liminf_{j\to\infty}|f_{n_j}|\right) \leq \liminf_{j\to\infty} \bar{\mu}\left(|f_{n_j}|\right) \leq \sup_{n\in\mathbb{N}} \bar{\mu}(|f_n|) < \infty$$

where the last inequality follows by Proposition 3.4.24, which shows $f \in \mathcal{L}^1(\mu)$. Since $\{f_n\}_{n\in\mathbb{N}}, \{f\}$ are uniformly integrable, $g_n := f_n - f$ is uniormly integrable by Proposition 3.4.26. Note that by uniform absolute continuity, for any $\epsilon > 0$, there exists some $\delta > 0$ such that for any $A \in \mathcal{F} : P(A) < \delta \implies \bar{\mu}(|g_n|\mathbb{1}_A) < \frac{\epsilon}{2}$. Note that by convergence in probability, there exists some $n_0 \in \mathbb{N}$ $\mu\left(|g_n| > \frac{\epsilon}{2}\right) < \delta$ and so

$$\bar{\mu}(|g_n|) = \bar{\mu}\left(|g_n|\mathbb{1}_{\{|g_n|\leq\frac{\epsilon}{2}\}}\right) + \bar{\mu}\left(|g_n|\mathbb{1}_{\{|g_n|>\frac{\epsilon}{2}\}}\right) \leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

which shows L^1 convergence. □

3.5. The Riemann integral

The theory of integration that we have developed so far is only useful if (1) it is, in some sense, a suitable generalization of Riemann's theory of integration, and (2) if it obeys the fundamental theorem of calculus. We leave (2) to section **??** and in this section review and characterize the Riemann integral. In particular, we give a precise characterization of Riemann integrable functions using the Lebesgue theory we have developed so far, and show that for such functions, their Lebesgue integrals (with respect to the Lebesgue measure) and Riemann integrals coincide.

We first consider the classical construction of the Riemann integral, as articulated by Darboux and Riemann. After we have developed this baisc material carefully, we can present the Riemann integral under the more modern Lebesgue formulation. We do this by, a more primitive version of the Lebesgue measure, called the *Jordan content* or *Jordan measure*, which coincides with the Lebesgue measure for sufficiently well behaved sets.

3.5.1. The classical construction of the Riemann integral.

Definition 3.5.1. Let $[a,b] \subset \mathbb{R}$ be a closed and bounded interval. A *partition* π of $[a,b]$ is a finite set

$$\pi = \{t_i \mid t_0 = a, t_k = b, t_i < t_{i+1}\}$$

where $k \in \mathbb{N}$ is the *size* of the partition. The size of a partition $\pi$is often denoted $k(\pi)$.

A second notion of size is given by th concept of a mesh (which is closely related to, and a special case of, the Lebesgue measure).

Definition 3.5.2. Given a partition $\pi$of $[a,b]$, the *mesh* of $\pi$is given by

$$\operatorname{mesh}(\pi) := \max_{1 \le i \le k} t_i - t_{i-1}.$$

As one can see, the mesh is basically given by $\max_i \lambda([t_{i-1}, t_i])$ where λ is the Lebesgue measure. We now introduce the notion of a *Darboux* sum, which are the building blocks of Riemann integrals much in the way that integrals of simple functions are in the Lebesgue theory. In fact, Darboux sums are the Lebesgue integrals of a special type of signed simple function called a step function.

Definition 3.5.3. Let $f : [a,b] \to \mathbb{R}$ be a bounded function and let π be a partition of $[a,b]$ with size k. Then the *upper Darboux sum* is given by

$$U(f,\pi) := \sum_{i=1}^{k} \sup_{x \in [t_{i-1}, t_i)} f(x)(t_i - t_{i-1})$$

whereas the *lower Darboux sum* is given

$$L(f,\pi) := \sum_{i=1}^{k} \inf_{x \in [t_{i-1}, t_i)} f(x)(t_i - t_{i-1}).$$

Proposition 3.5.4. *Let $\pi' \supseteq \pi$ be a refinement of π; that is a partition of $[a,b]$ that contains the coarser partition π as a subset. Then*

$$L(f,\pi) \le L(f,\pi') \le U(f,\pi') \le U(f,\pi).$$

Proof. Note that the second inequality is obvious and so we focus on proving the first (the third inequality is analagous). First suppose $\pi' \setminus \pi = \{t^*\}$; that is, the refinement contains only one additional element t^*. Then, there is some $1 \le i \le k$ such that $t_{i-1} < t^* < t_i$. Then

$$\begin{aligned}\inf_{x \in [t_{i-1}, t_i)} f(x)(t_i - t_{i-1}) &\le \min\left\{\inf_{x \in [t_{i-1}, t^*)} f(x), \inf_{x \in [t^*, t_i)} f(x)\right\}(t_i - t_{i-1}) \\ &\le \inf_{x \in [t_{i-1}, t^*)} f(x)(t^* - t_{i-1}) + \inf_{x \in [t^*, t_i)} f(x)(t_i - t^*)\end{aligned}$$

and so the first inequality follows since

$$\begin{aligned}L(f,\pi) &= \sum_{j\neq i,1\leq j\leq k}\inf_{x\in[t_{j-1},t_j)} f(x)(t_j-t_{j-1}) + \inf_{x\in[t_{i-1},t_i)} f(x)(t_i-t_{i-1}) = L(f,\pi').\\ &\leq \sum_{j\neq i,1\leq j\leq k}\inf_{x\in[t_{j-1},t_j)} f(x)(t_j-t_{j-1}) + \inf_{x\in[t_{i-1},t^*)} f(x)(t^*-t_{i-1}) + \inf_{x\in[t^*,t_i)} f(x)(t_i-t^*)\\ &= L(f,\pi')\end{aligned}$$

The general case then follows by forming nested sequence $\pi\subset\pi_1\subset\ldots\subset\pi_m\subset\pi'$, each of which contain one additional element, and then applying the above result in sequence. □

DEFINITION 3.5.5. A bounded real-valued function $f:[a,b]\to\mathbb{R}$is said to be *Riemann integrable* (or just *integrable,* in context) if

$$L^*(f) := \sup_\pi L(f,\pi) = \inf_\pi U(f,\pi) =: U^*(f)$$

where the supremum and infimum are taken over the collection of all finite partitions π of $[a,b]$. The *Riemann integral* of an integrable f on $[a,b]$ is then denoted

$$\int_a^b f(x)\,dx := L^*(f) = U^*(f).$$

The collection of Riemann integrable functions on $[a,b]$ is denoted $\mathcal{R}[a,b]$.

PROPOSITION 3.5.6. *Let $f:[a,b]\to\mathbb{R}$ be a bounded function. Then*

$$L^*(f)\leq U^*(f)$$

and

$$L^*(f) = -U^*(-f).$$

PROOF. For the first claim, note that for any partitions $\pi$and π' , Proposition 3.5.4 implies that

$$L(f,\pi)\leq L(f,\pi'\cup\pi)\leq U(f,\pi'\cup\pi)\leq U(f,\pi')$$

and so

$$L(f,\pi)\leq\inf_{\pi'} U(f,\pi')$$

and similarly

$$\sup_\pi L(f,\pi)\leq\inf_{\pi'} U(f,\pi').$$

For the second claim, notice that $\inf_{x\in[t_{i-1},t_i)} f(x) = -\sup_{x\in[t_{i-1},t_i)} -f(x)$ and so for any partition π

$$L(f,\pi) = -U(-f,\pi)$$

and the result follows by noting that

$$L^*(f) = \sup_\pi L(f,\pi) = \sup_\pi -U(-f,\pi) = -\inf_\pi U(-f,\pi) = -U^*(-f).$$

□

THEOREM 3.5.7. *The following are equivalent for any bounded function $f:[a,b]\to\mathbb{R}$*

(i) $f\in\mathcal{R}[a,b]$

(ii) For every $\epsilon>0$ there exists some partition π such that

$$U(f,\pi)-L(f,\pi)<\epsilon.$$

(iii) For every $\epsilon>0$ there exists some $\delta>0$ such that for any partition π with $\operatorname{mesh}(\pi)<\delta$

$$U(f,\pi)-L(f,\pi)<\epsilon$$

(iv) The limit

$$I := \lim_{\text{mesh}(\pi)\to 0} \sum_{i:t_i\in\pi} f(c_i)(t_i - t_{i-1})$$

exists for any choice of intermediate values $t_{i-1} \leq c_i \leq t_i$*. Further, this limit is equal to* $I = L^*(f) = U^*(f)$*.*

Proof. We first prove that (i) $\implies$ (ii). Suppose that $f \in \mathcal{R}[a,b]$, and so $L^*(f) = U^*(f)$. Fix $\epsilon > 0$ and observe that by the definition of supremum and infimum, there exist partitions π and π' such that $L^*(f) - L(f,\pi) < \frac{\epsilon}{2}$ and $U(f,\pi') - L^*(f) < \frac{\epsilon}{2}$. Then clearly, for the common refinement $\pi \cup \pi'$

$$|L(f,\pi\cup\pi') - U(f,\pi\cup\pi')| \leq |L(f,\pi\cup\pi') - L^*(f)| + |L^*(f) - U(f,\pi\cup\pi')| < \epsilon.$$

Next, we prove (ii) $\implies$ (iii). Fix $\epsilon > 0$ and note that by (ii) there exists some partition $\pi := \{a = t_0 < t_1 < \ldots < t_k = b\}$such that

$$U(f,\pi) - L(f,\pi) = \sum_{i=1}^{k}\left(\sup_{x\in[t_{i-1},t_i)} f(x) - \inf_{x\in[t_{i-1},t_i)} f(x)\right)(t_i - t_{i-1}) < \frac{\epsilon}{2}. \tag{5}$$

Now $\pi' := \{a = t'_0 < t'_1 < \ldots < t'_l = b\}$ be a partition such that $\text{mesh}(\pi) < \delta := \frac{\epsilon}{4k\|f\|_\infty}$ where $\|f\|_\infty = \sup_{x\in[a,b]}|f(x)|$. Then for each $t_i \in \pi$, there can be at most one interval $\left[t'_{j-1}, t'_j\right)$ with $1 \leq j \leq l$ that contains t_i since these intervals form a partition of $[a,b]$. Thus the set $S := \left\{j \in \{1,\ldots,l\} \mid \exists i \in \{0,1,\ldots,k\} \text{ s.t } t_i \in \left[t'_{j-1}, t'_j\right)\right\}$ has size at most k. Now consider the difference of Darboux sums

$$\begin{aligned}
U(f,\pi') - L(f,\pi') &= \sum_{j=1}^{l}\left(\sup_{x\in\left[t'_{j-1},t'_j\right)} f(x) - \inf_{x\in\left[t'_{j-1},t'_i\right)} f(x)\right)\left(t'_j - t'_{j-1}\right) \\
&= \sum_{j\in S}\left(\sup_{x\in\left[t'_{j-1},t'_j\right)} f(x) - \inf_{x\in\left[t'_{j-1},t'_i\right)} f(x)\right)\left(t'_j - t'_{j-1}\right) \\
&\quad + \sum_{j\in S^C}\left(\sup_{x\in\left[t'_{j-1},t'_j\right)} f(x) - \inf_{x\in\left[t'_{j-1},t'_i\right)} f(x)\right)\left(t'_j - t'_{j-1}\right).
\end{aligned}$$

We control each of these sums separately. For the first sum, note that there at most k terms in the sum, where $\left(\sup_{x\in\left[t'_{j-1},t'_j\right)} f(x) - \inf_{x\in\left[t'_{j-1},t'_i\right)} f(x)\right) \leq 2\|f\|_\infty$ and $\left(t'_j - t'_{j-1}\right) < \delta$ for every $j \in S$. Thus the first term is bounded above by $\frac{\epsilon}{2}$. Next, note that for the second sum, by definition for every $j \in S^C$, there exists some $i_j \in \{1,\ldots,k\}$ such that $\left[t'_{j-1}, t'_j\right) \subset \left[t_{i_j-1}, t_{i_j}\right)$ and that

$$\sum_{j\in S^C}\left(\sup_{x\in\left[t'_{j-1},t'_j\right)} f(x) - \inf_{x\in\left[t'_{j-1},t'_i\right)} f(x)\right)\left(t'_j - t'_{j-1}\right) \leq \sum_{j\in S^C}\left(\sup_{x\in\left[t_{i_j-1},t_{i_j}\right)} f(x) - \inf_{x\in\left[t_{i_j-1},t_{i_j}\right)} f(x)\right)\left(t_{i_j} - t_{i_j-1}\right) < \frac{\epsilon}{2}$$

where the final inequality is due to (5). This means

$$|U(p,\pi') - L(f,\pi')| < \epsilon$$

which completes this part of the proof.

Next, assume that (iii) holds and let $\epsilon > 0$. Observe that there exists some $\delta > 0$ such that for any partition $\pi := \left\{a = t_0 < t_1 < \ldots t_{k(\pi)} = b\right\}$ with mesh less than δ,

$$0 \leq \sum_{i=1}^{k(\pi)} f(c_i)(t_i - t_{i-1}) - L(f,\pi) \leq U(f,\pi) - L(f,\pi) < \frac{\epsilon}{2}$$

where $c_i \in [t_{i-1}, t_i)$. Now note that since $L(f,\pi_n) \leq L^*(f) \leq U^*(f) \leq U(f,\pi_n)$, we have that $L^*(f) - L(f,\pi_n) < \frac{\epsilon}{2}$ (which shows that $L(f,\pi_n) \to L^*(f)$). Then

$$\left|\sum_{i=1}^{k_n} f(c_i^n)\left(t_i^n - t_{i-1}^n\right) - L^*(f)\right| \leq \left|\sum_{i=1}^{k_n} f(c_i^n)\left(t_i^n - t_{i-1}^n\right) - L(f,\pi_n)\right| + |L(f,\pi) - L^*(f)| < \epsilon.$$

Similarly, we know tht

$$U(f,\pi) - \sum_{i=1}^{k(\pi)} f(c_i)(t_i - t_{i-1}) \leq U(f,\pi) - L(f,\pi) < \frac{\epsilon}{2}$$

and so $U(f,\pi) - U^*(f) < \frac{\epsilon}{2}$ and

$$\left|\sum_{i=1}^{k_n} f(c_i^n)\left(t_i^n - t_{i-1}^n\right) - U^*(f)\right| \leq \left|\sum_{i=1}^{k_n} f(c_i^n)\left(t_i^n - t_{i-1}^n\right) - U(f,\pi_n)\right| + |U(f,\pi) - U^*(f)| < \epsilon$$

and so the uniqueness of limits shows (iv).

Finally, Suppose that (iv) hods i.e. that $I = \lim_{\text{mesh}(\pi)\to 0} \sum_{i=1}^{k(\pi)} f(c_i^\pi)\left(t_i^\pi - t_{i-1}^\pi\right)$ exists and is independent of any choice $t_{i-1}^\pi \leq c_i^\pi < t_i^\pi$. Fix $\epsilon > 0$ and notice that by assumption there exists some $\delta > 0$ such that for any partition π with $\text{mesh}(\pi) < \delta$

$$\left|I - \sum_{i=1}^{k(\pi)} f(c_i^\pi)\left(t_i^\pi - t_{i-1}^\pi\right)\right| < \frac{\epsilon}{3}.$$

Crucially, the choice of intermediate values c_i^π *does not* influence our choice of δ. Then, by Proposition 3.5.4 and the definition of a supremum, we can actually choose a partition π_0 fine enough, such that its mesh is less than δ and

$$|L(f,\pi_0) - L^*(f)| < \frac{\epsilon}{3}.$$

Finally, notice that by the definition of infimum, we can choose the intermediates $c_i^{\pi_0}$ such that that $f(c_i^{\pi_0}) - \inf_{x\in\left[t_{i-1}^{\pi_0}, t_i^{\pi_0}\right)} f(x) < \frac{\epsilon}{3k(\pi_0)\text{mesh}(\pi_0)}$ and so

$$|I - L^*(f)| \leq \left|I - \sum_{i=1}^{k(\pi_0)} f(c_i^{\pi_0})\left(t_i^{\pi_0} - t_{i-1}^{\pi_0}\right)\right| + \left|\sum_{i=1}^{k(\pi_0)} f(c_i^{\pi_0})\left(t_i^{\pi_0} - t_{i-1}^{\pi_0}\right) - L(f,\pi_0)\right| + |L(f,\pi_0) - L^*(f)|$$
$$<\epsilon.$$

Since $\epsilon$is arbitrary, we have $I = L^*(f)$. A similar argument shows that $I = U^*(f)$ which completes the proof. ☐

Note that the existence of *particular* sums of the form $\sum f(c_i^\pi)\left(t_i^\pi - t_{i-1}^\pi\right)$ that converge to a limit when the mesh goes to zero does not imply the integrability of f. We need the convergence to hold for *any* choice of intermediate values and *any* set of partitions with mesh going to zero. This is highlighted in the next example, which is also the canonical example of a function that is not Riemann integrable.

Example 3.5.8. Let $f : [0,1] \to \mathbb{R}$ be given by $f(x) = \mathbb{1}\{x \in [0,1] \cap \mathbb{Q}\}$.Then $L(f,\pi) = 0$ for any partition πgiven the density of the rationals and the irrationals in $[0,1]$. Therefore $L^*(f) = 0$. A similar argument shows that $U^*(f) = 1$ and so f is not integrable. On the other hand, the Riemann sums given by

$$\sum_{i=1}^{k} f(c_i)\left(\frac{i}{k} - \frac{i-1}{k}\right) = \frac{1}{k}\sum_{i=1}^{k} f(c_i)$$

can be made to converge to 0, for instance, by choosing all intermediate points as irrational (this is possible to do again by the density of irrationals). Nevertheless the integral does not exist.

EXAMPLE 3.5.9. For $k \geq 1$, let

$$a_k = \lim_{n \to \infty} \frac{1}{n} \sum_{m=1}^{kn} \exp\left(-\frac{1}{2}\frac{m^2}{n^2}\right).$$

Find $\lim_{k \to \infty} a_k$.TODO

THEOREM 3.5.10. *Suppose $f : [a, b] \to \mathbb{R}$ is Riemann integrable then $f \in L^1([a, b], \mathscr{B}([a, b]), \lambda)$ and*

$$\bar{\lambda}(f) = L^*(f) = U^*(f).$$

PROOF. Let $\{\pi_i\}_{i=1}^{\infty}$ be a sequences of partitions such that $\pi_i \subset \pi_{i+1}$ and $\operatorname{mesh}(\pi_i) \to 0$. Notice we can define the Darboux functions – which are step functions – as

$$l(f, \pi_i)(x) := \sum_{j=1}^{k(\pi_i)} \inf_{c \in \left[t_{j-1}^{\pi_i}, t_j^{\pi_i}\right)} f(c) \mathbb{1}\left\{x \in \left[t_{j-1}^{\pi_i}, t_j^{\pi_i}\right)\right\}$$

and

$$u(f, \pi_i)(x) := \sum_{j=1}^{k(\pi_i)} \sup_{c \in \left[t_{j-1}^{\pi_i}, t_j^{\pi_i}\right)} f(c) \mathbb{1}\left\{x \in \left[t_{j-1}^{\pi_i}, t_j^{\pi_i}\right)\right\}.$$

For brevity we will write $l_i := l(f, \pi_i)$ and $u_i := u(f, \pi_i)$. Note then that for all $i \in \mathbb{N}$

$$l_i \leq l_{i+1} \leq \sup_i l_i \leq f \leq \inf_i u_i \leq u_{i+1} \leq u_i$$

pointwise by our construction and the nested nature of the partitions π_i (see Proposition 3.5.4 for the argument). Notice further that step functions are simple functions on a finite measure space and hence integrable. Moreover, since f is bounded (Riemann integrable functions are bounded by definition), $\max\{|u_i|, |l_i|\} \leq \|f\|_\infty$ and so by the dominated convergence theorem

$$\bar{\lambda}(l_i) \to \bar{\lambda}(\sup l_i)$$

and

$$\bar{\lambda}(u_i) \to \bar{\lambda}(\inf u_i).$$

But notice that $\bar{\lambda}(l_i) = L(f, \pi_i)$ and $\bar{\lambda}(u_i) = U(f, \pi_i)$ and so by Riemann integrability

$$\bar{\lambda}(\sup l_i) = L^*(f) = U^*(f) = \bar{\lambda}(\inf u_i).$$

Note further that $\inf_i u_i - \sup_i l_i \geq 0$ and so $\bar{\lambda}(\inf_i u_i - \sup_i l_i) = 0$ implies that $\inf_i u_i \overset{\text{a.e}}{=} \sup_i l_i$ by Proposition 3.3.9. Of course, then $f \overset{\text{a.e}}{=} \inf_i u_i$ and so f is measurable (and Lebesgue integrable since its bounded in a finite measure space). Moreover, Proposition 3.3.7 implies

$$\bar{\lambda}(f) = L^*(f) = U^*(f)$$

which completes the proof. □

COROLLARY 3.5.11. *Let $f : [a, b] \to \mathbb{R}$ be a bounded functon. Then f is Riemann integrable if and only if for any sequence of nested partitions $\{\pi_i\}_{i=1}^{\infty}$ with $\pi_i \subseteq \pi_{i+1}$ and $\operatorname{mesh}(\pi_i) \to 0$*

$$\sup_i l(f, \pi_i) \overset{a.e}{=} \inf_i u(f, \pi_i).$$

PROOF. Note that the "only if" direction follows from Proposition 3.5.10. To see the converse, fix $\epsilon > 0$ and let $N := \{x \in [a, b] \mid \sup_i l_i(x) \neq \inf_i u_i(x)\}$. Note that for any $x \in [a, b] \setminus N$ we have pointwise convergence

$$\lim_{i \to \infty} l_i(x) = \sup_i l_i(x)$$

and

$$\lim_{i \to \infty} u_i(x) = \inf_i u_i(x).$$

Then

$$\lim_{i\to\infty}\underbrace{u_i(x)-l_i(x)}_{=:g_i(x)}=0$$

for $x\in[a,b]\setminus N$. But notice that $|g_i|\le 2\|f\|_\infty$ and so by (generalized) dominated convergence

$$U^*(f)=L^*(f)$$

which completes the proof. □

Now we are finally ready to characterize Riemann integrability in terms of continuity.

THEOREM 3.5.12. *Let $f:[a,b]\to\mathbb{R}$ be a bounded function. Then f is Riemann integrable if and only if f is continuous almost everywhere.*

PROOF. Let $\{\pi_i\}_{i=1}^{\infty}$, l_i, and u_i be defined as before and let $c\in[a,b]$ be arbitrary. First, suppose that f is Riemann integrable. Fix $\epsilon>0$ and notice that by the definition of a partition, for every $i\in\mathbb{N}$ there exists some $1\le j_i\le k(\pi_i)$ such that $c\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)$. Further, note that there is some $i_c\in\mathbb{N}$ such that for all $i\ge i_c$

$$u_i(c)-\inf_i u_i(c)<\frac{\epsilon}{2}$$

and

$$\sup_i l_i(c)-l_i(c)<\frac{\epsilon}{2}.$$

Adding the two inequalities and re-arranging, we have that

$$u_i(c)-l_i(c)<\epsilon+\inf_i u_i(c)-\sup_i l_i(c).$$

For any x such that $|x-c|<\delta:=\frac{\min\left\{t_{j_{i_c}}^{\pi_{i_c}}-c,\,c-t_{j_{i_c}-1}^{\pi_{i_c}}\right\}}{2}$ we have that $x,c\in\left[t_{j_{i_c}-1}^{\pi_{i_c}},t_{j_{i_c}}^{\pi_{i_c}}\right)$ and so

$$|f(x)-f(c)|\le u_{i_c}(c)-l_{i_c}(c)<\epsilon+\inf_i u_i(c)-\sup_i l_i(c).$$

Thus f is continuous at c if $c\notin N$ which is a null set if and only if f is Riemann integrable. So we have proved that a Riemann integrable function is continuous almost everywhere.

Conversely, suppose that the set of discontinuities $D:=\{x\in[a,b]\mid f \text{ is not continuous at } x\}$ is such that $\lambda(D)=0$. Then, for any $c\in[a,b]\setminus D$ and any $\epsilon>0$, there is a $\delta_c>0$ such that $|x-c|<\delta_c\implies|f(x)-f(c)|<\frac{\epsilon}{4}$. Let $\{\pi\}_{i=1}^{\infty}$ be as before, and note that

$$\inf_i u_i(c)-\sup_i l_i(c)\le u_i(c)-l_i(c)=\sup_{x\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)}f(x)-\inf_{x\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)}f(x)$$

for any $i\in\mathbb{N}$where j_i and $t_{j_i}^{\pi_i}$ etc are defined as before. Note that there exists some i_0 such that for all $i\ge i_0$ $\operatorname{mesh}(\pi_i)<\delta_c$. Then, note that we can find $y,z\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)$ such that

$$\sup_{x\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)}f(x)-f(y)<\frac{\epsilon}{4}$$

and

$$f(z)-\inf_{x\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)}f(x)<\frac{\epsilon}{4}.$$

Further, $|y-c|<\delta_c$ and $|z-c|<\delta_c$ and so

$$\begin{aligned}\inf_i u_i(c)-\sup_i l_i(c)&\le\sup_{x\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)}f(x)-\inf_{x\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)}f(x)\\&\le\sup_{x\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)}f(x)-f(y)+|f(y)-f(c)|+|f(c)-f(z)|+f(z)-\inf_{x\in\left[t_{j_i-1}^{\pi_i},t_{j_i}^{\pi_i}\right)}f(x)\\&<\epsilon.\end{aligned}$$

Since ϵ and c were arbitrary,

$$\inf_i u_i \overset{\text{a.e}}{=} \sup_i l_i$$

which implies, by Corollary 3.5.11, that f is Riemann integrable. □

3.5.2. The improper Riemann integral.

3.5.3. The Jordan content.

CHAPTER 4

General topology

CHAPTER 5

Spaces of functions

5.1. $\mathcal{L}^p$ spaces as almost Banach spaces over $\mathbb{R}$

The central objects of study in analysis are functions, and the study of sets of functions is what broadly characterizes functional analysis. Often, the types of function spaces studied in analysis are *vector spaces* over some field (usually $\mathbb{R}$ or $\mathbb{C}$). We have already seen one such space in Chapter 3: the $\mathcal{L}^1$ space or the space of integrable functions. $\mathcal{L}^1$ spaces can be suitably generalized to allow for $p-$th power integrability

Definition 5.1.1. Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space. For any $p \in [1, \infty)$, the spaces

$$\mathcal{L}^p(\mathcal{X}, \mathcal{F}, \mu) := \{f \in \mathcal{M}(\mathcal{X}, \mathcal{F}) \mid \bar{\mu}(|f|^p) < \infty\}$$

are called $\mathcal{L}^p$ spaces over $\mathbb{R}$.

For $p = 1$ we get our original integrable functions. We will soon show that $\mathcal{L}^p$ spaces are vector spaces; in-fact, they are (semi)-normed vector spaces with the (semi)-norm given by the function

$$\|f\|_p := \bar{\mu}(|f|^p)^{\frac{1}{p}}.$$

It should be clear that the function defined above satisfies the absolute homogeneity aspect of norms i.e. for any $\alpha \in \mathbb{R}$

$$\|\alpha f\|_p = |\alpha| \|f\|_p.$$

However, our function in question fails being positive definite in that $\|f\|_p = 0 \not\Longrightarrow f = 0$. In fact, all functions that are almost everywhere equal to zero are mapped to zero under this function, which means that it cannot be a bonafide norm. Hence, if we can show that the function satisfies the triangle inequality, we can show that $\mathcal{L}^p$ is a semi-normed vector space. To do this, we will establish a series of results that will be important in their own right, but shall also help us get the triangle inequality

Lemma 5.1.2. *Let $\{a_i\}_{i=1}^k \in [0, \infty)$ and let $\theta_1, \dots, \theta_k > 0$ be real numbers such that*

$$\sum_{i=1}^k \theta_i = 1.$$

Then,

$$\prod_{i=1}^k a_i^{\theta_i} \le \sum_{i=1}^k a_i \theta_i.$$

Proof. Note that since the $\log(\cdot)$ function is concave, by induction on the definition of concavity, we have that

$$\log\left(\prod_{i=1}^k a_i^{\theta_i}\right) = \sum_{i=1}^k \theta_i \log(a_i) \le \log\left(\sum_{i=1}^k \theta_i a_i\right).$$

Since e^x is a monotonic function, the inequality is preserved under exponentiation which yields are result. $\square$

Remark. This lemma is often referred to as the inequality between arithmetic and geometric means as the right hand side is the standard weighted average i.e. the arithmetic mean whereas the left hand side is the geometric mean.

PROPOSITION 5.1.3. *Let functions* $f_1, \ldots, f_k \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ *and let* $\theta_1, \ldots, \theta_k > 0$ *be real numbers such that*

$$\sum_{i=1}^{k} \theta_i = 1.$$

Then, for any measure μ *on* $\mathcal{X}$

$$\bar{\mu}\left(\prod_{i=1}^{k} f_i^{\theta_i}\right) \leq \prod_{i=1}^{k} \bar{\mu}(f_i)^{\theta_i}.$$

PROOF. First notice that if $\bar{\mu}(f_i) = 0$ for any $i \in \{1, \ldots, k\}$ then by Proposition 3.3.3, $f_i = 0$ almost everywhere, which would imply that the left hand side is identically zero. This would make the inequality hold trivially. Conversely, if any $\bar{\mu}(f_i) = \infty$ and all $\mu(f_i) > 0$ then the right hand side is identically ∞ which would again let the inequality hold trivially.

Thus, without loss of generality, assume that $0 < \bar{\mu}(f_i) < \infty$ for all i and define

$$f_i^* = \frac{f_i}{\bar{\mu}(f_i)}.$$

Clearly, $\bar{\mu}(f_i^*) = 1$ and moreover if

$$\bar{\mu}\left(\prod_{i=1}^{k} f_i^{*\theta_i}\right) = \frac{\bar{\mu}\left(\prod_{i=1}^{k} f_i^{\theta_i}\right)}{\prod_{i=1}^{k} \bar{\mu}(f_i)^{\theta_i}} \leq \prod_{i=1}^{k} \bar{\mu}(f_i^*)^{\theta_i} = 1$$

then our claim follows. To show this, note that by Lemma 5.1.2

$$\prod_{i=1}^{k} f_i^{*\theta_i} \leq \sum_{i=1} \theta_i f_i^*$$

pointwise. Integrating both sides, we have

$$\begin{aligned}
\bar{\mu}\left(\prod_{i=1}^{k} f_i^{*\theta_i}\right) &\leq \bar{\mu}\left(\sum_{i=1}^{k} \theta_i f_i^*\right) \\
&= \sum_{i=1}^{k} \theta_i \bar{\mu}(f_i^*) \\
&= \sum_{i=1}^{k} \theta_i \\
&= 1
\end{aligned}$$

where the first inequality follows from the monotonicty of the integral and the first equality from the linearity of integration. This completes the proof. □

COROLLARY 5.1.4 (Hölder's inequality). *Let* $(\mathcal{X}, \mathcal{F}, \mu)$ *be a measure space. For any real numbers* $p, q \in (1, \infty)$ *such that* $\frac{1}{p} + \frac{1}{q} = 1$ *and functions* $g, h \in \mathcal{M}(\mathcal{X}, \mathcal{F})$, *we have that*

$$\|gh\|_1 \leq \|g\|_p \|h\|_q.$$

PROOF. Let $k = 2$, $f_1 = |g|^p, f_2 = |h|^q$, $\theta_1 = \frac{1}{p}$, and $\theta_2 = \frac{1}{q}$ and apply Proposition 5.1.3. □

Now we can finally establish the triangle inequality for the so-called $p-$norms.

THEOREM 5.1.5 (Minkowski's inequality). *Let* $f, g \in \mathcal{L}^p(\mathcal{X}, \mathcal{F}, \mu)$ *for some* $p \in [1, \infty)$. *Then,* $f + g \in \mathcal{L}^p(\mu)$ *and*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p.$$

PROOF. Note that if $\bar{\mu}(|f+g|^p) = 0$ then the inequality follows trivially so let's assume that $\bar{\mu}(|f+g|^p) > 0$. Then

$$\begin{aligned}\bar{\mu}(|f+g|^p) &= \bar{\mu}\left(|f+g||f+g|^{p-1}\right)\\ &\le \bar{\mu}\left(|f||f+g|^{p-1} + |g||f+g|^{p-1}\right)\\ &= \bar{\mu}\left(|f||f+g|^{p-1}\right) + \bar{\mu}\left(|g||f+g|^{p-1}\right)\\ &= \|f|f+g|^{p-1}\|_1 + \|g|f+g|^{p-1}\|_1\\ &\le \|f\|_p\||f+g|^{p-1}\|_{\frac{p}{p-1}} + \|g\|_p\||f+g|^{p-1}\|_{\frac{p}{p-1}}\end{aligned}$$

where the first inequality follows from the triangle inequality of $|\cdot|$ and the monotonicty of integration and the second inequality from Hölder's inequality above. Dividing both sides by $\||f+g|^{p-1}\|_{\frac{p}{p-1}} = \bar{\mu}(|f+g|^p)^{1-\frac{1}{p}} = \frac{\bar{\mu}(|f+g|^p)}{\|f+g\|_p}$ yields the result. Of course, this then shows that $f+g \in \mathcal{L}^p(\mu)$. □

For finite measures , $\mathcal{L}^p$ spaces enjoy a nesting property

PROPOSITION 5.1.6. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space such that $0 < \mu(\mathcal{X}) < \infty$. Then for $1 \le q < p < \infty$*

$$\mathcal{L}^p(\mu) \subseteq \mathcal{L}^q(\mu)$$

and there exists some $C > 0$ such that for any $f \in \mathcal{L}^p(\mu)$

$$C\|f\|_p \ge \|f\|_q.$$

PROOF. By Hölder's inequality, we have for any$f \in \mathcal{L}^p(\mu)$

$$\||f|^q\|_1 \le \||f|^q\|_s\|\mathbb{1}_{\mathcal{X}}\|_{\frac{s}{s-1}}$$

for any $s \in (1, \infty)$. In particular, for $s = \frac{p}{q}$, the inequality is

$$\||f|^q\|_1 \le \bar{\mu}(|f|^p)^{\frac{q}{p}}\,\mu(\mathcal{X})^{\frac{s-1}{s}}.$$

Taking the qth root yields

$$\|f\|_q \le \|f\|_p \mu(\mathcal{X})^{\frac{s-1}{sq}} < \infty$$

which completes the proof with $C = \mu(\mathcal{X})^{\frac{s-1}{sq}}$. □

DEFINITION 5.1.7. A normed vector space $(V, \|\cdot\|)$ is called a *Banach space* if it's complete with respect to the metric induced by its norm.

We would like to prove that our $\mathcal{L}^p$ spaces are actually Banach spaces but the problem is that they are not normed spaces to begin with; as we noted earlier, the $p-$norms map non-zero functions to zero, violating the definiteness condition for norms. This does not actually turn out to be a major impediment in practice, as it is easy to transform our $\mathcal{L}^p$ spaces into actual normed spaces.

To see this, define a relation $\sim$ on $\mathcal{L}^p(\mathcal{X}, \mathcal{F}, \mu)$ such that $f \sim g$ if $f = g$ on all but a null set. It's straightforward to verify that this is in fact an equivalence relation and so the quotient space

$$L^p(\mathcal{X}, \mathcal{F}, \mu) := \mathcal{L}^p(\mathcal{X}, \mathcal{F}, \mu) / \sim$$

consisting of all equivalence classes in $\mathcal{L}^p$ generated by $\sim$ is actually a normed space, where

$$\|[u]\|_p := \inf\{\|w\|_p \mid w \in \mathcal{L}^p \text{such that } w \sim u\} = \|u\|_p \;\forall u \in \mathcal{L}^p$$

Here the norm of an equivalence class is simply the norm of a element of the equivalence class since the norm is invariant if the function changes only on a null set. In this case

$$\|[u]\|_p = 0 \implies [u] = [0].$$

While this construction is useful to illustrate the fact that $p-$norms can be transformed into proper norms, in practice people do not think of spaces of functions as collections of equivalence classes of functions. For now, shall adopt the more pragmatic approach of not worrying about whether our norm is a semi norm or a proper norm and explore the more substantive questions. When we begin our investigation of continuous time stochastic processes, the distinction between spaces of functions and spaces of equivalence classes of functions shall become more important.

In order to discuss completeness, we need a good notion of limits. In the context of a normed vector space, limits can be defined naturally like in Euclidean spaces. However, in $\mathcal{L}^p$ spaces, limits may not be unique unless we adopt the quotient space construction above.

THEOREM 5.1.8. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space. For $p \in [1, \infty)$, $\mathcal{L}^p(\mathcal{X}, \mathcal{F}, \mu)$ is a semi-normed vector space and $L^p(\mathcal{X}, \mathcal{F}, \mu)$ is a normed vector space.*

DEFINITION 5.1.9 (Convergence in $\mathcal{L}^p$). A sequence of functions $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{L}^p(\mathcal{X}, \mathcal{F}, \mu)$ converges to a function $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ in $\mathcal{L}^p$ if

$$\lim_{n\to\infty} \|f_n - f\|_p = 0.$$

In this case, we write

$$f_n \xrightarrow{\mathcal{L}^p} f.$$

Note that this definition strictly subsumes Definition 3.4.1 which discussed limits in $\mathcal{L}^1$. There, we had implictly assumed that the limiting function f was also $\mathcal{L}^1$; now, we shall show that this is in fact always true for all $\mathcal{L}^p$ spaces, that is to say, $\mathcal{L}^p$ spaces contain their limit points. But we can show more, as every Cauchy sequence converges to some limit that is $\mathcal{L}^p$. This is the main result of this section.

THEOREM 5.1.10 (Completeness of $\mathcal{L}^p$). *Let $\{f_n\}_{n\in\mathbb{N}} \in \mathcal{L}^p(\mathcal{X}, \mathcal{F}.\mu)$ be a Cauchy sequence in $\mathcal{L}^p$; that is to say, for any $\epsilon > 0$ there exists some $n_\epsilon \in \mathbb{N}$ such that for all $m, n \geq n_\epsilon$*

$$\|f_m - f_n\|_p < \epsilon.$$

Then, there exists some function $f \in \mathcal{L}^p(\mathcal{X}, \mathcal{F}, \mu)$ such that

$$f_n \xrightarrow{\mathcal{L}^1} f.$$

PROOF. Note that by the definition of a Cauchy sequence and the well ordering principle of natural numbers, for any $k \in \mathbb{N}$ there exists some smallest natural number n_k such that for all $m, n \geq n_k$

$$\|f_m - f_n\|_p < 2^{-k}.$$

In particular, this implies that

$$\|f_{n_{k+1}} - f_{n_k}\|_p < 2^{-k}$$

as $n_{k+1} \geq n_k$. Further, observe that we can rewrite the elements of our subsequence of functions as

$$f_{n_k} = \sum_{i=0}^{k-1} \left(f_{n_{i+1}} - f_{n_i}\right)$$

where $f_{n_0} = 0$. Then, note that

$$\begin{aligned}\|\sum_{i=0}^{k-1} |f_{n_{i+1}} - f_{n_i}|\|_p &\leq \sum_{i=0}^{k-1} \|f_{n_{i+1}} - f_{n_i}\|_p \\ &\leq \|f_{n_1}\|_p + \sum_{i=1}^{k-1} 2^{-i}\end{aligned}$$

where the first inequality follows from Theorem 5.1.5. Then, applying limits, we have

$$\begin{aligned}\lim_{k\to\infty} \|\sum_{i=0}^{k-1} |f_{n_{i+1}} - f_{n_i}|\|_p &\leq \|f_{n_1}\|_p + \sum_{i=1}^{\infty} 2^{-i} \\ &= \|f_{n_1}\|_p + 1. \end{aligned} \tag{6}$$

Observe that the limit on the left hand side exists because the sequence is increasing and bounded above. Finally, note that the sequence $g_k := \sum_{i=0}^{k-1} |f_{n_{i+1}} - f_{n_i}|$ is a sequence of non-negative and increasing measurable functions and since $p \geq 1$, g_k^p is also non-negative, increasing and measurable. Therefore, we can apply the monotone convergence theorem to deduce

$$\begin{aligned} \lim_{k\to\infty} \|g_k\|_p &= \lim_{k\to\infty} \bar{\mu}\left(g_k^p\right)^{\frac{1}{p}} \\ &= \bar{\mu}\left(\lim_{k\to\infty} g_k^p\right)^{\frac{1}{p}} \\ &= \bar{\mu}\left(\left(\lim_{k\to\infty} g_k\right)^p\right)^{\frac{1}{p}} \\ &= \|\lim_{k\to\infty} g_k\|_p \end{aligned} \tag{7}$$

where we have also used the fact that the maps $x \to x^p$ and $x \to x^{\frac{1}{p}}$ are continuous for $p \geq 1$. Together, equations (6) and (7) tell us that

$$\|\lim_{k\to\infty} g_k\|_p < \infty$$

which, by Proposition 3.3.6 shows that $\lim_{k\to\infty} g_k^p < \infty$ $\mu-$a.e and so $\lim_{k\to\infty} g_k < \infty$ $\mu-$a.e. In other words, since absolute summability implies summability

$$\sum_{i=0}^{\infty} |f_{n_{i+1}} - f_{n_i}| < \infty \implies f := \sum_{i=0}^{\infty} \left(f_{n_{i+1}} - f_{n_i}\right) < \infty \ \mu - \text{a.e}$$

Now we have a candidate function $f \in \mathcal{L}^p(\mu)$ such that

$$f = \lim_{k\to\infty} f_{n_k}$$

where the limit is taken pointwise. If we can show that $f_{n_k} \xrightarrow{\mathcal{L}^p} f$ then it will imply that our original sequence $f_n \xrightarrow{\mathcal{L}^p} f$ since by "Cauchyness" and Minkowski's inequality

$$\begin{aligned} \|f_n - f\|_p &\leq \|f_n - f_{n_k}\|_p + \|f_{n_k} - f\| \\ &\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon \end{aligned}$$

for any $\epsilon > 0$ and appropriately large values of n and k. To show subsequential convergence, observe that

$$\begin{aligned} \|f - f_{n_k}\|_p &= \|\sum_{i=k}^{\infty} \left(f_{n_{i+1}} - f_{n_i}\right)\|_p \\ &\leq \|\sum_{i=k}^{\infty} |f_{n_{i+1}} - f_{n_i}|\|_p \\ &\leq \sum_{i=k}^{\infty} \|f_{n_{i+1}} - f_{n_i}\|_p \\ &\leq \sum_{i=k}^{\infty} 2^{-i} \\ &= 1 - \sum_{i=1}^{k-1} 2^{-i} \end{aligned}$$

where the second inequality follows from the monotone convergence argument from earlier. Taking the limit in k yields the result. $\square$

5.1.1. Convexity. We take a little detour in this section to establish an important result in analysis and probability theory: Jensen's inequality. The proof of this theorem is remarkably simple once we develop a reasonable understanding on the behavior of convex functions; in particular, we need to show that convex functions always have support lines. This is intuitively obvious, but showing this rigorously requires a bit of work, which we do here.

Definition 5.1.11. Let X be a vector space. A function $f : X \to \mathbb{R}$ is called convex if for any $\lambda \in [0,1]$ and any $x, y \in X$

$$f(\lambda x + (1-\lambda) y) \leq \lambda f(x) + (1-\lambda) f(y).$$

In elementary calculus, we learnt that convex functions are those that have a graph shaped like a smile (concave functions are frowns), and that the second derivative of convex functions are positive. However, convex functions in general need not be differentiable ($|x|$ is convex but not differentiable at zero). Convex functions do have the property of *subdifferentiability,* which is basically captures the fact that convex functions with "corners" can have tangent lines, even if they are not unique.

Lemma 5.1.12. *Let $f : [a,b] \to \mathbb{R}$ be a convex function. Then for any $x \in (a,b)$*

$$\frac{f(x) - f(a)}{x-a} \leq \frac{f(b) - f(a)}{b-a} \leq \frac{f(b) - f(x)}{b-x}.$$

Proof. Let $\lambda = \frac{b-x}{b-a}$ and so $1-\lambda = \frac{x-a}{b-a}$ and

$$\begin{aligned} f(x) - f(a) &= f(\lambda a + (1-\lambda) b) - f(a) \\ &\leq \lambda f(a) + (1-\lambda) f(b) - f(a) \\ &= (1-\lambda)(f(b) - f(a)) \\ &= \frac{x-a}{b-a}(f(b) - f(a)) \end{aligned}$$

so rearranging yields the first inequality. The second inequality is similarly deduced by noticing that $f(b) - f(x) \geq f(b) - \lambda f(a) - (1-\lambda) f(b) = \lambda (f(b) - f(a))$. ☐

Now let $x_0 \in (a,b)$ be fixed and define

$$\begin{aligned} m^-(x_0) &:= \sup_{\eta \in (a,b), \eta < x_0} \frac{f(x_0) - f(\eta)}{x_0 - \eta} \\ m^+(x_0) &:= \inf_{\xi \in (a,b), \xi > x_0} \frac{f(\xi) - f(x_0)}{\xi - x_0}. \end{aligned}$$

Proposition 5.1.13. *For any $x_0 \in (a,b)$ and any convex function $f : [a,b] \to \mathbb{R}$, $m^-(x_0)$ and $m^+(x_0)$ are finite and $m^-(x_0) \leq m^+(x_0)$ with equality holding if and only if f is differentiable at x_0 in which case*

$$m^-(x_0) = m^+(x_0) = f'(x_0).$$

Proof. Fix $x_0 \in (a,b)$ and note that by our Lemma [1], for $a < \eta < x_0 < \xi < b$

$$\frac{f(x_0) - f(\eta)}{x_0 - \eta} \leq \frac{f(b) - f(x_0)}{b - x_0}$$

and so its supremum

$$m^-(x_0) \leq \frac{f(b) - f(x_0)}{b - x_0}$$

and similarly

$$\frac{f(\xi) - f(x_0)}{\xi - x_0} \geq \frac{f(x_0) - f(a)}{x_0 - a}$$

[1]in our lemma a and b are arbitrary. Here we are applying the Lemma to $x_0 \in (\eta, b)$ and $x_0 \in (a, \xi)$

and so

$$m^{+}(x_0) \geq \frac{f(x_0) - f(a)}{x_0 - a}.$$

Next, note that applying the Lemma to $x_0 \in (\eta, \xi)$ tells us that

$$\frac{f(x_0) - f(\eta)}{x_0 - \eta} \leq \frac{f(\xi) - f(x_0)}{\xi - x_0}$$

and so

$$m^{-}(x_0) \leq m^{+}(x_0).$$

Finally, apply the Lemma to $\eta_2 \in (\eta_1, x_0)$ to deduce that

$$\frac{f(x_0) - f(\eta_2)}{x_0 - \eta_2} \geq \frac{f(x_0) - f(\eta_1)}{x_0 - \eta_1}$$

which means that $\frac{f(x_0)-f(\eta)}{x_0-\eta}$ is increasing in $\eta < x_0$ and is bounded above and so

$$\lim_{\eta \to x_0^-} \frac{f(x_0) - f(\eta)}{x_0 - \eta} = m^{-}(x_0).$$

Similarly,

$$\lim_{\xi \to x_0^+} \frac{f(\xi) - f(x_0)}{\xi - x_0} = m^{+}(x_0)$$

and so our $m^{-}$and m^{+} are simply left and right hand derivatives and our claim follows. □

Now we can construct tangent lines to convex functions in the following sense

PROPOSITION 5.1.14. *Let $f : [a, b] \to \mathbb{R}$ be a convex function. For any $x_0 \in (a, b)$, for any $m \in [m^{-}(x_0), m^{+}(x_0)]$, we have that*

$$l(x) := f(x_0) + m(x - x_0) \leq f(x).$$

PROOF. Note that for $x > x_0$

$$\frac{f(x) - f(x_0)}{x - x_0} \geq m^{+}(x_0) \geq m$$

and if $x < x_0$ then

$$\frac{f(x_0) - f(x)}{x_0 - x} \leq m^{-}(x_0) \leq m.$$

The case of $x = x_0$ is trivial. □

REMARK. The interval $[m^{-}(x_0), m^{+}(x_0)]$ is called the set of *subderivatives* of f at x_0. It should be clear that f is differentiable at x_0 if and only if the set is a singleton. The function l is called a *supporting line* at x_0. More generally, a supporting line l at x_0 is any affine function that satisfies $l(x_0) = f(x_0)$ and $l(x) \leq f(x)$ for every $x \in (a, b)$. A convex function on $[a, b]$ has at least one supporting line at every point in (a, b)

Note that the existence of left and right derivatives at a point implies continuity at that point. This implies that convex functions on an interval are continuous.

PROPOSITION 5.1.15. *Let $f : [a, b] \to \mathbb{R}$ be a function such that for a point $x_0 \in (a, b)$ the left and right hand derivatives*

$$\lim_{x \to x_0^-} \frac{f(x) - f(x_0)}{x - x_0} = L$$

and

$$\lim_{x \to x_0^+} \frac{f(x) - f(x_0)}{x - x_0} = M$$

then f is continuous x_0.

PROOF. Note that

$$\lim_{x\to x_0^-} f(x) - f(x_0) = L \lim_{x\to x_0^-} x - x_0 = 0$$

and

$$\lim_{x\to x_0^+} f(x) - f(x_0) = M \lim_{x\to x_0^+} x - x_0 = 0$$

which completes the proof. □

COROLLARY 5.1.16. *A convex function $f : [a,b] \to \mathbb{R}$ is continuous on (a,b).*

We have proved that a convex function on $[a,b]$ has supporting lines; it turns out that if a function on $[a,b]$ has supporting lines everywhere on (a,b) then it is convex as well.

PROPOSITION 5.1.17. *Let $f : [a,b] \to \mathbb{R}$ be a function. Then f is convex if and only if it has a supporting line at each point $x_0 \in (a,b)$.*

PROOF. The "only if" part is Proposition 5.1.14. For the converse, let $\lambda \in [0,1]$ be arbitrary and let $x, y \in (a,b)$. Let $x_0 = \lambda x + (1-\lambda) y$ and note that

$$\begin{aligned}\lambda f(x) + (1-\lambda) f(y) &\geq \lambda l(x) + (1-\lambda) f(y) \\ &= l(x_0) \\ &= f(x_0)\end{aligned}$$

where the first equality follows from the fact that affine functions are both convex and concave. □

Finally, we recover the result that a twice continuously differentiable convex function has non-negative second derivative, a fact that follows from Taylor's theorem with remainder.

PROPOSITION 5.1.18. *Let $f : [a,b] \to \mathbb{R}$ be a twice continuously differentiable function. Then f is a convex function if and only if at every point $x_0 \in (a,b)$ $f''(x_0) \geq 0$.*

PROOF. Note that by Taylor's theorem

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + f''(c) \frac{(x - x_0)^2}{2} \tag{8}$$

where c is between x and x_0. Then if $f''(c) \geq 0$ then

$$f(x) \geq f(x_0) + f'(x_0)(x - x_0)$$

and so by Proposition 5.1.17 our f is convex. Conversely, if f is convex then □

We are finally ready to present the main result from this section.

THEOREM 5.1.19 (Jensen's inequality). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space wth $\mu(\mathcal{X}) = 1$. For any integrable function $f \in \mathcal{L}^1(\mu)$ and a convex function $g : \mathbb{R} \to \mathbb{R}$, we have that*

$$\bar{\mu}(g \circ f) \geq g(\bar{\mu}(f)).$$

PROOF. First note that for any restriction of g to any interval $[-n,n]$,g has supporting lines everywhere on $(-n,n)$ by Proposition 5.1.17. Since $n$can be arbitrarily large, g admits a supporting line at $x_0 = \bar{\mu}(f)$.That is to say, there exists an affine function l such that

$$l(\bar{\mu}(f)) = g(\bar{\mu}(f))$$

and

$$l(f(x)) \leq g(f(x))$$

for all $x \in \mathbb{R}$. Any affine function $l : \mathbb{R} \to \mathbb{R}$ can be written as $l(x) = a + bx$ where $a, b \in \mathbb{R}$. Therefore, by the monotonicty of integration

$$\begin{aligned}\bar{\mu}(g \circ f) &\geq \bar{\mu}(l \circ f) \\ &= a + b\bar{\mu}(f) \\ &= l(\bar{\mu}(f)) \\ &= g(\bar{\mu}(f))\end{aligned}$$

where in the second line we have used the linearity of integration along with the fact that $\mu(\mathcal{X}) = 1$. $\square$

EXAMPLE 5.1.20 (ISI 2005 Sample PSB 4). Let f be a non-decreasing, integrable function defined on $[0,1]$. Show that

$$\left(\int_0^1 f(x)dx\right)^2 \leq 2\int_0^1 x(f(x))^2 dx.$$

Note that Jensen's inequality immediately implies that

$$\left(\int_0^1 f(x)dx\right)^2 \leq \int_0^1 (f(x))^2 dx.$$

Further, note that $f(x)^2\, 2x\mathbb{1}\left\{\frac{1}{2} \leq x \leq 1\right\} \geq f(x)^2\, \mathbb{1}\left\{\frac{1}{2} \leq x \leq 1\right\}$ but $f(x)^2\, 2x\mathbb{1}\left\{0 \leq x \leq \frac{1}{2}\right\} \leq f(x)^2\, \mathbb{1}\left\{0 \leq x \leq \frac{1}{2}\right\}$. However, since f is non decreasing, we know that

$$|f(x)^2(2x-1)\mathbb{1}\left\{\frac{1}{2} \leq x \leq 1\right\}| \geq |f(x)^2(2x-1)\mathbb{1}\left\{0 \leq x \leq \frac{1}{2}\right\}|$$

and so

$$\begin{aligned}\int_0^1 (f(x))^2 dx &= \int_0^{\frac{1}{2}} (f(x))^2 dx + \int_{\frac{1}{2}}^1 (f(x))^2 dx \\ &\leq \int_0^{\frac{1}{2}} 2x(f(x))^2 dx + \int_{\frac{1}{2}}^1 2x(f(x))^2 dx \\ &= \int_0^1 2x(f(x))^2 dx.\end{aligned}$$

5.1.2. The space L^∞. In the context of L^p spaces, we replace the notion of boundedness (i.e. the quality of a function in a space being bounded) with the related notion of *essentially boundedness.* A function is said to be essentially bounded if it is bounded except on a null set. Then the space of essentially bounded functions from a measure space $(\mathcal{X}, \mathcal{F}, \mu)$ to $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$is denoted $\mathcal{L}^\infty(\mathcal{X}, \mathcal{F}, \mu)$. This space comes equipped with a (semi) norm $\|\cdot\|_\infty$ so that for $f \in \mathcal{L}^\infty(\mu)$,we have that

$$\|f\|_\infty := \inf\{C \in \mathbb{R} \mid \mu(\{|f| > C\}) = 0\}.$$

To see this is indeed a semi-norm, we can verify that non-negativity is satisfied by definition. To see homogeneity, observe that for any $\alpha \neq 0$, the collection $\{C \in \mathbb{R} \mid \mu(|\alpha f| > C)\} = \{|\alpha| C \in \mathbb{R} \mid \mu(|f| > C)\}$ and so we have homogeneity. For the triangle inequality, consider the fact that for any $C \in \mathbb{R}$ and any $f, g \in \mathcal{L}^\infty(\mu)$, we have $|f(x)| \leq \|f\|_\infty, |g(x)| \leq \|g\|_\infty$ for almost all $x \in \mathcal{X}$ and so

$$|f(x) + g(x)| \leq |f(x)| + |g(x)| \leq \|f\|_\infty + \|g\|_\infty$$

almost everywhere. In words, $\|f\|_\infty + \|g\|_\infty$ is an *essential upper-bound* for $|f(x) + g(x)|$ and so $\|f\|_\infty + \|g\|_\infty \in \{C \in \mathbb{R} \mid \mu(\{|f+g| > C\}) = 0\}$. The inequality then follows since $\|f+g\|_\infty$ is the infimum of that set. The notation $\mathcal{L}^\infty(\mu)$ of course deserves some scrutiny; one potential justification for the choice of notation is the following result.

PROPOSITION 5.1.21. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and let $f \in \cap_{p \in \mathbb{N}} \mathcal{L}^p(\mu)$ .Then*

$$\lim_{p \to \infty} \|f\|_p = \|f\|_\infty.$$

PROOF. First, suppose that $\|f\|_\infty < \infty$. Then, for any sequence $q_n \to \infty$, we have that

$$\begin{aligned}\|f\|_{p+q_n} &= \bar{\mu}\left(|f|^{p+q_n}\right)^{\frac{1}{p+q_n}}\\ &\leq \bar{\mu}\left(\|f\|_\infty^{q_n}|f|^p\right)^{\frac{1}{p+q_n}}\\ &= \|f\|_\infty^{\frac{q_n}{p+q_n}}\bar{\mu}\left(|f|^p\right)^{\frac{1}{p+q_n}}\end{aligned}$$

where we used the fact that $|f| \leq \|f\|_\infty$ almost everywhere and the monotonicity and linearity of the integral. Taking lim-sups[2] on both sides we have

$$\limsup_{q_n\to\infty}\|f\|_{p+q_n} \leq \|f\|_\infty.$$

Conversely, note that for any $\epsilon > 0$, we have the fact that $|f| \geq (\|f\|_\infty - \epsilon)\mathbb{1}\{|f| > \|f\|_\infty - \epsilon\}$ and so since p-norms respect monotonicty, we have that

$$\begin{aligned}\|f\|_p &\geq \|(\|f\|_\infty - \epsilon)\,\mathbb{1}\{|f| > \|f\|_\infty - \epsilon\}\|_p\\ &= |\|f\|_\infty - \epsilon|\mu\left(|f| > \|f\|_\infty - \epsilon\right)^{\frac{1}{p}}\end{aligned}$$

where we have used homogeneity in the second line. Note that the right hand side is finite by Markov's inequality since $f \in \mathcal{L}^1(\mu)$ and $\|f\|_\infty < \infty$[3] and positive by the fact that $\|f\|_\infty$is an essential supremum so any smaller real number cannot be an essential upper bound. Letting $p \to \infty, \epsilon \to 0$ on the right hand side (while taking a lim-inf on the left), we have

$$\liminf_{p\to\infty}\|f\|_p \geq \|f\|_\infty$$

which completes the proof for the essentially bounded case. □

5.1.2.1. *Uniform convergence and L^∞.* Recall from basic analysis (or our discussion on Egorov's theorem) the notion of uniform convergence. We say

DEFINITION 5.1.22. A sequence of functions $f_n \in \mathcal{M}(\mathcal{X}, \mathcal{F})$ converges uniformly to a function $f \in \mathcal{M}(\mathcal{X}, \mathcal{F})$

5.2. Hilbert spaces over ℝ

5.2.1. Introduction to inner product spaces. The inner product is a generalization of the dot product of Euclidean spaces. Recall from calculus that the dot product is defined $\bullet : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}$ where

$$x \bullet y := \sum_{i=1}^{n} x_i y_i$$

for any $x, y \in \mathbb{R}^n$. We can generalize this with the following definition.

DEFINITION 5.2.1. Let V be a vector space over $\mathbb{R}$. The function $\langle\cdot,\cdot\rangle : V \times V \longrightarrow \mathbb{R}$ is called an *inner product* if

(i) $\langle v, v\rangle \geq 0$ for all $v \in V$
(ii) $\langle v, v\rangle = 0 \implies v = 0$ for all $v \in V$
(iii) For any $v, w \in V : \langle v, w\rangle = \langle w, v\rangle$
(iv) For any $\alpha, \beta \in \mathbb{R}$ and any $v, w, u \in V$

$$\langle \alpha v + \beta w, u\rangle = \alpha\langle v, u\rangle + \beta\langle w, u\rangle.$$

If $\langle\cdot,\cdot\rangle$ satsifies these properties then $(V, \langle\cdot,\cdot\rangle)$ is called an *inner product space.*

---

[2]We don't know if the limit on the left side exists; the right-hand side has a proper limit and so is equal to its lim-sup.

[3]$\mu\left(|f| > \|f\|_\infty - \epsilon\right) \leq \frac{\bar{\mu}(|f|)}{\|f\|_\infty - \epsilon}$

The first and second properties are together called the *definiteness* condition of inner products, analagous to the one for norms. The third property is called *symmetry,* and the the final property is *linearity in the first argument.* Note that symmetry and linearity in the first argument together imply linearity in the second argument. Since we are ultimately concerned with basic probability theory, we shall always assume the field over which V is defined is $\mathbb{R}$; if the field were $\mathbb{C}$ then the second property above would be replaced by *skew-symmetry* i.e. $\langle v, w\rangle = \overline{\langle w, v\rangle}$ where the $\overline{c}$ for any $c \in \mathbb{C}$ denotes the complex conjugate. In this case, proper linearity in the second argument would not hold.

We can immediately use these properties to derive a familiar result from a different context.

PROPOSITION 5.2.2 (Cauchy-Schwarz inequality). *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space. For any $v, w \in V$*

$$\langle v, w\rangle^2 \leq \langle v, v\rangle\langle w, w\rangle$$

and moreover, strict equality holds if and only if $v = \alpha w$ for some $\alpha \in \mathbb{R}$.

PROOF. The claim is trivially true if either $v = 0$ or $w = 0$ as $\langle 0, w\rangle = 0\langle v, w\rangle$ by linearity. Thus we can assume without loss of generality both are non-zero. Write $v = (v - \alpha w) + \alpha w$ where $\alpha = \frac{\langle v,w\rangle}{\langle w,w\rangle}$. Then,

$$\begin{aligned} \langle v, v\rangle &= \langle (v - \alpha w) + \alpha w, (v - \alpha w) + \alpha w\rangle \\ &= \langle v - \alpha w, v - \alpha w\rangle + 2\langle \alpha w, v - \alpha w\rangle + \langle \alpha w, \alpha w\rangle \end{aligned} \tag{9}$$

where the second equality follows from an application of linearity and symmetry. Note that the term

$$\begin{aligned} \langle \alpha w, v - \alpha w\rangle &= \alpha\langle w, v\rangle - \alpha^2\langle w, w\rangle \\ &= \alpha\left(\langle v, w\rangle - \alpha\langle w, w\rangle\right) \\ &= 0. \end{aligned} \tag{10}$$

Note that equations (9) and (10), together with the definiteness of inner products, imply that

$$\begin{aligned} \langle v, v\rangle &\geq \langle \alpha w, \alpha w\rangle \\ &= \alpha^2\langle w, w\rangle \\ &= \frac{\langle v, w\rangle^2}{\langle w, w\rangle} \end{aligned}$$

and rearranging yields the inequality.

To see the equality result, note that $\langle v, v\rangle = \langle \alpha w, \alpha w\rangle$ in equation (9) when $\langle v - \alpha w, v - \alpha w\rangle = 0 \implies v = \alpha w$ which shows the necessity. To show sufficiency, let $v = \beta w$ for some $\beta \in \mathbb{R}$ and note that

$$\begin{aligned} \langle v, w\rangle^2 &= \langle \beta w, w\rangle^2 \\ &= \beta^2\langle w, w\rangle^2 \\ &= \beta^2\langle w, w\rangle\langle w, w\rangle \\ &= \langle \beta w, \beta w\rangle\langle w, w\rangle \\ &= \langle v, v\rangle\langle w, w\rangle \end{aligned}$$

where the second equality follows from linearity in the first argument and the fourth equality follows from linearity in both arguments. $\square$

The relationship between inner-products and norms is a tight one; every inner-product induces a norm.

PROPOSITION 5.2.3. *Let $(V, \langle\cdot,\cdot\rangle)$ be an inner product space and let the function $\|\cdot\| : V \longrightarrow \mathbb{R}$ be defined by*

$$\|v\| = \sqrt{\langle v, v\rangle}$$

for any $v \in V$. Then the function $\|\cdot\|$ is a norm.

PROOF. Note that the definiteness condition of norms corresponds to the definiteness condition of inner-products, and so is trivially satisfied. Next, observe that for any $\alpha \in \mathbb{R}$

$$\begin{aligned}\|\alpha v\| &= \sqrt{\langle \alpha v, \alpha v\rangle} \\ &= \sqrt{\alpha^2 \langle v, v\rangle} \\ &= |\alpha|\|v\|\end{aligned}$$

which gives us absolute homogeneity. Finally, for any $v, w \in V$

$$\begin{aligned}\|v+w\|^2 &= \langle v+w, v+w\rangle \\ &= \langle v, v\rangle + 2\langle v, w\rangle + \langle w, w\rangle \\ &\leq \|v\|^2 + 2\|v\|\|w\| + \|w\|^2 \\ &= (\|v\| + \|w\|)^2\end{aligned}$$

where the inequality follows from the Cauchy-Schwarz inequality. This completes the proof. $\square$

We call such a norm a norm *induced by* an inner product. It turns out that one can recover a norm from an inner product precisely when a particular identity is satisfied.

DEFINITION 5.2.4. Let $(V, \|\cdot\|)$ be a normed vector space. The norm $\|\cdot\|$ is said to satisfy the *parallelogram identity* if for any $v, w \in V$

$$\|\frac{v+w}{2}\|^2 + \|\frac{v-w}{2}\|^2 = \frac{1}{2}\left(\|v\|^2 + \|w\|^2\right).$$

Next, we need a few lemmas to aid in the proof of the main result.

LEMMA 5.2.5. *Let $(V, \|\cdot\|)$ be a normed vector space where the norm $\|\cdot\|$ satisfies the parallelogram identity. Then, for any $u, v, w \in V$*

$$\|u+v+w\|^2 = \|u+v\|^2 + \|v+w\|^2 + \|u+w\|^2 - \|u\|^2 - \|v\|^2 - \|w\|^2.$$

PROOF. Note that by the parallelogram identity

$$\|(u+v)+w\|^2 = 2\|u+v\|^2 + 2\|w\|^2 - \|(u+v)-w\|^2 \tag{11}$$

and

$$\|u+(v+w)\|^2 = 2\|u\|^2 + 2\|v+w\|^2 - \|u-v-w\|^2. \tag{12}$$

Adding equations (11) and (12), then dividing by two, we have that

$$\begin{aligned}\|u+v+w\|^2 &= \|u+v\|^2 + \|w\|^2 - \frac{1}{2}\|u+v-w\|^2 + \|u\|^2 + \|v+w\|^2 - \frac{1}{2}\|u-v-w\|^2 \\ &= \|u+v\|^2 + \|w\|^2 + \|u\|^2 + \|v+w\|^2 - \|u-w\|^2 - \|v\|^2 \\ &= \|u+v\|^2 + \|w\|^2 + \|u\|^2 + \|v+w\|^2 + \|u+w\|^2 - 2\|u\|^2 - 2\|w\|^2 - \|v\|^2 \\ &= \|u+v\|^2 + \|v+w\|^2 + \|u+w\|^2 - \|u\|^2 - \|v\|^2 - \|w\|^2\end{aligned}$$

where the second equality follows from an application of the parallelogram identity on $\frac{1}{2}\|u+v-w\|^2 + \frac{1}{2}\|u-v-w\|^2$ and the third equality follows from another application of the identity to $\|u-w\|^2$. $\square$

LEMMA 5.2.6. *For any norm $\|\cdot\|$ on some vector space V, the map $\phi_{v,w} : \mathbb{R} \longrightarrow \mathbb{R}$ given by*

$$\phi_{v,w}(t) := \frac{1}{4}\left(\|tv+w\|^2 - \|tv-w\|^2\right)$$

is continuous for any $v, w \in V$.

PROOF. Note that for any $v, w \in V$, our function $\phi_{v,w}$ is the composition of continous functions and thus the result follows.[4] $\square$

[4]The norm is continuous by the reversed triangle inequality $|\|x\| - \|y\|| \leq \|x-y\|$

THEOREM 5.2.7. *Let $(V, \|\cdot\|)$ be a normed vector space. The norm $\|\cdot\|$ is induced by an inner product if and only if the parallelogram identity is satisfied.*

PROOF. First suppose that the norm $\|\cdot\|$ is induced by the inner product $\langle\cdot,\cdot\rangle$. Then,

$$\begin{aligned}\|\frac{v+w}{2}\|^2+\|\frac{v-w}{2}\|^2 &= \langle\frac{v+w}{2},\frac{v+w}{2}\rangle+\langle\frac{v-w}{2},\frac{v-w}{2}\rangle\\ &= \frac{1}{4}\left(\langle v,v\rangle+2\langle v,w\rangle+\langle w,w\rangle\right)+\frac{1}{4}\left(\langle v,v\rangle-2\langle v,w\rangle+\langle w,w\rangle\right)\\ &= \frac{1}{2}\langle v,v\rangle+\frac{1}{2}\langle w,w\rangle\\ &= \frac{1}{2}\left(\|v\|^2+\|w\|^2\right).\end{aligned}$$

The harder part is showing that if a norm that satisfies the parallelogram identity, there exists an inner product that induces such a norm. To do so, first define a map

$$(v,w):=\frac{1}{4}\left(\|v+w\|^2-\|v-w\|^2\right)$$

and observe that

$$(v,v)=\|v\|^2\geq 0$$

and that $(v,v)=0$ if and only if $v=0$ by the definiteness condition of norms. Moreover, observe that $(v,w)=(w,v)$ since$\|v-w\|=\|w-v\|$. Next, note that for any $v,u,w\in V$

$$\begin{aligned}(v+u,w) &= \frac{1}{4}\left(\|v+u+w\|^2-\|v+u-w\|^2\right)\\ &= \frac{1}{4}\big(\|u+v\|^2+\|v+w\|^2+\|u+w\|^2-\|u\|^2-\|v\|^2-\|w\|^2\\ &\qquad -\|u+v\|^2-\|v-w\|-\|u-w\|^2+\|u\|^2+\|v\|^2+\|w\|^2\big)\\ &= \frac{1}{4}\left(\|v+w\|^2-\|v-w\|^2\right)+\frac{1}{4}\left(\|u+w\|^2-\|u-w\|^2\right)\\ &= (v,w)+(u,w)\end{aligned}$$

where the second equality follows from Lemma 5.2.5 which proves additive linearity.

In order to establish multiplicative linearity on $\mathbb{R}$, we shall first demonstrate it on natural numbers $\mathbb{N}$, then on integers $\mathbb{Z}$,then on the rationals $\mathbb{Q}$ and finally the entire real line. First, we show that for any $n\in\mathbb{N}:(nv,w)=n(v,w)$. This follows from induction on additive linearity since

$$\begin{aligned}(nv,w) &= \left(\sum_{i=1}^{n}v,w\right)\\ &= \sum_{i=1}^{n}(v,w)\\ &= n(v,w).\end{aligned}$$

Next, to show that multiplicative linearity holds over $\mathbb{Z}$, all we have to show is that

$$(0v,w)=0$$

and

$$(-v,w)=-(v,w).$$

The first follows on inspection; for the second, note

$$\begin{aligned}(-v,w) &= \frac{1}{4}\left(\|w-v\|^2-\|-(v+w)\|^2\right)\\ &= \frac{1}{4}\left(\|v-w\|^2-\|v+w\|^2\right)\\ &= -(v,w)\end{aligned}$$

where the second equality follows from absolute homogeneity of norms. To extend the multiplicative linearity to $\mathbb{Q}$, consider an arbitrary $q \in \mathbb{Q}$ and note that by definition $q = \frac{a}{b}, (a, b) \in \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$ and so

$$\begin{aligned}
(qv, w) &= \left(\frac{a}{b}v, w\right) \\
&= a\left(\frac{v}{b}, w\right) \\
&= \frac{a}{4}\left(\|\frac{v}{b} + w\|^2 - \|\frac{v}{b} - w\|^2\right) \\
&= \frac{a}{4}\left(\|\frac{b}{b}\left(\frac{v}{b} + w\right)\|^2 - \|\frac{b}{b}\left(\frac{v}{b} - w\right)\|^2\right) \\
&= \frac{a}{4b^2}\left(\|v + bw\|^2 - \|v - bw\|^2\right) \\
&= \frac{a}{b^2}(v, bw) \\
&= \frac{a}{b}(v, w) \\
&= q(v, w)
\end{aligned}$$

where we used the linearity on $\mathbb{Z}$ (and symmetry) in the second and second-to-last equalities, and the absolute homogeneity of norms in the fifth equality. Finally, to extend this linearity to all of $\mathbb{R}$, let $\alpha \in \mathbb{R}$ be unspecified and observe that by the density of rational numbers in $\mathbb{R}$, there exists a sequence $\{q_n\} \in \mathbb{Q}$ such that $\lim_{n\to\infty} q_n = \alpha$ and so,

$$\begin{aligned}
(\alpha v, w) &= \left(\lim_{n\to\infty} q_n v, w\right) \\
&= \lim_{n\to\infty}(q_n v, w) \\
&= \lim_{n\to\infty} q_n (v, w) \\
&= \alpha (v, w)
\end{aligned}$$

where the second equality follows by Lemma 5.2.6 and the third by our linearity result on $\mathbb{Q}$.

Together, we have shown that our function $(\cdot, \cdot) : V \times V \longrightarrow \mathbb{R}$ satisfies all the properties of the inner product, thus completing the proof. □

Corollary 5.2.8. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space. The space $\mathcal{L}^2(\mathcal{X}, \mathcal{F}, \mu)$ is an inner-product space with inner product $\langle\cdot,\cdot\rangle : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ given by*

$$\langle x, y\rangle = \bar{\mu}(xy)$$

Proof. Let us verify that $\mathcal{L}^2(\mu)$ satisfies the parallelogram identity. COMPLETE LATER □

5.2.2. Hilbert spaces. Corollary 5.2.8 showed that $\mathcal{L}^2(\mu)$ is an inner-product space but we also know, from Theorem 5.1.10 that $\mathcal{L}^2(\mu)$ is a complete metric space with respect to the metric induced by its norm. These types of spaces play a special role in analysis and are important objects of study in of themselves in functional analysis. In the context of probability theory, they play a key role in the development of the theory of conditional expectations and the existence of probability density functions.

Definition 5.2.9. An inner product space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ is called a *Hilbert space* if it is complete with respect to the norm induced by the inner product.

Theorem 5.2.10 (Projection). *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space and let $G \subseteq \mathcal{H}$ be a nonempty closed convex subset. Then, for any $h \in \mathcal{H}$, there exists a unique $g_0 \in G$ such that*

$$\|h - g_0\| = \inf_{g\in G}\|h - g\|$$

where $\|\cdot\|$ is the norm induced by $\langle\cdot,\cdot\rangle$.

PROOF. Let $\delta_h := \inf_{g\in G}\|h-g\|$ and note that by the definition of the infimum, for every $n \in \mathbb{N}$, there exists some $g_n \in G$ such that

$$\delta_h \leq \|h-g_n\| < \delta_h + \frac{1}{n}$$

and so $\lim_{n\to\infty}\|h-g_n\| = \delta_h$. It turns out that the sequence $\{g_n\}_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathcal{H}$. To see this, fix $\epsilon > 0$ observe that by the definition of a limit and the continuity of $x \to x^2$, there exists some n_0 such that $\|h-g_n\|^2 - \delta_h^2 < \frac{\epsilon^2}{4}$ for every $n \geq n_0$. By the parallelogram identity 5.2.4with $v = h - g_n$ and $w = h - g_m$ where $m, n \geq n_0$

$$\|h - \left(\frac{g_m+g_n}{2}\right)\|^2 + \|\frac{g_n-g_m}{2}\|^2 = \frac{1}{2}\left(\|h-g_n\|^2 + \|h-g_m\|^2\right).$$

Since G is convex, $\frac{g_m+g_n}{2} \in G$ and so, again by the definition of an infimum

$$\delta_h^2 \leq \|h - \left(\frac{g_m+g_n}{2}\right)\|^2$$

which implies that

$$\begin{aligned}
\delta_h^2 + \|\frac{g_n-g_m}{2}\|^2 &\leq \frac{1}{2}\left(\|h-g_n\|^2 + \|h-g_m\|^2\right)\\
\implies \|g_n-g_m\|^2 &\leq 2\left(\|h-g_n\|^2 - \delta_h^2 + \|h-g_m\|^2 - \delta_h^2\right)\\
&< 2\left(\frac{\epsilon^2}{4} + \frac{\epsilon^2}{4}\right)\\
&= \epsilon^2
\end{aligned}$$

which then shows that g_n is Cauchy and so by the completeness of $\mathcal{H}$ converges to a limit in $\mathcal{H}$. However, since G is closed, it contains this limit and thus $g_0 := \lim_{n\to\infty} g_n \in G$. Finally, by Minkowski's inequality

$$\|h-g_0\| \leq \|h-g_n\| + \|g_n-g_0\|$$

and by taking limits on the RHS we have

$$\|h-g_0\| \leq \delta_h.$$

The definition of the infimum then implies that

$$\delta_h \leq \|h-g_0\|$$

which gives equality.

Now suppose there exists some $g_1 \in G$ such that $\|h-g_1\| = \delta_h$. Then, applying the parallelogram identity 5.2.4 again with $v = h - g_0$ and $w = h - g_1$ we have

$$\begin{aligned}
\|h - \left(\frac{g_0+g_1}{2}\right)\|^2 + \|\frac{g_0-g_1}{2}\|^2 &= \frac{1}{2}\left(\|h-g_0\|^2 + \|h-g_1\|^2\right)\\
&= \delta_h^2
\end{aligned}$$

Note again by convexity, $\frac{g_0+g_1}{2} \in G$ and so $\|h - \left(\frac{g_0+g_1}{2}\right)\|^2 \geq \delta_h^2$ which implies that

$$\|\frac{g_0-g_1}{2}\|^2 \leq 0$$

and since norms cant be negative, we have that $\|\frac{g_0-g_1}{2}\|^2 = 0 \implies g_0 = g_1$ by the definiteness of norms. $\square$

REMARK. The unique vector $g_0 \in G$ described above is called the *projection* of h into G and is often denoted as $P_G h$.

Note that the projection theorem holds in general for any nonempty, closed, and convex subset G of any Hilbert space $\mathcal{H}$ but in particular it holds for any nonempty closed *subspace* of $\mathcal{H}$, since every subspace is automatically convex. However, in the case, of a subspace, the projection is *orthogonal.* We make this precise with the following result.

COROLLARY 5.2.11. *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space and let $\mathcal{G} \subset \mathcal{H}$ be a closed subspace. Then for any $h \in \mathcal{H}$, $k \in \mathcal{G}$ we have that*

$$\langle h - k, g\rangle = 0$$

for every $g \in \mathcal{G}$ if and only if $k = P_{\mathcal{G}}h$ and so

$$\|h\|^2 = \|P_{\mathcal{G}}h\|^2 + \|h - P_{\mathcal{G}}h\|^2.$$

PROOF. Fix $h \in \mathcal{H}$ and observe that $\tilde{g} = P_{\mathcal{G}}h + tg \in \mathcal{G}$ for any $g \in \mathcal{G}$ and $t \in \mathbb{R}$ since $\mathcal{G}$ is a subspace and so by the projection theorem 5.2.10

$$\begin{aligned}
\|h - P_{\mathcal{G}}h\|^2 &\leq \|h - \tilde{g}\|^2 \\
&= \langle h - \tilde{g}, h - \tilde{g}\rangle \\
&= \langle (h - P_{\mathcal{G}}h) - tg, (h - P_{\mathcal{G}}h) - tg\rangle \\
&= \|h - P_{\mathcal{G}}h\|^2 - 2t\langle h - P_{\mathcal{G}}h, g\rangle + t^2\|g\|^2 \qquad (13)
\end{aligned}$$

where the last equality follows by linearity of inner products. Now if $\|g\| = 0$ then $g = 0$ by definiteness and so $\langle h - P_{\mathcal{G}}h, g\rangle = 0$ by linearity. Thus, assume that $\|g\| > 0$ and so notice that the right hand side of (13) is a convex polynomial in t which is minimized at $t = \frac{\langle h - P_{\mathcal{G}}h, g\rangle}{\|g\|^2}$ with minimum

$$\|h - P_{\mathcal{G}}h\|^2 - \frac{\langle h - P_{\mathcal{G}}h, g\rangle^2}{\|g\|^2}.$$

But since the inequality in (13) holds for any $t \in \mathbb{R}$, we have

$$\|h - P_{\mathcal{G}}h\|^2 \leq \|h - P_{\mathcal{G}}h\|^2 - \frac{\langle h - P_{\mathcal{G}}h, g\rangle^2}{\|g\|^2}$$

and so

$$\langle h - P_{\mathcal{G}}h, g\rangle = 0$$

since all the terms are non-negative. Conversely, assume that $\langle h - k, g\rangle = 0$ for some $k \in \mathcal{G}$ and every $g \in \mathcal{G}$. Then, we have that

$$\begin{aligned}
\|h - g\|^2 &= \|h - k + k - g\|^2 \\
&= \langle (h - k) + (k - g), (h - k) + (k - g)\rangle \\
&= \|h - k\|^2 + \|k - g\|^2 + 2\langle h - k, k - g\rangle.
\end{aligned}$$

where we have again used the linearity and symmetry of inner products. But notice that $k - g \in \mathcal{G}$, so by our assumption, we have $\langle h - k, k - g\rangle = 0$ and so for every $g \in \mathcal{G}$

$$\begin{aligned}
\|h - g\|^2 &= \|h - k\|^2 + \|k - g\|^2 \\
&\geq \|h - k\|^2
\end{aligned}$$

by the non-negativity of norms. Then by the uniqueness clause of Theorem 5.2.10 $k = P_{\mathcal{G}}h$.

Finally, observe that

$$\begin{aligned}
\|h\|^2 &= \|P_{\mathcal{G}}h + (h - P_{\mathcal{G}}h)\|^2 \\
&= \|P_{\mathcal{G}}h\|^2 + 2\langle P_{\mathcal{G}}h, h - P_{\mathcal{G}}h\rangle + \|h - P_{\mathcal{G}}h\|^2 \\
&= \|P_{\mathcal{G}}h\|^2 + \|h - P_{\mathcal{G}}h\|^2
\end{aligned}$$

since $\langle h - P_{\mathcal{G}}h, P_{\mathcal{G}}h\rangle = 0$. □

The following result establishes some standard properties of Hilbert projections.

PROPOSITION 5.2.12 (Properties of projections). *Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a Hilbert space. For any closed subspace $\mathcal{G} \subseteq \mathcal{H}$, the projection operator $P_{\mathcal{G}}$ has the following properties*

(i) (Linearity) For any $f, g \in \mathcal{H}$ and any $\alpha, \beta \in \mathbb{R}$

$$P_{\mathcal{G}}(\alpha f + \beta g) = \alpha P_{\mathcal{G}} f + \beta P_{\mathcal{G}}.$$

(ii) (Tower) For any closed subspaces $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \mathcal{H}$ and any $h \in \mathcal{H}$

$$P_{\mathcal{G}_1} P_{\mathcal{G}_2} h = P_{\mathcal{G}_1} h$$

PROOF. For (i), observe that for any $k \in \mathcal{G}$

$$\begin{aligned}\langle \alpha f + \beta g - (\alpha P_{\mathcal{G}} f + \beta P_{\mathcal{G}} g), k\rangle &= \alpha\langle f - P_{\mathcal{G}} f, k\rangle - \beta\langle g - P_{\mathcal{G}} g, k\rangle \\ &= 0\end{aligned}$$

where the first equality follows from the linearity of inner products (in the first argument) and the second from Corollary5.2.11. Then, an application of the uniqueness clause of the same result furnishes the linearity result.

Next, pick an arbitrary $k \in \mathcal{G}_1$ and note that for any $h \in \mathcal{H}$

$$\begin{aligned}\langle P_{\mathcal{G}_2} h - P_{\mathcal{G}_1} h, k\rangle &= \langle P_{\mathcal{G}_2} h - h + h - P_{\mathcal{G}_1} h, k\rangle \\ &= \langle h - P_{\mathcal{G}_1} h, k\rangle - \langle h - P_{\mathcal{G}_2} h, k\rangle \\ &= 0\end{aligned}$$

where the second equality is due to linearity of inner products and the third is due the fact that both the inner products on the second line are zero due to the same uniqueness of orthogonal projections (the second inner product is zero because of $k \in \mathcal{G}_1 \implies k \in \mathcal{G}_2$). This is sufficient to deduce (ii) by yet another application of the uniqueness of orthogonal projections. □

The projection theorem for subspaces has important consequences in the theory of linear functionals on Hilbert spaces.

PROPOSITION 5.2.13. *Let $(V, \|\cdot\|)$ be a normed vector space and let*

$$\Gamma : V \to \mathbb{R}$$

be a linear functional on V. The functional Γ is continuous (with respect to the usual topologies) if and only if there exists a constant $C \in \mathbb{R}$ such that

$$|\Gamma(h)| \leq C\|h\| \tag{14}$$

for every $h \in V$.

PROOF. First, suppose that a functional Γ on satisfies the condition in(14). Fix $\epsilon > 0$ and notice that for any $h, \tilde{h} \in V$, $h - \tilde{h} \in V$ and so if $\|h - \tilde{h}\| < \frac{\epsilon}{C}$, then by (14)

$$|\Gamma\left(h - \tilde{h}\right)| = |\Gamma(h) - \Gamma\left(\tilde{h}\right)| < \epsilon$$

which proves continuity.

For the converse, assume that Γ is continuous and fix $\epsilon = 1$. By continuity at $h = 0$, there exists some $\delta_{\epsilon,0} > 0$ such that

$$\|h\| \leq \delta_{\epsilon,0} \implies |\Gamma(h)| \leq 1$$

for any nonzero $h \in V$. Let $g = \frac{\delta_{\epsilon,0} h}{\|h\|}$ and notice that since $\|g\| = \delta_{\epsilon_0}$,

$$|\Gamma(g)| \leq 1$$

which by linearity of Γ implies that

$$\frac{\delta_{\epsilon,0}}{\|h\|}|\Gamma(h)| \leq 1 \Longleftrightarrow |\Gamma(h)| \leq \frac{1}{\delta_{\epsilon,0}}\|h\|$$

for every nonzero $h \in V$. Of course, if $h = 0$ then the final (in) equality follows trivially since both sides are identically zero. This completes our proof with $C = \frac{1}{\delta_{\epsilon,0}}$. □

COROLLARY 5.2.14. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a finite measure space i.e. $\mu(\mathcal{X}) < \infty$ and define*

$$\Gamma : \mathcal{L}^2(\mu) \longrightarrow \mathbb{R}$$

as

$$\Gamma(f) := \bar{\mu}(f)$$

is a continuous linear functional.

PROOF. Note that by the Cauchy-Schwarz (or Hölder's inequality)

$$|\Gamma(f)| \leq \bar{\mu}(|f|) = \|f\|_1 \leq \|f\|_2 \|\mathbb{1}_{\mathcal{X}}\|_2 = \sqrt{\mu(\mathcal{X})}\|f\|_2$$

which completes the proof. □

It turns out that every continuous linear functional on a Hilbert space can be recovered as an inner product on the space. This result is key to proving the existence of conditional expectations and density functions in probability theory.

THEOREM 5.2.15 (Riesz representation theorem). *Let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be a Hilbert space and let*

$$\Gamma : \mathcal{H} \to \mathbb{R}$$

be a continuous linear functional on $\mathcal{H}$. Then there exists a unique element $k \in \mathcal{H}$ such that

$$\Gamma(h) = \langle h, k \rangle$$

for every $h \in \mathcal{H}$.

PROOF. Let $\mathcal{G} := \Gamma^{-1}(\{0\})$ and notice that $\mathcal{G}$ is a closed subspace[5] of $\mathcal{H}$ by the continuity and linearity of Γ and the fact that singletons are closed in the usual topology on $\mathbb{R}$. Note that if Γ is zero everywhere, then $k = 0$ uniquely satisfies the condition of the theorem. Thus, assuming that Γ is not identically zero, we can find some $h_0 \in \mathcal{H}$ such that $\Gamma(h_0) \neq 0$. Define $\tilde{h} := \frac{h_0}{\Gamma(h_0)}$ so that $\Gamma\left(\tilde{h}\right) = 1$ and observe that by Corollary 5.2.11 that

$$\langle \tilde{h} - P_{\mathcal{G}}\tilde{h}, g \rangle = 0$$

for all $g \in \mathcal{G}$. Let $k_0 := \tilde{h} - P_{\mathcal{G}}\tilde{h}$ and note that by linearity

$$\begin{aligned} \Gamma(k_0) &= \Gamma\left(\tilde{h}\right) - \Gamma\left(P_{\mathcal{G}}\tilde{h}\right) \\ &= 1 \end{aligned}$$

since $P_{\mathcal{G}}\tilde{h} \in \mathcal{G}$. By continuity and Proposition 5.2.13

$$1 = \Gamma(k_0) \leq C\|k_0\|$$

for some real C which implies that $\|k_0\| > 0$.

Next, observe that for any $h \in \mathcal{H}$,

$$h - \Gamma(h)\, k_0 \in \mathcal{G}$$

since

$$\begin{aligned} \Gamma(h - \Gamma(h)\, k_0) &= \Gamma(h) - \Gamma(h)\,\Gamma(k_0) \\ &= \Gamma(h) - \Gamma(h) \\ &= 0 \end{aligned}$$

where the second equality follows from the fact that $\Gamma(k_0) = 1$. This implies that

$$\langle k_0, h - \Gamma(h)\, k_0 \rangle = 0$$

[5]This type of a subspace is called the *kernel* of the operator and is often denotedAdd macro

which by linearity of inner products reduces to

$$\langle k_0, h\rangle = \Gamma(h)\,\|k_0\|^2.$$

Recall that $\|k_0\| > 0$ and so we can rearrange and apply linearity once again to deduce

$$\Gamma(h) = \langle h, \frac{k_0}{\|k_0\|^2}\rangle$$

which completes the existence part of the proof.

To see that the representation is unique, note that if $k_1, k_2 \in \mathcal{H}$ are both valid representers then

$$\begin{aligned} 0 &= \langle h, k_1\rangle - \langle h, k_2\rangle \\ &= \langle h, k_1 - k_2\rangle \end{aligned}$$

for any $h \in \mathcal{H}$. In particular, this holds $h = k_1 - k_2$ in which case

$$\|k_1 - k_2\|^2 = 0 \implies k_1 = k_2$$

completing the proof. □

This is the classical Hilbert-space theorem that describes the duality of linear operators and the vector spaces on which they act. This theorem has considerable power in probability because it allows us to construct conditional expectations as orhogonal projections of functions into lower dimensional subspaces, tbe full power of which will become apparent in Chapter 11

5.2.3. Orthonormal bases. Use proposition 19.14 in Bass for proof of existence. Use Schilling for other parts

5.3. Banach spaces over $\mathbb{R}$

We have discussed some important special cases of complete normed vector spacs in the form of the L^p spaces and Hilbert spaces. In this section, we discuss some of the properties of complete normed vector spaces over the reals that hold without reference to any measure or any notion of orthogonality. Without the additional structure afforded by these ideas, the study of normed spaces becomes considerably more complicated. Nevertheless, there are some important results in the theory of general Banach spaces that serve as important tools in analysis, probability, statistics, and economics. As we have already discussed the notion of a normed vector space in the previous sections, we first start with a discussion on the basic properties of such spaces. Most of these should be already known to the reader (indeed, we have implicitly used these concepts throughout this chapter).

5.3.1. Review of the basic properties of normed vector spaces.

Definition 5.3.1. Let V be a vector space over $\mathbb{R}$. A function $\|\cdot\| : V \times V \to [0, \infty]$ is called a norm if for any $u, v \in V$ and any $\alpha \in \mathbb{R}$

(1) $\|v\| = 0 \implies v = 0$ (Positive definiteness)
(2) $\|\alpha v\| = |\alpha|\|v\|$ (Absolute homogeneity)
(3) $\|u + v\| \leq \|u\| + \|v\|$ (Triangle inequality).

Proposition 5.3.2. *Let $(V, \|\|)$ be a normed vector space over $\mathbb{R}$. Then (V, d) where $d(u, v) := \|u - v\|$ is a metric space.*

Proof. Note that $d(u, v) = 0 \implies \|u - v\| = 0 \implies u = v$ by definiteness. Symmetry is follows from absolute homogeneity and the fact that $v - u = -1(u - v)$. The triangle inequality follows from the fact that for any $u, v, w \in V$

$$\begin{aligned} d(u, w) &= \|u - w\| \\ &= \|u - v + v - w\| \\ &\leq \|u - v\| + \|v - w\| \\ &= d(u, v) + d(v, w) \end{aligned}$$

where the inequality is the triangle inequality for norms. □

Definition 5.3.3. Let V be a vector space over $\mathbb{R}$. Given a finite collection of vectors $v_1, \dots v_n$ and scalars $\alpha_1, \dots, \alpha_n$, the sum

$$\sum_{i=1}^{n} a_i v_i \in V$$

is called a *linear combination* of V. For any subset $S \subseteq V$, the span (S) is the set of all linear combinations of S.

Proposition 5.3.4. *Let V be a vector space over $\mathbb{R}$ and let $S \subseteq V$ be a subset. Then* $\operatorname{span}(S) \subseteq V$ *and is a vector space i.e.* $\operatorname{span}(S)$ *is a subspace of V.*

Proof. Note that we only need to show closure under addition and scalar multiplication, along the existence of the additive identity; the other properties are inherited from V. Let $u, v \in \operatorname{span}(S)$ .Then $u = \sum_{i=1}^{m} a_i u_i$ and $v = \sum_{i=1}^{n} b_i v_i$ where $u_i, v_i \in S$ and $a_i, b_i \in \mathbb{R}$. Then $u + v = \sum_{i=1}^{m} a_i u_i + \sum_{i=1}^{n} b_i v_i$ which is another linear combination of vectors in S and so $u + v \in \operatorname{span}(S)$. Similarly, $\alpha v = \sum_{i=1}^{m} (\alpha a_i) v_i$ which is another linear combination of vectors in S. The additive identity $\mathbf{0} \in \operatorname{span}(S)$ because $0 \times v = \mathbf{0}$ where v □

Definition 5.3.5. Let V be a vector space over $\mathbb{R}$. Two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ are considered equivalent if there exist $c, C > 0$ such that for any $v \in V$

$$c\|v\|_a \leq \|v\|_b \leq C\|v\|_b$$

5.3.2. Finite-dimensional normed vector spaces.

5.4. Duality

CHAPTER 6

Differentiation

In elementary calculus, the theory of the Riemann integral was developed with a corresponding theory of differentiation, and these two operations were shown to be the inverse of each other for a suitable class of functions, namely the differentiable functions; this is the fundamental theorem of calculus. The Lebesgue theory also implies a fundamental theorem of calculus for a broader class of functions. The goal of this chapter is to lay the groundwork for this result – the full proof of which will have to be deferred to Chapter 7– and in order to do so, we will develop some theory that is very important in its own right. Before delving into new material, it might be worthwhile to review some basic facts about differentiation in the spirit of our discussion on the Riemann integral. In addition, I will take this opportunity to suitably generalize the notion of a derivative by defining differentiability as a property of functions on arbitrary normed vector spaces.

6.1. Differentiation in normed vector spaces

6.1.1. Review of derivatives on the line. We first review some basic material from single variable calculus for completeness; note that this section is technically presumed knowledge and so results appearing previously in the text use some of the facts established here. We will avoid circularity by keeping this section self-contained without any reference to the material we have developed so far in the body of the text (we will use results from the appendices).

Definition 6.1.1. A function $f:[a,b]\to\mathbb{R}$ is said to be *differentiable at a point* $c\in(a,b)$ if

$$\lim_{x\to a}\frac{f(x)-f(a)}{x-a}$$

exists as a real number. The function is said to be *differentiable* if it is differentiable at every point in the interior of its domain. If a function f is differentiable everywhere, the derivative function is denoted f'as in

$$f'(y)=\lim_{x\to y}\frac{f(x)-f(y)}{x-y}.$$

Remark. The derivative is also denoted $\frac{df}{dy}=f'(y)$ which captures the heuristic that the derivative can be thought as the ratio of small quantities.

Example 6.1.2. It should be immediately clearly that for any constant function $f(x)=a$, its derivative is zero everywhere, and for the function $f(x)=x$, its derivative is 1 is everywhere.

Proposition 6.1.3. *Let* $f:[a,b]\to\mathbb{R}$ *be differentiable at a point* $c\in(a,b)$ *. Then it is continuous at* c*.*

Proof. Recall that the limit of the product of two functions is the product of the limit, provided those individual limits exist.

$$\begin{aligned}\lim_{x\to c}f(x)-f(c)&=\lim_{x\to c}\frac{f(x)-f(c)}{x-c}x-c\\&=\lim_{x\to c}\frac{f(x)-f(c)}{x-c}\lim_{x\to c}x-c\\&=f'(c)\,0\\&=0\end{aligned}$$

which completes the proof. □

REMARK. The converse of this theorem is not true; for instance, consider that $|x|$ is continuous (this follows by the "reverse" triangle inequality $||x|-|y|| \leq |x-y|$). That it is not differentiable can be seen by looking at

$$\lim_{x\to 0^-} \frac{|x|}{x} = -1 \neq 1 = \lim_{x\to 0^+} \frac{|x|}{x}.$$

PROPOSITION 6.1.4. *Let f,g be real-valued functions on $[a,b]$ that are differentiable at some $x \in [a,b]$. Then $f+g$ and fg are differentiable at x. If $g(x) \neq 0$ then $\frac{f}{g}$is also differentaible. The derivatives are given*

$$(f+g)'(x) = f'(x) + g'(x) \tag{15}$$

$$(fg)'(x) = f'(x)\, g(x) + g'(x)\, f(x) \tag{16}$$

$$\left(\frac{f}{g}\right)'(x) = \frac{f'(x)\, g(x) - g'(x)\, f(x)}{(g(x))^2} \tag{17}$$

PROOF. The first one, called the sum rule, follows by algebra of limits. The second, known as the product rule, follows by the same trick that is used to prove that the product of two sequences converges to the product of their limits. To reiterate, note that

$$\begin{aligned}\frac{f(y)\,g(y) - f(x)\,g(x)}{y-x} &= \frac{f(y)\,g(y) - f(y)\,g(x) + f(y)\,g(x) - f(x)\,g(x)}{x-c} \\ &= \frac{f(y)\,(g(y)-g(x))}{y-x} + \frac{g(x)\,(f(y)-f(x))}{y-x}.\end{aligned}$$

Taking limits and applying Proposition 6.1.3 gets the result.

Finally, the third result, called the quotient rule, follows as

$$\begin{aligned}\frac{\frac{f(y)}{g(y)} - \frac{f(x)}{g(x)}}{y-x} &= \frac{1}{g(x)\,g(y)}\left(\frac{f(y)\,g(x) - f(x)\,g(y)}{y-x}\right) \\ &= \frac{1}{g(x)\,g(y)}\left(\frac{f(y)\,g(x) - f(y)\,g(y) + f(y)\,g(y) - f(x)\,g(y)}{y-x}\right) \\ &= \frac{1}{g(x)\,g(y)}\left(\frac{f(y)\,[g(x)-g(y)]}{y-x} + \frac{g(y)\,[f(y)-f(x)]}{y-x}\right)\end{aligned}$$

and taking limits and again applying the continuity of differentiable functions completes the proof. □

COROLLARY 6.1.5 (Power rule). *The derivative of a real-valued function on $\mathbb{R}$ defined $f(x) = x^n$ where $n \in \mathbb{Z}$ is given by $f'(x) = nx^{n-1}$.*

PROOF. First we examine the case when $n \in \mathbb{N}$. For the base cases of $n=0$ and $n=1$ look at the remark above. Now suppose that the rule holds for $n \in \mathbb{N}$. Then, writing $x^{n+1} = xx^n$, we apply the product rule to yield

$$\frac{dx^{n+1}}{dx} = xnx^{n-1} + x^n = (n+1)\, x^n$$

which completes the proof.

To extend the result to negative integers, we can apply the quotient rule. Note that for a negative integer m, we have some positive integern n such that $m = -n$ and so $f(x) = x^m = \frac{1}{x^n}$ which has derivative (by applying the power rule for positive integers and the quoteint rule)

$$f'(x) = \frac{-nx^{n-1}}{x^{2n}} = mx^{m-1}.$$

□

Remark. The power rule also holds for rational exponents; we shall establish this fact after we state the chain rule. For real exponents, the power rule holds if $x \geq 0$ since we define $x^a := \exp(a \log(x))$ when $a \in \mathbb{R}$. Since the logarithm is only defined on the positive reals, the result only extends to that case (the case of $x = 0$ is trivial). We shall prove this after we construct the exponential and natural logarithm functions in Appendix section 6.1.2.

Proposition 6.1.6. *Let $f \in P_n(x)$ the space of $n-$degree polynomials. Then f is differentiable and its derivative $f' \in P_{n-1}(x)$.*

Proof. Let $f(x) = \sum_{i=0}^{n} a_i x^i$. By the “sum rule”,”product rule” and the “power rule”,

$$f'(x) = \sum_{i=0}^{n} a_i i x^{i-1} \in P_{n-1}(x).$$

□

Remark. An immediate corrolary is that every polynomial is infinitely differentiable. Infinitely differentiable functions are called *smooth.*

Theorem 6.1.7 (Chain rule). *Let $g : C \subseteq \mathbb{R} \to \mathbb{R}$ be differentiable at some $c \in D$ and let $f : D \subseteq \mathbb{R} \to \mathbb{R}$ be differentiable at $g(c) \in D$. Then the composition $f \circ g$ is differentiable at c with derivative*

$$(f \circ g)'(c) = f'(g(c))\, g'(c).$$

We omit the proof of this theorem right now; we will prove it in the more general setting of derivatves on functions between Banach spaces in the next section.

Proposition 6.1.8. *A function $f(x) = x^r$ where $r \in \mathbb{Q}$ and $x \in \mathbb{R}$ is differentiable with derivative $f'(x) = rx^{r-1}$.*

Proof. We will establish the result when $x \neq 0$ (when $x = 0$, the derivative is 0 as established in an earlier remark). Let $r = \frac{p}{q}$ where $p, q \in \mathbb{Z}$ and note that $g(x) := x^q$ has derivative $g'(x) = qx^{q-1}$ by the integral power rule6.1.5. Then, observe that $g(f(x)) = x^p$ and so applying the chain rule on the left hand hand side and the integral power rule on the right hand side, we have

$$qf(x)^{q-1} f'(x) = px^{p-1}$$

which can be re-arranged to yield

$$\begin{aligned} f'(x) &= \frac{px^{p-1}}{qf(x)^{q-1}} \\ &= \left(\frac{p}{q}\right) x^{p-1-\frac{p}{q}(q-1)} \\ &= \frac{p}{q} x^{\frac{p}{q}-1} \\ &= rx^{r-1} \end{aligned}$$

which completes the proof. □

Example 6.1.9 (ISI 2013 PSB 2). Let $a_1 < a_2 < \cdots < a_m$ and $b_1 < b_2 < \cdots < b_n$ be real numbers such that

$$\sum_{i=1}^{m} |a_i - x| = \sum_{j=1}^{n} |b_j - x| \text{ for all } x \in \mathbb{R}.$$

We can use differentiability to show that $m = n$ and $a_j = b_j$ for1 $\leq j \leq n$. To see this, note that $f(x) := \sum_{i=1}^{m} |a_i - x|$ is not differentiable exactly on the set $F := \{a_1, \ldots, a_n\}$. This is because eacj component $|a_i - x|$ is non-differentiable only at $a_i$and the sum of differentiable and non-differentiable functions are not differentiable. Similarly, the function $g(x) := \sum_{i=1}^{n} |b_i - x|$ is exactly not differentiable on the set G which implies $F = G$.

6.1.1.1. *Mean value theorems.*

DEFINITION 6.1.10. A function $f : [a, b] \to \mathbb{R}$ is set to have a local maximum at $c \in [a, b]$ if there exists some $\delta > 0$ such that for every $x \in (c - \delta, c + \delta) \cap [a, b]$

$$f(c) \geq f(x).$$

Local minima are defined analagously.

PROPOSITION 6.1.11. *Let $f : [a, b] \to \mathbb{R}$ have a local maximum or minimum at $c \in (a, b)$. If $f'(c)$ exists then $f'(c) = 0$.*

PROOF. We consider the case of the local maximum; the case of the local minimum follows by applying the maximum result to $-f$. Suppose $c \in (a, b)$ is a local maximum for f, in which case we can choose some $\delta > 0$ small enough that

$$(c - \delta, c + \delta) \subseteq (a, b)$$

and $f(c) \geq f(x)$ on $(c - \delta, c + \delta)$. Note that for $c - \delta < x < c$ we have that

$$\frac{f(x) - f(c)}{x - c} \geq 0$$

and so

$$f'\left(c^{-}\right) := \lim_{x \to c^{-}} \frac{f(x) - f(c)}{x - c} \geq 0.$$

Similarly, for $c < x < c + \delta$

$$\frac{f(x) - f(c)}{x - c} \leq 0$$

and so

$$f'\left(c^{+}\right) := \lim_{x \to c^{+}} \frac{f(x) - f(c)}{x - c} \leq 0.$$

But of course, $f'(c) = f'(c^{-}) = f'(c^{+})$ since left and right hand limits are always equal to the limit when it exists and thus

$$f'(c) = 0.$$

□

THEOREM 6.1.12 (Cauchy's mean value theorem). *If f, g are real-valued continuous functions on $[a, b]$ such that the derivative functions f', g' exist on (a, b), there exists some $x \in (a, b)$ such that*

$$(f(b) - f(a))\, g'(x) = (g(b) - g(a))\, f'(x)$$

PROOF. Write

$$h(t) := (f(b) - f(a))\, g(t) - (g(b) - g(a))\, f(t)$$

and observe that since it's built out of differeniable functions by adding and multiplying them, h is continuous and differentiable. Since $[a, b]$ is compact, h has a minimum and maximum by Weierstrass'. If the minimum and maximum are both at the end points a and b, then the function is constant and the derivative is zero everywhere, which yields the result. If at least one of the minimum or maximum is interior, say at $x \in (a, b)$, then by Proposition 6.1.11, $f'(x) = 0$ which again yields the result. □

COROLLARY 6.1.13 (Mean value theorem). *Let $f : [a, b] \to \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b). Then there exists some $x \in (a, b)$ such that*

$$f'(c) = \frac{f(b) - f(a)}{b - a}.$$

PROOF. Apply Cauchy's mean value theorem with $g(x) = x$. □

PROPOSITION 6.1.14. *If $f:[a,b]\to\mathbb{R}$ is differentiable on (a,b) then*

$$\forall x\in(a,b): f'(x)\geq 0 \implies (x_1\geq x_2 \implies f(x_1)\geq f(x_2))$$
$$\forall x\in(a,b): f'(x)\leq 0 \implies (x_1\geq x_2 \geq f(x_1)\leq f(x_2))$$

PROOF. Applying the mean value theorem, we have

$$f(x_1)-f(x_2)=f'(c)(x_1-x_2)$$

for $c\in(a,b)$ and if $f'(c)\geq 0$ then $x_1\geq x_2$ implies that $f(x_!)\geq f(x_2)$ and the other implication follows in the same way. □

REMARK. An obvious corollary of the above result is that if $\forall x\in(a,b)$: $f'(x)=0 \implies f$ is constant on (a,b).

6.1.1.2. *Continuity of derivatives.* For a function f on $[a,b]$ that is differentiable everywhere on (a,b), it need not be the case that the derivative function f' is continuous everywhere. In fact, the points of discontinuity of a derivative can be dense in the domain and have positive Lebesgue measure. We postpone the discussion of these more complicated examples until we prove the fundamental theorem of calculus. The simplest canonical example of a discontinuous derivative is the following

EXAMPLE 6.1.15. Let $f(x)=x^2\sin\left(\frac{1}{x}\right)\mathbb{1}\{x\neq 0\}$.We take on faith the fact that the function $\sin(x)$ is bounded between -1 and 1, is infinitely differentiable and that its derivative is $\cos(x)$. We construct these functions from first principles in Appendix section 6.1.2. Note that the derivative of this function away from zero is given by the chain and product rules as

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

The derivative at zero can be computed from first principles by looking at the limit

$$\lim_{x\to 0}\frac{x^2\sin\left(\frac{1}{x}\right)\mathbb{1}\{x\neq 0\}}{x}=\lim_{x\to 0}x\sin\left(\frac{1}{x}\right)\mathbb{1}\{x\neq 0\}=0$$

since $\left|\sin\left(\frac{1}{x}\right)\right|\leq 1$. Then we have that

$$f'(x)=\begin{cases}2x\sin\left(\frac{1}{x}\right)-\cos\left(\frac{1}{x}\right), & x\neq 0\\ 0, & x=0\end{cases}$$

but $\lim_{x\to 0}f'(x)$ does not exist since $\cos\left(\frac{1}{x}\right)$ oscillates rapidly near zero.

While derivatives can be discontinuous, they share the intermediate value property with continuous functions.

PROPOSITION 6.1.16 (Darboux's theorem). *Let $f:[a,b]\to\mathbb{R}$ be differentiable on (a,b). Then the derivative function $f':(a,b)\to\mathbb{R}$ has the intermediate value property that the image $f[(a,b)]$ is an interval.*

PROOF. Let $c,d\in(a,b)$ be such that, without loss generality, $f'(c)>f'(d)$. Take any $y\in(f'(c),f'(d))$. The function $g(x)=f(x)-yx$ is continuous on $[\min\{c,d\},\max\{c,d\}]$ and thus achieves a maximum and a minimum by Weierstrass' theorem. If both the minimum and maximum occur at the end points then the function g is constant on $[\min\{c,d\},\max\{c,d\}]$ and so its derivative is zero everywhere, completing the proof. If there's at least one interior extrema, say at $t\in(\min\{c,d\},\max\{c,d\})$, then

$$f'(t)=y$$

by Proposition 6.1.11. □

LEMMA 6.1.17. *Any function that satisifies the intermediate value property cannot have removable or jump discontinuities.*

PROOF. Let $f : [a, b] \to \mathbb{R}$ be a function that satisfies the intermediate value property and let $c \in (a, b)$ such that $\lim_{x \to c^-} f(x) = L$ and $\lim_{x \to c^+} f(x) = M$. Suppose, without loss of generality, that $L \neq f(c)$ and let $\epsilon = \frac{|L - f(c)|}{2}$. Note that by definition, there exists some $\delta > 0$ such that for any x with $0 < c - x < \delta \implies |f(x) - L| < \epsilon$ which implies that $|f(x) - f(c)| \geq \epsilon$. By the intermediate value property, for any y in between $f(x)$ and $f(c)$, there exists some $z \in (x, c)$ such that $f(z) = y$. In particular, this holds true for any y.such that $|y - f(c)| < \epsilon$. But then $0 < c - z < \delta$ which implies that $|y - f(c)| \geq \epsilon$ which is a contradiciton. □

6.1.1.3. *Differentiability classes and local approximation by polynomials.*

EXAMPLE 6.1.18. Let f be a function such that $f(0) = 0$ and f has derivatives of all order. Show that

$$\lim_{h \to 0} \frac{f(h) + f(-h)}{h^2} = f''(0)$$

where $f''(0)$ is the second derivative of f at 0 .TODO

6.1.2. Special functions.

6.1.3. Derivatives of functions between Banach spaces.

6.2. Decomposition of measures

Recall that in Proposition 3.2.10, we showed that for a measure space $(\mathcal{X}, \mathcal{F}, \mu)$ and a non-negative measurable function $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$, the function $\nu(A) := \bar{\mu}(f \mathbb{1}_A)$is a measure. Moreover, in the remark following the proposition, we observed that such a measure ν was *absolutely continuous* with respect to μ, that is to say, for any $A \in \mathcal{F}, \mu(A) = 0 \implies \nu(A) = 0$. Further, we also claimed that every absolutely continuous measure could be represented this way i.e as an integral with respect to the dominating measure. In other words, we claimed that there was always an $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ such that this relation would hold. It turns out we can establish a slightly stronger result, which in turn is key to establishing the fundamental theorem of calculus for the Lebesgue measure. What we can show, using the Hilbert space machinery we developed earlier, is that we can decompose a measure (relative to another measure) into two measures that are in some sense orthogonal. This is in direct correspondence with the decomposition of Hilbert spaces given to us by the projection theorem (Theorem 5.2.10).

To motivate how the Hilbert space theory can be useful in this context, we can begin with a change of perspective in the spirit of Theorem 3.2.16. Note that saying $\nu(A) = \bar{\mu}(f \mathbb{1}_A)$ for every $A \in \mathcal{F}$ is equivalent to saying $\bar{\nu}(g) = \bar{\mu}(fg) \forall g \in \mathcal{L}^1(\mathcal{X}, \mathcal{F}, \nu)$. The right hand side is the inner product associated with the space $\mathcal{L}^2(\mu)$ and so we wish to show that a linear functional $g \to \bar{\nu}(g)$ can be represented as an inner product on $\mathcal{L}^2(\mu)$. At this juncture, one can guess that the Riesz representation theorem for Hilbert spaces is clearly right tool to finish off the proof; however, several hurdles remain. For one, $g \to \bar{\nu}(g)$ a linear functional on $\mathcal{L}^1(\nu)$ and it is not $\mathcal{L}^2(\mu)$. Further, it is unclear whether our functional is bounded, which is a necessary condition for the theorem to apply. The following lemma provides the conditions under which the theorem finds purchase.

LEMMA 6.2.1. *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space. If μ, ν are two $\sigma-$finite measures on $\mathcal{F}$, then there exists some non-negative (almost everywhere with respect to μ) measurable function $k \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ and a set $G \in \mathcal{F}$ such that $\mu(G) = 0$ and*

$$\nu(A) = \bar{\mu}(k \mathbb{1}_A) + \nu(A \cap G)$$

for every $A \in \mathcal{F}$. Moreover, the function k is unique $\mu-$almost-everywhere and the set G is unique $\mu + \nu$-almost everywhere.

PROOF. Suppose first that the measures $\mu$and $\nu$are finite. Define the measure $\psi := \mu + \nu$ and observe that by Proposition 3.2.15

$$\bar{\psi}(f) = \bar{\mu}(f) + \bar{\nu}(f).$$

Next define an operator

$$\Gamma : \mathcal{L}^2(\psi) \longrightarrow \mathbb{R}$$

by

$$\Gamma(f) := \bar{\nu}(f)$$

and notice that Γ is a continuous linear functional by a slight variant of the argument presented in the proof of Corollary 5.2.14, using the fact that $\mu(\mathcal{X}), \nu(\mathcal{X}) < \infty$. Since $\mathcal{L}^2(\psi)$ is a Hilbert space by Corollary 5.2.8 and Theorem 5.1.10, the Riesz representation theorem tells us there exists an (almost $\psi-$everywhere unique) function $h \in \mathcal{L}^2(\psi)$ such that

$$\Gamma(f) = \langle f, h\rangle$$

for every $f \in \mathcal{L}^2(\psi)$. Of course, in the context of $\mathcal{L}^2$,

$$\langle f, h\rangle = \|fh\|_1 = \bar{\psi}(fh) = \bar{\mu}(fh) + \bar{\nu}(fh). \tag{18}$$

Next consider the following measurable partition of $\mathcal{X}$

$$N := \{x \in \mathcal{X} \mid h(x) < 0\}, M := \{x \in \mathcal{X} \mid 0 \leq h(x) < 1\}, G := \{x \in \mathcal{X} \mid h(x) \geq 1\}.$$

Note that by the fact that $0 \geq h\mathbb{1}_N$ and the monotonicity of integration

$$\begin{aligned}0 &\geq \bar{\psi}(h\mathbb{1}_N)\\ &= \bar{\mu}(h\mathbb{1}_N) + \bar{\nu}(h\mathbb{1}_N)\\ &= \bar{\nu}(\mathbb{1}_N)\end{aligned}$$

where the last equality is due to (18). But since $\bar{\nu}(\mathbb{1}_N) = \nu(N)$,we have $\nu(N) = 0$ by non-negativity of measures. Then

$$\bar{\mu}(h\mathbb{1}_N) + \bar{\nu}(h\mathbb{1}_N) = 0$$

where the second term is automatically zero since $h\mathbb{1}_N \overset{\nu-\text{a.e}}{=} 0$. Therefore,

$$\bar{\mu}(h\mathbb{1}_N) = 0$$

which by Proposition 3.3.3 implies that

$$h\mathbb{1}_N \overset{\mu-\text{a.e}}{=} 0 \implies \mu(N) = 0$$

and so both $\mu(N) = \nu(N) = 0$.

On the other hand, observe that

$$\begin{aligned}\Gamma(\mathbb{1}_G) &= \nu(G)\\ &= \bar{\mu}(h\mathbb{1}_G) + \bar{\nu}(h\mathbb{1}_G)\\ &\geq \mu(G) + \nu(G)\end{aligned}$$

where the inequality follows from the fact that $h\mathbb{1}_G \geq \mathbb{1}_G$ and the monotonicity of the integral. Since $\nu(G) < \infty$ as ν is a finite measure, we can subtract if from both sides of the inequality to deduce

$$\mu(G) \leq 0$$

which reduces to equality by the non-negativity of measures.

To control the final piece of the partition, observe that we can use the definition $\Gamma(f) = \bar{\nu}(f)$ and Eq. (18) together to deduce (since all terms are finite) that

$$\bar{\nu}(f) - \bar{\nu}(fh) = \bar{\mu}(fh)$$

for $f \in \mathcal{L}^2(\psi)$. By linearity this reduces to

$$\bar{\nu}((1-h)f) = \bar{\mu}(hf). \tag{19}$$

Next we define the increasing sequence of sets

$$M_n := \left\{x \in \mathcal{X} \mid 0 \leq h(x) \leq 1 - \frac{1}{n}\right\}$$

and note that

$$M = \bigcup_{n\in\mathbb{N}} M_n.$$

Further, observe that for any $A \in \mathcal{F}$ the functions

$$f_n := \frac{\mathbb{1}_{M_n} \mathbb{1}_A}{1-h} \leq n \mathbb{1}_{M_n} \mathbb{1}_A$$

and

$$0 \leq f_n \leq f_{n+1}$$

pointwise for all $n \in \mathbb{N}$. This shows us that our functions $f_n \in \mathcal{L}^2(\psi)$[1] and are monotonically increasing. Therefore,

$$\begin{aligned}
\nu(M \cap A) &= \bar{\nu}\left(\mathbb{1}_M \mathbb{1}_A \frac{1-h}{1-h}\right) \\
&= \lim_{n\to\infty} \bar{\nu}\left(\frac{\mathbb{1}_{M_n}\mathbb{1}_A}{1-h} 1-h\right) \\
&= \lim_{n\to\infty} \bar{\mu}\left(\frac{\mathbb{1}_{M_n}\mathbb{1}_A}{1-h} h\right) \\
&= \bar{\mu}\left(\mathbb{1}_M \mathbb{1}_A \frac{h}{1-h}\right) + \bar{\mu}\left(\mathbb{1}_{N\cup G}\mathbb{1}_A \frac{h}{1-h}\right) \\
&= \bar{\mu}\left(\frac{h}{1-h}\mathbb{1}_A\right)
\end{aligned}$$

where the second equality follows by the monotone convergence theorem, the third equality is due to (19), the fourth equality is due to the fact that $\mu(G) = \mu(N) = 0$ and the last equality is due to the linearity of integration.

Letting $k = \frac{h}{1-h}$ we have that

$$\mu(\{x \in \mathcal{X} \mid k(x) < 0\}) = 0$$

since $\mu(N) = \mu(G) = 0$. Therefore

$$\begin{aligned}
\nu(A) &= \nu((A \cap N) \cup (A \cap M) \cup (A \cap G)) \\
&= \nu(A \cap N) + \nu(A \cap M) + \nu(A \cap G) \\
&= \bar{\mu}(k\mathbb{1}_A) + \nu(A \cap G)
\end{aligned}$$

where the first equality follows from the fact that N, M, and G form a partition of $\mathcal{X}$, the second equality by finite additivity, and the final equality by the fact that N is a $\nu-$null set.

The extension to $\sigma-$finite measures is relatively straightforward. Let $\{E_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ be a partition of $\mathcal{X}$ such that $\mu$and $\nu$are finite on each E_i (this is possible by Propositions 1.3.9 and 1.3.10). Then the measures given by

$$\begin{aligned}
\nu_i(A) &:= \nu(A \cap E_i), \\
\mu_i(A) &:= \mu(A \cap E_i)
\end{aligned}$$

are finite on $\mathcal{F}$ and so the finite version of theorem implies the existence of functions $k_i \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ and μ_i-null sets B_i such that

$$\nu_i(A) = \bar{\mu_i}(k_i \mathbb{1}_A) + \nu_i(A \cap B_i)$$

which is equivalent to saying that

$$\nu(A \cap E_i) = \bar{\mu}(k_i \mathbb{1}_A \mathbb{1}_{E_i}) + \nu(A \cap B_i \cap E_i)$$

---

[1] $n\mathbb{1}_{M_n}$is square integrable since ψ is a finite measure; f_n then is integrable since both functions are non-negative and so

$$\bar{\psi}(|f_n|^2) \leq \bar{\psi}(|n\mathbb{1}_{M_n}\mathbb{1}_A|^2) \leq n^2\psi\left(\left\{0 \leq h \leq 1 - \frac{1}{n}\right\}\right)$$

for each $i \in \mathbb{N}$. Summing over i we have

$$\begin{aligned}\nu(A) &= \sum_{i=1}^{\infty} \nu(A \cap E_i) \\ &= \sum_{i=1}^{\infty} \bar{\mu}(k_i \mathbb{1}_{E_i} \mathbb{1}_A) + \sum_{i=1}^{\infty} \nu(A \cap B_i \cap E_i) \\ &= \bar{\mu}\left(\sum_{i=1}^{\infty} k_i \mathbb{1}_{E_i} \mathbb{1}_A\right) + \nu\left(A \cap \left(\cup_{i\in\mathbb{N}} (B_i \cap E_i)\right)\right).\end{aligned}$$

where the last equality follows by monotone convergence and countable additivity. Note that $\mu\left(\bigcup_{i\in\mathbb{N}} (B_i \cap E_i)\right) = \sum_{i=1}^{\infty} \mu(B_i \cap E_i) = \sum_{i=1}^{\infty} \mu_i(B_i) = 0$. This completes the proof of existence with $k = \sum_{i=1}^{\infty} k_i \mathbb{1}_{E_i}$ and $B = \bigcup_{i\in\mathbb{N}} (B_i \cap E_i)$.

Finally, for uniqueness, consider two pairs (k_1, G_1) and (k_2, G_2) such that $k_1, k_2 \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$, and $G_1, G_2 \in N_\mu$, and

$$\nu(A) = \bar{\mu}(k_1 \mathbb{1}_A) + \nu(A \cap G_1) = \bar{\mu}(k_2 \mathbb{1}_A) + \nu(A \cap G_2)$$

for every $A \in \mathcal{F}$. Then,

$$\begin{aligned}\nu\left(G_2 \cap G_1^C\right) &= \bar{\mu}\left(k_1 \mathbb{1}_{G_2} \mathbb{1}_{G_1^C}\right) + \nu\left(\underbrace{G_2 \cap G_1^C \cap G_1}_{=\emptyset}\right) \\ &= 0\end{aligned}$$

where the first term is zero because $G_1 \in N_\mu$. Similarly, we can show that

$$\begin{aligned}\nu\left(G_1 \cap G_2^C\right) &= \bar{\mu}\left(k_2 \mathbb{1}_{G_1} \mathbb{1}_{G_2^C}\right) + \nu\left(\underbrace{G_1 \cap G_2^C \cap G_2}_{=\emptyset}\right) \\ &= 0.\end{aligned}$$

In other words, $\nu(G_1 \Delta G_2) = 0$ which by Proposition 3.3.4 implies that $G_1 \overset{\nu-\text{a.e}}{=} G_2$. Of course, since $G_1, G_2 \in N_\mu$, we have that $\mu(A\Delta B) = 0$ and so $G_1 \overset{\mu+\nu-\text{a.e}}{=} G_2$. Note that since $G_1^C \Delta G_2^C = G_1 \Delta G_2$, we have that $G_1^C \overset{\mu+\nu-\text{a.e}}{=} G_2^C$ and so for any $A \in \mathcal{F}$

$$\nu\left(A \cap G_1^C\right) = \nu\left(A \cap G_2^C\right).$$

But

$$\nu\left(A \cap G_1^C\right) = \bar{\mu}\left(k_1 \mathbb{1}_A \mathbb{1}_{G_1^C}\right) + \nu\left(\underbrace{A \cap G_1^C \cap G_1}_{=\emptyset}\right)$$

$$\nu\left(A \cap G_2^C\right) = \bar{\mu}\left(k_2 \mathbb{1}_A \mathbb{1}_{G_2^C}\right) + \nu\left(\underbrace{A \cap G_2^C \cap G_2}_{=\emptyset}\right)$$

This combined with the fact that G_1 and G_2 are $\mu-$null show that for every $A \in \mathcal{F}$ we have $\bar{\mu}(k_1 \mathbb{1}_A) = \bar{\mu}(k_2 \mathbb{1}_A)$ which by Proposition 3.3.9 implies that $k_1 \overset{\mu-\text{a.e}}{=} k_2$. $\square$

This is a powerful lemma as it gives us two very important results as simple corollaries.

Definition 6.2.2. Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let μ, ν be two measures on this space. We say ν is absolutely continuous with respect to μ if for any $A \in F : \mu(F) = 0 \implies \nu(F) = 0$. We denote this relation by $\nu << \mu$.

Why is this relation labeled continuity? There are two good reasons for this: first, at least for finite measures, the above definition is equivalent to an $\epsilon - \delta$ definition that resembles the continuity definitions that we have seen before. More importantly, we will see that measures that are absolutely continuous with respect to the Lebesgue measure give rise to a particular class of real-valued *functions* with the property of absolute continuity. Such functions are exactly the class of functions for which the Lebesgue fundamental theorem of calculus applies.

PROPOSITION 6.2.3. *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let μ, ν be two measures on this space such that $\nu(X) < \infty$. Then $\nu << \mu$ if and only if for every $\epsilon > 0$ there exists a $\delta > 0$ such that for any $A \in \mathcal{F}$, $\mu(A) < \delta \implies \nu(A) < \epsilon$*

PROOF. Suppose that $\epsilon - \delta$ characterization holds. Take a sequence $\epsilon_n > 0$ such that $\epsilon_n \to 0$ and find the corresponding $\delta_n > 0$ such that $\mu(A) < \delta_n \implies \nu(A) < \epsilon_n$. If $\mu(A) = 0$ then $\mu(A) < \delta_n$ for all $n \in \mathbb{N}$. Then $\nu(A) < \epsilon_n$ for every n. Taking limits yields $\nu(A) = 0$.

Conversely, assume that $\epsilon - \delta$ property does not hold. Then, there exists some $\epsilon_0 > 0$ such that for any $\delta > 0$, there's some $A \in \mathcal{F}$ such that $\mu(A) < \delta$ but $\nu(A) \geq \epsilon$. Let $\delta_n = \frac{1}{2^n} \to 0$ and let $A_n \in \mathcal{F}$ be the corresponding sequence of sets such that $\mu(A_n) < \delta_n$ but $\nu(A_n) \geq \epsilon$. Note then $\sum_{n=1}^{\infty} \mu(A_n) < 1$ and so by the Borel-Cantelli lemma,

$$\mu(\limsup_{n\to\infty} A_n) = 0.$$

where $\limsup_{n\to\infty} A_n = \cap_{n=1}^{\infty} \cup_{i=1}^{n} A_i \in \mathcal{F}$. Since $\nu(X) < \infty$, by the reverse Fatou lemma

$$\nu(\limsup_{n\to\infty} A_n) \geq \limsup_{n\to\infty} \nu(A_n) \geq \epsilon$$

which completes the proof. □

While this is one way to characterize absolutely continuous measures, a different characterization provides the shortest route to the fundamental theorem of calculus.

THEOREM 6.2.4 (Radon-Nikodym). *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space. If μ, ν are two σ-finite measures on $\mathcal{F}$ then $\nu << \mu$ if and only if there exists almost-everywhere unique (with respect to both measures) non-negative function $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ such that*

$$\nu(A) = \bar{\mu}(f \mathbb{1}_A)$$

for every $A \in \mathcal{F}$

PROOF. Note that the existence of f implies absolute continuity by Proposition 3.2.10. The converse is far more challenging; fortunately for us, Lemma 6.2.1 does all of the work. To see this, note that by our lemma

$$\nu(A) = \bar{\mu}\left(\tilde{f}\mathbb{1}_A\right) + \nu(A \cap B)$$

where $A \in \mathcal{F}$, $B \in N_\mu$ and $\tilde{f}$ is $\mu-$almost everywhere non-negative and unique. If $\nu << \mu$ then $\nu(A \cap B) = 0$. Letting $f = \tilde{f}\mathbb{1}_{\tilde{f}\geq 0}$, our result follows by Proposition 3.3.7. □

COROLLARY 6.2.5. *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space. If μ, ν are two σ-finite measures on $\mathcal{F}$ such that $\mu << \nu$ then there exists a unique $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$ such that for any $g \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$*

$$\bar{\mu}(g) = \bar{\nu}(gf).$$

PROOF. (Sketch) This is the standard approximation argument. For indicator functions the result follows by the Radon-Nikodym theorem. For any simple function, it follows by the linearity of the integral (and the result for indicators). Finally for non-negative measurable function it follows by monotone convergence (and the result for simple functions). Uniqueness follows by Proposition 3.3.9. □

The Radon-Nikodym theorem builds on the Hilbert space theory we described earlier in Chapter 5 and is the proper converse of the comparatively simple result on constructing new measures using non-negative measurable functions that we saw in Proposition 3.2.10. This theorem is also fundamental to developing conditional expectations; indeed the existence and uniqueness of conditional expectations is an almost trivial corollary of the Radon-Nikodym theorem. In the context of probability theory, the functions described by the Radon-Nikodym theorem are probability density functions. Recall from undergraduate probability that probability density functions could be recovered as derivatives of cumulative distribution functions, provided those distribution functions were sufficiently well behaved. It turns out such densities can be thought of as derivatives generally in an admittedly contrived sense: if we write the Radon Nikodym theorem in traditional notation, then we have can write for $\nu << \mu$

$$\nu(A) = \int_A f d\mu.$$

Then, in Leibniz notation, we could write f as $\frac{d\nu}{d\mu}$ and refer to f as the *Radon-Nikodym derivative* of ν with respect to μ. Of course, for function to actually be some sort of derivative of measures, we need to be able to represent it as a limiting ratio of the two measures; in the special case of the Lebesgue measure, we can indeed do this. However, the notation is useful more generally, partly due to the following facts

PROPOSITION 6.2.6. *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let $\mu, \nu,$ and γ be $\sigma-$finite measures on the space. Then*

(i) If $\mu << \nu$ and $\nu << \gamma$ then $\mu << \gamma$ and

(Chain rule) $$\frac{d\mu}{d\gamma} \overset{a.e}{=} \frac{d\mu}{d\nu}\frac{d\nu}{d\gamma}$$

(ii) If $\mu << \gamma$ and $\nu << \gamma$ then the sum measure $\mu + \nu$ is $\sigma-$finite, $\mu + \nu << \gamma$, and

(Sum Rule) $$\frac{d\mu + \nu}{d\gamma} \overset{\gamma-a.e}{=} \frac{d\mu}{d\gamma} + \frac{d\nu}{d\gamma}$$

(iii) If $\mu << \nu$ then $\nu << \mu$ if and only if $\frac{d\mu}{d\nu} > 0$ $\nu-a.e$ and then

(Inverse "function" rule) $$\frac{d\nu}{d\mu} = \frac{1}{\frac{d\mu}{d\nu}}.$$

PROOF. To show *(i)*, let $A \in \mathcal{F}$ be such that $\gamma(A) = 0$.Then $\nu(A) = 0$ since $\nu << \gamma$. Then $\mu(A) = 0$ since $\mu << \nu$ which establishes that absolute continuity is transitive. Note that since the measures are $\sigma-$finite, by the Radon Nikodym theorem there exists some almost everywhere unique non-negative measurable functions f, g, h such that for any $A \in \mathcal{F}$

$$\mu(A) = \bar{\gamma}(h\mathbb{1}_A)$$
$$\mu(A) = \bar{\nu}(f\mathbb{1}_A)$$
$$\nu(A) = \bar{\gamma}(g\mathbb{1}_A)$$

and so by Corollary 3.2.12, we have that - for any $A \in \mathcal{F}$

$$\mu(A) = \bar{\nu}(f\mathbb{1}_A) = \bar{\gamma}(fg\mathbb{1}_A).$$

Since the function fg is non-negative measurable, we have by the uniqueness of Radon-Nikodym derivatives that

$$h \overset{\text{a.e}}{=} fg$$

which completes the proof.

For *(ii),* note that $\mu+\nu$ is $\sigma-$finite by Proposition 1.3.10. Absolute continuity follows trivially. To show the almost sure equivalence of the Radon-Nikodym derivatives, let $f,g,h\in\mathcal{M}^{+}(\mathcal{X},\mathcal{F})$ be Radon-Nikodym derivatives such that for every $A\in\mathcal{F}$

$$\begin{aligned}\mu(A)+\nu(A)&=\bar{\gamma}(h\mathbb{1}_{A})\\ \mu(A)&=\bar{\gamma}(f\mathbb{1}_{A})\\ \nu(A)&=\bar{\gamma}(g\mathbb{1}_{A}).\end{aligned}$$

By linearity of the Lebesgue integral, we have that for every $A\in\mathcal{F}$

$$\bar{\gamma}((f+g)\mathbb{1}_{A})=\bar{\gamma}(h\mathbb{1}_{A}).$$

By Proposition 3.3.9 we have that

$$h\overset{\gamma-\text{a.e}}{=}f+g.$$

Finally, to show *(iii),* let f denote $\frac{d\mu}{d\nu}$ and suppose that that $f\overset{\nu-\text{a.e}}{>}0$ and that $\mu(A)=0$ for some $A\in\mathcal{F}$. The Radon-Nikodym theorem implies then that $\bar{\nu}(f\mathbb{1}_{A})=0$. Since f is positive almost everywhere, Proposition 3.3.3 implies that $\mathbb{1}_{A}\overset{\nu-\text{a.e}}{=}0\iff\nu(A)=0$. Conversely, suppose $\nu<<\mu$ and there's a set of positive (with respect to both measures) mass $B\in\mathcal{F}$ such that $f=0$ on B. Then $\mu(B)=\bar{\nu}(f\mathbb{1}_{B})=0$ which is a contradiction. To conclude, observe that when $\frac{d\mu}{d\nu}\overset{\text{a.e}}{>}0$, both measures are *mutually* absolutely continuous (or equivalent) and so $\frac{d\nu}{d\mu}$ exists and by part *(i)* above

$$1=\frac{d\mu}{d\mu}=\frac{d\mu}{d\nu}\frac{d\nu}{d\mu}.$$

Taking reciprocals yields the result. □

Absolute continuity is a special property between pairs of measures characterized by the notion that the null sets of a measure are a subset of the other; that is to say $\mu<<\nu\iff N_{\nu}\subseteq N_{\mu}$. Of course, if the two measures μ,ν are mututally absolutely continuous (or equivalent), then $N_{\mu}=N_{\nu}$. Dual to this notion of absolute continuity, we can define the concept of mutual singularity, which corresponds to the situation where N_{ν} and N_{μ} can together cover $\mathcal{X}$

Definition 6.2.7. Let $(\mathcal{X},\mathcal{F},\mu)$ be a measure space. The measure μ is said to *concentrate* on a set $A\in\mathcal{F}$ if $\mu(E)=\mu(A\cap E)$. Equivalenty, one can say that μ concentrates on A if $\mu(E)=0$ if and only if $E\cap A=\emptyset$.

Definition 6.2.8. Let $(\mathcal{X},\mathcal{F})$ be a measurable space. Two measures μ,ν on $\mathcal{F}$ are said to be *mutually singular* if there exists two disjoint sets $A,B\in\mathcal{F}$ such that μ concentrates on A and ν concentrates on B. We can then write $\mu\perp\nu$.

In the next result we summarize some basic properties of mutual singularity and its relationship with absolute continuity.

Proposition 6.2.9. *Let $(\mathcal{X},\mathcal{F})$ be a measurable space and let $\mu,\nu,$ and γ be three measures on $\mathcal{F}$. Then*

(i) *$\mu\perp\nu$ if and only if there exist two disjoint sets $A,B\in\mathcal{F}$ such that $A\cup B=\mathcal{X}$ and $\mu(A)=0,\nu(A)=0$.*
(ii) *If $\mu\perp\gamma$ and $\nu\perp\gamma$ then $\mu+\nu\perp\gamma$.*
(iii) *If $\mu<<\gamma$ and $\nu\perp\gamma$ then $\mu\perp\nu$.*
(iv) *If $\mu<<\nu$ and $\mu\perp\nu$ then $\mu=0$.*

Proof. For *(i),* observe that if $\mu\perp\nu$ then there exist disjoint sets $C,D\in\mathcal{F}$ such that μ concentrates on C and ν concentrates on D. By definition, we have that $\mu\left(C^{C}\right)=\mu\left(C\cap C^{C}\right)=0$ and $\nu\left(D^{C}\right)=\nu\left(D\cap D^{C}\right)=0$. Note that $\nu(C)=\nu(C\cap D)=0$ and so $A=C^{C}$and $B=D^{C}$ does the trick. Conversely, let A,B be as given in the hypothesis. Then $\mu(F)=\mu(F\cap A)+\mu(F\cap B)=\mu(F\cap B)$ and similarly $\nu(F)=\nu(F\cap A)$ for every $F\in\mathcal{F}$.

For *(ii),* let $A_1, B_1 \in \mathcal{F}$ be a partition of $\mathcal{X}$ such that $\mu(A_1) = 0$ and $\gamma(B_1) = 0$. Similarly, let $A_2, B_2 \in \mathcal{F}$ be a partition of $\mathcal{X}$ such that $\nu(A_2) = 0$ and $\gamma(B_2) = 0$. Let $C = A_1 \cap A_2$ and $D = (A_1 \cap A_2)^C = B_1^C \cup B_2^C$. Clearly, C and D form a partition of $\mathcal{X}$ such that $\mu + \nu(A_1 \cap A_2) = \mu(A_1 \cap A_2) + \nu(A_1 \cap A_2) = 0$ by subadditivity. Similarly, $\gamma(B_1 \cup B_2) \leq \gamma(B_1) + \gamma(B_2) = 0$.

Next, for *(iii)* let $A \in \mathcal{F}$ such that $\nu(A) = 0$ and $\gamma\left(A^C\right) = 0$. Absolute continuity implies that $\mu\left(A^C\right) = 0$. This shows $\mu \perp \nu$.

Finally, suppose that $A \in \mathcal{F}$ such that $\mu(A) = 0$ and $\nu\left(A^C\right) = 0$. By absolute continuity, $\mu(A) = 0$. Then, $\mu(\mathcal{X}) = \mu(A) + \mu\left(A^C\right) = 0$. $\square$

It turns out that absolute continuity and singularity represent a sort of *orthogonality* notion for measures. Considering that the spaces of measures is a subset of the dual space of L^1 functions, this does not correspond to the canonical notion of orthogonality. MAKE PRECISE USING TARCSAY 2014. Again, this result falls trivially out of Lemma 6.2.1.

THEOREM 6.2.10 (Lebesgue Decomposition). *Let $(\mathcal{X}, \mathcal{F})$ be a measurable space and let μ and ν be two $\sigma-$finite measures on $\mathcal{F}$. Then there exist two unique measures μ_1, μ_2 such that*

$$\mu = \mu_1 + \mu_2$$

and $\mu_1 << \nu$ and $\mu_2 \perp \nu$. Such a pair (μ_1, μ_2) is called a Lebesgue decomposition of μ with respect to ν.

PROOF. Note that existence follows from Lemma 6.2.1. To see this, write

$$\mu(A) = \bar{\nu}(k\mathbb{1}_A) + \mu(A \cap B)$$

where $B \in N_\nu$ and k is non-negative almost-everywhere with respect to ν. Then setting $\mu_1(A) := \bar{\nu}(k\mathbb{1}_A)$we have that $\mu_1 << \nu$. Similarly, setting $\mu_2(A) := \mu(A \cap B)$ which is a measure since it satisfies countable additivity and is null on the empty set. Then note that $\nu(B) = 0$ and $\mu_2\left(B^C\right) = 0$ which implies that $\mu_2 \perp \nu$. $\square$

Note that we postpone the proof of the uniqueness of the decomposition to the following subsection, since we shall need the concept of a signed measure.

6.2.1. Signed measures, Duality of L^p spaces, and the Riesz representation theorem. Tao (epsilon of room vol 1, section 1.3.2)?

6.3. Absolutely continuous functions

The absolute continutiy of measures is in some sense a generalization of the notion of absolute continuity of a real-valued function on $[a, b] \subset \mathbb{R}$.

DEFINITION 6.3.1. A function $f : [a, b] \to \mathbb{R}$ is said to be *absolutely continuous* if for every $\epsilon > 0$, there exists some $\delta > 0$ such that for any finite collection of disjoint open intervals $\{(a_i, b_i)\}_{i=1}^n \subset [a, b]$,

$$\sum_{i=1}^n (b_i - a_i) < \delta \implies \sum_{i=1}^n |f(b_i) - f(a_i)| < \epsilon.$$

It should be immediately clear that every absolutely continuous function is uniformly continuous (and hence continuous). The converse is not true, however: the Cantor function is a canonical example of a function that is uniformly continuous but not absolutely continuous. The Cantor function is continuous on $[0, 1]$ by Proposition 1.4.22 and hence uniformly continuous since $[0, 1]$ is compact (see TheoremD.2.11). One the other hand, the Cantor function "concentrates" on C, in that the image of the Cantor set under the Cantor function is all of $[0, 1]$, and so for any finite collection $\{(a_i b_i)\}_{i=1}^n$ of disjoint intervals that covers the entire Cantor set, the corresponding image sequence $\sum_{i=1}^n |f(b_i) - f(a_i)| \geq 1$. Of course, since the Cantor set has measure zero, it can be covered by countably many open intervals whose total length is arbitrarily small. Moreover, since the Cantor set is compact, we can extract a finite subcollection of such intervals (disjointly, without loss of generality) and so absolute continuity fails for any $0 < \epsilon < 1$.

The fact that the Cantor function concentrates on C should be telling; the measure that the Cantor function corresponds to is singular with respect to the Lebesgue measure. As we shall see, the duality of singular and absolutely continuous functions mimics the duality of singular and absolutely continuous measures with respect to the Lebesgue measure. We start with the following fact that links absolute continuity of Stieljes functions to the absolute continuity of the measures induced by such functions with respect to the Lebesgue measure.

PROPOSITION 6.3.2. *Let $F : [a,b] \to \mathbb{R}$ be a Stieljes function. Then the measure $\mu_F((x,y]) := F(y) - F(x)$ for $a \le x \le y \le b$ on $\mathscr{B}([a,b])$ is absolutely continuous with respect to the Lebesgue measure if and only if F is absolutely continuous as a function.*

PROOF. Note that both μ_F and λ are finite measures on $\mathscr{B}([a,b])$ and so if $\mu_F \ll \lambda$ then (by Proposition 6.2.3) for any $\epsilon > 0$ there exists some $\delta > 0$ such that for any $B \in \mathscr{B}([a,b]) : \lambda(B) < \delta \implies \mu_F(B) < \epsilon$. In particular, for $B = \bigcup_{i=1}^{n}(a_i, b_i)$ where $(a_i, b_i) \subset [a,b]$ are disjoint, we have that

$$\lambda\left(\bigcup_{i=1}^{n}(a_i,b_i)\right) > \delta \implies \mu_F\left(\bigcup_{i=1}^{n}(a_i,b_i)\right) < \epsilon$$

and applying finite additivity yields one side of the result. Conversely, assume F is absolutely continuous as a function and so for any $\epsilon > 0$ there exists some $\delta > 0$ such that for any finite collection of disjoint open intervals $\{(a_i,b_i)\}_{i=1}^{n} \subset [a,b]$

$$\sum_{i=1}^{n}(b_i - a_i) < \delta \implies \sum_{i=1}^{n} f(b_i) - f(a_i) < \epsilon.$$

Taking limits, we can extend this result to countable disjoint collections so that for any $\{(a_i,b_i)\}_{i\in\mathbb{N}}$ disjoint

$$\sum_{i=1}^{\infty}(b_i - a_i) < \delta \implies \sum_{i=1}^{\infty} f(b_i) - f(a_i) < \epsilon.$$

Note that by Lemma 1.2.14, each open set $O \subset [a,b]$ can be written as a countable union of disjoint open intervals and by Proposition 1.4.8, every Borel set $B \in \mathscr{B}([a,b])$ (except sets that contain the endpoints) is contained in some open set $O \subset \mathbb{R}$ such that

$$\lambda(O \setminus B) = \lambda(O) - \lambda(B) < \frac{\delta}{2}.$$

Choosing B such that $\lambda(B) < \frac{\delta}{2}$, we can assume without loss of generality that $O \subset [a,b]$ and so

$$\mu_F(B) \le \mu_F(O) < \epsilon$$

where the inequality is due to monotonicity of measures. This completes the proof by yet another application of Proposition 6.2.3. $\square$

COROLLARY 6.3.3. *A Stieljes function $F : [a,b] \to \mathbb{R}$ is absolutely continuous if and only if there exists some unique $f \in \mathcal{M}^+([a,b], \mathscr{B}([a,b]))$ such that for any $x \in [a,b]$*

$$F(x) = F(a) + \bar{\lambda}\left(f\mathbb{1}_{[a,x]}\right).$$

PROOF. First suppose that F is absolutely continuous as a function. Then by Proposition 6.3.2, the measure μ_F extended from $\mu_F((a,x]) := F(x) - F(a)$ is absolutely continuous with respect to the Lebesgue measure on $[a,b]$. By the Radon-Nikodym theorem, there exists some $f \in \mathcal{M}^+([a,b], \mathscr{B}([a,b]))$ such that

$$\mu_F(A) = \bar{\lambda}(f\mathbb{1}_A)$$

for any $A \in \mathscr{B}([a,b])$. In particular this works for $A = [a,x]$ where $x \in [a,b]$ is a arbitrary. Conversely, if F has this integral representation then μ_F and $\gamma(A) := \bar{\lambda}(f\mathbb{1}_A)$ agree on all sets of the form $[a,x]$ which means they agree on intersections of such sets which is the collection of all closed intervals in $[a,b]$. This collection is a $\pi-$system that generates $\mathscr{B}([a,b])$ and so by Theorem 2.4.7 $\mu_F = \gamma$ on all

Borel sets which implies absolute continuity of the measure μ_F and thus of the function F by the previous Proposition.

Uniqueness of f follows by the standard argument we used to prove that Radon Nikodym derivatives are unique. □

REMARK. Note that the function f here actually turns out to be the (almost everywhere) derivative of F, a fact that requires considerable effort to prove. We leave the proof of this result (and thus the fundamental theorems of calculus) to the chapter on product spaces, where we would be able to show some remarkable results related to the geometry of the Euclidean space $\mathbb{R}^n$, and use those results to complete the proof started here.

Next we extend this representation result to absolutely continuous functions that are not necessarily non-decreasing. As usual, we do this by decomposing an arbitrary absolutely continuous function as a difference of absolutely continuous Stieljes functions and apply the above result separately to each component.

LEMMA 6.3.4. *Let $f : [a,b] \to \mathbb{R}$ be absolutely continuous. Then there exist non-negative, non-decreasing, and absolutely continuous functions $f_1, f_2 : [a,b] \to \mathbb{R}$ such that*

$$f(x) - f(a) = f_1(x) - f_2(x)$$

for all $x \in [a,b]$.

PROOF. Define

$$f_1(x) := \sup_{\pi[a,x]} \sum_{i=1}^{k(\pi)} \left(f(t_i) - f(t_{i-1})\right)^+$$
$$f_2(x) := \sup_{\pi[a,x]} \sum_{i=1}^{k(\pi)} \left(f(t_i) - f(t_{i-1})\right)^-$$

where $\pi[a,x] : \left\{t_i \mid t_0 = a, t_{k(\pi)} = x, t_i < t_{i+1}\right\}$ is a partition of $[a,x]$. Note that the non-negativity of these functions follows by definition. Next, notice that for $x_2 \geq x_1$, $f_j(x_2) \geq f_j(x_1)$ for $j \in \{1,2\}$ .To see this, notice that any partition π of $[a,x_1]$ can be extended to a partition π' of $[a,x_2]$ by adding $t_{k(\pi')} = t_{k(\pi)+1} = x_2$. Then,

$$\sum_{i=1}^{k(\pi)} \left(f(t_i) - f(t_{i-1})\right)^{\pm} \leq \sum_{i=0}^{k(\pi)+1} \left(f(t_i) - f(t_{i-1})\right)^{\pm}$$

since the summands are non-negative. Next, observe that for any partition π of $[a,x] \subset [a,b]$, we have that

$$\begin{aligned} f(x) - f(a) &= \sum_{i=1}^{k(\pi)} \left(f(t_i) - f(t_{i-1})\right) \\ &= \sum_{i=1}^{k(\pi)} \left(f(t_i) - f(t_{i-1})\right)^+ - \sum_{i=1}^{k(\pi)} \left(f(t_i) - f(t_{i-1})\right)^- \end{aligned}$$

and so taking supremums we get the equality $f(x) - f(a) = f_1(x) - f_2(x)$. Next, we shall show absolute continuity for f_1; the proof for f_2 is identical. Fix $\epsilon > 0$ and note that by the absolute continuity of f we have that there exists some $\delta > 0$ such that for any disjoint collection $\{(a_i,b_i)\}_{i=1}^n \subset [a,b]$,

$\sum_{i=1}^{n} b_i - a_i < \delta$ implies $\sum_{i=1}^{n} |f(b_i) - f(a_i)| < \epsilon$. Note that

$$\begin{aligned}\sum_{i=1}^{n} |f_1(b_i) - f_1(a_i)| &= \sum_{i=1}^{n} |\sup_{\pi[a,b_i]} \sum_{j=1}^{k(\pi)} \left(f\left(t_j^{\pi}\right) - f\left(t_{j-1}^{\pi}\right)\right)^{+} - \sup_{\pi'[a,a_i]} \sum_{j=1}^{k(\pi')} \left(f\left(t_j^{\pi'}\right) - f\left(t_{j-1}^{\pi'}\right)\right)^{+}| \\ &= \sum_{i=1}^{n} \sup_{\pi[a_i,b_i]} \sum_{j=1}^{k(\pi)} \left(f\left(t_j^{\pi}\right) - f\left(t_{j-1}^{\pi}\right)\right)^{+} \\ &\leq \sum_{i=1}^{n} \sup_{\pi[a_i,b_i]} \sum_{j=1}^{k(\pi)} \left(f\left(t_j^{\pi}\right) - f\left(t_{j-1}^{\pi}\right)\right)^{+} + \left(f\left(t_j^{\pi}\right) - f\left(t_{j-1}^{\pi}\right)\right)^{-} \\ &= \sum_{i=1}^{n} \sup_{\pi[a_i,b_i]} \sum_{j=1}^{k(\pi)} |f\left(t_j^{\pi}\right) - f\left(t_{j-1}^{\pi}\right)| \\ &\leq \epsilon\end{aligned}$$

where in the second equality we have used the fact that we can enlarge any partition πon $[a, b_i]$ to contain any given partition π' on $[a, a_i]$ since $[a, a_i] \subset [a, b_i]$. The final inequality then follows since for any partition π_i of $[a_i, b_i]$, entire collection $\left\{\left(t_{j-1}^{\pi_i}, t_j^{\pi_i}\right)\right\}_{1 \leq j \leq k(\pi_i), 1 \leq i \leq n}$ is a disjoint collection with total length

$$\sum_{i=1}^{n} \sum_{j=1}^{k(\pi_i)} t_j^{\pi_i} - t_{j-1}^{\pi_i} = \sum_{i=1}^{n} b_i - a_i < \delta$$

and so by absolute continuity

$$\sum_{i=1}^{n} \sum_{j=1}^{k(\pi_i)} |f\left(t_j^{\pi}\right) - f\left(t_{j-1}^{\pi}\right)| \leq \epsilon.$$

Taking supremums over partitions preserves the inequality. □

THEOREM 6.3.5. *Let $f : [a, b] \to \mathbb{R}$ be a measurable function. Then f is absolutely continuous if and only if there exists an (almost-everywhere) unique function $h \in \mathcal{L}^1([a, b], \mathscr{B}[a, b], \lambda)$ such that for each $x \in [a, b]$,*

$$f(x) = f(a) + \lambda\left(h \mathbb{1}_{[a,x]}\right).$$

PROOF. First suppose that f is absolutely continuous and so by Lemma 6.3.4 we can write

$$f(x) - f(a) = f_1(x) - f_2(x)$$

where f_i are non-negative, non-decreasing, and absolutely continuous (and thus Stieljes functions). By Corollary 6.3.3 there exist functions $h_1, h_2 \in \mathcal{M}^{+}([a, b], \mathscr{B}([a, b]))$ such that

$$f_i(x) = f_i(a) + \lambda\left(h_i \mathbb{1}_{[a,x]}\right)$$

and so by the linearity of integration we have that

$$f(x) - f(a) = \underbrace{f_1(a) - f_2(a)}_{=0} + \lambda\left((h_1 - h_2) \mathbb{1}_{[a,x]}\right).$$

Note that $h := h_1 - h_2$ is integrable since $f(b) - f(a) < \infty$. Uniqueness is a generating class argument as usual. To spell it out, notice that for $h, g \in \mathcal{L}^1([a, b], \mathscr{B}([a, b]), \lambda)$, where the representation result holds we have that

$$\mathcal{D} := \left\{B \in \mathscr{B}([a, b]) \mid \bar{\lambda}(h \mathbb{1}_B) = \bar{\lambda}(g \mathbb{1}_B)\right\}$$

is a $\lambda-$system: $[a,b]$ is clearly in $\mathcal{D}$, and if $B \in \mathcal{D}$ then $B^C \in \mathcal{D}$ because $\mathbb{1}_{B^C} = \mathbb{1}_{[a,b]} - \mathbb{1}_B$, the linearity of integration, and the fact that $[a,b] \in \mathcal{D}$. Finally, if $\{B_i\}_{i\in\mathbb{N}} \in \mathcal{D}$ are disjoint then

$$\begin{aligned}
\bar{\lambda}\left(h\mathbb{1}_{\bigcup_{i\in\mathbb{N}} B_i}\right) &= \bar{\lambda}\left(\sum_{i=1}^{\infty} h\mathbb{1}_{B_i}\right) \\
&= \sum_{i=1}^{\infty} \bar{\lambda}\left(h\mathbb{1}_{B_i}\right) \\
&= \sum_{i=1}^{\infty} \bar{\lambda}\left(g\mathbb{1}_{B_i}\right) \\
&= \bar{\lambda}\left(\sum_{i=1}^{\infty} g\mathbb{1}_{B_i}\right) \\
&= \bar{\lambda}\left(g\mathbb{1}_{\bigcup_{i\in\mathbb{N}} B_i}\right)
\end{aligned}$$

where in the second and fourth equalities leverage dominated convergence. Note that $\mathcal{E}$ the collection of all closed interval subsets of $[a,b]$ generates $\mathscr{B}([a,b])$ and the equality of representations holds on $\mathcal{E}$ by definition. By the $\pi - \lambda$ theorem, the equality extends to the entirety of $\mathscr{B}([a,b])$. By Proposition 3.3.9 we have that $h \overset{\text{a.e}}{=} g$.

Finally, suppose that f has the integral representation

$$f(x) - f(a) = \bar{\lambda}\left(h\mathbb{1}_{[a,x]}\right)$$

and fix $\epsilon > 0$. Note that by the linearity of integration, for any $[a_i, b_i] \subseteq [a,b]$

$$f(b_i) - f(a_i) = \bar{\lambda}\left(h\mathbb{1}_{[a_i,b_i]}\right)$$

and so for any disjoint collection $\{[a_i,b_i]\}_{i=1}^n$

$$\begin{aligned}
\sum_{i=1}^{n} |f(b_i) - f(a_i)| &= \sum_{i=1}^{n} |\bar{\lambda}\left(h\mathbb{1}_{[a_i,b_i]}\right)| \\
&\leq \sum_{i=1}^{n} \bar{\lambda}\left(|h|\mathbb{1}_{[a_i,b_i]}\right)
\end{aligned}$$

by Corollary 3.2.3. Then, using Proposition 6.2.3, we have that there exists some $\delta > 0$ such that for $\lambda\left(\bigcup_{i=1}^n [a_i,b_i]\right) < \delta$ we have $\bar{\lambda}\left(|h|\mathbb{1}_{\bigcup_{i=1}^n [a_i,b_i]}\right) < \epsilon$ which completes the proof. $\square$

6.4. Optimization

EXAMPLE 6.4.1. Maximize $x + y$ subject to the condition that $2x^2 + 3y^2 \leq 1$.

CHAPTER 7

Measures on product spaces

In calculus, we learnt that the theory of integration readily extends from from real valued functions on $\mathbb{R}$ to real valued functions on the Euclidean space $\mathbb{R}^n$. The extension is usually motivated geometrically by studying the volume under the surface of a sufficiently smooth function $f : \mathbb{R}^2 \to \mathbb{R}$. The idea is that if a function is sufficiently well behaved, then one can recover the volume under the surface by looking at the areas under various "slices" of the function and then summing up those areas. Importantly, under the requisite smoothness conditions, the "slices" could have been made horizontally or vertically, and we would get the same result. This intuition leads to the Fubini theorem for multiple integration in the Riemann setting:

$$\int_a^b \int_c^d f(x,y)dydx = \int_c^d \int_a^b f(x,y)dxdy.$$

In our linear functional notation, this can be written as

$$\lambda^x \left(\lambda^y \left(f(x,y)\, \mathbb{1}_{[c,d]}\right) \mathbb{1}_{[a,b]}\right) = \lambda^y \left(\lambda^x \left(f(x,y)\, \mathbb{1}_{[a,b]}\right) \mathbb{1}_{[c,d]}\right).$$

Here λ is the Lebesgue measure as usual, and the superscripts denote the variable of integration. Note that we have dropped the bars on top of the λ to denote integration to clean up notation; for comfort, you can just think of integration as an extension of a measure from the space of measurable indicators to the space of non-negative measurable or integrable functions, in the spirit of Theorem 3.2.16.

The main goal of this chapter is to recover this result for the general measures. While important in its own right for the study of analysis on Euclidean spaces, in the context of probability theory, this result takes a far more important role. In particular, the ability to write a multiple integral as an iterated integral corresponds directly with the ability to factor the joint distribution of random variables into their marginal distributions; that is, it underpins the theory of *independent* random variables. Independence, and the departures from independence, constitute the central concepts of probabiliy theory.

7.1. Product measures on finite product spaces

7.1.1. Iterated integrals. Let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be measure spaces. Our interest is in defining measurable functions on the product space $\mathcal{X} \times \mathcal{Y}$. The principle hurdle that is immediately apparent here is that the product of the $\sigma-$algebras $\mathcal{F} \times \mathcal{G} := \{F \times G \mid F \in \mathcal{F}, G \in \mathcal{G}\}$ is not necessarily a $\sigma-$algebra. ADD CE. It is easy to see, however, that it is a $\pi-$system (this fact will turn out to be important!). Any $\sigma-$algebra we use should certainly *contain* $\mathcal{F} \times \mathcal{G}$ and so the canonical choice is given by $\mathcal{F} \otimes \mathcal{G} := \sigma(\mathcal{F} \times \mathcal{G})$. A series of natural questions follow if we want iterated integrals to make sense. In particular, for a measurable map $f \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$,we want to know if the projections $x \to f(x,y)$ and $y \to \mu^x(f(x,y))$ are $\mathcal{F}$ and $\mathcal{G}$ measurable, respectively. Thankfully, this turns out to be the case. We will use a long lost result from Chapter 2: the $\pi - \lambda$ theorem for functions. Go over the hypotheses of this theorem before reading the following results.

LEMMA 7.1.1. *For every $f \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$, the maps $x \to f(x,y)$ and $y \to f(x,y)$ are $\mathcal{F}/\mathscr{B}(\mathbb{R})$ and $\mathcal{G}/\mathscr{B}(\mathbb{R})$ measurable for every $y \in \mathcal{Y}$ and $x \in \mathcal{X}$, respectively.*

PROOF. First we note, due to the fact that $(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \times B_2)$,that $\mathcal{E} := \mathcal{F} \times \mathcal{G}$ is a $\pi-$system. Next, we claim that the space

$$\mathcal{H} := \{f \in \mathcal{M}_{\text{bdd}}(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G}) \mid \forall y \in \mathcal{Y} : x \to f(x,y) \text{ is } \mathcal{G}/\mathscr{B}(\mathbb{R}) \text{ measurable }\}$$

is a $\lambda-$space of functions. Note that $\mathbb{1}_{\mathcal{X}\times\mathcal{Y}}$is constant and bounded (and evidently partially measurable in our sense). Further, our space $\mathcal{H}$ is a vector space, since linear combinations of bounded measurable functions is bounded, and the partial measurability condition is also preserved under linear combinations (both results of Proposition 2.2.6). Finally, $\mathcal{H}$ is closed under monotone limits (if they exist) since measurability (resp. partial measurability) is preserved under limits.

Now, observe that $\{\mathbb{1}_A \mid A \in \mathcal{E}\} \subseteq \mathcal{H}$, since $\mathbb{1}_{F\times G} = \mathbb{1}_F \mathbb{1}_G$which is clearly bounded, measurable, and partially measurable in our sense. Thus applying the $\pi - \lambda$ theorem as discussed, we have that

$$\mathcal{M}_{\text{bdd}}\left(\mathcal{X}\times\mathcal{Y}, \mathcal{F}\otimes\mathcal{G}\right) \subseteq \mathcal{H}$$

which completes the proof for bounded functions.

Now we can simply take a function $f \in \mathcal{M}\left(\mathcal{X}\times\mathcal{Y}, \mathcal{F}\otimes\mathcal{G}\right)$ and construct the bounded monotone sequences $f_n^{\pm} := \min\left\{f^{\pm}, n\right\} \in \mathcal{H}$ by our results. We complete the proof by taking limits and noting that partial measurability is preserved. The same argument, of course, holds for $y \to f(x,y)$. □

Note that the converse of this result is not necessarily true; that is, for a function $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ with measurable sections $x \to f(x,y)$ and $y \to f(x,y)$, it need not be the case that f is product measurable.

Example 7.1.2. Let V be a Vitali set in $\mathbb{R}$ and let $E = \{(x,x) \mid x \in V\}$. Then let $f(x,y) = \mathbb{1}_E(x,y)$ and notice that

$$x \to f(x,y) = \begin{cases} 0, & y \notin V \\ \mathbb{1}\{x = y\} & y \in V. \end{cases}$$

In the first case, the section $x \to f(x,y)$ is constantly zero and hence measurable. In the second case, its an indicator function for the singleton $\{y\}$ which is measurable since singletons are in $\mathscr{B}(\mathbb{R})$ .Therefore, for every $y \in \mathbb{R}$ the section $x \to f(x,y)$ is $\mathscr{B}(\mathbb{R})/\mathscr{B}(\mathbb{R})$ measurable. A similar argument shows that for every $x \in \mathbb{R}$, $y \to f(x,y)$ is Borel measurable. But notice that f is measurable if and only if E is measurable, but the map $g(x) = (x,x)$is measurable and $g^{-1}[E] = V$ is not measurable and so E is not measurable.

Lemma 7.1.3. *Let $(\mathcal{X},\mathcal{F},\mu)$ and $(\mathcal{Y},\mathcal{G},\nu)$ be $\sigma-$finite measure spaces. For every function $f \in \mathcal{M}^+\left(\mathcal{X}\times\mathcal{Y},\mathcal{F}\otimes\mathcal{G}\right)$, the maps $x \to \nu^y(f(x,y))$ and $y \to \mu^x(f(x,y))$ are $\mathcal{F}/\mathscr{B}(\mathbb{R})$ and $\mathcal{G}/\mathscr{B}(\mathbb{R})$ measurable respectively.*

Proof. We will prove that $x \to \nu^y f(x,y)$ is measurable; the other case follows by symmetry. First, let $F_i \in \mathcal{F}$ be an increasing sequence of sets with finite measure such that $\bigcup_{i\in\mathbb{N}} F_i = \mathcal{X}$. Observe that for the indicator of any rectangle in $F \times G \in \mathcal{E}$,we have that $\nu^y(F\times G) = \mathbb{1}_F \nu(G)$ which is measurable. Next, define

$$\mathcal{D}_i := \left\{D \in \mathcal{F}\otimes\mathcal{G} \mid x \to \nu^y\left(\mathbb{1}_D \mathbb{1}_{F_i}\right) \text{ is } \mathcal{F}/\mathscr{B}(\mathbb{R}) \text{ measurable}\right\}$$

using Lemma 7.1.1 to ensure that the expression is well defined. Notice that $\mathcal{E} \subseteq \mathcal{D}$ (and so in particular, $\mathcal{X}\times\mathcal{Y} \in \mathcal{D}$). Now suppose $D_1, D_2 \in \mathcal{D}$ such that $D_1 \subseteq D_2$. Then, letting ν_i denoting the measure that has density $\mathbb{1}_{F_i}$ with respect to ν

$$\nu_i^y\left(\mathbb{1}_{D_2\setminus D_1}\right) = \nu_i^y\left(\mathbb{1}_{D_2}\right) - \nu_i^y\left(\mathbb{1}_{D_1}\right)$$

which is measurable by Proposition 2.2.6. Note that the expression on the right is well defined since ν_i is a finite measure and indicator functions are bounded (and so no $\infty - \infty$ situations arise). Similarly, suppose we have an increasing sequence of sets $D_n \in \mathcal{D}_i$, with $D := \bigcup_{n\in\mathbb{N}} D_n$, we have

$$\begin{aligned} \nu_i^y\left(\mathbb{1}_D\right) &= \nu_i^y\left(\sup_{n\in\mathbb{N}} \mathbb{1}_{D_n}\right) \\ &= \nu_i^y\left(\lim_{n\to\infty} \mathbb{1}_{D_n}\right) \\ &= \lim_{n\to\infty} \nu_i^y\left(\mathbb{1}_{D_n}\right) \end{aligned}$$

where we have used Proposition 2.1.3 in the first equality, the nested nature of D_n in the second equality, and the monotone convergence theorem in the final equality. By Corollary 2.2.10, the limit is measurable and so $D \in \mathcal{D}_i$. Thus we have shown that each $\mathcal{D}_i$ is a $\lambda-$system containing $\mathcal{E}$, a $\pi-$system, and so by the $\pi - \lambda$ theorem, $\mathcal{F} \otimes \mathcal{G} = \sigma(\mathcal{E}) \subseteq \mathcal{D}_i$ for all $i \in \mathbb{N}$. Finally, for any $D \in \mathcal{F} \otimes \mathcal{G}$, we know that $x \to \nu_i^y(\mathbb{1}_D)$ is measurable, and by yet another application of monotone convergence and Corollary 2.2.10, $x \to \nu^y(\mathbb{1}_D)$ is also measurable.

This completes the proof for indicator functions. We can then show that measurability holds for non-negative measurable simple functions by using the linearity of integration and the fact that measurability is preserved under linear combinations. Finally, we show the result for general non-negative measurable functions by approximating them by simple functions from below, applying monotone convergence, and using Corollary 2.2.10 yet again. □

Add remarks under each measurability lemma to extend to n dimensions via induction

With the technicalities out of the way, we can show that iterated integrals give the same result under certain conditions. This is a generalization of the result we saw all the way back in Lemma 1.2.16.

THEOREM 7.1.4 (Tonelli). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be measure spaces. Then, for any $f \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$, the functions*

$$\gamma_1(f) := \nu^y \mu^x(f)$$

and

$$\gamma_2(f) : \mu^x \nu^y(f)$$

are integrals on $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$. Moreover, if $\mu$and ν are $\sigma-$finite, $\gamma_1(f) = \gamma_2(f)$ for every $f \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$.

PROOF. We show that γ_1 is an integral; the argument for γ_2 is the analagous. First, observe that $\gamma_1(0) = \nu^y \mu^x(0) = \nu^y(0) = 0$, given that ν and μ are integrals. Second, note that for $\alpha, \beta \geq 0$, and $f, g \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$

$$\begin{aligned}
\gamma_1(\alpha f + \beta g) &= \nu^y \mu^x(\alpha f + \beta g) \\
&= \nu^y(\alpha \mu^x f + \beta \mu^x g) \\
&= \alpha \nu^y \mu^x(f) + \beta \nu^y \mu^x(g) \\
&= \alpha \gamma_1(f) + \beta \gamma_1(g).
\end{aligned}$$

Finally, observe that for $f_n \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$ such that $f_n \leq f_{n+1}$ and $f_n \to f$, we have

$$\begin{aligned}
\lim_{n\to\infty} \gamma_1(f_n) &= \lim_{n\to\infty} \nu^y \mu^x(f_n) \\
&= \nu^y \mu^x\left(\lim_{n\to\infty} f_n\right) \\
&= \gamma_1(f)
\end{aligned}$$

by applying monotone convergence twice.

Therefore, by Theorem 3.2.16, γ_1 and γ_2 are integrals on $\mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$ with respect to measures defined by integrating indicator functions in $\mathcal{F} \otimes \mathcal{G}$. If μ and ν are $\sigma-$finite, the two integrals can be shown to be equal by showing that the corresponding measures are equal on a generating $\pi-$system that can approximate the full space (courtesy of our uniqueness theorem). Of course, since $\mathcal{F} \times \mathcal{G}$ is a $\pi-$system, for $F \in \mathcal{F}, G \in \mathcal{G}$

$$\begin{aligned}
\gamma_1(\mathbb{1}_{F \times G}) &= \nu^y \mu^x(\mathbb{1}_{F \times G}(x, y)) \\
&= \nu^y \mu^x(\mathbb{1}_F(x) \mathbb{1}_G(y)) \\
&= \nu(G) \mu(F) \\
&= \mu^x \nu^y(\mathbb{1}_F(x) \mathbb{1}_G(y)) \\
&= \gamma_2(\mathbb{1}_{F \times G})
\end{aligned}$$

completing the proof. Since our measures μ, ν are $\sigma-$finite, we know that there exist sets $E_i \in \mathcal{F} \times \mathcal{G}$ such that $\bigcup E_i = \mathcal{X} \times \mathcal{Y}$. This completes the proof. $\square$

Note that the $\sigma-$finiteness condition is actually necessary for the uniqueness of the integrals, as Example 7.1.14 illustrates.

Corollary 7.1.5 (Fubini's Theorem). *Let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be $\sigma-$finite measure spaces and let $f \in \mathcal{M}(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$. If one of $\gamma_1(|f|) := \nu^y \mu^x(|f|)$ or $\gamma_2(|f|) : \mu^x \nu^y(|f|)$ is finite, then*

(i) $x \to f(x,y) \in \mathcal{L}^1(\mu)$
(ii) $y \to f(x,y) \in \mathcal{L}^1(\nu)$
(iii) $x \to \nu^y(f(x,y)) \in \mathcal{L}^1(\mu)$
(iv) $y \to \mu^x(f(x,y)) \in \mathcal{L}^1(\nu)$
(v) $\gamma_1(f) = \gamma_2(f)$

Proof. Note that since $|f| \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$, by Tonelli's theorem,

$$\gamma_1(|f|) = \gamma_2(|f|) < \infty$$

which implies, by Proposition 3.3.6, that

$$\nu^y(|f(x,y)|) < \infty$$

and

$$\mu^x(|f(x,y)|) < \infty$$

which gives us (i) and (ii). Next, note that

$$|\nu^y(f(x,y))| \leq \nu^y(|f(x,y)|) < \infty$$

and

$$|\mu^x(f(x,y))| \leq \mu^x(|f(x,y)|) < \infty$$

for all $y \in \mathcal{Y}$ by (i) , (ii) and Corollary 3.2.3. The measurabiltiy of $x \to |\nu^y(f(x,y))|$ and $y \to |\mu^x(f(x,y))|$ follows from the fact that $x \to \nu^y(f^{\pm})$ and $y \to \mu^x(f^{\pm})$ are measurable (and finite) by Lemma 7.1.3 and so $x \to \nu^x(f(x,y)) = x \to \nu^x(f^+) - \nu^x(f^-)$ and $y \to \mu^x(f(x,y)) = y \to \mu^x(f^+) - \mu^x(f^-)$ are measurable by Corollary 2.2.7. Since the absolute value is continuous, our measurability results hold. Then, monotonicity of integration implies that

$$\mu^x|\nu^y(f(x,y))| < \infty \text{ and } \nu^y|\mu^x(f(x,y))| < \infty$$

which gives is (iii) and (iv). Finally, to see (v), you note that $\gamma_1(f^+) = \gamma_2(f^+)$ and $\gamma_1(f^-) = \gamma_2(f^-)$ by Tonelli. Then the result follows by subtracting these two equalities and apply linearity. $\square$

7.1.2. Product sigma algebras and measures.

Definition 7.1.6. Let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be measure spaces and let $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$ be the product measurable space. Then a measure $\mu \otimes \nu : \mathcal{F} \otimes \mathcal{G} \to [0, \infty]$ is called a *product measure* if

$$\mu \otimes \nu(F \times G) = \mu(F)\nu(G)$$

for all $F \in \mathcal{F}, G \in \mathcal{G}$.

Of course, the way we have set this up, Theorem 7.1.4(Tonelli) guarantees existence of this measure, since measures and integrals are equivalent. Uniqueness, when μ and ν are $\sigma-$finite, follows from the uniqueness theorem, as outlined in the proof of Tonelli. Note also the fact that the product of sigma finite measures is sigma finite. Tonelli (or Fubini) also tell us that for non-negative measurable (or integrable) functions, the integral with respect to the product measure coincides with iterated integrals, at least when the measure spaces are $\sigma-$finite. That is to say

$$\mu \bar{\otimes} \nu(f) = \gamma_1(f) = \gamma_2(f)$$

for any $f \in \mathcal{M}^+(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$ $(f \in \mathcal{L}^1(\mu \otimes v))$.

To get some of the standard results on multiple integration in $\mathbb{R}^n$, we first need to do some legwork to characterize products of Borel $\sigma-$algebras. We begin first by generalizing the concept of a $\sigma-$algebra generated by a measurable function.

DEFINITION 7.1.7. Let $\mathcal{X}$ be a set and let $(\mathcal{Y}_i, \mathcal{G}_i)$ be a measurable spaces, for $i \in I$ where I is an index set. Further, let $\{f_i\}_{i\in I}$, be a collection of functions $f_i : \mathcal{X} \to \mathcal{Y}_i$. Then, the $\sigma-$algebra generated by $\{f_i\}_{i\in I}$ is

$$\sigma\left(\{f_i\}_{i\in I}\right) := \sigma\left(\bigcup_{i\in I} \sigma(f_i)\right)$$

where $\sigma(f_i)$ is the usual definition of the sigma algebra generated by a function.

REMARK. When the I is finite or countable, we simply write $\sigma(f_1, f_2, \ldots)$.

PROPOSITION 7.1.8. *Let $\mathcal{X}$ be a set and let $(\mathcal{Y}, \mathcal{G})$ be a measurable space. For any collection $\{f_i\}_{i\in I}$ where I is just some index set, $\sigma\left(\{f_i\}_{i\in I}\right)$ is the smallest $\sigma-$algebra that makes every f_i measurable.*

PROOF. Note that if $\mathcal{F}$ is some $\sigma-$algebra that makes f_i measurable for all $i \in I$, then $\sigma(f_i) \subseteq \mathcal{F}$ which implies that $\bigcup_{i\in I} \sigma(f_i) \subseteq \mathcal{F}$ and the result follows. $\square$

REMARK. We can use the definition above to provide an alternate characterization of product $\sigma-$algebras. First note for measurable spaces $(\mathcal{X}_1, \mathcal{F}_1)$ and $(\mathcal{X}_2, \mathcal{F}_2)$ with the product space $(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$, we have that the coordinate projections π_1 and π_2 defined as

$$\pi_i : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}_i$$

where $\pi_i(x_1, x_2) = x_i$ for any $i = 1, 2$.

PROPOSITION 7.1.9. *Let $(\mathcal{X}_1, \mathcal{F}_1), (\mathcal{X}_2, \mathcal{F}_2)$, and $(\mathcal{Y}, \mathcal{G})$ be measurable spaces. Then*

(i) $\sigma(\pi_1, \pi_2) = \mathcal{F}_1 \otimes \mathcal{F}_2$

(ii) $T : (\mathcal{Y}, \mathcal{G}) \to (\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ *is measurable if and only if the maps* $\pi_i \circ T : (\mathcal{Y}, \mathcal{G}) \to (\mathcal{X}_i, \mathcal{F}_i)$ *are measurable for* $i = 1, 2$.

(iii) If $S : (\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{F}_1 \otimes \mathcal{F}_2) \to (\mathcal{Y}, \mathcal{G})$ *is measurable then the maps* $x_1 \to S(x_1, x_2)$ *and* $x_2 \to S(x_1, x_2)$ *are measurable for all* $x_2 \in \mathcal{X}_2$ *and* $x_1 \in \mathcal{X}_1$*, respectively.*

PROOF. For (i), note that for any $F \in \mathcal{F}_1$, $\pi_1^{-1}[F] = F \times \mathcal{X}_2 \in \mathcal{F}_1 \otimes \mathcal{F}_2$ and so π_1 is measurable with respect to $\mathcal{F}_1 \otimes \mathcal{F}_2$ and so by definition $\sigma(\pi_1) \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2$. Similarly, $\sigma(\pi_2) \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2$ and so $\sigma(\pi_1, \pi_2) \subseteq \mathcal{F}_1 \otimes \mathcal{F}_2$. Conversely, for any $F_1 \in \mathcal{F}_1$ and $F_2 \in \mathcal{F}_2$,

$$F_1 \times F_2 = (F_1 \times \mathcal{X}_2) \cap (\mathcal{X}_1 \times F_2) \in \sigma(\pi_1, \pi_2)$$

and so $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\{F_1 \times F_2 \mid F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2\}) \subseteq \sigma(\pi_1, \pi_2)$.

Next, for (ii), first note that if T is measurable, then $\pi_i \circ T$ is measurable since $\pi_i \in \mathcal{M}(\mathcal{X}_1 \times \mathcal{X}_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ by part (i) and the composition of measurable functions is measurable. Conversely, if both π_i and $\pi \circ T$ are measurable, then for any $F_1 \in \mathcal{F}_1$ and $F_2 \in \mathcal{F}_2$

$$\begin{aligned} T^{-1}[F_1 \times F_2] &= T^{-1}[(F_1 \times \mathcal{X}_2) \cap (\mathcal{X}_1 \times F_2)] \\ &= T^{-1}\left[\pi_1^{-1}[F_1] \cap \pi_2^{-1}[F_2]\right] \\ &= (\pi_1 \circ T)^{-1}[F_1] \cap (\pi_2 \circ T)^{-1}[F_2] \end{aligned}$$

where both terms in the intersection are in $\mathcal{G}$ by the measurability of $\pi_i \circ T$. Then a standard generating class argument (Theorem 2.2.3) yields the result.

Finally, note that $x_1 \to S(x_1, x_2)$ for some fixed x_2 can be thought of as the composition $S \circ i_{x_2}$ where $i_{x_2} : \mathcal{X}_1 \to \mathcal{X}_1 \times \mathcal{X}_2$ is given by $i_{x_2}(x_1) = (x_1, x_2)$. S is measurable by assumption and $i_{x_2}^{-1}[F_1 \times F_2] = F_1$ and so i_{x_2} is measurable by the standard generating class argument. A composition of measurable maps is measurable and the result follows. The argument for $x_2 \to S(x_1, x_2)$ is analagous. $\square$

This result allows us to show that our definition of the $\sigma-$algebra generated by multiple functions is quite natural.

COROLLARY 7.1.10. *Let $\mathcal{X}$ be a set, let $(\mathcal{Y}_i, \mathcal{G}_i)$ be a measurable spaces, for $i \in \{1,2\}$, and let $(\mathcal{Y}_1 \times \mathcal{Y}_2, \mathcal{G}_1 \otimes \mathcal{G}_2)$ be the product measurable space. For any functions $f_1 : \mathcal{X} \to \mathcal{Y}_1$ and $f_2 : \mathcal{X} \to \mathcal{Y}_2$, along with a function $h(x) = (f_1(x), f_2(x))$, we have that*

$$\sigma(h) = \sigma(f_1, f_2).$$

PROOF. Note by Proposition 7.1.9, h is measurable with respect to $\sigma(f_1, f_2)$ since $f_1 = \pi_1 \circ h$ and $f_2 = \pi_2 \circ h$ are measurable with respect to $\sigma(f_1, f_2)$ and so $\sigma(h) \subseteq \sigma(f_1, f_2)$. Conversely, again by Proposition 7.1.9, both f_1 and f_2 are measurable with respect to $\sigma(h)$ since h is measurable with respect to $\sigma(h)$. Thus $\sigma(f_1, f_2) \subseteq \sigma(h)$. □

PROPOSITION 7.1.11. *Let $(\mathcal{X}_i, \mathcal{F}_i)$ $i = 1, 2$ be measurable spaces where there exist some collection of sets $\mathcal{E}_i \subset \mathcal{F}_i$ such that $\mathcal{X}_i \in \mathcal{E}_i$ and*

$$\sigma(\mathcal{E}_i) = \mathcal{F}_i.$$

Then, letting $\mathcal{E} := \{E_1 \times E_2 \mid E_1 \in \mathcal{E}_1, E_2\}$, we have that

$$\sigma(\mathcal{E}) = \mathcal{F}_1 \otimes \mathcal{F}_2.$$

PROOF. Let $\mathcal{F} := \{F_1 \times F_2 \mid F_1 \in \mathcal{F}_1, F_2 \in \mathcal{F}_2\}$ .Obviously, $\mathcal{E} \subseteq \mathcal{F}$ and so $\sigma(\mathcal{E}) \subseteq \sigma(\mathcal{F}) = \mathcal{F}_1 \otimes \mathcal{F}_2$. To see the reverse inclusion, take the projections $\pi_i : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathcal{X}_i$. We wish to show these functions are $\sigma(\mathcal{E}) / \mathcal{F}_i$ measurable. To that end, note that for any $E \in \mathcal{E}_1$, $\pi_1^{-1}[E] = E \times \mathcal{X}_2 \in \mathcal{E} \subseteq \sigma(\mathcal{E})$ and so a generating class argument tells us that π_1 is measurable as needed. An analagous argument establishes the measurability of π_2 and so

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\pi_1, \pi_2) \subseteq \sigma(\mathcal{E})$$

where the equality is due to Proposition 7.1.9. □

EXAMPLE 7.1.12. In the proposition above, the requirement that $\mathcal{X}_i \in \mathcal{E}_i$ is necessary. Let $\mathcal{X}_1 = \mathcal{X}_2 = \{1, 2, 3\}$. Suppose then that $\mathcal{E}_1 = \{\{1\}\}$and $\mathcal{E}_2 = \{\{2\}\}$ and so

$$\mathcal{F}_1 := \sigma(\mathcal{E}_1) = \{\{1,2,3\}, \emptyset, \{1\}, \{2,3\}\}$$
$$\mathcal{F}_2 := \sigma(\mathcal{E}_2) = \{\{1,2,3\}, \emptyset, \{2\}, \{1,3\}\}.$$

Meanwhile, $\mathcal{E} = \{\{(1,2)\}\}$ and so

$$\sigma(\mathcal{E}) = \left\{\{1,2,3\} \times \{1,2,3\}, \emptyset, \{(1,2)\}, \{(1,2)\}^C\right\}$$

which doesnt contain $\{1\} \times \{1,2,3\} \in \mathcal{F}_1 \otimes \mathcal{F}_2$.

PROPOSITION 7.1.13. *Let $(\mathcal{X}_1, \tau_1)$ and $(\mathcal{X}_2, \tau_2)$ be topological spaces with Borel sigma algebras $\mathscr{B}(\mathcal{X}_1) = \sigma(\tau_1)$ and $\mathscr{B}(\mathcal{X}_2) = \sigma(\tau_2)$. Then, letting $\tau_1 \otimes \tau_2$ denote the box topology*[1]*i.e. the topology generated by $\{O_1 \times O_2 \mid O_1 \in \tau_1, O_2 \in \tau_2\}$, we have that*

$$\mathscr{B}(\mathcal{X}_1) \otimes \mathscr{B}(\mathcal{X}_2) \subseteq \mathscr{B}(\mathcal{X}_1 \times \mathcal{X}_2) := \sigma(\tau_1 \otimes \tau_2).$$

Moreover, in the case of Euclidean spaces

$$\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}) = \mathscr{B}\left(\mathbb{R}^2\right).$$

PROOF. Observe that since $\mathcal{X}_i \in \tau_i$, $i = 1, 2$, $\sigma(\{O_1 \times O_2 \mid O_1 \in \tau_1, O_2 \in \tau_2\}) = \mathscr{B}(\mathcal{X}_1) \otimes \mathscr{B}(\mathcal{X}_2)$ by Proposition 7.1.11. Note that the box topology $\tau_1 \otimes \tau_2 = \tau(\{O_1 \times O_2 \mid O_1 \in \tau_1, O_2 \in \tau_2\})$; that is, it is the *topology generated by* $\{O_1 \times O_2 \mid O_1 \in \tau_1, O_2 \in \tau_2\}$ (exactly analagous to a $\sigma-$algebra generated by a collection of sets). Since $\{O_1 \times O_2 \mid O_1 \in \tau_1, O_2 \in \tau_2\} \subseteq \tau(\{O_1 \times O_2 \mid O_1 \in \tau_1, O_2 \in \tau_2\}) = \tau_1 \otimes \tau_2$,

$$\mathscr{B}(\mathcal{X}_1) \otimes \mathscr{B}(\mathcal{X}_2) = \sigma(\{O_1 \times O_2 \mid O_1 \in \tau_1, O_2 \in \tau_2\}) \subseteq \sigma(\tau_1 \otimes \tau_2) = \mathscr{B}(\mathcal{X}_1 \times \mathcal{X}_2)$$

which is the first claim.

The second claim makes use of the separability of $\mathbb{R}$. We extend the argument from Lemma 1.2.14 to $\mathbb{R}^2$ to show that any open set in $\mathbb{R}^2$ can be written as a countable union of open rectangles in $\mathbb{R}^2$.

[1] which is equivalent to the product topology for finite products

Take any open set $U \subseteq \mathbb{R}^2$. By definition, for any norm $\|\cdot\|$ on $\mathbb{R}^2$, for any $x := (x_1, x_2) \in U$ there exists some $r > 0$ such that

$$B_{\|\cdot\|}(x, r) \subset U.$$

Since norms on $\mathbb{R}^2$ are all equivalent, let $\|\cdot\| = \|\cdot\|_\infty$ and notice that $y \in B_{\|\cdot\|_\infty}(x, r)$ if and ony if $y \in (x_1 - r, x_1 + r) \times (x_2 - r, x_2 + r)$. Now note that since the rationals are dense in $\mathbb{R}$, we can find rational endpoints $a_{i,x}, b_i,$ such that $(a_{i,x}, , b_{i,x}) \subseteq (x_i - r, x_i + r)$ and $x_i \in (a_{i,x}, b_{i,x})$. Then,

$$U = \bigcup_{x \in U} (a_{1,x}, , b_{1,x}) \times (a_{2,x}, , b_{2,x}) .$$

Since the endpoints are rational, there can only be countably many distinct such rectangles and so our claim follows. Now note that open rectangles are in $\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})$ by definition, and so every open set in $\mathbb{R}^2$ is a countable union of sets in $\mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})$. Let the collection of open sets in $\mathbb{R}^2$ be denoted $\mathcal{O}$. Then

$$\mathcal{O} \subseteq \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}) \implies \mathscr{B}\left(\mathbb{R}^2\right) \subseteq \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R})$$

which completes the proof. □

We can use the above result to construct classic counterexamples to Tonelli and Fubini when one of their hypotheses is not satsfied.

EXAMPLE 7.1.14. Let $([0,1], \mathscr{B}([0,1]), \lambda)$ and $\left([0,1], 2^{[0,1]}, \mu_0\right)$ where λ and μ_0 are the Lebesgue and counting measures, respectively. Now consider the product space $\left([0,1]^2, \mathscr{B}([0,1]) \otimes 2^{[0,1]}, \lambda \otimes \mu_0\right)$ and observe that by a variant of the argument in Proposition 7.1.13, the diagonal set $D = \left\{(x, y) \in [0,1]^2 \mid x = y\right\} \in \mathscr{B}\left([0,1]^2\right) = \mathscr{B}([0,1]) \otimes \mathscr{B}([0,1]) \subseteq \mathscr{B}([0,1]) \otimes 2^{[0,1]}$ since it is closed. Then

$$\begin{aligned}
\lambda^x \mu_0^y (\mathbb{1}_D) &= \lambda^x (1 \{x \in [0,1]\}) \\
&= 1 \\
&\neq 0 \\
&= \mu_0^y (0 \mathbb{1} \{y \in [0,1]\}) \\
&= \mu_0^y \lambda^x (\mathbb{1}_D) .
\end{aligned}$$

Note that μ_0 is not $\sigma-$finite.

7.1.3. Radon-Nikodym derivatives.

PROPOSITION 7.1.15. *Let $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ be measurable spaces with product space $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$. Let μ_1, μ_2 be $\sigma-$finite measures on $\mathcal{F}$ such that $\mu_1 << \mu_2$. Similarly, let ν_1, ν_2 be measures on $\mathcal{G}$ such that $\nu_1 << \nu_2$. Then $\mu_1 \otimes \nu_1$ and $\mu_2 \otimes \nu_2$ are $\sigma-$finite, $\mu_1 \otimes \mu_2 << \nu_1 \otimes \nu_2$ and*

$$\frac{d(\mu_1 \otimes \nu_1)}{d(\mu_2 \otimes \nu_2)} \overset{a.e}{=} \frac{d\mu_1}{d\mu_2} \times \frac{d\nu_1}{d\nu_2}.$$

REMARK. Note that the almost everywhere equality here holds with respect to both product measures.

PROOF. Note that the $\sigma-$finiteness of the products follows from the fact that if $\{F_i\}_{i \in \mathbb{N}} \in \mathcal{F}$ such that $\bigcup_{i \in \mathbb{N}} F_i = \mathcal{X}$ and $\mu_1(F_i) < \infty$ (and similarly for $\{G_j\}_{j \in \mathbb{N}} \in \mathcal{G}$) then $\bigcup_{i \in \mathbb{N}} \bigcup_{j \in \mathbb{N}} F_i \times G_j = \mathcal{X} \times \mathcal{Y}$and $\mu_1(F_i) \nu_1(G_j) < \infty$. The same argument works for the other pair of measures. Then, notice that for any $F \in \mathcal{F} \otimes \mathcal{G}$,

$$\begin{aligned}
\mu_1 \otimes \nu_1 (F) &= \mu_1^x \nu_1^y (\mathbb{1}_F (x,y)) \\
&= \mu_1^x \nu_2^y \left(\mathbb{1}_F (x,y) \frac{d\nu_1}{d\nu_2} \right) \\
&= \mu_2^x \left(\nu_2^y \left(\mathbb{1}_F (x) \frac{d\nu_1}{d\nu_2} \right) \frac{d\mu_1}{d\mu_2} \right) \\
&= \mu_2^x \nu_2^y \left(\mathbb{1}_F (x) \frac{d\nu_1}{d\nu_2} \frac{d\mu_1}{d\mu_2} \right)
\end{aligned}$$

where the first equality is Tonelli (and the uniquenes of product measures), the second is due the fact that $y \to \mathbb{1}_F (x,y)$ is a non-negative measurable function for every $x \in \mathcal{X}$ (by Lemma 7.1.1) and so Corollary 6.2.5 applies. The third equality is similar: now we have the function $x \to \nu_2^y \left(\mathbb{1}_F (x,y) \frac{d\nu_1}{d\nu_2} \right)$ which is non-negative since the the integrand is non-negative function for every $x \in \mathcal{X}$. It's also measurable by Lemma 7.1.3 and so Corollary 6.2.5 again applies. Finally, we apply linearity (since $\frac{d\mu_1}{d\mu_2}$ is constant in y) in the last line. This completes the proof. □

This result has a converse, in that for any $\sigma-$finite measure ν on $\mathcal{F} \otimes \mathcal{G}$ which is absolutely continuous with respect to a $\sigma-$finite product measure and whose Radon Nikodym derivative can be written as a product is itself a product measure. We make this precise with the following result.

Proposition 7.1.16. *Let $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ be measurable spaces with product $(\mathcal{X} \times \mathcal{Y}, \mathcal{F} \otimes \mathcal{G})$. Let ν_1 and μ_1 be $\sigma-$finite measures on $\mathcal{F}$ such that $\nu_1 << \mu_1$. Similarly, let ν_2 and μ_2 be $\sigma-$finite measures on $\mathcal{G}$ such that $\nu_2 << \mu_2$. Finally, let γ be a measure on $\mathcal{F} \otimes \mathcal{G}$ such that $\gamma << \mu_1 \otimes \mu_2$ and*

$$\frac{d\gamma}{d\mu_1 \otimes \mu_2} \overset{a.e}{=} \frac{d\nu_1}{d\mu_1} \times \frac{d\nu_2}{d\mu_2}.$$

Then $\gamma = \nu_1 \otimes \nu_2$.

Proof. Note that for any $A \in \mathcal{F} \otimes \mathcal{G}$

$$\begin{aligned}
\gamma (A) &= \mu_1^x \mu_2^y \left(\mathbb{1}_A (x,y) \frac{d\nu_1}{d\mu_1} (x) \frac{d\nu_2}{d\mu_2} (y) \right) \\
&= \mu_1^x \nu_2^y \left(\mathbb{1}_A (x,y) \frac{d\nu_1}{d\mu_1} (x) \right) \\
&= \nu_2^y \mu_1^x \left(\mathbb{1}_A (x,y) \frac{d\nu_1}{d\mu_1} (x) \right) \\
&= \nu_2^y \nu_1^x (\mathbb{1}_A (x,y)) \\
&= \nu_2 \otimes \nu_1 (A)
\end{aligned}$$

where the second equality follows by Corollary6.2.5, the third equality by Tonelli, the fourth again by Corollary 6.2.5. □

Remark. As we shall see, this results in this section are a measure theoretic justification for the probabilistic fact that two *absolutely continuous* random variables X and Y are *independent* if and only if we can factor their probability density functions.

7.1.4. Integration by parts. Tonelli (or Fubini) gives us a generalization of integration by parts from elementary calculus. To see this, let F and G be Stieljes (that is, real valued, nondecreasing, and right continuous) functions on a compact set $[a,b]$. By Theorem 1.4.3, we have that the set functions

$$\begin{aligned}
\mu ((x,y]) &:= F(y) - F(x) \\
\nu ((x,y]) &:= G(y) - G(x)
\end{aligned}$$

for any $a \leq x \leq y \leq b$ extend to measures on $\mathscr{B}([a,b])$. We can show that under mild regularity conditions,

$$\bar{\mu}(G) + \bar{\nu}(F) = F(b)G(b) - F(a)G(a).$$

This implies the "usual" integration by parts formula since complete after finishing chapter on differentiation.

THEOREM 7.1.17 (Integration by Parts). *Let F and G be Stieljes functions on a compact set $[a,b]$. If the set of discontinuities of F and G are disjoint, then*

$$\bar{\mu}(G) + \bar{\nu}(F) = F(b)G(b) - F(a)G(a).$$

PROOF. Note that

$$\begin{aligned}\bar{\mu}(G) &= \mu^x(G(x)) \\ &= \mu^x(\nu((a,x]) + G(a)) \\ &= \mu^x\nu^y(\mathbb{1}\{a < y \leq x \leq b\}) + G(a)(F(b) - F(a))\end{aligned}$$

where the third equality uses linearity and the fact that $\mu((a,b]) = F(b) - F(a)$. Similarly, we can write

$$\begin{aligned}\bar{\nu}(F) &= \nu^y(F(y)) \\ &= \nu^y(\mu((a,y]) + F(a)) \\ &= \nu^y\mu^x(\mathbb{1}\{a < x \leq y \leq b\}) + F(a)(G(b) - G(a)).\end{aligned}$$

Now note that

$$\begin{aligned}\bar{\mu}(G) + \bar{\nu}(F) &= \mu^x\nu^y(\mathbb{1}\{a < y \leq x \leq b\}) + \nu^y\mu^x(\mathbb{1}\{a < x \leq y \leq b\}) + G(a)(F(b) - F(a)) + F(a)(G(b) - G(a)) \\ &= \mu^x\nu^y(\mathbb{1}\{a < y \leq x \leq b\} + \mathbb{1}\{a < x \leq y \leq b\}) + G(a)(F(b) - F(a)) + F(a)(G(b) - G(a))\end{aligned} \tag{20}$$

where the second equality follows by Tonelli and linearity. We can expand upon the term inside the integral by noticing that $\{a < y \leq x \leq b\} = \underbrace{\{a < y = x \leq b\}}_{=:A} \cup \underbrace{\{a < y < x \leq b\}}_{=:B}$ where the union is disjoint and so by Fact 2.1.2, we know that

$$\mathbb{1}\{a < y \leq x \leq b\} = \mathbb{1}_A + \mathbb{1}_B.$$

Further, $\{a < x \leq y \leq b\} = (a,b]^2 \setminus B$. Then

$$\begin{aligned}\mu^x\nu^y(\mathbb{1}\{a < y \leq x \leq b\} + \mathbb{1}\{a < x \leq y \leq b\}) &= \mu^x\nu^y\left(\mathbb{1}_A + \mathbb{1}_B + \mathbb{1}_{(a,b]^2\setminus B}\right) \\ &= \mu^x\nu^y(\mathbb{1}_A) + \mu\otimes\nu\left((a,b]^2\right) \\ &= \mu^x\nu(\{x\}) + (F(a) - F(b))(G(a) - G(b)) \\ &= (F(a) - F(b))(G(a) - G(b))\end{aligned} \tag{21}$$

where the third equality is due to the fact that $\mu\otimes\nu$ is a product measure, and the last equality follows by our assumption that F and G contain no common points of discontinuity and Proposition 1.4.4explain this better. Then, substituting (21) into (20) and doing some algebra yields the result. $\square$

An interesting application of this result is showing that the integral of a Stieljes function F under the measure induced by F is simply $\frac{F(b)^2 - F(a)^2}{2}$, which follows simply by applying the above theorem where $F = G$. In probability theory, this is a special case of the fact that a cumulative distribution function of a random variable is itself uniformly distributed on $[0,1]$. We shall establish this fact in Part II of these notes.

7.1.5. Area under the graph of a function. In elementary calculus, we understood the ordinary (Riemann) integral of a bounded function f on a compact interval $[a,b]$ as representing the area of under the graph of f on $[a,b]$. The Riemann theory formalized this by partitioning the interval $[a,b]$ into subintervals and computing the integral of step functions which were constant on each subinterval in the partition. We can do this more generally with Lebesgue integrals as well. That is to say, we can show that for a non-negative measurable function $f : \mathbb{R} \to \mathbb{R}$

$$\bar{\lambda}(f) = \lambda^2 \left(\left\{(x,t) \in \mathbb{R}^2 \mid 0 \le t < f(x)\right\}\right)$$

where λ^2 is the 2-dimensional Lebesgue measure $\lambda \otimes \lambda$ on $\mathscr{B}\left(\mathbb{R}^2\right)$. In fact, we can establish this result more generally for other $\sigma-$finite measures using Tonelli's theorem.

THEOREM 7.1.18. *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a $\sigma-$finite measure space and let $f \in \mathcal{M}^+(\mathcal{X}, \mathcal{F})$. Then $\{(x,t) \in \mathcal{X} \times \mathbb{R} \mid f(x) > t \ge 0\} \in \mathcal{F} \otimes \mathscr{B}(\mathbb{R})$ and*

$$\bar{\mu}(f) = \lambda^t \left(\mu^x \left(\{f(x) > t\}\right) \mathbb{1}\{t \ge 0\}\right) = \lambda \otimes \mu \left(\{\{f(x) > t \ge 0\}\}\right)$$

PROOF. First notice that function $T(x,t) = f(x) - t$ is $\mathcal{F} \otimes \mathscr{B}(\mathbb{R}) / \mathscr{B}(\mathbb{R})$ measurable. To see this, define $g(x,t) = f(x)$ and notice that for any $B \in \mathscr{B}(\mathbb{R})$:$g^{-1}[B] = f^{-1}[B] \times \mathbb{R} \in \mathcal{F} \otimes \mathscr{B}(\mathbb{R})$. Similarly, letting $h(x,t) = t$, we have that $h^{-1}[B] = \mathbb{R} \times B \in \mathcal{F} \otimes \mathscr{B}(\mathbb{R})$. Then, using the fact that linear combinations of Borel measurable functions are Borel measurable, the function T is $\mathcal{F} \otimes \mathscr{B}(\mathbb{R}) / \mathscr{B}(\mathbb{R})$ measurable. Notice then that $T^{-1}[(0,\infty)] = \left\{(x,t) \in \mathbb{R}^2 \mid f(x) > t\right\}$ is measurable. Further, the set $\left\{(x,t) \in \mathbb{R}^2 \mid t \ge 0\right\} = \mathcal{X} \times [0,\infty) \in \mathcal{F} \otimes \mathscr{B}(\mathbb{R})$. Then,

$$\{(x,t) \in \mathcal{X} \times \mathbb{R} \mid f(x) > t \ge 0\} = T^{-1}[(0,\infty)] \cap (\mathcal{X} \times [0,\infty)) \in \mathcal{F} \otimes \mathscr{B}(\mathbb{R})$$

and the first result follows. Then,

$$\begin{aligned}
\bar{\mu}(f) &= \mu^x \left(\lambda^t \left(\mathbb{1}\{0 \le t < f(x)\}\right)\right) \\
&= \lambda^t \left(\mathbb{1}\{t \ge 0\} \mu^x \left(\mathbb{1}\{f(x) > t\}\right)\right) \\
&= \lambda^t \left(\mu^x \left(\{f(x) > t\}\right) \mathbb{1}\{t > 0\}\right)
\end{aligned}$$

where in the second equality we used Tonelli (7.1.4), linearity, and the fact that $\mathbb{1}_{A \cap B} = \mathbb{1}_A \mathbb{1}_B$. Observe that Tonelli can be applied here since $\mathbb{1}\{0 < t < f(x)\} \in \mathcal{M}^+(\mathcal{X} \times \mathbb{R}, \mathcal{F} \otimes \mathscr{B}(\mathbb{R}))$ and both μ and λ are $\sigma-$finite. Note that Tonelli's theorem also implies that

$$\lambda^t \left(\mu^x \left(\mathbb{1}\{f(x) > t \ge 0\}\right)\right) = \lambda \otimes \mu \left(\{\{f(x) > t \ge 0\}\}\right).$$

$\square$

7.2. The Lebesgue measure on $\mathbb{R}^n$

Formalize extension of Fubini etc to n dimension

It should be easy to guess that by induction, we can extend the construction of product measures to products of $n\sigma-$finite measure spaces. That is, for spaces $\{(\mathcal{X}_i, \mathcal{F}_i, \mu_i)\}_{i=1}^n$, we can define the product space $\left(\prod_{i=1}^n \mathcal{X}_i, \otimes_i \mathcal{F}_i\right)$ and show that for sigme finite measures μ_i on $\mathcal{F}_i$, there exists a unique product measure $\otimes_{i=1}^n \mu_i$ such that for $\prod F_i$, $F_i \in \mathcal{F}_i$

$$\otimes_{i=1}^n \mu_i \left(\prod_{i=1}^n F_i\right) = \prod_{i=1}^n \mu_i(F_i).$$

This requires us to use the canonical identification between $\prod_{i=1}^n \mathcal{X}_i$ and $\left(\prod_{i=1}^{n-1} \mathcal{X}_i\right) \times \mathcal{X}_n$ and the fact that the product of $\sigma-$finite measures is $\sigma-$finite. We can also similarly extend the Fubini-Tonelli theorems, so that for any permutation $\sigma : \{1,2,\ldots,n\} \to \{1,2,\ldots,n\}$ and $f \in \mathcal{M}^+\left(\prod_{i=1}^n \mathcal{X}_i, \otimes_i \mathcal{F}_i\right)$ (or $f \in \mathcal{L}^1\left(\otimes_{i=1}^n \mu_i\right)$)

$$\otimes_{i=1}^n \mu_i(f) = \mu_{\sigma(1)}\left(\mu_{\sigma(2)} \ldots \left(\mu_{\sigma(n)}(f)\right)\right).$$

Using the above facts, we can easily construct the Lebesgue measure λ^n on $\mathscr{B}(\mathbb{R}^n) = \otimes_{i=1}^n \mathscr{B}(\mathbb{R})$ as the product measure $\otimes_{i=1}^n \lambda$, where uniqueness is guaranteed by the $\sigma-$finiteness of λ. Equivalently, we can construct the Lebesgue measure using the Caratheodory approach, by defining $\mathcal{L}^n = \{\prod_{i=1}^n (a_i, b_i] \mid (a_i, b_i] \in \mathcal{L}\}$, and restricting the canonical outer measure

$$\lambda_*^n(A) = \inf \left\{ \sum_{j=1}^{\infty} \lambda_1^n(L_j) \mid A \subseteq \bigcup_{j \in \mathbb{N}} L_j, L_j \in \mathcal{L}^n \right\}$$

to the Lebesgue $\sigma-$algebra $\mathcal{C}(\lambda_*^n)$. This works because the collection $\mathcal{L}^n$ can be shown to be a semi-ring and $\lambda_1^n : \mathcal{L}^n \to [0, \infty]$ (here $\lambda_1^n (\prod_{i=1}^n L_i) = \prod_{i=1}^n \lambda_1(L_i)$where $L_i \in \mathcal{L}$) can be shown to satisfy the requirements of Theorem 1.2.26.

7.2.1. Basic properties of the Lebesgue measure. The first property of the Lebesgue measure that we establish is a generalization of Lemma 1.4.6 to $n-$dimensions.

PROPOSITION 7.2.1. *For $\delta \neq 0$ and any $A \in \mathscr{B}(\mathbb{R}^n)$ we have $\delta B \in \mathscr{B}(\mathbb{R}^n)$. Further, for any $A \subseteq \mathbb{R}^n$*

$$\lambda_*^n(\delta A) = |\delta|^n \lambda_*^n(A).$$

PROOF. Measurability follows by a simple generating class argument as in the Lemma (or alternatively, let $f(x) = \frac{x}{\delta}$ and notice that $f^{-1}$is a continuous function and $f^{-1}[B] = \delta B \in \mathscr{B}(\mathbb{R})$). Next, notice that for $n = 1$ this result is result is Lemma 1.4.6. Suppose the result holds for $n-1$. Let $B \in \mathscr{B}(\mathbb{R}^{n-1})$ and $A \in \mathscr{B}(\mathbb{R})$, then

$$\begin{aligned}
\lambda^n(\delta(B \times A)) &= \lambda^n((\delta B) \times (\delta A)) \\
&= \lambda^{n-1}(\delta B)\lambda(\delta A) \\
&= |\delta^n| \lambda^{n-1}(B)\lambda(A) \\
&= |\delta^n| \lambda^n(B \times A).
\end{aligned}$$

By the identification between $\mathbb{R}^{n-1} \times \mathbb{R}$ and $\mathbb{R}^n$, our scaling result holds on $\mathcal{L}^n$. Extending the proof to any arbitrary set $A \subseteq \mathbb{R}^n$proceeds exactly as in Lemma 1.4.6. $\square$

COROLLARY 7.2.2. *Let $\delta \neq 0$ be arbitrary. For any $f \in \mathcal{M}^+(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ (or $f \in \mathcal{L}^1(\lambda^n)$)*

$$\bar{\lambda^n}(f(\delta x)) = \frac{1}{|\delta^n|} \bar{\lambda^n}(f(x)).$$

PROOF. Let $g(x) = \delta x$ and notice that by Proposition 7.2.1, the image measure $\lambda^n g$ agrees with the measure $\frac{1}{|\delta^n|}\lambda^n$ on all Borel sets. By the change of variables formula in Corollary 3.2.14, $\bar{\lambda^n}(f(\delta x)) = \bar{\lambda^n g}(f) = \frac{1}{|\delta^n|}\bar{\lambda^n}(f).$ $\square$

We also have translation invariance, which is the analogue of Lemma 1.4.5.

PROPOSITION 7.2.3. *Let $A \subseteq \mathbb{R}^n$ be an arbitrary set. Then the translated set $A+t := \{a + t \mid a \in A\}$ has Lebesgue outer measure*

$$\lambda_*^n(A + t) = \lambda_*^n(A)$$

for all $t \in \mathbb{R}$. Moreover, for any set $A \in \mathscr{B}(\mathbb{R}^n)$,$A + t \in \mathscr{B}(\mathbb{R}^n)$.

PROOF. Again, the measurability result is a generating class argument (or let $f(x) = x - t...$). The result holds for $n = 1$ by the Lemma. Suppose it holds for $n-1$. Letting $B \in \mathscr{B}(\mathbb{R}^{n-1})$ and $A \in \mathscr{B}(\mathbb{R})$, we have

$$\begin{aligned}
\lambda^n((B \times A) + t) &= \lambda^n((B + t) \times (A + t)) \\
&= \lambda^{n-1}(B + t)\lambda(A + t) \\
&= \lambda^{n-1}(B)\lambda(A) \\
&= \lambda^n(B \times A).
\end{aligned}$$

Extending this to the general case follows exactly as in the Lemma. $\square$

COROLLARY 7.2.4. *Let $t \in \mathbb{R}$ be arbitrary. For any $f \in \mathcal{M}^+(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ (or $f \in \mathcal{L}^1(\lambda^n)$)*

$$\bar{\lambda^n}(f(x+t)) = \bar{\lambda^n}(f(x)).$$

PROOF. Let $g(x) = x + t$ and notice that by Proposition 7.2.3, the image measure $\lambda^n g$ and λ^n agree on all Borel sets. But recall that by the general change of variables formula in Corollary3.2.14, $\bar{\lambda^n}(f(x+t)) = \bar{\lambda^n} g(f) = \bar{\lambda^n}(f)$. $\square$

These two results are rather banal because they don't offer much additional insight over and above the corresponding results for the usual Lebesgue measure on the line. In order to derive some interesting properties of λ^n, we need to exploit the rich geometric structure of the Euclidean space and the functions that act on it. In order to do so, we need to review some basic linear algebra. Appendix B.1 on determinants is key here; please re-read this section carefully before we embark on the next result.

LEMMA 7.2.5. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an elementary map and let $f \in \mathcal{M}^+(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ (or $f \in \mathcal{L}^1(\lambda^n)$). Then $f \circ T \in \mathcal{M}^+(\mathbb{R}, \mathcal{F})$ ($f \circ T \in \mathcal{L}^1(\lambda^n)$ and*

$$|\det(T)|\bar{\lambda^n}(f \circ T) = \bar{\lambda^n}(f).$$

PROOF. The measurability is trivial and so first, suppose T is the row scaling operator, and note that if $T(x_1, \ldots, x_k, \ldots, x_n) = (x_1, \ldots, cx_k, \ldots x_n)$,with $c \neq 0$,then $\det(T) = c$ and so by Tonelli (Fubini)

$$\begin{aligned}
\lambda^n(f \circ T) &= \lambda^{x_{-k}}\lambda^{x_k}(f(x_1, \ldots, cx_k, \ldots, x_n)) \\
&= \frac{1}{|c|}\lambda^{x_{-k}}\lambda^{x_k}(f(x_1, \ldots, x_k, \ldots x_n)) \\
&= \frac{1}{|c|}\lambda^n(f)
\end{aligned}$$

where in the second line we used Corollary 7.2.2 for $n = 1$. Next, suppose that T is the row switching operator which as $\det(T) = -1$ and clearly Tonelli-Fubini imply that the integral is invariant under any permutation of indices and so the result follows. Finally, suppose that T is the row replacement operator i.e $T(x_1, \ldots, x_l, \ldots x_k, \ldots, x_n) = (x_1, \ldots, x_l + cx_k, \ldots x_k, \ldots, x_n)$. Again, by Tonelli (Fubini)

$$\begin{aligned}
\lambda^n(f \circ T) &= \lambda^{x_{-l}}\lambda^{x_l}(f(x_1, \ldots, x_l + cx_k, \ldots, x_n)) \\
&= \lambda^{x_{-l}}\lambda^{x_l}(f(x_1, \ldots, x_l, \ldots x_k, \ldots, x_n)) \\
&= \lambda^n(f)
\end{aligned}$$

which establishes the result since $\det(T) = 1$. $\square$

LEMMA 7.2.6. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertbe linear map and let $f \in \mathcal{M}^+(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ (or $f \in \mathcal{L}^1(\lambda^n)$). Then $f \circ T \in \mathcal{M}^+(\mathbb{R}, \mathcal{F})$ ($f \circ T \in \mathcal{L}^1(\lambda^n)$ and*

$$|\det(T)|\lambda^n(f \circ T) = \lambda^n(f).$$

PROOF. The measurability is again trivial. Since the matrices of arbitrary invertible maps are the product of elementary matrices, all we need to show is that for two elementary matrices E_1, E_2 , the product matrix $E = E_1E_2$ satisfies the conclusion of the theorem. The general claim then follows by induction. Clearly,

$$\begin{aligned}
|\det(E)|\lambda^n(f \circ E) &= |\det(E_1)||\det(E_2)|\lambda^n((f \circ E_1) \circ E_2) \\
&= |\det(E_1)|\lambda^n(f \circ E_1) \\
&= \lambda^n(f)
\end{aligned}$$

where the first equality uses the associativity of composition and Proposition B.1.4, the second equality uses Lemma 7.2.5 applied to $T = E_2$ and $f = f \circ E_1$. The last equality is yet another application of the previous lemma. $\square$

Note that a corollary of this result is the fact that the Lebesgue measure is rotationally invariant.

COROLLARY 7.2.7. *Let $R : \mathbb{R}^n \to \mathbb{R}^n$ be a rotation. Then for any Borel set $B \in \mathscr{B}(\mathbb{R}^n)$*

$$\lambda^n(R(B)) = \lambda^n(B).$$

PROOF. Note that rotations are orthogonal linear maps in that $RR^T = I$ and so

$$\det(R) = \frac{1}{\det(R^T)}.$$

But recall from Proposition B.1.8 that $\det(R) = \det\left(R^T\right)$ and so

$$|\det(T)| = 1.$$

We then apply 7.2.6 with $f = \mathbb{1}_B$ and $T = R$. □

7.2.2. The Lebesgue differentiation theorem and the fundamental theorem of calculus. Establishing the change of variables theorem in full requires a difficult measure theoretic result due to Lebesgue called the Lebesgue Differentiation Theorem. It turns out that this result is also key to establishing the fundamental theorem of calculus, and so we shall kill two birds with one stone and prove the fundamental theorem of calculus and the change of variables theorem.

The Lebesgue Differentiation Theorem is built on top of lemmata that are important and interesting in their own right. We start with a *covering lemma* which is a geometric argument that shows that a cover of a set E that consists of bounded balls can be reduced to an enlargement of a countable subcollection of disjoint balls. Throughout, we shall be working in $\mathbb{R}^n$ where the standard notation $B_r(c)$ denotes a ball of radius r centered at c under the standard Euclidean norm. We shall use superscripts to denote indices; that is to say, a ball B^a – where A is an index set – is short hand for $B_r^a(c)$ where r is the radius and c is the center. We will use the notation $r(B^a)$ to denote the radius of the ball B^a and $c(B^a)$ to denote its center.

LEMMA 7.2.8. *Let $E \subset \mathbb{R}^n$ be a subset of the Euclidean space such that there exists a (possibly uncountable) collection $\{B^a\}_{a\in A}$ of balls of radius at most R that covers E. Then there exists a disjoint countable*[2] *subcollection $\{B^{a_i}\}_{i\in\mathbb{N}}$ of balls such that*

$$\lambda^n(E) \leq 5^n \sum_{i=1}^{\infty} \lambda^n(B^{a_i}).$$

PROOF. We can construct our subcollection inductively as follows: first, let B^{a_1} be any ball in $\{B_a\}_{a\in A}$ such that its radius $r(B^{a_1}) \geq \frac{1}{2}\sup_{a\in A} r(B^a)$. This is possible because radii of the balls are bounded above by R. Then, we can define B^{a_k} , $a_k \in A$ such that

$$r(B^{a_k}) \geq \frac{1}{2}\sup\left\{r(B^a) \mid B^a \cap B^{a_i} = \emptyset \text{ for } 1 \leq i \leq k-1\right\}.$$

This process can either terminate or continue indefinitely; this is

Next, note that if $\sum_{k=1}^{\infty} \lambda^n(B^{a_k}) = \infty$ then we are done. In the case that this sum is finite, we have that $\lim_{k\to\infty} r(B^{a_k}) = 0$ and so for any $a^* \in A$ there exists some smallest $k^* \in \mathbb{N}$ such that for all $k \geq k^*$

$$r(B^{a_k}) < \frac{1}{2} r\left(B^{a^*}\right).$$

But by the construction of our sequence B^{a_k}, this means that there must exist some $i < k^*$ such that $B^{a_i} \cap B^{a^*} \neq \emptyset$. Let $x \in B^{a^*}$ be arbitrary, and pick some $y \in B^{a_i} \cap B^{a^*}$. By our results so far, we have

[2] or finite

that

$$\begin{aligned}\|x - c(B^{a_i})\| &\leq \|x-y\| + \|y - c(B^{a_i})\| \\ &\leq 2r\left(B^{a^*}\right) + r(B^{a_i}) \\ &\leq 4r(B^{a_i}) + r(B^{a_i}) \\ &= 5r(B^{a_i})\end{aligned}$$

where we have used the fact that for any $x, y \in B_r(c)$, $\|x-y\| \leq 2r$ in the first inequality, and the fact that $r(B^{a_i}) \geq \frac{1}{2} r\left(B^{a^*}\right)$ in the second inequality. Since x was arbitrary, we have concluded that $B^{a^*} \subseteq 5B^{a_i}$. Since $a^* \in A$ was itself arbitrary, we have that

$$E \subseteq \bigcup_{a \in A} B^a \subseteq \bigcup_{k=1}^{\infty} 5B^{a_k}.$$

Then, by Proposition 7.2.1, the monotonicity of measures, and countable subadditivity, our result follows. □

Remark. The bound is not sharp and the constant 5 is not necessarily optimal. Instead of choosing $r(B^{a_i}) \geq \frac{1}{2}\sup_{a \in A} r(B^a)$, we could have chosen $r(B_i^a) \geq c \sup_{a \in A} r(B^a)$ where $c \in (0,1)$. In this case, we get the inequality that

$$\lambda^n(E) \leq \left(\frac{2+c}{c}\right)^n \sum_{i=1}^{\infty} \lambda^n(B^{a_i}).$$

Since $c \to \frac{c+2}{c}$ is decreasing on $(0,1)$, it doesnt achieve a minimum, leaving us unable to optimize bound this way.

This geometric result appears to be a non-sequitur in our development of the fundamental theorem of calculus but it turns out to be indispensable to a key intermediate result called the *Hardy-Littlewood weak inequality.* Before we get to it, we need a special case of a result we shall later establish in Chapter Measure and Topology (?)

Definition 7.2.9. A function $f \in \mathcal{M}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ is called *locally integrable* if for any compact $K \subset \mathbb{R}^n$

$$\lambda^n(|f|\mathbb{1}_K) < \infty.$$

For any locally integrable function f, its *maximal function* Mf is defined

$$Mf(x) := \sup_{r>0} \frac{1}{\lambda^n(B_r(x))} \lambda^n\left(|f|\mathbb{1}_{B_r(x)}\right).$$

REVIEW

Lemma 7.2.10. *Let $B_r(x) \subset \mathbb{R}^n$ be an open ball. Then $\lambda^n(\partial B_r(x)) = 0$ where ∂E denotes the boundary of set E.*

Proposition 7.2.11. *Let $f \in \mathcal{M}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$ be locally integrable. Then $Mf : \mathbb{R}^n \to \mathbb{R}$ is Borel-measurable.*

Proof. First, for a fixed $r > 0$, observe that $x \to \lambda^n\left(|f|\mathbb{1}_{B_r(x)}\right)$ is continuous. To see this, let $\{x_n\}_{n \in \mathbb{N}} \in \mathbb{R}^n$ be a sequence converging to x. Eventually, for large enough n (say $n \geq n^*$), the sequence is in $B_r(x)$ and so for arbitrary $y \in B_r(x_n)$, $n \geq n^*$, we have that

$$\begin{aligned}\|y - x\| &\leq \|y - x_n\| + \|x_n - x\| \\ &\leq 2r\end{aligned}$$

and so for large enough n, $B_r(x_n) \subseteq B_{2r}(x)$. Writing $g_n := |f|\mathbb{1}_{B_r(x_n)}$, we claim that $g_n \to g := |f|\mathbb{1}_{B_r(x)}$ almost everywhere. Indeed, notice that for any $y \in \mathbb{R}^n$, if $\|y-x\| > r$ then we can find some $\epsilon > 0$ such that $\|y-x\| = r - \epsilon$. Since $x_n \to x$, for large enough n, we have that $\|x - x_n\| < \frac{\epsilon}{2}$ and so

$$\begin{aligned}\|y-x_n\| &\leq \|y-x\| + \|x-x_n \\ &\leq r - \epsilon + \frac{\epsilon}{2} \\ &< r.\end{aligned}$$

Similarly, if $\|y-x\| > r$ we know there's some $\epsilon > 0$ such that $\|y-x\| = r + \epsilon$ and so with the same type of argument, we have for large n

$$\begin{aligned}\|y-x_n\| &\geq \|y-x\| - \|x_n - x\| \\ &\geq r + \epsilon - \frac{\epsilon}{2} \\ &> \epsilon.\end{aligned}$$

Thus $\{y \in \mathbb{R}^n \mid \lim g_n(y) \neq g(y)\} \subseteq \{y \in \mathbb{R}^n \mid \|y-x\| = r\}$ where this latter set has measure zero since .... Since $\overline{B_{2r}(x)}$ is compact, local integrability implies that the sequence of functions g_n is dominated by the integrable function $h := |f|\mathbb{1}_{\overline{B_{2r}(x)}}$ and so by dominated convergence

$$\lambda^n(g_n) \to \lambda^n(g).$$

and so the continuity result follows.

By a similar argument, $\frac{1}{\lambda^n(B_r(x))}$ is continuous and thus so is their product. In other words, for any $a \in \mathbb{R}$ and $r > 0$, $\{x \in \mathbb{R}^n \mid \frac{1}{\lambda^n(B_r(x))}\lambda^n\left(|f|\mathbb{1}_{B_r(x)}\right) > a\}$ is open. Note that for any $x \in \mathbb{R}^n$ and fixed $a \in \mathbb{R}$.

$$Mf(x) > a \iff \exists r > 0 \text{ s.t } \frac{1}{\lambda^n(B_r(x))}\lambda^n\left(|f|\mathbb{1}_{B_r(x)}\right) > a$$

and so

$$\{x \in \mathbb{R}^n \mid Mf(x) > a\} = \bigcup_{r>0}\left\{x \in \mathbb{R}^n \mid \frac{1}{\lambda^n(B_r(x))}\lambda^n\left(|f|\mathbb{1}_{B_r(x)}\right) > a\right\}$$

which is open since arbitrary unions of open sets are open. Then, applying the standard generating class argument (Lemma 2.2.8) yields the result. □

REVIEW

THEOREM 7.2.12 (Hardy-Littlewood). *If $f \in \mathcal{L}^1(\lambda^n)$ then*

$$\lambda^n(\{\ x \in \mathbb{R}^n \mid Mf(x) > a\}) \leq \frac{5^n}{a}\lambda^n(|f|)$$

for any $a > 0$.

PROOF. Fix $a > 0$ and consider, any $x_0 \in \{Mf(x) > a\}$. By the definition of the maximal function, there exists some $r_0 > 0$ such that

$$\frac{1}{\lambda^n(B_{r_0}(x_0))}\lambda^n\left(|f|\mathbb{1}_{B_{r_0}(x_0)}\right) > a$$

which is equivalent to saying

$$\begin{aligned}\lambda^n(B_{r_0}(x_0)) &< \frac{\lambda^n\left(|f|\mathbb{1}_{B_{r_0}(x_0)}\right)}{a} \\ &\leq \frac{\bar{\lambda^n}(|f|)}{a}.\end{aligned}$$

Therefore $\{x \in \mathbb{R}^n \mid Mf(x) > a\}$ can be covered by balls which are bounded above in measure by $R := \frac{\bar{\lambda^n}(|f|)}{a}$. We can therefore apply Lemma 7.2.8 and extract a countable subcollection of such balls $\{B_i\}_{i\in\mathbb{N}}$ which are pairwise disjoint such that

$$\begin{aligned}\lambda^n\left(\{x \in \mathbb{R}^n \mid Mf(x) > a\}\right) &\leq 5^n \sum_{i=1}^{n} \lambda^n\left(B_i\right)\\ &\leq \frac{5^n}{a}\sum_{i=1}^{n} \lambda^n\left(|f|\mathbb{1}_{B_i}\right)\\ &= \frac{5^n}{a}\lambda^n\left(|f|\right)\end{aligned}$$

where the equality at the end is via monotone convergence. □

Example 7.2.13. Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded continuous function. Define $g : [0, \infty) \to \mathbb{R}$ by,

$$g(x) = \int_{-x}^{x} (2xt + 1)f(t)dt.$$

Show that g is differentiable on $(0, \infty)$ and find the derivative of g.TODO

7.2.3. Non-linear change-of-variables.

7.2.4. Integration with polar coordinates.

7.2.5. Symmetry of partial derivatives.

7.2.6. Convolutions. When measure spaces have a vector space structure such as $\mathbb{R}^n$, we can define some interesting image measures which play an important role in probability, fourier analysis, compressed sensing, signal processing etc.

Definition 7.2.14. Let μ, ν be $\sigma-$finite Borel measures on $\mathbb{R}^n$ (i.e. measures on $\mathscr{B}(\mathbb{R}^n)$). The *convolution* $\mu \star \nu$ is is the image measure of the product $\mu \otimes \nu$ under the map $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ given by $T(x, y) = x + y$.

Since the product measure always exists and is unique, the convolution always exists since T is a continuous (and hence measurable) map. Therefore for any bounded, continuous $f \in \mathcal{M}(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, we have that

$$\begin{aligned}\mu \bar{\star} \nu (f) &= \mu \bar{\otimes} \nu (f \circ T)\\ &= \mu^x \nu^y (f(x + y))\end{aligned}$$

where we have used Corollary 3.2.14 and Fubini's theorem. The definition implies that the convolution is symmetric in that $\mu \star v = v \star \mu$. It is also possible to define the convolution of a measure with a measurable function or of two measurable functions on their own, as is illustrated in the following result.

Proposition 7.2.15. *Let μ, ν be $\sigma-$finite measures on $\mathscr{B}(\mathbb{R})$ such that μ is absolutely continuous with respect to the Lebesgue measure with Radon-Nikodym derivative $g \in \mathcal{M}^+(\mathbb{R}, \mathscr{B}(\mathbb{R}))$. Then, for any bounded, continuous $f \in \mathcal{M}(\mathbb{R}, \mathscr{B}(\mathbb{R}))$*

$$\mu \bar{\star} \nu (f) = \lambda^x \nu^y (f(x) g(x - y))$$

and so $\mu \star \nu$ is absolutely continuous with respect to the Lebesgue measure with Radon-Nikodym derivative

$$g \star \nu := \nu^y (g(x - y)).$$

PROOF. Note that

$$\begin{aligned}\mu \bar{\star} \nu (f) &= \mu \otimes \nu (f (x+y)) \\ &= \nu^y \mu^x (f (x+y)) \\ &= \nu^y \lambda^x (f (x+y) g (x)) \\ &= \nu^y \lambda^x (f (x) g (x-y)) \\ &= \lambda^x (f (x) \nu^y (g (x-y)))\end{aligned}$$

where we have used the translation invariance of the Lebesgue measure in the fourth equality. □

The Radon-Nikodym derivative $g \star \nu$ is called the convolution of the function g with the measure ν. If ν too was absolutely continuous and had some density h then $\mu \star \nu$ would have density $g \star h := \lambda^y (g (x-y) h (y))$. Note that this definition can be expanded in the usual way to define convolutions of functions that are not necessarily non-negative (as Radon-Nikodym derivatives of positive measures are). We characterize the symmetry and other properties of convolutions in this next result

PROPOSITION 7.2.16. *Let μ, ν be $\sigma-$finite measures on $\mathscr{B}(\mathbb{R}^n)$ and let $f, g, h : \mathbb{R}^n \to \mathbb{R}$ be measurable functions. Then*

(i) *The convolution of two functions is symmetric in that*

$$f \star g = g \star f$$

.

(ii) *The convolution of two measures (or of a measure and a function, or of two functions) is bilinear in that*

$$(\alpha f + \beta g) \star h = \alpha f \star h + \beta h \star h$$

with a similar result for the second argument.

PROOF. For (i), note that $f \star g = \lambda_n^y (f (x-y) g (y)) = \lambda^y (f (y) g (x-y)) = g \star f$ by using $T (y) = x - y$ TODO AFTER CHANGE OF VARIABLES □

PROPOSITION 7.2.17.

7.2.7. Laplace's method.

THEOREM 7.2.18 (Laplace's method).

7.3. Kernels and disintegration

So far we have looked at constructing measures on product spaces by taking individual measures and combining them in some way to create a measure on the product space. Often we are interested in the reverse process; that is, we are interested in taking a measure on a product space and *disintegrating* it into measures on the individual spaces. It turns out that this process is more complicated than the process we have seen for constructing product measures. As discussed earlier, in the context of probability theory, the process of constructing product measures corresponds to the process of constructing independent random variables. In contrast, the process of disintegrating a measure on a product space into component measures corresponds to the process of computing conditional distributions. If you have some familiarity with probability, you would know that conditioning is more complicated than independence l

The central mathematical objects that underpin the theory of disintegration are kernels. Kernels are important as they wear two hats: on one hand, they act as measures, on the other they act as measurable functions.

Definition 7.3.1. Let $(\mathcal{X},\mathcal{F})$ and $(\mathcal{Y},\mathcal{G})$ be measure spaces A kernel κ from $(\mathcal{X},\mathcal{F})$ to $(\mathcal{Y},\mathcal{G})$ is a function $\kappa:\mathcal{G}\times\mathcal{X}\to[0,\infty]$ such that for any fixed $G_0\in\mathcal{G}$, the map

$$x\to\kappa\left(G_0,x\right)$$

is $\mathcal{F}/\mathscr{B}\left(\mathbb{R}\right)$ measurable and for every fixed $x_0\in\mathcal{X}$, the map

$$G\to\kappa\left(G,x_0\right)$$

is a measure on $\mathcal{G}$.

Remark. It is sometimes more intuitive to think of a kernel from $(\mathcal{X},\mathcal{F})$ to $(\mathcal{Y},\mathcal{G})$ as a collection of measures on $\mathcal{G}$ indexed by $\mathcal{X}$. So we can write

$$K:=\{\kappa_x:x\in\mathcal{X}\}$$

where each κ_x is a measure on $\mathcal{G}$ and $x\to\kappa_x\left(G\right)$ is measurable for each $G\in\mathcal{G}$. When these measures are restricted to the range $[0,1]$ the kernel is called a *Markov kernel.*

A generalization of the Tonelli theorem allows us to work with kernels.

Theorem 7.3.2. *Let $K=\{\kappa_x:x\in\mathcal{X}\}$ be a kernel from $(\mathcal{X},\mathcal{F})$ and $(\mathcal{Y},\mathcal{G})$ such that κ_x is $\sigma-$finite for every $x\in\mathcal{X}$. Let μbe a $\sigma-$finite measure on $\mathcal{F}$. Then, for $f\in\mathcal{M}^+\left(\mathcal{X}\times\mathcal{Y},\mathcal{F}\otimes\mathcal{G}\right)$*

(i) $y\to f\left(x,y\right)$ *is* $\mathcal{G}/\mathscr{B}\left(\mathbb{R}\right)$ *measurable*
(ii) $x\to\kappa_x\left(f\left(x,y\right)\right)$ *is* $\mathcal{F}/\mathscr{B}\left(\mathbb{R}\right)$ *measurable*
(iii) $\left(\mu\otimes K\right):=\mu^x\left(\kappa_x^y\left(f\left(x,y\right)\right)\right)$ *is an integral on* $\mathcal{F}\otimes\mathcal{G}$.

Proof. Note that (i) is simply Lemma 7.1.1. For (ii), we can treat □

Tatikonda
Chang and Pollard, 1997

7.4. Extension to infinite product spaces

ASH or KLENKE?

CHAPTER 8

Measure and Topology

8.1. Topological approximation theorems

Use Tao (epsilon of room) + Royden (chap 12). Background in appendix chapter on topological spaces

8.2. Convergence of measures

Ash + Klenke + Royden background on weak* convergence covered in the section on Duality

Part 2

Probability

CHAPTER 9

Independence and random variables

Formal probability theory is often described as measure theory (or more generally, analysis) on measure spaces $(\mathcal{X}, \mathcal{F}, \mu)$ where the $\mu(\mathcal{X}) = 1$ and to a certain extent the foundational theory does resemble this characterization. However probability itself is distinct from foundational probability *theory*, in that its goals and aims are to solve problems that are of a fundamentally different character than problems seen in analysis.

Probability as as subject has a strong combinatorial flavor, since it owes its origins to gambling and games of chance considered by amateur mathematicians in the 17th and 18th centuries. These classical ideas still permeate the modern *probabilistic way of thinking.* As such, it would be useful to revise the basic combinatorial tools that are indispensable when tackling problems of this nature. These are discussed in Appendix C.

In general, we can say that probability theory "has a right hand and a left hand". The right hand is the rigorous measure-theoretic axiomatization that one encounters in a course on probability theory. The left hand is the probablistic intuition that one develops through trying to model stochastic phenomena with the aforementioned probabilistic way of thinking. Probabilists aim to be ambidextrous in this regard, using the axiomatic framework of probability theory to deduce facts about processes whose parts and subparts can be reduced to "naïve" probabilistic concepts.

In order to cover both parts of probability adequately, we will use the background in analysis developed in Part I of these notes to prove general theorems, while using numerous examples to develop probabilistic intuition. I hope that the combination of theory and examples prove sufficient in painting a vivid picture of probability theory, with a view towards applications in statistics and economics.

9.1. Probability spaces and probability measures

While formal probability theory is based on measure theory, the language of probability theory is different. To start our exploration into probability theory, we shall first have to translate a lot of the basic terminology of measure theory into the langauge of probability. To begin with, we specialize the notion of a measure space $(\mathcal{X}, \mathcal{F}, \mu)$ to a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ where $\mathbb{P}(\Omega) = 1$. Measurable sets $A \in \mathcal{F}$ are called *events* in the language of probability. The basic properties of probability measures carry over from Chapter 1; we list a few more for completeness.

PROPOSITION 9.1.1 (Inclusion-Exclusion). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A_1, A_2, \ldots, A_n \in \mathcal{F}$. Then*

$$\mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} (-1)^{i-1} \sum_{J \subset \{1,2,\ldots n\}, |J|=i} \mathbb{P}\left(\bigcap_{j \in J} A_j\right).$$

PROOF. Integrate the equality (25) in Lemma (C.2.1). □

EXAMPLE 9.1.2 (ISI 2017 PSA 23). Three numbers are chosen at random from $\{1, 2, \ldots, 10\}$ without replacement. What is the probability that the minimum of the chosen numbers is 3 or their maximum is 7 ? Let A be the event that we select three numbers with minimum 3 and B be the event that we select three numbers with maximum 7. We want $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. We need to compute each component. To compute the first term, note that there are $\binom{7}{2}$ ways to pick 2 numbers without replacement after fixing the first one at 3 since we are picking only from the remaining numbers which are greater than three. Therefore $\mathbb{P}(A) = \binom{7}{2} / \binom{10}{3}$. Similarly, we have that $\mathbb{P}(B) = \binom{6}{2} / \binom{10}{3}$. To find the intersection, note that after fixing 3 and 7 there are only 4 choices of the middle number, leaving us with $\mathbb{P}(A \cap B) = 4 / \binom{10}{3}$. In sum,

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \frac{\binom{7}{2} + \binom{6}{2} - 4}{\binom{10}{3}} \\
&= \frac{21 + 15 - 4}{120} \\
&= \frac{4}{15}.
\end{aligned}$$

Proposition 9.1.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A_1, A_2, \dots, A_n \in \mathcal{F}$. Then*

$$\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) \geq \sum_{i=1}^{n} \mathbb{P}(A_i) - (n-1).$$

Proof. First note that for $n = 2$, the result follows due to the fact that for any $A, B \in \mathcal{F}$

$$\begin{aligned}
\mathbb{P}(A \cup B) &= \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \\
&\leq \mathbb{P}(\Omega) \\
&= 1.
\end{aligned}$$

Now assume the induction hypothesis and note that

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^{n} A_i\right) &= \mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i \cap A_n\right) \\
&\geq \mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right) + \mathbb{P}(A_n) - 1 \\
&\geq \sum_{i=1}^{n-1} \mathbb{P}(A_i) - (n-2) + \mathbb{P}(A_n) - 1 \\
&= \sum_{i=1}^{n} \mathbb{P}(A_i) - (n-1).
\end{aligned}$$

□

9.2. Independent events

In Chapter 7 we discussed the notion of product measures; the analagous concept in probability is that of independence. Informally, we think of independent events as those where the occurence or non-occurenece of one event does not impact the occurence or non-occurence of another. We can formalize this idea with the following definition-

Definition 9.2.1. Let I be an arbitrary index set. Then a collection of events $\{A_i\}_{i \in I} \subset \mathcal{F}$ are independent if for every finite subset $J \subset I$

$$\mathbb{P}\left[\bigcap_{j \in J} A_j\right] = \prod_{j \in J} \mathbb{P}[A_j].$$

Two independent events $A, B \in \mathcal{F}$ are often denoted $A \perp\!\!\!\perp B$.

Notice how the index set I was arbitrary and so we can have countable or uncountable collections of independent events. Of course, this notion isn't particualrly interesting unless we have some rich examples of such events and the probability spaces they live in. More generally, we have a question of existence: do independent events always exist, no matter the probability space? The answer to this question is trivially yes since $\mathbb{P}(\Omega \cap \emptyset) = \mathbb{P}(\Omega)\mathbb{P}(\emptyset) = 0$. So then the question reduces to asking whether *non-trivial* independent events always exist. Here of course, the answer is "no", since for $\mathcal{F} = \left\{\emptyset, A, A^C, \Omega\right\}$ where $0 < \mathbb{P}(A) < 1$, we have that $\mathbb{P}\left(A \cap A^C\right) = 0$ but $\mathbb{P}(A)\mathbb{P}\left(A^C\right) = \mathbb{P}(A)(1 - \mathbb{P}(A)) \neq 0$. Nevertheless, we have a rich collection of examples of independent events: think about rolling a dice twice and observing a six in each roll; these event of seeing a six in the first roll is independent of the event of seeing one in the next roll.

Example 9.2.2 (ISI 2017 PSA 15). Suppose that the events A, B, and C are pairwise independent such that each of them occurs with probability p. Assume that all three of them cannot occur simultaneously. What is $P(A \cup B \cup C)$? Well, we apply independence and inclusion exclusion to note that

$$\begin{aligned}
\mathbb{P}(A \cup B \cup C) &= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(A \cap C) - \mathbb{P}(B \cap C) \\
&= \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A)\mathbb{P}(B) - \mathbb{P}(A)\mathbb{P}(C) - \mathbb{P}(B)\mathbb{P}(C) \\
&= 3p(1-p).
\end{aligned}$$

The canonical example of the an infinite collection of independent events is given by the following description of an infinitely repeated experiment.

Example 9.2.3. Let E consist of a finite set of outcomes and let $\Omega = E^{\mathbb{N}}$ be the collection of E-valued sequences. Thus for any $\omega \in \Omega$, we can write

$$\omega = (\omega_1, \omega_2, \ldots)$$

where $\omega_i \in E$. We can then construct the collection

$$[\omega_1^*, \ldots, \omega_n^*] := \{\omega \in \Omega \mid \omega_i = \omega_i^*, 1 \leq i \leq n\}$$

TODO

Lemma 9.2.4. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $A, B \in \mathcal{F}$ be events. Then the claims that A, B are independent, A, B^C are independent, A^C, B are independent, and A^C, B^C are independent are equivalent.*

Proof. Note that

$$\begin{aligned}
\mathbb{P}\left(A \cap B^C\right) &= \mathbb{P}(A \setminus B) \\
&= \mathbb{P}(A \setminus A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A \cap B) \\
&= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\
&= \mathbb{P}(A)(1 - \mathbb{P}(B)) \\
&= \mathbb{P}(A)\mathbb{P}\left(B^C\right)
\end{aligned}$$

where the fourth equality uses independence. Then it should be clear that A^C and B are independent exactly for the same reason. Finally, we can apply the logic above using A^C and B instead of A and B to get that A^C and B^C are independent. $\square$

Proposition 9.2.5. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let I be an arbitrary index set. Let $\{A_i\}_{i \in I} \in \mathcal{F}$ be a collection of events and define $B_i^1 = A_i^C$ and $B_i^0 = A_i$. Then, the following statements are equivalent*

(i) *$\{A_i\}_{i \in I} \in \mathcal{F}$ are independent events.*

(ii) *There exists some* $\alpha \in \{0,1\}^I$ *such that* $\{B_i^{a(i)}\}$ *are independent.*

(iii) *For every* $\alpha \in \{0,1\}^I$, $\{B_i^{a(i)}\}$ *are independent.*

Proof. (Sketch) Note that if *(i)* holds, then *(ii)* holds automatically with $a(i) = 0$ for all $i \in I$. Similarly, if (iii) holds then (i) holds trivially. Thus we need to prove that $(ii) \implies (iii)$. First, fix $\alpha \in \{0,1\}^I$ such that our claim holds. Notice that by Lemma 9.2.4, for any $J \subset I$ such that $|J| = 2$,

$$\mathbb{P}\left(\prod_{j\in J} B_j^{\gamma(j)}\right) = \prod_{j\in J} \mathbb{P}\left(B_j^{\gamma(j)}\right)$$

for any $\gamma \in \{0,1\}^I$. For induction, suppose that the claim holds for any $J \subset I$ such that $|J| = n$ and the consider a subset $J' \subset I$ with $|J'| = n+1$. Note that for any $i \in J'$, we can define $B_{-i} := \bigcap_{j\in J'\setminus\{i\}} B_j^{\alpha(j)}$ and observe that

$$\mathbb{P}\left(B_{-i} \cap B_i^{\alpha(i)}\right) = \mathbb{P}(B_{-i})\,\mathbb{P}\left(B_i^{\alpha(i)}\right)$$

and so $B_{-i} \perp\!\!\!\perp B_i^{\alpha(i)}$ and so by our Lemma, $B_{-i} \perp\!\!\!\perp B_i^{\gamma(i)}$. The induction hypothesis then implies that $\left\{B_j^{\alpha(j)}\right\}, B_i^{\gamma(i)}$ are all mutually independent. We can yet again repeat this process with some $i' \in J'$ where now $B_{-i'} := \bigcap_{j\in J'\setminus\{i,i'\}} B_j^{\alpha(j)} \cap B_i^{\gamma(i)}$ and so on until we have replaced all αs with γs. $\square$

This idea of mutual independence finds purchase in some unexpected contexts as well. For instance, we can use it to prove Euler's prime number formula.

Theorem 9.2.6. *Let* $\mathcal{P}$ *denote the set of primes. The Riemann zeta function*

$$\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$$

has the representation

$$\zeta(s) = \prod_{p\in\mathcal{P}} \left(1 - p^{-s}\right)^{-1}.$$

Proof. Let $(\Omega, \mathcal{F}, \mathbb{P}) = \left(\mathbb{N}, 2^{\mathbb{N}}, \mathbb{P}_s\right)$ where $\mathbb{P}_s(\{n\}) = \frac{n^{-s}}{\zeta(s)}$ for $s > 1$. Define $p\mathbb{N} := \{pn \mid n \in \mathbb{N}\}$ and let for any $p \in \mathcal{P}$ and notice that $\mathbb{P}_s(p\mathbb{N}) = \sum_{n=1}^{\infty} \mathbb{P}(pn) = p^{-s}$ by countable additivity. Then, for any distinct collection $p_1, \ldots, p_k \in \mathcal{P}$ we have that

$$\begin{aligned}
\mathbb{P}_s\left(\bigcap_{i=1}^{k} p_i\mathbb{N}\right) &= \sum_{n=1}^{\infty} \mathbb{P}_s\left(\prod_{i=1}^{k} pn\right) \\
&= \prod_{i=1}^{k} p^{-s} \\
&= \prod_{i=1}^{k} \mathbb{P}_s(p_i\mathbb{N}).
\end{aligned}$$

In other words, the events $\{p\mathbb{N}\}_{p\in\mathcal{P}}$are mutually independent and so by Proposition 9.2.5

$$\begin{aligned}
\frac{1}{\zeta(s)} &= \mathbb{P}_s(\{1\}) \\
&= \mathbb{P}_s\left(\bigcap_{p\in\mathcal{P}} (p\mathbb{N})^C\right) \\
&= \lim_{n\to\infty} \mathbb{P}_s\left(\bigcap_{p\in\mathcal{P},p\leq n} (p\mathbb{N})^C\right) \\
&= \lim_{n\to\infty} \prod_{p\in\mathcal{P},p\leq n} (1-\mathbb{P}_s(p\mathbb{N})) \\
&= \lim_{n\to\infty} \prod_{p\in\mathcal{P},p\leq n} \left(1-p^{-s}\right) \\
&= \prod_{p\in\mathcal{P}} \left(1-p^{-s}\right)
\end{aligned}$$

where the second equality is due to the fact that 1 is neither prime nor a product of primes, the third due to the continuity of measures (see Propositions 1.3.4 and 1.3.6), and the fourth due the fact that $\mathbb{P}_s(\mathbb{N}) = 1$. □

We are finally ready to present the second Borel-Cantelli lemma, which we promised back in Chapter 2. The idea here is that if you an infinite sequence of independent events, like coin-tosses where the outcomes can either be head or tails, the probability that you will only see a finite number of heds is intuitively zero.

Theorem 9.2.7 (Second Borel-Cantelli lemma). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ be mutually independent events. If*

$$\sum_{i=1}^{\infty} \mathbb{P}(A_i) = \infty$$

then

$$\mathbb{P}\left(\limsup_{i\to\infty} A_i\right) = 1.$$

PROOF. Note that

$$\begin{aligned}
\mathbb{P}\left((\limsup A_i)^C\right) &= \mathbb{P}\left(\left(\bigcap_{n\in\mathbb{N}}\bigcup_{i\geq n} A_i\right)^C\right) \\
&= \mathbb{P}\left(\bigcup_{n\in\mathbb{N}}\bigcap_{i\geq n} A_i^C\right) \\
&= \lim_{n\to\infty}\mathbb{P}\left(\bigcap_{i\geq n} A_i^C\right) \\
&= \lim_{n\to\infty}\lim_{m\to\infty}\mathbb{P}\left(\bigcap_{m\geq i\geq n} A_i^C\right) \\
&= \lim_{n\to\infty}\lim_{m\to\infty}\prod_{m\geq i\geq n}\left(1-\mathbb{P}\left(A_i\right)\right) \\
&= \lim_{n\to\infty}\lim_{m\to\infty}\exp\left(\sum_{m\geq i\geq n}\log\left(1-\mathbb{P}\left(A_i\right)\right)\right) \\
&\leq \lim_{n\to\infty}\lim_{m\to\infty}\exp\left(-\sum_{m\geq i\geq n}\mathbb{P}\left(A_i\right)\right) \\
&= \lim_{n\to\infty}\exp\left(-\sum_{i\geq n}\mathbb{P}\left(A_i\right)\right) \\
&= 0
\end{aligned}$$

where in the second equality we have used DeMorgan's laws, the third and fourth follow from the continuity of measures, the fifth by independence and the fact that $\mathbb{P}(\Omega) = 1$, and the inequality by the fact that $\log(1-x) \leq -x$ for $x \in [0,1)$.[1] □

Note that we can't dispense with the independence assumption on this version of the Borel Cantelli lemma, since for identical events A_i, $\limsup A_i = A_1$ and so if $\mathbb{P}(A_1) < 1$ then $\mathbb{P}(\limsup A_i) < 1$ even though $\sum \mathbb{P}(A_i) = \infty$.

Differences between mutual and pariwise independence

9.3. Independent $\sigma-$algebras

The idea of independence of events can be extended to the idea of independence of structures of events. For a given probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we say that a collection of events $\{\mathcal{F}_i\}_{i\in I}$ – where I is an arbitrary index set and each $\mathcal{F}_i \subset \mathcal{F}$ – is mutually independent if for any finite $J \subset I$ and $F_j \in \mathcal{F}_j$

$$\mathbb{P}\left(\bigcap_{j\in J} F_j\right) = \prod_{j\in J}\mathbb{P}\left(F_j\right).$$

Usually, the structures we consider are $\sigma-$algebras, since those are the most important types of collections of events in probability theory. Immediately, our $\pi-\lambda$ theorem gives us some useful tools that allows us to use independence on generating classes to prove independence on the $\sigma-$algebras generated by them.

[1] This can be verified by noting that the function $g(x) = \log(1-x) + x$ is 0 at $x = 0$ and has derivative $g'(x) = 1 - \frac{1}{1-x} = \frac{-x}{1-x} < 0$ for $x \in (0,1)$.

PROPOSITION 9.3.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let I be an arbitrary index set. If a collection of $\pi-$systems $\{\mathcal{E}_i\}_{i\in I} \subset \mathcal{F}$ are independent, then $\{\sigma(\mathcal{E}_i)\}_{i\in I}$ are independent.*

PROOF. Pick some $J \subset I$ such that $|J| = n$. Without loss of generality, we can assume that $J = \{1, 2, \ldots, n\}$. Define

$$\mathcal{D}_1 := \{F_1 \in \sigma(\mathcal{E}_1) \mid \mathbb{P}(F_1 \cap E_2 \cap \ldots \cap E_n) = \mathbb{P}(F_1)\mathbb{P}(E_2)\ldots\mathbb{P}(E_n) \forall E_i \in \mathcal{E}_i, 2 \le i \le n\}$$

and note that we can show that $\mathcal{D}_1$ is a $\lambda-$system. To see this, first note that $\Omega \in \mathcal{D}_1$ since

$$\begin{aligned}\mathbb{P}(\Omega \cap E_2 \cap \ldots \cap E_n) &= \mathbb{P}(E_2 \cap \ldots \cap E_n) \\ &= \mathbb{P}(E_2)\ldots\mathbb{P}(E_n) \\ &= \mathbb{P}(\Omega)\mathbb{P}(E_2)\ldots\mathbb{P}(E_n)\end{aligned}$$

where in the second equality we use the independence of the generators. Next, suppose that $F_1, F_2 \in \mathcal{D}_1$ such that $F_2 \subseteq F_1$ and observe that

$$\begin{aligned}\mathbb{P}((F_1 \setminus F_2) \cap E_2 \cap \ldots \cap E_n) &= \mathbb{P}(F_1 \cap E_2 \cap \ldots \cap E_n) - \mathbb{P}(F_2 \cap E_2 \cap \ldots \cap E_n) \\ &= \mathbb{P}(F_1)\mathbb{P}(E_2)\ldots\mathbb{P}(E_n) - \mathbb{P}(F_2)\mathbb{P}(E_2)\ldots\mathbb{P}(E_n) \\ &= (\mathbb{P}(F_1) - \mathbb{P}(F_2))\mathbb{P}(E_2)\ldots\mathbb{P}(E_n) \\ &= \mathbb{P}(F_1 \setminus F_2)\mathbb{P}(E_2)\ldots\mathbb{P}(E_n)\end{aligned}$$

which shows that $F_1 \setminus F_2 \in \mathcal{D}_1$. Finally, suppose that $\{F_i\}_{i\in\mathbb{N}} \in \mathcal{D}_1$ such that $F_i \subset F_{i+1}$ and note that

$$\begin{aligned}\mathbb{P}\left(\left(\bigcup_{i\in\mathbb{N}} F_i\right) \cap E_2 \cap \ldots \cap E_n\right) &= \lim_{n\to\infty} \mathbb{P}(F_n \cap E_2 \cap \ldots \cap E_n) \\ &= \lim_{n\to\infty} \mathbb{P}(F_n)\mathbb{P}(E_2)\ldots\mathbb{P}(E_n) \\ &= \mathbb{P}\left(\bigcup_{i\in\mathbb{N}} F_i\right)\mathbb{P}(E_2)\ldots\mathbb{P}(E_n)\end{aligned}$$

which shows that $\bigcup_{i\in\mathbb{N}} F_i \in \mathcal{D}_1$. Thus $\mathcal{D}_1$ is a $\lambda-$system containing $\mathcal{E}_1$ and so $\sigma(\mathcal{E}_1) \subseteq \mathcal{D}_1$. Next, we can define

$$\mathcal{D}_2 := \{F_2 \in \sigma(\mathcal{E}_2) \mid \mathbb{P}(F_1 \cap F_2 \cap \ldots \cap E_n) = \mathbb{P}(F_1)\mathbb{P}(F_2)\ldots\mathbb{P}(E_n)\, F_1 \in \sigma(\mathcal{E}_1), E_i \in \mathcal{E}_i, 3 \le i \le n\},$$

observe that $\mathcal{E}_2 \subseteq \mathcal{D}_2$ by the result on $\mathcal{D}_1$, and show that $\mathcal{D}_2$ is a $\lambda-$system and so $\sigma(\mathcal{E}_2) \subseteq \mathcal{D}_2$ and so on. This completes the argument. □

COROLLARY 9.3.2. *Let $\{\mathcal{F}_i\}_{i\in I}$ be a collection of independent $\sigma-$algebras. If $J, K \subset I$ are disjoint, then*

$$\sigma\left(\bigcup_{j\in J} \mathcal{F}_j\right) \perp\!\!\!\perp \sigma\left(\bigcup_{k\in K} \mathcal{F}_k\right).$$

PROOF. Let $\mathcal{E}_J$ be the collection of all finite intersections of sets in $\bigcup_{j\in J} \mathcal{F}_j$ and note that $\bigcup_{j\in J} \mathcal{F}_j \subset \mathcal{E}_J$ for any $j \in J$ since we can take all sets one at a time. Thus $\sigma\left(\bigcup_{j\in J} \mathcal{F}_j\right) \subseteq \sigma(\mathcal{E}_J)$. Conversely, $\mathcal{E}_J \subseteq \sigma\left(\bigcup_{j\in J} \mathcal{F}_j\right)$ and so $\sigma\left(\bigcup_{j\in J} \mathcal{F}_j\right) = \sigma(\mathcal{E}_J)$. Note that $\mathcal{E}_J$ and $\mathcal{E}_K$ (where $\mathcal{E}_K$ is defined analogously) are $\pi-$systems and so if $\mathcal{E}_J \perp\!\!\!\perp \mathcal{E}_K$ then $\sigma(\mathcal{E}_J) \perp\!\!\!\perp \sigma(\mathcal{E}_K)$ by Proposition 9.3.1. Now for any $E_J \in \mathcal{E}_J$, we can write $E_J = \bigcap_{j\in J} F_j$ where $F_j \in \mathcal{F}_j$ (where for some js $F_j = \Omega$).

Similarly $E_K \in \mathcal{E}_K$ could be written $E_K = \bigcap_{k\in K} F_k$. Therefore,

$$\begin{aligned}\mathbb{P}(E_J \cap E_K) &= \mathbb{P}\left(\bigcap_{j\in J} F_j \cap \bigcap_{k\in K} F_k\right)\\ &= \mathbb{P}\left(\bigcap_{j\in J} F_j\right)\mathbb{P}\left(\bigcap_{k\in K} F_k\right)\end{aligned}$$

where we have used the independence of $\{\mathcal{F}_i\}_{i\in I}$. $\square$

These results yield as our first *zero-one law* which is a result that characterizes some type of $\sigma-$algebra that only contain events iwth probability zero or one.

Definition 9.3.3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose $\{\mathcal{F}_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$ is a collection of sub $\sigma-$algebras on this space. Suppose further that $\mathcal{F} = \sigma\left(\bigcup_{i\in I}\mathcal{F}_i\right)$ and $\mathcal{H}_n := \sigma\left(\bigcup_{i\geq n}\mathcal{F}_i\right)$ and so $\mathcal{H}_{n+1} \subseteq \mathcal{H}_n$. The *tail $\sigma-$algebra* with respect to $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$ is given

$$\mathcal{H}_\infty = \bigcap_{n\in\mathbb{N}} \mathcal{H}_n.$$

Theorem 9.3.4 (Kolmogorov's Zero-One Law). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose $\{\mathcal{F}_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$ is a collection of mutually independent sub $\sigma-$algebras on this space. Then the tail $\sigma-$algebra $\mathcal{H}_\infty$ is trivial in that for any $A \in \mathcal{H}_\infty$*

$$\mathbb{P}(A) \in \{0,1\}.$$

Proof. Note that $\mathcal{F}_1, \ldots, \mathcal{F}_{n-1}$ and $\mathcal{H}_n$ are mutually independent by an inductive extension of Corollary 9.3.2. Of course, since $\mathcal{H}_\infty \subseteq \mathcal{H}_n$ so $\mathcal{F}_1, \ldots, \mathcal{F}_n$ and $\mathcal{H}_\infty$ are all mutually independent. Of course, since n is arbitrary, we have that $\mathcal{H}_\infty$ and $\{\mathcal{F}_i\}_{i\in\mathbb{N}}$are all mutually independent. Again, applying Corollary 9.3.2, we have that $\mathcal{H}_\infty \perp\!\!\!\perp \sigma\left(\bigcup_{i\in\mathbb{N}}\mathcal{F}_i\right)$.Note that since $\mathcal{H}_\infty \subset \mathcal{F} = \sigma\left(\bigcup_{i\in\mathbb{N}}\mathcal{F}_i\right)$ we have that $\mathcal{H}_\infty \perp\!\!\!\perp \mathcal{H}_\infty$and so for any $A \in \mathcal{H}_\infty$

$$\mathbb{P}(A) = \mathbb{P}(A)^2 \implies \mathbb{P}(A) \in \{0,1\}.$$

$\square$

9.4. Random variables

9.4.1. General description and terminology.

Definition 9.4.1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A *random variable* is a real-valued Borel-measurable map $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$. A *random vector* is a measurable map $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$. A *random element* is a measurable map $X : (\Omega, \mathcal{F}) \to (S, \mathcal{B})$ where $(S, \mathcal{B})$ is a Polish space with its Borel $\sigma-$algebra.

Our definitions are increasingly general in that random variables, are random vectors, which in turn are random elements. Much of our language about measurable functions has analogues in probability. For instance, the image measure $X\mathbb{P}$ (which we shall denote as $\mathbb{P}_X$ from now on) of X under $\mathbb{P}$ is called the *distribution* of X. The Radon-Nikodym derivative of this distribution with respect to the Lebesgue measure (if it exists) is called the *density* of X, which is always non-negative (like all RN-derivatives) and integrates to 1. The Stieljes function $F_X(x) := \mathbb{P}_X((-\infty, x])$ is called the *cumulative distribution function (CDF)* of X. Theorem 1.4.3 tells us that CDFs completely characterize the distribution of a random variable X and so random variables with the same CDF are called *identically distributted.* Moreover, Proposition 6.3.2 tells us that the absolute continuity of CDFs (as functions on the real line) is equivalent to the absolute continuity of the distribution with respect to the Lebesgue measure. Of course, not all random variables have absolutely continuous CDFs. The canonical pathological example of a random variable which has a continuous but not absolutely continuous CDF is the Cantor random variable, whose CDF is the Cantor function. More importantly, many random variables have countable supports[2] and can be thought of as *discrete.* A reasonably comprehensive accounting of various distributions that we can find in the wild can be found in Appendix E.

The CDF has two additional properties that we did not see for Stieljes functions in general.

PROPOSITION 9.4.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X : \Omega \to \mathbb{R}$ be a random variable. Then its cumulative distribution function F_X has the property that*

$$\lim_{x \to \infty} F_X(x) = 1$$

and

$$\lim_{x \to -\infty} F_X(x) = 0.$$

PROOF. Note that the first limit is equivalent to

$$\begin{aligned}\lim_{n \to \infty} \mathbb{P}_X((-\infty, n]) &= \mathbb{P}_X\left(\bigcup_{n \in \mathbb{N}} (-\infty, n]\right) \\ &= \mathbb{P}_X((-\infty, \infty)) \\ &= 1.\end{aligned}$$

where we used the upper continuity of measures. Similarly, the other limit can be written as

$$\begin{aligned}\lim_{n \to \infty} \mathbb{P}_X((-\infty, -n]) &= \mathbb{P}_X\left(\bigcap_{n \in \mathbb{N}} (-\infty, -n]\right) \\ &= \mathbb{P}_X(\emptyset) \\ &= 0.\end{aligned}$$

$\square$

EXAMPLE 9.4.3. Let n be a positive integer and suppose that

$$F(x) = \frac{\lfloor x \rfloor}{n} \mathbb{1}\{0 \leq x \leq n\} + \mathbb{1}\{x > n\}$$

and observe that F is clearly non-decreasing, it is right continuous as the floor function is right continuous. The limiting behavior is obvious. To construct the probability distribution that generates this CDF, one can define a random variable X such that

$$\begin{aligned}\mathbb{P}(X = x) &= F(x) - F(x-1) \\ &= \frac{1}{n}\end{aligned}$$

which is the *discrete uniform distribution* over $\{1, 2, \ldots, n\}$.

EXAMPLE 9.4.4. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X be a random variable with distribution characterized by

$$\mathbb{P}_X(\{x\}) = \left(\frac{1}{2}\right)^x \mathbb{1}\{x \in \mathbb{N}\}.$$

Since

$$\mathbb{P}_X(\mathbb{N}) = \sum_{i=1}^{\infty} \frac{1}{2^i} = 1,$$

[2]This is again a shorthand! What we mean here is that the CDF has a countable support as a real valued function on $\mathbb{R}$.

for this to be a valid distribution, it must be that $\mathbb{P}_X(\mathbb{R}\setminus\mathbb{N}) = 0$. What is the CDF of such a distribution? Well, we know that

$$\begin{aligned} F_X(x) &= \mathbb{P}_X((-\infty, x]) \\ &= \mathbb{P}_X((-\infty, x]\cap\mathbb{N}) \\ &= \sum_{i=1}^{\lfloor x\rfloor}\frac{1}{2^i} \\ &= 1-\frac{1}{2^{\lfloor x\rfloor}}. \end{aligned}$$

The convex combination of CDFs remains a CDF , which gives rise to mixture distributions.

Proposition 9.4.5. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and let X, Y be random variables with CDFs F_X and F_Y. For any $\lambda\in[0,1]$, the function $F(x) := \lambda F_X(x) + (1-\lambda)F_Y(x)$ satisfies all the properties of CDF.*

Proof. Note that convex combinations of increasing functions is increasing and the convex combinations of right continuous functions is right continuous. The limiting properties also follow easily since limits behave linearly. □

Example 9.4.6 (ISI 2023 PSB 3). Do there exist CDFs on $\mathbb{R}$ such that $F(x) = F(x^n)$ where $n > 1$? First, note that if $n$is even, then for any $x < 0$ $F(x) = F(x^n)$ which implies that

$$\begin{aligned} 0 &= \lim_{x\to-\infty}F(x) \\ &= \lim_{x\to-\infty}F(x^n) \\ &= \lim_{x\to\infty}F(x^n) \\ &= 1 \end{aligned}$$

which is a contradiction and so no such distribution exists. If n is odd then, we have that for any $x, y > 1$, we $F(x) = F(y)$ since if $x < y$ there exists some $k\in\mathbb{N}$ such $x^{n^k} > y$ and so the non-decreasing nature of F yields the result. Of course, since $\lim_{x\to\infty}F(x) = 1$, we must have that $F(x) = 1$ for every $x > 1$. A similar result shows that $F(x) = F(y) = 0$ for $x < -1$. For the intermediate values of x, we know that for any $x_1, x_2\in[-1, 0)$ and $y_1, y_2\in(0, 1]$ that $0\le F(x_1) = F(x_2)\le F(0)\le F(y_1) = F(y_2)\le 1$.

Example 9.4.7. Let F and G be (one dimensional) distribution functions. Decide which of the following are distribution functions. (a) F^2, (b) H where $H(t) = \max\{F(t), G(t)\}$.
Justify your answer.

The integral of a random variable is called an *expectation* (provided it exists). That is, $\bar{\mathbb{P}}(X)$ is the expectation of X. In the probability literature, this is often denoted $\mathbb{E}[X]$. Since random variables are defined on measures spaces with unit measure, the integral captures the average value of the function X. If $X\in\mathcal{L}^p$ for $p > 1$ then the random variable is said to have finite pth moment. Of course, by Proposition 5.1.6 we know that if a random variable has a finite pth moment, it has a finite qth moment for $1\le q\le p$. Some moments have special names. So, for instance, for the centered random variable $Y := X - \mathbb{E}[X]$, the second moment $\mathbb{V}\mathrm{ar}[X]{:=}\mathbb{E}\left[Y^2\right] = \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2$ is called the *variance* of X. It is often denoted as σ^2 and it captures the degree to which a random variable deviates from its mean value. The third *standardized* moment $\mathbb{E}\left[\left(\frac{Y}{\sigma}\right)^3\right]$ is called the *skewness;* it captures the degree to which a random variables distribution is asymmetric. Here σ is the square root of the variance, and is referred to as the *standard deviation* of X. The fourth standardized moment $\mathbb{E}\left[\left(\frac{Y}{\sigma}\right)^4\right]$ is called the kurtosis. Of course, random variables need not even by $\mathcal{L}^1$, as the following two examples illustrate. A more detailed discussion on moments can be found in the Appendix on common probability distributions

EXAMPLE 9.4.8. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probabiltiy space and let $X : \Omega \to \mathbb{R}$ be a random variable whose distribution is absolutely continuous with respect to the Lebesgue measure with density $f_X(x) = \frac{1}{\pi(1+x^2)}$. Note that this is a valid density since

$$\begin{aligned}\bar{\lambda}\left(\frac{1}{\pi(1+x^2)}\right) &= \int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)} \\ &= \int_{-\infty}^{0} \frac{1}{\pi(1+x^2)} + \int_{0}^{\infty} \frac{1}{\pi(1+x^2)} \\ &= \frac{2}{\pi}\int_{0}^{\infty} \frac{1}{(1+x^2)} \\ &= \frac{2}{\pi}\lim_{b\to\infty} \tan^{-1}(b) - \tan^{-1}(0) \\ &= \frac{2}{\pi}\frac{\pi}{2} \\ &= 1\end{aligned}$$

where in the third equality we used the the symmetry of the density (and hence distribution)[3] and in the fourth the fact that the derivative of $\tan^{-1}(x)$ is $\frac{1}{1+x^2}$ and the fundamental theorem of calculus. To see that the first moment doesn't exist, observe that

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[X^+\right] - \mathbb{E}\left[X^-\right] \\ &= \int_0^{\infty} \frac{x}{\pi(1+x^2)}dx - \int_0^{\infty} \frac{x}{\pi(1+x^2)}dx\end{aligned}$$

where in the second equality we have used Corollaries 3.2.14 and 3.2.12 along with the Radon-Nikodym theorem. Thus the expectation is zero if the integral of the parts is finite; otherwise it is not defined. To this end, note that

$$\begin{aligned}\int_0^{\infty} \frac{x}{\pi(1+x^2)}dx &= \frac{1}{2\pi}\int_1^{\infty} \frac{1}{u}du \\ &= \frac{1}{2\pi}\left[\lim_{b\to\infty} \log(b) - \log(1)\right] \\ &= \infty\end{aligned}$$

which shows that $\mathbb{E}[X]$ is not defined.

EXAMPLE 9.4.9. Consider the unit interval $(0, 1)$ that is divided into two sub-intervals by picking a point at random from inside the interval. Denoting by Y and Z the lengths of the longer and the shorter sub-intervals respectively, we can show that Y/Z does not have finite expectation. Indeed,

[3]The symmetry of the distribution of a random variable is an important property that characterizes many of the common probability distributions we hear about in probability and statistics. For instance, the normal distribution, and the uniform distribution, are both symmetric distributions. A probability distribution of a random variable X is said to be symmetric about 0 if for any $x \in \mathbb{R}$: $F_X(x) = 1 - F_X(-x)$. One can show that the skewness of such a random variable is zero by

letting X be the uniformly distributed random variable in $(0,1)$, we have that

$$\begin{aligned}\mathbb{E}\left[\frac{Y}{Z}\right] &= \mathbb{E}\left[\frac{1-X}{X}\mathbb{1}\{0 \le X \le 0.5\} + \frac{X}{1-X}\mathbb{1}\{0.5 \le X \le 1\}\right] \\ &= \mathbb{E}\left[\frac{1-X}{X}\mathbb{1}\{0 \le X \le 0.5\}\right] + \mathbb{E}\left[\frac{X}{1-X}\mathbb{1}\{0.5 \le X \le 1\}\right] \\ &= \int_0^{0.5} \frac{1-x}{x}dx + \int_{0.5}^1 \frac{x}{1-x}dx\end{aligned}$$

where neither integral is finite.

The following is an example of a discrete probability distribution where the expectation does exist.

EXAMPLE 9.4.10. 18 boys and 2 girls are made to stand in a line in a random order. Let X be the number of boys standing in between the girls. Whati s $P(X = k)$ and $\mathbb{E}(X)$? Appendix C will be useful here. Note that if there are b boys and 2 girls, there are $(b+2)!$ ways to arrange them and $b+2-k-1$ ways to sandwich k boys between two girls, with $b!$ ways to arrange the boys and $g!$ ways to arrange the girls. Therefore

$$\begin{aligned}\mathbb{P}(X = k) &= \frac{(b+2-k-1)\, b!2!}{(b+2)!} \\ &= \frac{b+2-k-1}{(b+2)!}\end{aligned}$$

To find the expectation, note that

$$\begin{aligned}\mathbb{E}[X] &= \frac{2}{(b+2)(b+1)} \sum_{k=0}^{b} k(b+2-k-1) \\ &= \frac{2}{(b+2)(b+1)} \sum_{k=1}^{b} k(b+1-k) \\ &= \frac{2}{(b+2)(b+1)} \binom{b+2}{3} \\ &= \frac{2}{(b+2)(b+1)} \frac{b(b+1)(b+2)}{6} \\ &= \frac{b}{3}\end{aligned}$$

where in the third equality we used Lemma C.1.12.

We can characterize *joint* moments of multiple random variables just as easily. For instance, the *covariance* of two random variables X and Y on the same probability space is simply their centered inner product $\mathbb{C}\text{ov}[X,Y] := \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. Notice that this is a generalization of the variance which is essentially the covariance of a random variable with itself. The bilinearity of inner products lends itself very nicely to covariances.

EXAMPLE 9.4.11. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X and Y be $\mathcal{L}^2$ random variables such that $\mathbb{V}\text{ar}[X+Y] = 3$, and $\mathbb{V}\text{ar}[X-Y] = 1$. What is $\mathbb{C}\text{ov}[X,Y]$? We use the bilinearity of covariances here. Note that $\mathbb{V}\text{ar}[X+Y] = \mathbb{V}\text{ar}[X] + \mathbb{V}\text{ar}[Y] + 2\mathbb{C}\text{ov}[X,Y]$ and $\mathbb{V}\text{ar}[X-Y] = \mathbb{V}\text{ar}[X] + \mathbb{V}\text{ar}[Y] - 2\mathbb{C}\text{ov}[X,Y]$. Subtracting the two equations yields $\mathbb{C}\text{ov}[X,Y] = \frac{1}{2}$.

EXAMPLE 9.4.12. Let Y_1, Y_2, Y_3 be i.i.d. continuous random variables. For $i = 1, 2$, define U_i as

$$U_i = \begin{cases} 1 & \text{if } Y_{i+1} > Y_i, \\ 0 & \text{otherwise} \end{cases}$$

What is the mean and variance of $U_1 + U_2$? The key here is that the random variables are continuous and so $\mathbb{P}(Y_i = Y_{i+1}) = 0$. Since then $\mathbb{P}(Y_i > Y_{i+1}) = \mathbb{P}(Y_i < Y_{i+1}) = \frac{1}{2}$, we have that $\mathbb{E}[U_1 + U_2] = \mathbb{P}(Y_2 > Y_1) + \mathbb{P}(Y_3 > Y_2) = 1$. For the variance, note that

$$\begin{aligned}
\mathbb{V}\mathrm{ar}[U_1 + U_2] &= \mathbb{V}\mathrm{ar}[U_1] + \mathbb{V}\mathrm{ar}[U_2] + 2\mathbb{C}\mathrm{ov}[U_1, U_2] \\
&= 2\left(\mathbb{E}\left[U_1^2\right] - \mathbb{E}[U_1]^2\right) + 2\left(\mathbb{E}[U_1 U_2] - \mathbb{E}[U_1]\,\mathbb{E}[U_2]\right) \\
&= 2\left(\frac{1}{2} - \frac{1}{4}\right) + 2\left(\frac{1}{6} - \frac{1}{4}\right) \\
&= \frac{1}{2} - \frac{1}{6} \\
&= \frac{1}{3}.
\end{aligned}$$

where in the second equality we have used the fact that U_1 and U_2 are identically distributed and in the third the fact that $\mathbb{P}(Y_3 > Y_2 > Y_1) = \frac{1}{6}$ since there are 3! ways to arrange those three random variables.

The *Pearson correlation coefficient* between X and Y, is defined

$$r_{X,Y} := \frac{\mathbb{C}\mathrm{ov}[X, Y]}{\sqrt{\mathbb{V}\mathrm{ar}[X]\,\mathbb{V}\mathrm{ar}[Y]}}$$

which captures the degree of linear association between X and Y. This plays a large role in the context of linear regression which we shall see in Chapter 18.

EXAMPLE 9.4.13. Suppose two teams play a series of games, each producing a winner and a loser, until one team has won two more games than the other. Let G be the total number of games played. Assume each team has a chance of 0.5 to win each game, independent of the results of the previous games. (a) Find the probability distribution of G. (b) Find the expected value of G.

SOLUTION. Note that for this series to end the last two games have to be won by the same time, with the previous games being won alternatively by one team and the other. For a series of n games, there are 2^n total possible outcomes and only two lead to termination, one where team 1 wins the last two games and the previous ones are won alternatingly and the other where team 2 wins the last two games etc. Therefore $\mathbb{P}(G = n) = \frac{n}{2^{n-1}} 1\mathbb{1}\{n \geq 2\}$. Then the expectation is

$$\mathbb{E}[G] = \sum_{n=2}^{\infty} \frac{n}{2^{n-1}}$$

which is arithmo-geometric series that can be shown to converge using a ratio test. We can note that for $|x| \leq 1$

$$\frac{1}{1-x} = \sum_{n=0}^{\infty} x^n$$

and so by the fact that the power series converges uniformly,

$$\frac{1}{(1-x)^2} = \sum_{n=1}^{\infty} n x^{n-1}$$

where we have interchanged derivatives and limitsAdd hyperrefs once differentiation section is complete. Finally, we have

$$\frac{1}{(1-x)^2} - 1 = \sum_{n=2}^{\infty} n x^{n-1}.$$

Letting $x = \frac{1}{2}$, we have that

$$\mathbb{E}[G] = 4 - 1 = 3.$$

9.4.1.1. *Quantile functions and coupling.* The expectation of a random variable X, if it exists, provides one measure of central tendency for a random variable. Other useful measures include the *median,* which is a real number c such that the probability that X less than or equal c is $\frac{1}{2}$. This is a useful measure of central tendency when the distribution of X is highly skewed. The median may not be unique, as is the case when the random variable X has a discrete distribution. More formally, we can define a median as any real number $c \in \left[\sup\left\{x \in \mathbb{R} \mid F(x) < \frac{1}{2}\right\}, \sup\left\{x \in \mathbb{R} \mid F(x) \leq \frac{1}{2}\right\}\right]$ where F is the CDF of X. In general, the pth quantile is any element of the set $Q_F(p) := [\sup\{x \in \mathbb{R} \mid F(x) < p\}, \sup\{x \in \mathbb{R} \mid F(x) \leq p\}]$. To prevent ambiguity, we often use the lowest value of this interval to construct a quantile function.

DEFINITION 9.4.14. A quantile function q_F for a given CDF F is defined

$$q_F(p) := \inf\{x \in \mathbb{R} \mid F(x) \geq p\}.$$

PROPOSITION 9.4.15. *Let q_F be the quantile function associated with a CDF F. Then q_F is the unique function such that $F(x) \geq p \Longleftrightarrow q_F(p) \leq x$.*

PROOF. First we prove that q_F satisfies these properties. Assume that $F(x) \geq p$ for some $x \in \mathbb{R}$. Then clearly $q_F(p) \leq x$ by the definition of an infimum. Conversely, assume that $q_F(p) \leq x$ and observe that since F is non-decreasing, $F(x) \geq p$. Next, suppose that some function $\tilde{q}$ satisfies this equivalence i.e.

$$F(x) \geq p \Longleftrightarrow \tilde{q}(p) \leq x. \tag{22}$$

Observe that by the right continuity of F, $F(q_F(p)) \geq p$ and so $\tilde{q}(p) \leq q_F(p)$. On the other hand, if there exists some $p^* \in [0,1]$ such that $\tilde{q}(p^*) < q_F(p^*)$ then $F(\tilde{q}(p^*)) < p^*$ which by (22) implies that $\tilde{q}(p^*) > \tilde{q}(p^*)$, a contradiction. Therefore,

$$q_F(p) = \tilde{q}(p)$$

for all $p \in [0,1]$. □

Note that the quantile function can also be written as $q_F(p) = \sup\{x \in \mathbb{R} \mid F(x) < p\}$, a fact that can be established by noting that the non-decreasing nature of F implies that $\sup\{x \in \mathbb{R} \mid F(x) < p\} \leq \inf\{x \in \mathbb{R} \mid F(x) \geq p\}$. If the inequality were strict, then there would be some $\sup\{x \in \mathbb{R} \mid F(x) < p\} < c < \inf\{x \in \mathbb{R} \mid F(x) \geq p\}$ and so $F(c) \not< p$ and $F(c) \not\geq p$ which is a contradiction.

We can use the quantile function to construct a standard probability space for random variables to live in. To begin, let $(\Omega, \mathcal{F}, \mathbb{P})$ be arbitrary and let X be a random variable on the space with CDF F_X. Define an auxiliary random variable on $\tilde{X} : ([0,1], \mathscr{B}[0,1], \lambda) \to (\mathbb{R}, \mathscr{B}(\mathbb{R}))$ by

$$\tilde{X}(p) := q_{F_X}(p)$$

and notice that its CDF

$$\begin{aligned} F_{\tilde{X}}(x) &:= \lambda\left(\tilde{X}^{-1}[(-\infty, x]]\right) \\ &= \lambda(\{p : q_{F_X}(p) \leq x\}) \\ &= \lambda(\{p : F_X(x) \geq p\}) \\ &= \lambda([0, F_X(x)]) \\ &= F_X(x). \end{aligned}$$

Thus we have taken an arbitrary random variable on X on an unspecified probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and constructed an auxiliary random variable $\tilde{X}$ on $([0,1], \mathscr{B}[0,1], \lambda)$ such that X and $\tilde{X}$ are identically distributed. Since probability is fundamentally concerned with *distributions* rather than random variables as functions themselves, we have found a canonical representation for all random variables on the same space $([0,1], \mathscr{B}[0,1], \lambda)$. The next result provides some evidence that this representation is indeed special.

Definition 9.4.16. Let X and Y be random variables, not necessarily on the same probability space. A *coupling* of X and $Y$are two new random variables $\tilde{X}$ and $\tilde{Y}$ on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\tilde{X}$ has the same distribution as X and $\tilde{Y}$ has the same distribution as Y. When we define $\tilde{X}(p) = q_{F_X}(p)$ and $\tilde{Y}(p) = q_{F_Y}(p)$ on $([0,1], \mathscr{B}[0,1], \lambda)$, then we call $\left(\tilde{X}, \tilde{Y}, ([0,1], \mathscr{B}[0,1], \lambda)\right)$ a *quantile coupling.*

Proposition 9.4.17. *Let X and Y be random variables, not necessarily on the same probability space. Then for any coupling $\left(\hat{X}, \hat{Y}, (\Omega, \mathcal{F}, \mathbb{P})\right)$, we have that*

$$\lambda\left(|\hat{X} - \hat{Y}|\right) \leq \mathbb{P}\left(|\tilde{X} - \tilde{Y}|\right).$$

Proof. Note that for any $s \in \mathbb{R}$

$$\begin{aligned}
\mathbb{P}\left(\hat{X} > s, \hat{Y} > s\right) &\leq \min\left\{\mathbb{P}\left(\hat{X} > s\right), \mathbb{P}\left(\hat{Y} > s\right)\right\} \\
&= \min\left\{1 - F_X(s), 1 - F_Y(s)\right\} \\
&= 1 - \max\left\{F_X(s), F_Y(s)\right\} \\
&= 1 - \lambda\left(p : 0 \leq p \leq \max\left\{F_X(s), F_Y(s)\right\}\right) \\
&= \lambda\left(p : \max\left\{F_X(s), F_Y(s)\right\} < p \leq 1\right) \\
&= \lambda\left(p : F_X(s) < p, F_Y(s) < p\right) \\
&= \lambda\left(p : q_{F_X}(p) > s, q_{F_Y}(p) > s\right) \\
&= \lambda\left(p : \tilde{X}(p) > s, \tilde{Y}(p) > s\right)
\end{aligned}$$

where in the second to last equality wee have used Proposition 9.4.15. Further, by the definition of a coupling

$$\begin{aligned}
\mathbb{P}\left(\hat{X} > s\right) &= \lambda\left(\tilde{X} > s\right) \\
\mathbb{P}\left(\hat{Y} > s\right) &= \lambda\left(\tilde{Y} > s\right)
\end{aligned}$$

and so

$$\bar{\mathbb{P}}\left(\mathbb{1}\left\{\hat{X} > s\right\} + \mathbb{1}\left\{\hat{Y} > x\right\} - 2\mathbb{1}\left\{\hat{X} > s, \hat{Y} > s\right\}\right) \geq \lambda\left(\mathbb{1}\left\{\tilde{X} > s\right\} + \mathbb{1}\left\{\tilde{Y} > x\right\} - 2\mathbb{1}\left\{\tilde{X} > s, \tilde{Y} > s\right\}\right).$$

Note that the expressions inside the integrals are indicators of the symmetric differences $\left\{\hat{X} > s\right\} \Delta \left\{\hat{Y} > s\right\}$ and $\left\{\tilde{X} > s\right\} \Delta \left\{\tilde{Y} > s\right\}$ and so we can write the above as

$$\begin{aligned}
&\mathbb{P}\left(\omega \in \Omega : \hat{X}(\omega) > s > \hat{Y}(\omega)\right) + \mathbb{P}\left(\omega \in \Omega : \hat{Y}(\omega) > s > \hat{X}(\omega)\right) \\
&\geq \lambda\left(p \in [0,1] : \tilde{X}(p) > s > \tilde{Y}(p)\right) + \lambda\left(p \in [0,1] : \tilde{Y}(p) > s > \tilde{X}(p)\right) \qquad (23)
\end{aligned}$$

Taking an integral with respect to the Lebesgue measure on $\mathbb{R}$ on the left hand side, we hav

$$\begin{aligned}
&\lambda^s\left(\mathbb{P}^\omega\left(\omega \in \Omega : \hat{X}(\omega) > s > \hat{Y}(\omega)\right) + \mathbb{P}^\omega\left(\omega \in \Omega : \hat{Y}(\omega) > s > \hat{X}(\omega)\right)\right) \\
&\qquad = \lambda^s\left(\mathbb{P}^\omega\left(\mathbb{1}\left\{\hat{X}(\omega) > s > \hat{Y}(\omega)\right\}\right)\right) + \lambda^s\left(\mathbb{P}^\omega\left(\mathbb{1}\left\{\hat{Y}(\omega) > s > \hat{X}(\omega)\right\}\right)\right) \\
&\qquad = \mathbb{P}^\omega\left(\lambda^s\left(\mathbb{1}\left\{\hat{X}(\omega) > s > \hat{Y}(\omega)\right\}\right)\right) + \mathbb{P}^\omega\left(\lambda^s\left(\mathbb{1}\left\{\hat{Y}(\omega) > s > \hat{X}(\omega)\right\}\right)\right) \\
&\qquad = \mathbb{P}^\omega\left(\hat{X} - \hat{Y}\right)^+ + \mathbb{P}^\omega\left(\hat{Y} - \hat{X}\right)^+ \\
&\qquad = \mathbb{P}^\omega\left(|\hat{X} - \hat{Y}|\right) \qquad .
\end{aligned}$$

where we have used Tonelli's theorem in the second equality and the fact that $f^- + f^+ = |f|$ in the last. A similar argument where we apply λ^s to the right side of (23) shows (by monotonicity of integration) that

$$\mathbb{P}^\omega\left(|\hat{X}-\hat{Y}|\right) \geq \lambda\left(|\tilde{X}-\tilde{Y}|\right).$$

□

9.4.2. Independence of random variables.

Definition 9.4.18. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let I be an arbitrary index set. Random variables $X_i : \Omega \to \mathbb{R}$ where $i \in I$ are independent if $\{\sigma(X_i)\}_{i\in I}$ are mutually independent $\sigma-$algebras.

Note that when I is finite, the existence of finite product measures tells us that we can always construct independent random variables with distributions identical to $\{X_i\}_{i\in I}$. For infinite products we use the Kolmogorov extension theorem. TODO: Flesh outA collection of random variables which are independent and have the same distribution are called *independent and identically distributed* or *i.i.d* for short.

Proposition 9.4.19. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let I be an index set. If $X_i : \Omega \to \mathbb{R}$ – where $i \in I$ – are independent random variables and $f_i : \mathbb{R} \to \mathbb{R}$ are Borel-measuarable maps then the random variables $f_i(X_i)$ are mutually independent.*

Proof. Note that by Proposition 2.2.12, $\sigma(f_i(X_i)) \subseteq \sigma(X_i)$ and so $\sigma(f_i(X_i))$ are all mutually independent by definition. □

Proposition 9.4.20. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X and Y be random variables. Then X and Y are independent if and only if*

$$F_{X,Y}(x,y) := \mathbb{P}(X \leq x \cap Y \leq y) = \mathbb{P}(X \leq x)\,\mathbb{P}(Y \leq y) = F_X(x)\,F_Y(y).$$

Proof. One direction of this is trivial. For the non-trivial direction, notice that – by Lemma 2.2.8 – sets like $(-\infty, x]$ can be used to generate $\mathscr{B}(\mathbb{R})$. Further, observe that the collection is a $\pi-$system, and so by Corollary 2.2.13, $\left\{X^{-1}((-\infty,x])\right\}_{x\in\mathbb{R}}$ and $\left\{Y^{-1}((-\infty,y])\right\}_{y\in\mathbb{R}}$ are $\pi-$systems (since intersections and preimages commute) that generate $\sigma(X)$ and $\sigma(Y)$, respectively. The result then follows by Proposition 9.3.1. □

The generalization of this result to the arbitrary collection of independent random variables proceeds in the obvious way.

Proposition 9.4.21. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probabiltiy space and let X and Y be non-negative (or integrable) independent random variables. Then,*

$$\mathbb{E}[XY] = \mathbb{E}[X]\,\mathbb{E}[Y].$$

Proof. First suppose $X, Y \geq 0$ almost surely. Note that

$$\begin{aligned}
\mathbb{E}[XY] &= \mathbb{E}_{X,Y}[xy] \\
&= \mathbb{E}_X[\mathbb{E}_Y[xy]] \\
&= \mathbb{E}_X[x]\,\mathbb{E}_Y[y] \\
&= \mathbb{E}[X]\,\mathbb{E}[Y]
\end{aligned}$$

where in the first equality we have moved to the image measure under the random vector (X,Y) through the standard change of variables (Corollary 3.2.14). Then, since the image measure under the pair of independent random vectors is a product measure, we can use Tonelli's theorem to turn it into an iterated integral. The final equality is yet another application of the change of variables formula. Now if $X, Y \in \mathcal{L}^1(\mathbb{P})$, then we can apply the result for non-negative random variables to $|X|, |Y|$ to show that

$$\mathbb{E}[|XY|] = \mathbb{E}[|X|]\,\mathbb{E}[|Y|] < \infty$$

and then applying the same argument (except replacing the use of Tonelli with the use of Fubini) yields the result. $\square$

EXAMPLE 9.4.22. Consider the i.i.d. sequence

$$X_1, X_2, X_3, X_4, X_5, X_6$$

where each X_i is one of the four symbols $\{a, t, g, c\}$. Further suppose that

$$\begin{array}{ll} P(X_1 = a) = 0.1 & P(X_1 = t) = 0.2 \\ P(X_1 = g) = 0.3 & P(X_1 = c) = 0.4. \end{array}$$

Let Z denote the random variable that counts the number of times that the subsequence cat occurs (i.e. the letters c, a and t occur consecutively and in the correct order) in the above sequence. Find $\mathbb{E}[Z]$.

SOLUTION. Let $Y_i = \mathbb{1}\{X_i = c\} \mathbb{1}\{X_{i+1} = a\} \mathbb{1}\{X_{i+2} = t\}$ for $1 \leq i \leq 4$ and notice that $Z = \sum_{i=1}^{4} Y_i$ and so

$$\begin{aligned} \mathbb{E}[Z] &= \sum_{i=1}^{4} \mathbb{E}[Y_i] \\ &= \sum_{i=1}^{4} \mathbb{P}(X_i = c) \mathbb{P}(X_{i+1} = a) \mathbb{P}(X_{i+1} = t) \\ &= 4 \times 0.4 \times 0.1 \times 0.2 \\ &= 0.032 \end{aligned}$$

where we used independence in the second equality.

EXAMPLE 9.4.23. Let X and Y be independent random variables with X having a binomial distribution with parameters n_1 and p_1 and Y having a binomial distribution with parameters n_2 and p_2. What is the probability that $|X - Y|$ is even? Well, we don't need to really worry about the absolute value sign since if $X - Y$ is even then $Y - X$ is even. Moreover, we know that $X - Y$ is even if and only if either both X and Y are even or both X and Y are odd. Therefore, formally letting $E := \{0, 2, 4, \dots\}$ and $O := \{1, 3, 5, \dots\}$

$$\begin{aligned} \mathbb{P}(|X - Y| \in E) &= \mathbb{P}(X \in E, Y \in E) + \mathbb{P}(X \in O, Y \in O) \\ &= \mathbb{P}(X \in E) \mathbb{P}(Y \in E) + \mathbb{P}(X \in O) \mathbb{P}(Y \in O) \\ &= \sum_{l \in E} \binom{n_1}{l} p_1^l (1 - p_1)^{n_1 - l} \sum_{k \in E} \binom{n_2}{k} p_2^k (1 - p_2)^{n_1 - k} \\ &\quad + \sum_{l \in O} \binom{n_1}{l} p_1^l (1 - p_1)^{n_1 - l} \sum_{k \in O} \binom{n_2}{k} p_2^k (1 - p_2)^{n_1 - k} \end{aligned}$$

9.4.3. Multiple random variables and random vectors. Most of the results that we have deduced for random variables hold *mutis mutandis* for random vectors, although some ideas need clarification in this more general context. For instance, the notion of a CDF is not immediately obvious since there is no canonical ordering on $\mathbb{R}^n$ for $n > 1$.

9.4.4. Transformations of random variables. A typical question in basic probability theory is trying to find the distribution of some transformation of a collection of random variables. More formally, given random variables $X_1, X_2, \dots, X_n$, on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we wish to find the distribution of $T(X_1, X_2, \dots, X_n)$ where $T : \mathbb{R}^n \to \mathbb{R}^m$ is a Borel-measurable map.

9.4.4.1. *Transformations on scalar random variables.*

9.4.4.2. *Adding independent random variables.* The simplest transformation we make on a collection of random variables is simply adding them. The theory of convolutions we developed in Chapter 7 gives us the tools to find the distribution of the sum of random variables.

PROPOSITION 9.4.24. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X and Y be $\mathcal{L}^1(\mathbb{P})$ (or non-negative) independent random variables on the space. Then, the distribution of $Z := X + Y$ is given by*

$$\mathbb{P}_Z(B) = \mathbb{E}_Y[\mathbb{P}_X(B - y)].$$

for any $B \in \mathscr{B}(\mathbb{R})$. In particular, the CDF

$$F_Z(z) = \mathbb{E}_Y[F_X(z - y)]$$

PROOF. Note that

$$\begin{aligned}
\mathbb{P}_Z(B) &= \mathbb{E}_Y[\mathbb{E}_X[\mathbb{1}_B(x + y)]] \\
&= \mathbb{E}_Y[\mathbb{E}_X[\mathbb{1}_{B-y}(x)]] \\
&= \mathbb{E}_Y[\mathbb{P}_X(B - y)]
\end{aligned}$$

since $x + y \in B \iff x \in B - y$. Of course, when $B = (-\infty, z]$ the result about CDFs follows. $\square$

COROLLARY 9.4.25. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let X and Y be $\mathcal{L}^1(\mathbb{P})$ (or non-negative) independent and discrete random variables supported on some countable set S. Then, the distribution of $Z := X + Y$ is given*

$$\mathbb{P}(Z = z) = \sum_{y \in S} \mathbb{P}(Y = y)\,\mathbb{P}(X = z - y).$$

PROOF. Let $B = \{z\}$ in Proposition 9.4.24. $\square$

EXAMPLE 9.4.26. Let X and Y be i.i.d. random variables, with $P(X = k) = 2^{-k}$ for $k = 1, 2, 3, \ldots$ Find $P(X > Y)$ and $P(X > 2Y)$. The trick here is to define $Z := X - \alpha Y$ for $\alpha > 0$ and note that

$$\begin{aligned}
\mathbb{P}(Z = z) &= \sum_{y \in \mathbb{N}} \mathbb{P}(Y = y)\,\mathbb{P}(X = z + \alpha y) \\
&= \sum_{y \in \mathbb{N}} 2^{-y} 2^{-(z+\alpha y)} \mathbb{1}\{z + \alpha y \geq 1\} \\
&= 2^{-z} \sum_{y \in \mathbb{N}} 2^{-(1+\alpha)y} \mathbb{1}\left\{y \geq \frac{1 - z}{\alpha}\right\} \\
&= 2^{-z} \sum_{y = \max\left\{1, \lceil \frac{1-z}{\alpha} \rceil\right\}}^{\infty} 2^{-(1+\alpha)y}.
\end{aligned}$$

For $\alpha = 1$, the above reduces to $\mathbb{P}(Z = z) = \frac{2^{-z}}{3}$ and so $\mathbb{P}(Z \geq 0) = \sum_{z=0}^{\infty} \frac{2^{-z}}{3} = \frac{2}{3}$. If $\alpha = 2$ then $\mathbb{P}(Z = z) = \frac{2^{-z}}{7}$ and so $\mathbb{P}(Z \geq 0) = \sum_{z=0}^{\infty} \frac{2^{-z}}{7} = \frac{2}{7}$.

More often than not, we are dealing with a situation where we are adding random variables that are not of the same "type", as is the case in the following example.

EXAMPLE 9.4.27. Suppose X and U are independent random variables with

$$P(X = k) = \frac{1}{N + 1}, \quad k = 0, 1, 2, \ldots, N,$$

and U having a uniform distribution on $[0,1]$. Let $Y = X + U$. What is the distribution of Y? What is the Pearson correlation coefficient $r_{Y,X}$? First note that by Proposition 9.4.24

$$\begin{aligned} F_Y(y) &= \mathbb{E}_X\left[F_U(y-x)\right] \\ &= \mathbb{E}_X\left[(y-x)\mathbb{1}\{y-1 \le x \le y\} + \mathbb{1}\{x < y-1\}\right] \\ &= \frac{1}{N+1}\sum_{x=0}^{N}(y-x)\mathbb{1}\{y-1\le x\le y\} + \frac{1}{N+1}\sum_{x=0}^{N}\mathbb{1}\{x<y-1\} \\ &= \frac{y-\lfloor y\rfloor}{N+1}\mathbb{1}\{0\le y\le N+1\} + \frac{\lfloor y\rfloor}{N+1}\mathbb{1}\{0\le y\le N+1\} + \mathbb{1}\{y>N+1\} \\ &= \frac{y}{N+1}\mathbb{1}\{0\le y\le N+1\} + \mathbb{1}\{y>N+1\}. \end{aligned}$$

For the correlation between X and Y, note that

$$\begin{aligned} \mathbb{C}\mathrm{ov}\left[X+U,X\right] &= \mathbb{V}\mathrm{ar}\left[X\right] \\ &= \end{aligned}$$

and $\mathbb{V}\mathrm{ar}\left[Y\right] = \mathbb{V}\mathrm{ar}\left[X\right] + \mathbb{V}\mathrm{ar}\left[U\right]$ and so

$$r_{X,Y} = \frac{\mathbb{V}\mathrm{ar}\left[X\right]}{\mathbb{V}\mathrm{ar}\left[X\right] + \mathbb{V}\mathrm{ar}\left[U\right]}$$

TODO

9.4.4.3. *Products of random variables.*
9.4.4.4. *Absolute values of random variables.*

9.4.5. Order statistics.

Example 9.4.28. Let $U_1, U_2, \ldots, U_n$ be i.i.d. uniform $(0,1)$ random variables and suppose

$$X = \max\left(U_1, U_2, \ldots, U_n\right) \text{ and } Y = \min\left(U_1, U_2, \ldots, U_n\right).$$

Find the distribution of $Z = X - Y$. TODO

9.4.6. Concentration inequalities.

Example 9.4.29. Consider the setting in Example 9.4.11 again. Can we provide a bound for $\mathbb{E}\left[|X+Y|\right]$ with the additional information that $\mathbb{E}\left[X+Y\right] = \mathbb{E}\left[X-Y\right] = 0$? Note that in this case $\mathbb{E}\left[X\right] = \mathbb{E}\left[Y\right] = 0$ and so $\mathbb{V}\mathrm{ar}\left[X+Y\right] = \mathbb{E}\left[(X+Y)^2\right] = 3$ and so by Jensen's inequality and the fact that $x \to \sqrt{x}$ is concave, we have that

$$\mathbb{E}\left[|X+Y|\right] \le \sqrt{\mathbb{E}\left[(X+Y)^2\right]} = \sqrt{3}.$$

Example 9.4.30 (ISI 2016 PSA 26). Two integers m and n are chosen at random with replacement from $\{1, 2, \ldots, 9\}$. What is the probabiliy that that $m^2 - n^2$ is even? First note that $m^2 - n^2 = (m+n)(m-n)$ so we need to either $m+n$ or $m-n$ to be even. For this to be true, both m and n need to be even, or both odd. The probability that they are both even is $\frac{4}{9} \times \frac{4}{9} = \frac{16}{81}$ since sampling with replacement leads to indepenent events. The probability that they are both odd is similarly $\frac{5}{9} \times \frac{5}{9} = \frac{25}{81}$. Thus the probability that they are either both odd or even is $\frac{16+25}{81} = \frac{41}{81}$.

Example 9.4.31. Two policemen are sent to watch a road that is 1 km long. Each of the two policemen is assigned a position on the road which is chosen according to a uniform distribution along the length of the road and independent of the other's position. Find the probability that the policemen will be less than 1/4 kilometer apart when they reach their assigned posts. TODO

EXAMPLE 9.4.32. Suppose X has a normal distribution with mean 0 and variance 25 . Let Y be an independent random variable taking values -1 and 1 with equal probability. Define$S = XY + \frac{X}{Y}$ and$T = XY - \frac{X}{Y}$. (a) Find the probability distribution of S. (b) Find the probability distribution of $\left(\frac{S+T}{10}\right)^2$.TODO

EXAMPLE 9.4.33. Let $X_1, X_2 \ldots$ be i.i.d. Bernoulli random variables with parameter $\frac{1}{4}$, let $Y_1, Y_2, \ldots$ be another sequence of i.i.d. Bernoulli random variables with parameter $\frac{3}{4}$ and Let N be a geometric random variable with parameter $\frac{1}{2}$ (i. e. $P(N = k) = \frac{1}{2^k}$for $k = 1, 2 \ldots$). Assume the X_i 's, Y_j's and N are all independent. Compute $\mathbb{C}\mathrm{ov}\left(\sum_{i=1}^N X_i, \sum_{i=1}^N Y_i\right)$.TODO

CHAPTER 10

Fourier transforms, Laplace transforms, and generating functions

CHAPTER 11

Conditioning

11.1. Elementary notions of conditional probability

The basic notion of conditional probabilities is clear when we are conditioning of events of positive probability. That is, on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with an event $B \in \mathcal{F}$ such that $\mathbb{P}(B) > 0$, we can easily define a conditional measure

$$\mathbb{P}_B(A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{24}$$

[1]for any $A \in \mathcal{F}$. It's easy to verify that $\mathbb{P}_B$ is in fact a probability measure on $\mathcal{F}$ and thus we can derive the corresponding integral

$$\mathbb{E}_B[f] = \frac{\mathbb{E}[f \mathbb{1}_B]}{\mathbb{P}(B)}.$$

This definition quickly yields some of the most foundational and basic probabilistic ideas: for instance with two events A, B where both occur with positive probability, they are independent if and only if $\mathbb{P}(A) = \mathbb{P}_B(A)$ .The following basic facts

Lemma 11.1.1. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $B \in \mathcal{F}$ be an event. Then for any countable partition $\{A_i\}_{i \in \mathbb{N}}$ of Ω such that $\mathbb{P}(A_i) > 0$, we have that*

$$\mathbb{P}(B) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)\, \mathbb{P}_{A_i}(B).$$

Proof. Note that by countable additivity

$$\begin{aligned} \mathbb{P}(B) &= \sum_{i=1}^{\infty} \mathbb{P}(B \cap A_i) \\ &= \sum_{i=1}^{\infty} \mathbb{P}(A_i)\, \mathbb{P}_{A_i}(B). \end{aligned}$$

This immediately yields the famous Bayes Theorem. □

Theorem 11.1.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{A_i\}_{i \in \mathbb{N}}$be a partition of Ω such that $\mathbb{P}(A_i) > 0$. Then, for any $A, B \in \mathcal{F}$ with positive probability,*

$$\mathbb{P}_B(A) = \frac{\mathbb{P}_A(B)\, \mathbb{P}(A)}{\sum_{i=1}^{\infty} \mathbb{P}(A_i)\, \mathbb{P}_{A_i}(B)}.$$

Of course, we are not satisfied with a theory of conditional probabilities where we restrict the conditioning event to one with positive probability. After all, we are often interested in conditioning on events like "$\{X = a\}$" where X is a continuous random variable and a is a real number. For instance, consider the problem of choosing a point uniformly at random on the unit square. What is the probability that the point has an $x-$coordinate greater than $\frac{1}{2}$ if it lies on the diagonal? We can formally model this problem by taking two independent random variables $X, Y \sim U[0,1]$ and computing $\mathbb{P}\left(X > \frac{1}{2} \mid X = Y\right)$. Of course, the event $X = Y$ has probability zero as can be verified by Tonelli's theorem. Intuitively, we can guess that the answer should be $\frac{1}{2}$ since we should effectively

[1]Sometimes we write $\mathbb{P}(A \mid B)$ instead of $\mathbb{P}_B(A)$ if the event B is actually some long expression like "$\{X = 1\}$".

have a uniform distribution on the diagonal in this setting. It turns out that the answer is not as straightforward as we would expect.

Example 11.1.3. Let

Example 11.1.4. Suppose a random vector (X, Y) has joint probability density function

$$f(x, y) = 3y$$

on the triangle bounded by the lines $y = 0, y = 1 - x$, and $y = 1 + x$. Compute $\mathbb{E}\left(Y \,\middle|\, X \leq \frac{1}{2}\right)$.TODO

Example 11.1.5. Let X and Y be i.i.d. exponentially distributed random variables with mean $\lambda > 0$. Define Z by:

$$Z = \begin{cases} 1 & \text{if } X < Y \\ 0 & \text{otherwise.} \end{cases}$$

Find the conditional mean $\mathbb{E}[X \mid Z = 1]$. TODO

Example 11.1.6. Let X and Y be exponential random variables with parameters 1 and 2 respectively. Another random variable Z is defined as follows.

A coin, with probability p of Heads (and probability $1 - p$ of Tails) is tossed. Define Z by

$$Z = \begin{cases} X & \text{if the coin turns Heads} \\ Y & \text{if the coin turns Tails} \end{cases}$$

Find $P(1 \leq Z \leq 2)$.TODO

11.2. Kolmogorov conditional expectations

The abstract formulation of the conditional expectation was quite unintuitive when it was first put forward by Kolmogorov (1933) add ref as the right way to think about conditioning.

Example 11.2.1. Let X and Y be two random variables with joint probability density function

$$f(x, y) = \begin{cases} 1 & \text{if } -y < x < y, 0 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find the regression equation of Y on X and that of X on Y.TODO

11.3. Regular conditional probabilities

CHAPTER 12

Martingales

CHAPTER 13

Ergodic theory and laws of large numbers

CHAPTER 14

Central limit theorems

Part 3

Statistics

CHAPTER 15

Point estimation

15.1. Population and sample

15.2. Sufficiency and completeness

EXAMPLE 15.2.1. Let X_1 and X_2 be i.i.d. random variables from Bernoulli(θ) distribution. Verify if the statistic $X_1 + 2X_2$ is sufficient for θ.TODO

EXAMPLE 15.2.2. Let Y be a random variable with probability density function

$$f_Y(y \mid \theta) = \begin{cases} \frac{1}{\theta}e^{-y/\theta} & \text{if } y > 0, \\ 0 & \text{otherwise} \end{cases}$$

with $\theta > 0$. Suppose that the conditional distribution of X given $Y = y$ is $N\left(y, \sigma^2\right)$, with $\sigma^2 > 0$. Both θ and σ^2 are unknown parameters. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from the joint distribution of X and Y. Find a nontrivial joint sufficient statistic for $\left(\theta, \sigma^2\right)$. TODO

EXAMPLE 15.2.3. Suppose X takes three values 1,2 and 3 with

$$P(X = k) = \begin{cases} (1-\theta)/2 & \text{if } k = 1 \\ 1/2 & \text{if } k = 2 \\ \theta/2 & \text{if } k = 3 \end{cases}$$

where $0 < \theta < 1$. Suppose that the following random sample of size 10 was drawn from the above distribution :

$$1, 3, 1, 2, 3, 1, 2, 2, 1, 1$$

Find the m.le. of θ based on the above sample. TODO

15.3. Unbiasedness, consistency, and efficiency

EXAMPLE 15.3.1. Let r be the number of successes in n Bernoulli trials with unknown probability p of success. Obtain the minimum variance unbiased estimator of $p - p^2$.

15.4. Properties of extremum estimators

EXAMPLE 15.4.1. Let Y_1, Y_2, Y_3 and Y_4 be four uncorrelated random variables with

$$\mathbb{E}(Y_i) = i\theta, \quad \mathbb{V}\text{ar}(Y_i) = i^2\sigma^2, \quad i = 1, 2, 3, 4,$$

where θ and$\sigma (> 0)$ are unknown parameters. Find the values of c_1, c_2, c_3 andc_4 for which $\sum_{i=1}^4 c_i Y_i$ is unbiased for θ and has least variance.TODO

EXAMPLE 15.4.2. Let $X_1, X_2, \ldots X_n$ be i. i. d. with common density $f(x; \theta)$ given by

$$f(x;\theta) = \frac{1}{2\theta}\exp(-|x|/\theta), \quad -\infty < x < \infty, \quad \theta \in (0, \infty).$$

In case of each of the statistics S and T defined below, decide (a) if it is an unbiased estimator of θ, (b) if it is an MLE for θ and (c) if it is sufficient for θ. Give reasons. TODO

$$S = \frac{1}{n}\sum_{i=1}^n X_i, \quad T = \frac{1}{n}\sum_{i=1}^n |X_i|$$

EXAMPLE 15.4.3. Suppose that $X_1, \ldots, X_n$ is a random sample of size $n \geq 1$ from a Poisson distribution with parameter λ. Find the minimum variance inbiased estimator of $e^{-\lambda}$. TODO

EXAMPLE 15.4.4. Suppose $X_1, \ldots, X_n$ constitute a random sample from a population with density

$$f(x, \theta) = \frac{x}{\theta^2} \exp\left(-\frac{x^2}{2\theta^2}\right), x > 0, \theta > 0.$$

Find the Cramer-Rao lower bound to the variance of an unbiased estimator of θ^2. TODO

EXAMPLE 15.4.5. Let X_1, X_2, X_3 be independent random variables such that X_i is uniformly distributed in $(0, i\theta)$ for $i = 1, 2, 3$. Find the maximum likelihood estimator of θ and examine whether it is unbiased for θ. TODO

EXAMPLE 15.4.6. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from the discrete distribution with joint probability mass function

$$f_{X,Y}(x, y) = \begin{cases} \frac{\theta}{4} & (x, y) = (0, 0) \text{ and } (1, 1), \\ \frac{2-\theta}{4} & (x, y) = (0, 1) \text{ and } (1, 0), \end{cases}$$

with $0 \leq \theta \leq 2$. Find the maximum likelihood estimator of θ. TODO

EXAMPLE 15.4.7. Let $X_1, X_2, \ldots$ be i.i.d. random variables with density $f_\theta(x), x \in \mathbb{R}, \theta \in (0, 1)$ being the unknown parameter. Suppose that there exists an unbiased estimator T of θ based on sample size 1, i. e. $\mathbb{E}_\theta (T(X_1)) = \theta$. Assume that $\mathbb{V}\mathrm{ar}(T(X_1)) < \infty$. (a) Find an estimator V_n for θ based on$X_1, \ldots X_n$ such that V_n is consistent for θ. (b) Let S_n be the MVUE (minimum variance unbiased estimator) of θ based on $X_1, \ldots, X_n$. Show that $\lim_{n\to\infty} \mathbb{V}\mathrm{ar}(S_n) = 0$.TODO

EXAMPLE 15.4.8. Let $X_1, \ldots, X_m$ be a random sample from a uniform distribution on $\{1, 2, \ldots, N\}$ where N is an unknown positive integer. Find the MLE $\widehat{N}$ of N and find its distribution function.TODO

CHAPTER 16

Hypothesis testing

Example 16.0.1. Let Y_1, Y_2, Y_3 and Y_4 be a random sample from a population with probability density function

$$f(y,\theta) = \begin{cases} \left(\frac{1}{2\theta^3}\right) y^2 \exp(-y/\theta) & \text{if } y > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$. Find the most powerful test for testing $H_0 : \theta = \theta_0$ against the hypothesis $H_1 : \theta = \theta_1$, where $\theta_1 > \theta_0$. Is the test uniformly most powerful for $\theta > \theta_0$?TODO

Example 16.0.2. On a particular day let X_1, X_2 and X_3 be the number of boys born before the first girl is born in hospitals 1, 2 and 3 respectively. If the observations are $X_1 = 0, X_2 = 3$ and $X_3 = 2$, find the most powerful test to test the null hypothesis that a girl and a boy are equally likely to be born against the alternative that a girl is less likely to be born than a boy.TODO

Example 16.0.3. Suppose that in 10 tosses of a coin we get 7 heads and 3 tails. Find a test at level $\alpha = 0.05$ to test that the coin is fair against the alternative that the coin is more likely to show up heads. Find the power function of this test.TODO

Example 16.0.4. Consider a possibly unbalanced coin with probability of heads in each toss being p, where$p$ is unknown. Let X be the number of tails before the first head occurs. Find the uniformly most powerful test of level α for testing $H_0 : p = \frac{1}{6}$ against $H_1 : p > \frac{1}{6}$.TODO

Example 16.0.5. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables from the exponential distribution with mean $\theta > 0$. Find the most powerful test for testing $H_0 : \theta = 2$ against $H_1 : \theta = 1$. Find the power of the test.TODO

Example 16.0.6. Consider a population with three kinds of individuals labelled 1,2 and 3. Suppose the proportion of individuals of the three types are given by $f(k,\theta), k = 1,2,3$ where $0 < \theta < 1$ and

$$f(k,\theta) = \begin{cases} \theta^2 & \text{if } k = 1 \\ 2\theta(1-\theta) & \text{if } k = 2 \\ (1-\theta)^2 & \text{if } k = 3 \end{cases}$$

Let $X_1, X_2, \ldots, X_n$ be a random sample from this population. Find the most powerful test for testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta = \theta_1$ $(\theta_0 < \theta_1 < 1)$ TODO

CHAPTER 17

Confidence sets

CHAPTER 18

Linear models

EXAMPLE 18.0.1. A straight line regression $\mathbb{E}(y) = \alpha + \beta x$ is to be fitted using four observations. Assume $\mathbb{V}\mathrm{ar}(y \mid x) = \sigma^2$ for all x. The values of x at which observations are to be made lie in the closed interval $[-1, 1]$. The following choices of the values of x where observations are to be made are available:

(1) two observations each at $x=-1$ and $x=1$,
(2) one observation each at $x=-1$ and $x=1$ and two observations at $x=0$,
(3) one observation each at $x=-1,-\frac{1}{2}, \frac{1}{2}, 1$.

If the interest is to estimate the slope with least variance, which of the above strategies would you choose and why? TODO

EXAMPLE 18.0.2. Let Y_1, Y_2 and Y_3 be uncorrelated random variables with common variance $\sigma^2 > 0$ such that

$$E(Y_1) = \beta_1 + \beta_2, E(Y_2) = 2\beta_1 \text{ and } E(Y_3) = \beta_1 - \beta_2$$

where β_1 and β_2 are unknown parameters. Find the residual (error) sum of squares under the above linear model.TODO

CHAPTER 19

Design of experiments

EXAMPLE 19.0.1. Here is a partial key-block of a 2^4 factorial experiment (with factors A, B, C, D) conducted in two blocks of size 8 each: Partial key-block: $ad \quad bd \quad c \quad \ldots$ Search out the other five treatment combinations for the key-block and also the confounded interaction. Also, give the treatment combination of the second block.TODO

EXAMPLE 19.0.2. For the data collected via a randomized block design with v treatments and b blocks, the following model is postulated:

$$\mathbb{E}\left(y_{ij}\right) = \mu + \tau_i + \beta_j, \quad 1 \leq i \leq v, 1 \leq j \leq b,$$

where t_i and β_j are the effects of the ith treatment and the jth block respectively, and μ is a general mean. For $1 \leq i \leq v$, define $Q_i = T_i - \frac{G}{v}$, where T_i is the total of observations under the ith treatment and $G = \sum_{i=1}^{v} T_i$. Show that

$$E\left(Q_i\right) = \left(b - \tfrac{b}{v}\right)\tau_i, \quad \operatorname{Var}\left(Q_i\right) = \sigma^2\left(b - \tfrac{b}{v}\right),$$
$$\operatorname{Cov}\left(Q_i, Q_j\right) = -\left(\tfrac{b}{v}\right)\sigma^2 \text{ for } i \neq j,$$

where σ^2 is the per observation variance.TODO

EXAMPLE 19.0.3. Consider a randomized block experiment with 4 treatments and 3 replicates (blocks) and let τ_i be the effect of the i th treatment $(1 \leq i \leq 4)$. Find all possible covariances between the least squares estimators of the following treatment contrasts: (a) $\tau_1 - \tau_2$ (b) $\tau_1 + \tau_2 - 2\tau_3$ (c) $\tau_1 + \tau_2 + \tau_3 - 3\tau_4$.TODO

Part 4

Appendices

APPENDIX A

Naive set theory

APPENDIX B

Finite Dimensional Vector Spaces

Definition B.0.1. A *vector space* is a non-empty set V over a field $\mathbb{K}$ together with two binary operations $+ : V \times V \to V$ and $\cdot : \mathbb{K} \times V \to V$, called *vector addition* and *scalar multiplication,* respectively, that satisfy

(1) Closure under vector addition and scalar multiplication: For any $\alpha \in K$ and any $u, v \in V$: $\alpha u \in V$ and $u + v \in V$.
(2) Associativity of vector addition: For any $u, v, w \in V : u + (v + w) = (v + u) + w$
(3) Commutativity of vector addition: For any $u, v \in V : u + v = v + u$
(4) Identity element of vector addition: There exists some element $\mathbf{0} \in V$, called the *zero vector* such that for any $v \in V : \mathbf{0} + v = v$.
(5) Inverse element of vector addition: For any $v \in V$, there exists a vector $-v \in V$ such that $v + (-v) = \mathbf{0}$
(6) Compatibility of scalar multiplication with field multiplication: For any $\alpha, \beta \in \mathbb{K}$ and $u \in V : (\alpha\beta) \cdot u = \alpha \cdot (\beta \cdot u)$.
(7) Identity element of scalar multiplication: There exists some element $1 \in \mathbb{K}$ such that for any $u \in V : 1v = v$.
(8) Distributivity of scalar multiplication with respect to vector addition: For any $\alpha \in \mathbb{K}$ and any $u, v \in V : \alpha \cdot (u + v) = \alpha \cdot u + \alpha \cdot v$
(9) Distributivity of scalar multiplication with respect to field addition: For all $\alpha, \beta \in \mathbb{K}$ and $u \in V : (\alpha + \beta) \cdot u = \alpha \cdot u + \beta \cdot u$.

Exercise B.0.2 (ISI 2013 PSB-1). Let $E = \{1, 2, \ldots, n\}$, where n is an odd positive integer. Let V be the vector space of all functions from E to $\mathbb{R}^3$, where the vector space operations are given by

$$(f + g)(k) = f(k) + g(k), \quad \text{for } f, g \in V, k \in E,$$
$$(\lambda f)(k) = \lambda f(k), \quad \text{for } f \in V, \lambda \in \mathbb{R}, k \in E.$$

(1) Find the dimension of V
(2) Let $T : V \to V$ be the map given by

$$T(f)(k) := \frac{1}{2}(f(k) + f(n + 1 - k)), \ k \in E$$

is linear.
(3) Find the null space of T.

B.1. Determinants

We shall construct the determinant of linear map $T : \mathbb{R}^n \to \mathbb{R}^n$axiomatically. Determinants are typically defined by a given matrix representation $M_T = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \ldots & v_n \\ | & | & & | \end{bmatrix}$ of T with respect to the standard basis on $\mathbb{R}^n$. Here v_i are vectors in $\mathbb{R}^n$ with respect to the standard basis. As we shall later see, the choice of basis does not matter and so we can think of the determinant as acting on the linear map itself rather the matrix representation.

B.1.1. Desiderata of the determinant. What properties should our determinant function have? Why do we even need such a function? What does it actually represent? Alas, the determinant is one of those concepts whose usefulness only becomes apparent *after* you are done constructing it and relaying its properties. In fact, the geometric interpretation of the determinant is often not even covered in a course on Linear Algebra; we shall in fact use the $n-$dimensional Lebesgue measure in Section 7.2.1to understand the geometry of determinants in the main text rather than this appendix.

Thus. without much motivation (for now), we set out on a goal to construct a map $\det : M_{n\times n} \to \mathbb{R}$ (where $M_{n\times n}$ is the vector space of all $n \times n$ real valued matrices) that satisfies the following properties:

(1) **Linearity in each argument:** Let $\{v_i\}_{i=1}^n, u \in \mathbb{R}^n$ and let $\alpha \in \mathbb{R}$. Then

$$\det\left(\begin{bmatrix} | & & | & & | \\ v_1 & \dots & v_k+\alpha u & \dots & v_n \\ | & & | & & | \end{bmatrix}\right) = \det\left(\begin{bmatrix} | & & | & & | \\ v_1 & \dots & v_k & \dots & v_n \\ | & & | & & | \end{bmatrix}\right) + \alpha\det\left(\begin{bmatrix} | & & | & & | \\ v_1 & \dots & u & \dots & v_n \\ | & & | & & | \end{bmatrix}\right).$$

(2) **Preservation under column replacement:** Let $\{v_i\}_{i=1}^n \in \mathbb{R}^n$ and let $\alpha \in \mathbb{R}$. Then

$$\det\left(\begin{bmatrix} | & & | & & | & & | \\ v_1 & \dots & v_l+\alpha v_k & \dots & v_k & \dots & v_n \\ | & & | & & | & & | \end{bmatrix}\right) = \det\left(\begin{bmatrix} | & & | & & | & & | \\ v_1 & \dots & v_l & \dots & v_k & \dots & v_n \\ | & & | & & | & & | \end{bmatrix}\right).$$

(3) **Antisymmetry:** Let $\{v_i\}_{i=1}^n \in \mathbb{R}^n$. Then

$$\det\left(\begin{bmatrix} | & & | & & | & & | \\ v_1 & \dots & v_l & \dots & v_k & \dots & v_n \\ | & & | & & | & & | \end{bmatrix}\right) = -\det\left(\begin{bmatrix} | & & | & & | & & | \\ v_1 & \dots & v_k & \dots & v_l & \dots & v_n \\ | & & | & & | & & | \end{bmatrix}\right).$$

(4) **Normalization:** $\det(I_n) = 1$ where I_n is the $n \times n$ identity matrix.

It turns out that these properties completely characterize the determinant, in that there exists a unique function that possesses these properties. We shall postpone the proof of existence and uniqueness to the end of this section and derive the properties of the determinant function.

B.1.2. Properties of the determinant. All the properties of the determinant can be recovered using the axioms above, although some results are harder to prove than others. We start with a few simple results.

PROPOSITION B.1.1. *If $A \in M_{n\times n}$ has a zero column, then* $\det(A) = 0$.

PROOF. This fact follows from linearity since a zero column can be written as $0.v$ for any $v \in \mathbb{R}^n$. ☐

PROPOSITION B.1.2. *If $A \in M_{n\times n}$ has linearly dependent columns then* $\det(A) = 0$.

PROOF. Let

$$A = \begin{bmatrix} | & & | & & | & & | \\ v_1 & \dots & v_l & \dots & v_k & \dots & v_n \\ | & & | & & | & & | \end{bmatrix}$$

and notice that if the columns of A are linearly dependent then there exist constants $\{\alpha_i\}_{i=1}^n \in \mathbb{R}$,, such that $\sum_{i=1 i\neq l}^n \alpha_i v_i = v_l$. Now, we can write

$$\begin{aligned}
\det(A) &= \det\left(\begin{bmatrix} | & & | & & | & & | \\ v_1 & \dots & \sum_{i=1 i\neq l}^n \alpha_i v_i & \dots & v_k & \dots & v_n \\ | & & | & & | & & | \end{bmatrix}\right) \\
&= \sum_{i\neq l} \alpha_i \det\left(\begin{bmatrix} | & & | & & | & & | \\ v_1 & \dots & v_i & \dots & v_k & \dots & v_n \\ | & & | & & | & & | \end{bmatrix}\right) \\
&= 0
\end{aligned}$$

where in the second equality we have used linearity and in the last equality we have used anti-symmetry and the fact that if $\det(B) = -\det(B)$ then $\det(B) = 0$. ☐

Now we have sufficiently many properties to describe the determinants of the most elementary types of matrices. Recall that the elementary linear maps $T:\mathbb{R}^n\to\mathbb{R}^n$ are

(1) **Row scaling:** $T(x_{1,},\ldots,x_k,\ldots,x_n)^T=(x_1,\ldots cx_k,\ldots x_n)^T$ for some $c\in\mathbb{R}$. The matrix with respect to the standard basis of this transformation is $\operatorname{diag}(1,1,\ldots,c,\ldots,1)$. By linearity and the normalization property of determinants, $\det(T)=c$. More generally, for any diagonal matrix $D=\operatorname{diag}(a_1,a_2,\ldots,d_n)$ we have that $\det(D)=\prod_{i=1}^n a_i$ by linearity and normalization. T is invertible with T^{-1} being represented by $\operatorname{diag}\left(1,1,\ldots,\frac{1}{c},\ldots,1\right)$. The tranpose of $T^T=T$
(2) **Row switching:** $T(x_1,\ldots x_l,\ldots x_k,\ldots,x_n)^T=(x_1,\ldots,x_k,\ldots,x_l,\ldots,x_n)^T$ for any $1\le l\le k\le n$. The matrix of this operator wrt to the standard basis is $[e_1,\ldots e_k,\ldots,e_l,\ldots,e_n]$ where e_i is the ith standard basis element. Clearly, $\det(T)=-1$ by antisymmetry. Again, T is invertible and is its own inverse. The tranpose of T is T itself.
(3) **Row replacement**$T(x_1,\ldots x_l,\ldots x_k,\ldots,x_n)^T=(x_1,\ldots,x_l+cx_k,\ldots,x_k,\ldots,x_n)^T$for any $1\le l<l\le n$ and any $c\in\mathbb{R}$. The matrix representation of this map is given by $[e_1,\ldots,e_l,\ldots,e_k+ce_l,\ldots,e_n]$ with determinant $\det(T)=1$ by linearity and Proposition B.1.2. This function is also clearly invertible and its inverse is an operations of the same type. The tranpose is also an operation of the same type.

PROPOSITION B.1.3. *Let $A\in M_{n\times n}$ and let $T\in M_{n\times n}$ be the standard basis representation of an elementary row operation as above. Then*

$$\det(AT)=\det(A)\det(T).$$

PROOF. First consider the case when T is the scaling operator for the kth coordinate. Then, letting $A=\begin{bmatrix} | & & | & & | & & | \\ v_1 & \ldots & v_l & \ldots & v_k & \ldots & v_n \\ | & & | & & | & & | \end{bmatrix}$, we have $ST=\begin{bmatrix} | & & | & & | & & | \\ v_1 & \ldots & v_l & \ldots & cv_k & \ldots & v_n \\ | & & | & & | & & | \end{bmatrix}$ and so

$$\det(AT)=c\det(A)=\det(A)\det(T).$$

Next, let T is the row switching operator. Then $AT=\begin{bmatrix} | & & | & & | & & | \\ v_1 & \ldots & v_k & \ldots & v_l & \ldots & v_n \\ | & & | & & | & & | \end{bmatrix}$ and by antisymmetry

$$\det(AT)=-\det(A)=\det(A)\det(T).$$

Finally, if T is the row replacement operator, then $AT=\begin{bmatrix} | & & | & & | & & | \\ v_1 & \ldots & v_l & \ldots & v_k+cv_l & \ldots & v_n \\ | & & | & & | & & | \end{bmatrix}$ and

$$\det(AT)=\det(A)+c\det\left(\begin{bmatrix} | & & | & & | & & | \\ v_1 & \ldots & v_l & \ldots & v_l & \ldots & v_n \\ | & & | & & | & & | \end{bmatrix}\right)=\det(A)$$

by linear dependence. □

Note that by our work on simultaneous linear equations and their solutions, every invertible matrix $A\in M_{n\times n}$ can be reduced to the identity matrix by multiplication with elementary matrices from the left i.e. $I=T_1T_2..T_kA$. Then,

$$A=T_k^{-1}T_{k-1}^{-1}\ldots T_1^{-1}$$

where each T_i^{-1} is an elementary operation and so $\det(A)=\prod_{k=1}^n\det\left(T_i^{-1}\right)$ by inducting on Proposition B.1.3.

PROPOSITION B.1.4. *Let $A,B\in M_{n\times n}$. Then,*

$$\det(AB)=\det(BA)=\det(A)\det(B).$$

Proof. First suppose that A, B are both invertible. Then A and B are products of elementary transformations and so $\det(AB) = \det(A)\det(B)$. If either, A or B is not invertible, then neither AB nor BA are invertible since neither operator is either injective or surjective (by rank-nullity) but have the same domains. Suppose without loss of generality that B is not invertible. Then by the rank nullity theorem, $\ker(B)$ is trivial, which is equivalent to the fact B has linearly dependent columns and so $\det(B) = 0$. The same argument applied to AB or BA shows that $\det(AB) = \det(BA) = 0$. This completes the proof. □

Corollary B.1.5. *A matrix $A \in M_{n\times n}$ is invertible if and only if* $\det(A) \neq 0$.

Proof. We have already shown that if A is not invertible, then A has linearly dependent columns and so $\det(A) = 0$. Conversely, suppose that $\det(A) = 0$ and A is invertible. Then $1 = \det(I) = \det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right) = 0$ which is a contradiction. □

Corollary B.1.6. *Let $A \in M_{n\times n}$ be invertible. Then*

$$\det\left(A^{-1}\right) = \frac{1}{\det(A)}.$$

Proof. Note that $1 = \det(I) = \det\left(AA^{-1}\right) = \det(A)\det\left(A^{-1}\right)$. □

Corollary B.1.7. *If $A, B \in M_{n\times n}$ are similar matrices, then*

$$\det(A) = \det(B).$$

Proof. Let $P \in M_{n\times n}$ be invertible such that

$$A = PBP^{-1}.$$

Then,

$$\begin{aligned}\det(A) &= \det\left(PBP^{-1}\right)\\ &= \det(P)\det(B)\det\left(P^{-1}\right)\\ &= \det(P)\det(B)\frac{1}{\det(P)}\\ &= \det(B).\end{aligned}$$

□

Note that this tells us that the determinant of a matrix doesn't change under a change of basis. In other words, we can think of the determinant as acting on the operator that the matrix represents rather than the matrix itself.

Proposition B.1.8. *For any $A \in M_{n\times n}$*

$$\det(A) = \det\left(A^T\right).$$

Proof. If A is not invertible then A^T is also not invertible and so both determinants are zero. On the other hand, if A is invertible then $A = T_1T_2\ldots T_k$ for some k where T_i are elementary matrices. Then, $A^T = T_k^T T_{k-1}^T \ldots T_1^T$, where the determinant of T_i is the same as the determinant of T_i^T. The result then follows by Proposition B.1.4. □

B.1.3. Existence and uniqueness.

B.1.4. The cofactor expansion.

EXAMPLE B.1.9 (ISI 2023 PSB 1). Let $A_n = ((a_{ij}))$ be the $n \times n$ matrix defined by

$$a_{ij} = \begin{cases} 0 & \text{if } |i-j| > 1 \\ 1 & \text{if } |i-j| = 1 \\ 2 & \text{if } i = j \end{cases}$$

What is the determinant of A_n for $n \geq 1$?TODO

EXAMPLE B.1.10. Is the following system of equations always consistent for real k ? Justify your answer.

$$\begin{aligned} x + y + kz &= 2, \\ 3x + 4y + 2z &= k, \\ 2x + 3y - z &= 1. \end{aligned}$$

Find the value of k for which this system admits more than one solution? Express the general solution for the system of equations for this value of k.TODO

EXAMPLE B.1.11. Let A be a $n \times n$ upper triangular matrix such that $AA^T = A^T A$. Show that A is a diagonal matrix.TODO

EXAMPLE B.1.12. Let A be a $n \times n$ orthogonal matrix, where n is even and suppose \$|A|=\$ -1, where $|A|$ denotes the determinant of A. Show that $|I - A| = 0$, where I denotes the $n \times n$ identity matrix.TODO

EXAMPLE B.1.13. Let A and B be two invertible $n \times n$ real matrices. Assume that $A + B$ is invertible. Show that $A^{-1} + B^{-1}$ is also invertible.TODO

EXAMPLE B.1.14. Let A be a 2×2 matrix with real entries such that$A^2 = 0$. Find the determinant of $I + A$ where I denotes the identity matrix.TODO

EXAMPLE B.1.15. Let A and B be $n \times n$ real matrices such that $A^2 = A$ and $B^2 = B$. Suppose that $I - (A + B)$ is invertible. Show that $\operatorname{rank}(A) = \operatorname{rank}(B)$.TODO

EXAMPLE B.1.16. 1. Let

$$A = \frac{1}{3} \begin{pmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{pmatrix}.$$

Which of the following statements are false. In each case, justify your answer. (a) A has only one real eigenvalue. (b) $\operatorname{Rank}(A) = \operatorname{Trace}(A)$. (c) Determinant of A equals the determinant of A^n for each integer $n > 1$.TODO

B.2. Eigenvalues and eigenvectors

APPENDIX C

Basic combinatorics

C.1. Binomial and multinomial coefficients

The following is an *algebraic* proof of Pascal's rule, a basic combinatorial identity. In the remark that follows, we provide a more combinatorial interpretation of the identity. Throughout this appendix, we will try to provide both algebraic and combinatorial arguments for results, unless the algebraic arguments are too cumbersome.

Proposition C.1.1 (Pascal's rule). *For $n \geq k \geq 1$*

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}.$$

Proof. Observe that

$$\begin{aligned}
\binom{n-1}{k-1} + \binom{n-1}{k} &= \frac{(n-1)!}{(k-1)!\,(n-k)!} + \frac{(n-1)!}{k!\,(n-k-1)!} \\
&= (n-1)!\left[\frac{1}{(k-1)!\,(n-k)!} + \frac{1}{k!\,(n-k-1)!}\right] \\
&= (n-1)!\left[\frac{k}{k!\,(n-k)!} + \frac{n-k}{k!\,(n-k)!}\right] \\
&= (n-1)!\frac{n}{k!\,(n-k)!} \\
&= \frac{n!}{k!\,(n-k)!} \\
&= \binom{n}{k}.
\end{aligned}$$

□

Remark. The combinatorial idea for the above result is simple. Suppose you want to select k items from a collection of n items, without any consideration for order. How does one do so? . First we arbitrarily pick some item and label it x. How many ways are ways are there to pick items excluding x? Well, there are $\binom{n-1}{k}$ ways to pick k items if we explicitly exclude x. What if we insist on including x? Then we have to pick $k-1$ items from $n-1$ remaining total items, that is $\binom{n-1}{k-1}$.

Example C.1.2 (ISI 2021 PSA-11). We can use Pascal's formula to recover a more succinct formula for the expression

$$\prod_{i=1}^{n}\left(\binom{n}{i} + \binom{n}{i-1}\right).$$

We can apply Pascal's rule to each term in the product so that

$$\begin{aligned}\prod_{i=1}^{n}\left(\binom{n}{i}+\binom{n}{i-1}\right) &= \prod_{i=1}^{n}\binom{n+1}{i} \\ &= k\prod_{i=1}^{n}\binom{n}{i}\end{aligned}$$

where $k = \frac{(n+1)^n}{n!}$.

Pascal's rule helps us establish the binomial theorem.

THEOREM C.1.3 (Binomial theorem). *Let $n \geq 1$ be an integer and let $x, y \in \mathbb{R}$. Then,*

$$(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}.$$

PROOF. The identity is easily verified for $n = 1$ and $n = 2$. Supopose that it holds for $n-1$ and observe

$$\begin{aligned}(x+y)^n &= x(x+y)^{n-1} + y(x+y)^{n-1} \\ &= x\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-1-k} + y\sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-1-k} \\ &= \sum_{k=0}^{n-1}\binom{n-1}{k}x^{k+1}y^{n-1-k} + \sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-k} \\ &= \sum_{k=1}^{n}\binom{n-1}{k-1}x^k y^{n-k} + \sum_{k=0}^{n-1}\binom{n-1}{k}x^k y^{n-k} \\ &= \binom{n-1}{n-1}x^n + \binom{n-1}{0}y^n + \sum_{k=1}^{n-1}\binom{n}{k}x^k y^{n-k} \\ &= \binom{n}{n}x^n + \binom{n}{0}y^n + \sum_{k=1}^{n-1}\binom{n}{k}x^k y^{n-k} \\ &= \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k}\end{aligned}$$

where in the third line we used a change of variables, and in the fourth we used Pascal's rule. □

REMARK. The combinatorial argument is again simple. We know that the coefficient for $x^k y^{n-k}$ in the expansion of $(x+y)^n$ involves choosing k out of n available x's and the number of ways to do that is $\binom{n}{k}$.

EXAMPLE C.1.4 (ISI 2016 PSA 1). How many terms in the binomial expansion of $\left(3x^2 + \frac{1}{x}\right)^5$ are independent of x? Note that the binomial expansion of this expression is of the form

$$\begin{aligned}\left(3x^2 + \frac{1}{x}\right)^5 &= \sum_{i=0}^{5}\binom{5}{i}3x^{2i}x^{i-5} \\ &= \sum_{i=0}^{5}\binom{5}{i}3x^{3i-5}.\end{aligned}$$

Note that $3i - 5 = 0 \implies i = \frac{5}{3}$ which is not an integer. Thus no terms are independent of x.

The binomial theorem gives us a simple formula for the sum of binomial coefficients.

Corollary C.1.5. *Let $n \geq 1$ be a fixed integer. Then,*

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n.$$

Proof. Let $x = y = 1$ in the binomial theorem. $\square$

Remark. Think about a set with n elements. You want to construct all possible subsets. You know the number of all possible subsets is 2^n. How do you count all possible subsets of a finite set? Well, order doesn't matter in a set so all you have to do is count the empty set, the sets with one element (singletons), the set with two elements, those with three and so on.

Example C.1.6 (ISI 2013 PSB 3). Let $S = \{1, 2, \ldots, n\}$. How many ways can we choose two subsets $B \subseteq A \subseteq S$ such that $B \neq \emptyset$? How many subsets $B \subsetneq A \subseteq S$ can we choose? For the first question, we know that for each subset A of size k we can select $2^k - 1$ nonempty subsets. Further, there are $\binom{n}{k}$ subsets of size k and so the number of such sets is $\sum_{k=1}^{n} \binom{n}{k} \left(2^k - 1\right)$.

There are a number of interesting identities regarding binomial coefficients which have simple combinatorial interpretations. We list a few here.

Proposition C.1.7. *Let n be a non-negative integer. Then,*

$$\binom{2n}{2} = 2\binom{n}{2} + n^2.$$

Proof. Note that

$$\begin{aligned}
\binom{2n}{2} &= \frac{(2n)!}{2!\,(2n-2)!} \\
&= \frac{2n\,(2n-1)}{2!} \\
&= 2n^2 - n \\
&= n(n-1) + n^2 \\
&= 2\binom{n}{2} + n^2.
\end{aligned}$$

$\square$

Remark. Suppose you have $n$distinguihsable red balls and n distinguishble green balls. How many ways are there to choose two balls from this collection? Well, you can get one red and one green and there are n^2 ways of getting such pairs. Alternatively, you can get two of red or two of green and there are $\binom{n}{2}$ possible ways to select two reds (or two greens).

Proposition C.1.8. *Let $m, n \geq 1$ be positive integers. Then,*

$$\binom{m+n}{2} - \binom{m}{2} - \binom{n}{2} = mn.$$

Proof. Note that

$$\begin{aligned}
\binom{m+n}{2} - \binom{m}{2} - \binom{n}{2} &= \frac{(m+n)!}{2!\,(m+n-2)!} - \frac{m!}{2!\,(m-2)!} - \frac{n!}{2!\,(n-2)!} \\
&= \frac{(m+n)\,(m+n-1) - m(m-1) - n(n-1)}{2} \\
&= \frac{m^2 + n^2 + 2mn - m^2 + m - n^2 + n}{2} \\
&= mn.
\end{aligned}$$

$\square$

REMARK. Suppose you want to count the number of ways in which you can choose 2 items from m distinguishable green balls and n distinguishable red balls and you want one of each. Then you can count the number of ways you can choose two of those in aggregate and subtract out the number of ways in which you could choose both of one color. On the other hand there are clearly mn such ways (by the product rule if you will).

PROPOSITION C.1.9. *Let $n \geq 1$ be an integer,. Then,*

$$\sum_{i=1}^{n} i \begin{pmatrix} n \\ i \end{pmatrix} = n2^{n-1}.$$

PROOF. Note that by the binomial theorem,

$$(1+x)^n = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x^k.$$

Taking derivatives on both sides, we have that

$$n(1+x)^{n-1} = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} kx^{k-1}.$$

Letting $x = 1$ yields the result. □

REMARK. Fill later

PROPOSITION C.1.10. *Let $n \geq 1$ be a fixed integer. Then,*

$$\sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix}^2 = \begin{pmatrix} 2n \\ n \end{pmatrix}.$$

PROOF. Consider the algebraic identity

$$[(1+x)^n]^2 = (1+x)^{2n}.$$

We can expand the left hand side as

$$\left(\sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x^k\right)^2 = \sum_{k=0}^{n}\sum_{j=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} \begin{pmatrix} n \\ j \end{pmatrix} x^{j+k}.$$

The coefficient to the term x^n is $\sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} \begin{pmatrix} n \\ n-k \end{pmatrix} = \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix}^2$.On the other hand, a simple application of binomial theorem tells us that the coefficient of x^n on the right hand side is $\begin{pmatrix} 2n \\ n \end{pmatrix}$ which completes the proof. □

REMARK. Suppose you have n distinguishable red balls and $n$distinguishable green balls. How many ways can you collect n items from this group? Well the answer is clearly $\begin{pmatrix} 2n \\ n \end{pmatrix}$, but we can break this up by considering that we can select 1 red ball and $n-1$ green balls, or 2 red balls and $n-2$ green balls...

The above result is a special case of Vandermonde's identity, which was known to mathematicians as early as the 14th century.

PROPOSITION C.1.11. *Let $m, n,$ and r be non-negative integers. Then,*

$$\begin{pmatrix} m+n \\ r \end{pmatrix} = \sum_{k=0}^{r} \begin{pmatrix} m \\ k \end{pmatrix} \begin{pmatrix} n \\ r-k \end{pmatrix}.$$

PROOF. The algebraic proof is similar to the one for the special case in Proposition C.1.10. The combinatorial argument is more illuminating and significantly less cumbersome. Consider a committee consisting of m men and n women. We want to form a subcommittee of r members. Clearly the number of such subcommittes is $\binom{m+n}{r}$. One way to form such a committee is to consider a committee with one man and $r-1$ women, or two men and $r-2$ women, or ... □

PROPOSITION C.1.12. *Let $n \geq 1$ be an integer. Then,*

$$\sum_{i=1}^{n} i\,(n+1-i) = \binom{n+2}{3}.$$

PROOF. For $n = 1$ both the left and right hand sides of the identity above are 1 and so the base case holds. Suppose the identity holds for n. Then,

$$\begin{aligned}
\sum_{i=1}^{n+1} i\,(n+1+1-i) &= \sum_{i=1}^{n+1} i\,(n+1-i) + \sum_{i=1}^{n+1} i \\
&= \sum_{i=1}^{n} i\,(n+1-i) + \binom{n+2}{2} \\
&= \binom{n+2}{3} + \binom{n+2}{2} \\
&= \binom{n+2}{3}
\end{aligned}$$

where the second equality is the Gaussian formula for the sum of consecutive natural numbers, the third is the induction hypothesis, and the last is Pascal's rule. □

PROPOSITION C.1.13. *Let $n \geq 2$ be an integer. Then*

$$\sum_{k=2}^{n} \binom{k}{2} = \binom{n+1}{3}.$$

PROOF. $n = 2$ the result follows easily since $\binom{2}{2} = \binom{3}{3} = 1$. For $n = 3$ we have $\binom{2}{2} + \binom{3}{2} = 4 = \binom{4}{3}$. Now suppose the result holds for $n-1$ and note that

$$\begin{aligned}
\sum_{k=2}^{n} \binom{k}{2} &= \sum_{k=2}^{n-1} \binom{k}{2} + \binom{n}{2} \\
&= \binom{n}{3} + \binom{n}{2} \\
&= \binom{n+1}{3}
\end{aligned}$$

where the second equality uses the induction hypothesis and the last uses Pascal's rule. □

We can generalize the notion of binomial coefficients to that of multinomial coefficients with the following result.

PROPOSITION C.1.14. *Let $r_1, r_2, \ldots, r_k$ be positive integers and let $n = \sum_{i=1}^{k} r_i$. Then the number of ways to split a set of size n into k ordered subsets where the ith subset contains r_i elements is given by*

$$\binom{n}{r_1, \quad r_2, \quad \ldots \quad r_k} := \frac{n!}{r_1! r_2! \ldots r_k!}.$$

PROOF. Note that first we choose r_1 items from n items, then r_2 items from $n - r_1$ remaining items, then r_3 from the remaining $n - r_1 - r_2$ and so on, yielding

$$\begin{aligned}\begin{pmatrix} & n & & \\ r_1, & r_2, & \dots & r_k \end{pmatrix} &= \prod_{i=1}^{k} \begin{pmatrix} n - \sum_{j=0}^{i-1} r_j \\ r_i \end{pmatrix} \\ &= \prod_{i=1}^{k} \frac{\left(n - \sum_{j=0}^{i-1} r_j\right)!}{r_i! \left(n - \sum_{j=0}^{i} r_j\right)!} \\ &= \frac{n!}{r_1! r_2! \dots r_k!}.\end{aligned}$$

The last equality follows because there's a "telescoping" product and so the all $\left(n - \sum_{j=0}^{i-1} r_j\right)!$ terms except the first (which is is $n!$) cancel. $\square$

This gives us the multinomial theorem.

THEOREM C.1.15 (Multinomial theorem). *Let $m, n \geq 1$ be fixed integers. Then, for $x_1, x_2, \dots, x_m \in \mathbb{R}$*

$$\left(\sum_{i=1}^{m} x_i\right)^n = \sum_{k_1+\ldots+k_m=n, k_i \geq 0} \begin{pmatrix} & n & & \\ k_1, & k_2, & \dots & k_m \end{pmatrix} \prod_{i=1}^{m} x_i^{k_i}.$$

PROOF. For $m = 1$ the result is trivial. For $m = 2$, we shall verify that it is in fact the binomial theorem. To see this, note that for $m = 2$, the binomial theorem implies that

$$\begin{aligned}(x_1 + x_2)^n &= \sum_{k=0}^{n} \begin{pmatrix} n \\ k \end{pmatrix} x_1^k x_2^{n-k} \\ &= \sum_{k=0}^{n} \begin{pmatrix} n \\ k, n-k \end{pmatrix} x_1^k x_2^{n-k} \\ &= \sum_{k_1+k_2=n, k_i \geq 0} \begin{pmatrix} n \\ k_1, k_2 \end{pmatrix} x_1^{k_1} x_2^{k_2}\end{aligned}$$

where the second equality uses Proposition C.1.14. To extend this result to arbitrary m, we assume that it holds for $m - 1$ and then note that

$$\begin{aligned}\left(\sum_{i=1}^{m-1} x_i + x_m\right)^n &= \sum_{K+k_m=n, K, k_m \geq 0} \begin{pmatrix} n \\ K, k_m \end{pmatrix} \left(\sum_{i=1}^{m-1} x_i\right)^K x_m^{k_m} \\ &= \sum_{K+k_m=n, K, k_m \geq 0} \begin{pmatrix} n \\ K, k_m \end{pmatrix} \sum_{k_1+\ldots+k_{m-1}=K, k_i \geq 0} \begin{pmatrix} K \\ k_1, \dots, k_{m-1} \end{pmatrix} \prod_{i=1}^{m} x_i^{k_i} \\ &= \sum_{K+k_m=n, K, k_m \geq 0} \sum_{k_1+\ldots+k_{m-1}=K, k_i \geq 0} \begin{pmatrix} n \\ K, k_m \end{pmatrix} \begin{pmatrix} K \\ k_1, \dots, k_{m-1} \end{pmatrix} \prod_{i=1}^{m} x_i^{k_i} \\ &= \sum_{K+k_m=n, K, k_m \geq 0} \sum_{k_1+\ldots+k_{m-1}=K, k_i \geq 0} \begin{pmatrix} & n & & \\ k_1, & k_2, & \dots & k_m \end{pmatrix} \prod_{i=1}^{m} x_i^{k_i} \\ &= \sum_{k_1+\ldots+k_m=n, k_i \geq 0} \begin{pmatrix} & n & & \\ k_1, & k_2, & \dots & k_m \end{pmatrix} \prod_{i=1}^{m} x_i^{k_i}\end{aligned}$$

which completes the proof.

There is an obvious generalization of Vandermonde's identity that should occur to you at this point. $\square$

PROPOSITION C.1.16 (Generalized Vandermonde identity). *Let $\{n_i\}_{i=1}^k$, r be non-negative integers. Then,*

$$\binom{\sum_{i=1}^k n_i}{r} = \sum_{i_1+i_2+\ldots+i_k=r} \prod_{j=1}^k \binom{n_j}{i_j}.$$

C.1.0.1. *Stars and bars.* Sums like $\sum_{i_1+i_2+\ldots+i_k=r}$ should trouble you, since it's not obvious how many terms are in such a sum. Fortunately, we have the tools we need to be able to answer such questions. To do so, we make an analogy with a different type of question: suppose you have k urns and r indistinguishable balls. How many ways could you allocate balls to urns? William Feller's famous textbook on probability provided an ingenious framework to think about this problem: suppose you have $k+1$ bars and r stars, you could then fix two bars at two ends and put all the stars and remaining bars in the middle. The gap between bars would act as urns and the stars would be balls. Suppose you have four bars and ten stars, one arrangment could be like $|**|*****|**|$. This corresponds to the solution where there are 3 urns and the first urn has 2 balls, the second has 5 balls, and the last again has 2 balls. How many other possibilities are there? Well we can rearrange the bars and stars in the interior of by looking at the number of slots in the interior (in this case 12) and choosing the slots for either the bars or the stars (in this case 2 or 10). So the answer is $\binom{12}{2} = \binom{12}{10}$. More generally, there are $k-1$ bars in the interior along with r stars, leading to $\binom{r+k-1}{k-1} = \binom{r+k-1}{r}$ ways to arrange the interior bars and stars. Equivalently, there are $\binom{r+k-1}{k}$ ways to put r balls in k urns. This simple idea has remarkable power in its ability to resolve combinatorial problems. For instance, if we apply the restriction that each urn needs to have at least 1 ball, then our result becomes $\binom{k-1+(r-k)}{r-k} = \binom{r-1}{k-1}$.

EXAMPLE C.1.17 (ISI 2018 PSA 13 (variant)). How many functions $f : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, m\}$ are strictly increasing? How many are non-decreasing? To answer questions like this, we can use some of the ideas from the discussion above. First, you should notice that there are $\binom{m}{n}$ ways to choose the range of the function (it has to be injective after all), and exactly one way for each of those ways to arrange them in increasing order. This gives us the total number of strictly increasing functions. To count non-decreasing functions, we

EXAMPLE C.1.18 (ISI 2016 PSA 21). Let $A = \{1, 2, \ldots n, \ldots, m\}$. How many functions $f : A \to A$ are there such that $f(1) < f(2) < \ldots, < f(n)$? Well there are $\binom{m}{n}$ ways of selecting the strictly increasing elements and $(m-n)^{(m-n)}$ ways of choosing how the remaining elements are arranged. Thus the total number of of such functions is $\binom{m}{n}(m-n)^{(m-n)}$.

EXAMPLE C.1.19 (ISI 2016 PSA 11). The number of ordered pairs $(a, b) \in \mathbb{N}^2$ such that $a+b \leq n$ where $n \in \mathbb{N}$ is $\sum_{i=2}^n \binom{i-1}{2-1} = 1+2+3+\ldots+n-1 = \frac{n(n-1)}{2}$

C.2. Inclusion-exclusion and its consequences

LEMMA C.2.1. *Let A denote the union of sets $A_1, A_2, \ldots, A_n$, all of which are subsets of some ambient set $\mathcal{X}$. Then,*

$$\mathbb{1}_A = \sum_{i=1}^n (-1)^{i-1} \sum_{J\subset\{1,2,\ldots n\}, |J|=i} \mathbb{1}_{\bigcap_{j\in J} A_j}. \tag{25}$$

PROOF. Consider the function

$$g(x) = \prod_{i=1}^{n} \left(\mathbb{1}_A(x) - \mathbb{1}_{A_i}(x)\right).$$

We claim that $g(x) = 0$ for all $x \in \mathcal{X}$. To see this, notice that if $x \in A_i$ for any $1 \leq i \leq n$ then that particular factor is zero. Conversely if $x \notin A$ then all the factors are zero. Rearranging the equation $g(x) = 0$ and using Fact (2.1.2)yields the result. □

THEOREM C.2.2 (Inclusion-Exclusion). *Let $A_1, A_2, \ldots, A_n$ be finite sets and let $A = \bigcup A_i$. Then,*

$$|A| = \sum_{i=1}^{n} (-1)^{i-1} \sum_{J \subset \{1,2,\ldots n\}, |J|=i} |\bigcap_{j \in J} A_j|.$$

PROOF. We are going to cheat here and use measure theory. We can take the integral with respect to the counting measure on (25)and recover the result by linearity. □

The inclusion-exclusion principle is very useful in counting the number of *derangements* of a given set. A derangement of a finite set A is a permutation on that set with no fixed points.

PROPOSITION C.2.3. *Let A be a finite set such that $|A| = n$. The number of derangements $\sigma : A \to A$ is given by*

$$!n := n! \sum_{i=0}^{n} \frac{(-1)^i}{i!}.$$

PROOF. Without loss of generality, we can assume that $A = \{1, 2, \ldots, n\}$. Let $S_k := \{\sigma \in \text{Perm}(A) \mid \sigma(k) = k\}$ for $1 \leq k \leq n$. That is, each S_k fixes k and may or may not fix any other elements. Then, for any $J \subset A$, we have that

$$\sum_{J \subset \{1,2,\ldots n\}, |J|=i} |\bigcap_{j \in J} S_j| = \binom{n}{i} (n-i)!$$

because the intersection of any i elements of $\{S_k\}_{1 \leq k \leq n}$ consists of permutations which fix at least i points and there are $\binom{n}{i}$ ways to pick i fixed points and $(n-i)!$ ways to permute all the other elements. Then, the number of ways in which you can have at least one fixed point is given by the inclusion-exclusion formula

$$\begin{aligned}
|\bigcup_{i=1}^{n} S_i| &= \sum_{i=1}^{n} (-1)^{i-1} \sum_{J \subset \{1,2,\ldots n\}, |J|=i} |\bigcap_{j \in J} A_j| \\
&= \sum_{i=1}^{n} (-1)^{i-1} \binom{n}{i} (n-i)! \\
&= \sum_{i=1}^{n} (-1)^{i-1} \frac{n!}{i!(n-i)!} (n-i)! \\
&= n! \sum_{i=1}^{n} \frac{(-1)^{i-1}}{i!}.
\end{aligned}$$

TABLE 1. Values of derangements

n	$n!$	$!n$
1	1	0
2	2	1
3	6	2
4	24	9
5	120	44
6	720	265
7	5040	1854
8	40320	14833
9	362880	133496
10	3628800	1334961

The number of derangements is then simply the difference between the total number of permutations and the number of permutations that has at least one fixed point i.e.

$$\begin{aligned} !n &= n! - n! \sum_{i=1}^{n} \frac{(-1)^{i-1}}{i!} \\ &= n! \left(1 - \sum_{i=1}^{n} \frac{(-1)^{i-1}}{i!}\right) \\ &= n! \left(1 + \sum_{i=1}^{n} \frac{(-1)^{i}}{i!}\right) \\ &= n! \sum_{i=0}^{n} \frac{(-1)^{i}}{i!}. \end{aligned}$$

□

REMARK. There's a recursive formulation for counting derangements as well. To see this, first think about Finish later

COROLLARY C.2.4. *Let A be a finite set with cardinality n. The number of permutations $\sigma : A \to A$ with exactly k fixed points, where $1 \leq k \leq n$. is*

$$D_{n,k} := \binom{n}{k} !\,(n-k)\,.$$

It's worth noting down the values of small derangements so that one isn't forced to compute these when solving problems (much in the way we often memorize small factorials).

EXAMPLE C.2.5 (JAM 2022 P-54). Suppose that five men go to a restaurant together and each of them orders a dish that is different from the dishes ordered by the other members of the group. However, the waiter serves the dishes randomly. Then what is the number of ways in which exactly one of them gets the dish he ordered? The answer is $D_{5,1} = \binom{5}{1} !4 = 5 \times 9 = 45$.

We can generalize the idea of derangements by considering permutations that don't have *cycles.* A cycle is itself a sort of generalization of a fixed point. A permutation f on a finite set A of size n is said to have a $k-$cycle if there exists some $x \in A$ such that $f^{k}(x) = x$ where the exponent is denoting repeated composition rather than multiplication. Of course, here k represents the *smallest* positive integer such that the equality holds true. It should be clear that it also holds true for any *multiple* of k. To understand cycles in permutations, it's useful to adopt a notation for describing permutations that can help clarify their cyclic structure. This is the so called *cyclic notation*. For example, we can write a permutation on $\{1,2,3,4,5\}$

$$f = (351)\,(24)$$

which basically tells us that the permutation consists of two cycles (351) and (24). The cycles here are *ordered*, in that the first cycle represents the fact that $f(3) = 5, f(5) = 1$ and $f(1) = 3$ and the second cycle tells us that $f(2) = 4$ and $f(4) = 2$. Every permutation can be decomposed into cycles, and *non-cyclic* permutations are those that consist of only the trivial cycle: for instance, a non-cyclic permutation on $\{1,2,3,4,5\}$ would be the permutation (14235). Note that two cycles of length k $(x_1x_2\dots x_k)$ and $(y_1y_2\dots y_k)$ are equivalent if there exists some $p \in \mathbb{N}$ such that $y_i = f^p(x_i)$. This defines an equivalence relation (as should be clear), so we can talk about *equivalence classes* of cycles, denoted $[(x_1x_2\dots x_k)]$. Such an equivalence class consists of exactly k distinct members since $f^k(x_i) = x_i$. Thus the total number of cycles of length k that can be formed out of k fixed elements is $\frac{k!}{k}$. We can use this to count the total number of permutations without any cycles using an inclusion-exclusion argument analagous to the one used for counting derangements in Proposition C.2.3.

Example C.2.6 (ISI 2021 PSA 14). What is the total number of permutations $\sigma : \{1,2,\dots,6\} \to \{1,2,\dots,6\}$ such that $\sigma(\sigma(i)) \neq i$ for any $i \in \{1,2,\dots,6\}$?TODO

C.3. Miscellaneous problems

Example C.3.1 (ISI 2019 PSA 11). What are the total number of divisors of $2^5 5^3 11^4$ that are perfect squares? The prime square divisors are $2^2, 5^2$, and 11^2, where the first and last divisors appear twice. Thus the first product can appear at most twice, the second at most once, and the third at most twice in any square divisor and so the number is $(2+1)(1+1)(2+1) = 18$. More generally, for any positive integer n with prime factorization $n = \prod_{i=1}^k p_i^{r_i}$, we can decompose the product into the largest square divisor

$$n = \prod_{i=1}^{k} \left(p_i^2\right)^{\lfloor \frac{r_i}{2} \rfloor} \prod_{i=1}^{k} p_i^{\mathbb{1}\left\{\lfloor \frac{r_i}{2} \rfloor \neq \frac{r_i}{2}\right\}}$$

where the second product is square-free and so the number of square divisors is given $\prod_{i=1}^k \left(\lfloor \frac{r_i}{2} \rfloor + 1\right)$.

Example C.3.2 (ISI 2019 PSA 18). Draw one observation N at random from the set $\{1,2,\dots,100\}$. What is the probability that the last digit of N^2 is 1 ? Well note that only if the units digit of N is 1 or 9 does the units digit of N^2 equal 1, which tells us that the probability is $\frac{1}{5}$.

Example C.3.3 (ISI 2019 PSA 6). How many times does the digit '2' appear in the set of integers $\{1,2,..,1000\}$? In the units digit, '2' appears $10 \times 10 = 100$ times; in the tens digit, it appears 10 times and in the 100s digit it appears once. Thus in total it appears 111 times.

Example C.3.4 (ISI 2019 PSA 5). Let the sum $3 + 33 + 333 + \cdots + \underbrace{33\dots3}_{200 \text{ times}}$ be $...zyx$ in the decimal system, i.e., x is the unit's digit, y the ten's digit, and so on. What is z ?

Example C.3.5 (ISI 2021 PSA 12). Let $\pi = (a_1, a_2, \cdots, a_{2021})$ be a permutation of $(1, 2, \cdots, 2021)$. For every such permutation π,

$$P(\pi) = \prod_{j=1}^{2021} (a_j - j)\,.$$

Is $P(\pi)$ always even for any permutation π? Yes. To see this, note that $\sum_{j=1}^{2021} (a_j - j) = 0$ and so

$$\sum_{j \neq i} (a_j - j) = -(a_i - i)$$

where for any $i \in \{1,2,\dots,2021\}$ the sum on the LHS has an even number of terms. If the terms are all odd, the sum (and therefore the RHS) is even and thus the product is even.

EXAMPLE C.3.6 (ISI 2020 PSA 14). Let S be the set of all 3×3 matrices$A$ such that among the 9 entries of A, there are exactly three 0 's, exactly three 1 's and exactly three 2 's. What is the number of matrices in S that have trace divisible by 3? TODO

EXAMPLE C.3.7 (ISI 2023 PSB 2). How many permutations of the numbers $1, 2, \ldots, n$ where $n$is even exist such that no two adjacent numbers have an odd product? Let's first count the number of such permutations where the first slot is taken up by an even number. The number of positive even integers less or equal to n is $\frac{n}{2}$ and so answer is for these is $\frac{n}{2}!^2$. The odd first answer is $\frac{n}{2}!^2$ so together the answer is $2\frac{n}{2}!^2$.

EXAMPLE C.3.8 (ISI 2015 PSB 4). Suppose 15 identical balls are placed in 3 boxes labeled A, B and C. What is the number of ways in which Box A can have more balls than Box C?TODO

APPENDIX D

Limits and continuity

For completeness, we state and prove all theorems from basic real analysis that are used in the main text.

D.1. Elementary analysis on the line

D.1.1. Sequences and series.

D.1.1.1. *Properties of Limit Inferiors and Superiors.* We collect here the following properties of the limit superiors and limit inferiors. In all of the following, $\{a_n\}$ and $\{b_n\}$ are real sequences. Recall that $\limsup_{n\to\infty} a_n = \inf_{n\geq 1}\left\{\sup_{i\geq n}\{a_i\}\right\}$ and $\liminf_{n\to\infty} a_n = \sup_{n\geq 1}\{\inf_{i\geq n}\{a_i\}\}$. Note that these always exist (in the extended reals) by the completeness of the real numbers.

PROPOSITION D.1.1. $\liminf_{n\to\infty} a_n = -\limsup_{n\to\infty} -a_n$.

PROOF. Note that $\inf_{i\geq n} a_i = -\sup_{i\geq n} -a_i$ and so taking limits completes the proof. □

PROPOSITION D.1.2. $\liminf_{n\to\infty} a_n \leq \limsup_{n\to\infty} a_n$.

PROOF. $\inf_{i\geq n} a_i \leq \sup_{i\geq n} a_i$ and so the result follows by taking limits. □

PROPOSITION D.1.3. *$\liminf_{n\to\infty} a_n$ and $\limsup_{n\to\infty} a_n$ are subsequential limits of a_n such that for any other subsequential limit L we have that*

$$\liminf_{n\to\infty} a_n \leq L \leq \limsup_{n\to\infty} a_n.$$

PROOF. Note that if $L < \liminf_{n\to\infty} a_n$, then there are infinitely many elements of $\{a_n\}$ in $(L-\epsilon, L+\epsilon)$where $\epsilon > 0$ is small enough that $L+\epsilon < \liminf_{n\to\infty} a_n$. Note that since $\{\inf_{i\geq n} a_i\}$ is a nondecreasing sequence, there exists some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $\inf_{i\geq n} a_i > L+\epsilon$. But then $a_i \geq \inf_{i\geq n_0} a_i > L+\epsilon$ for all every $i \geq n_0$ which contradicts the fact that infinitely many of our $\{a_n\}$ were in $[L, L+\epsilon)$. The proof for the limit superior case is analagous.

Next fix $\epsilon > 0$, let $c_n := \inf_{i\geq n} a_i.$, and let $I := \lim c_n$.By the definition of infimum, there exists some $k_n \geq n$ such that $|c_n - a_{k_n}| < \frac{\epsilon}{2}$. By definition of limit inferiors, we have that for large enough n, $|c_n - I| < \frac{\epsilon}{2}$ and so

$$|a_{k_n} - I| \leq |c_n - a_{k_n}| + |c_n - I| < \epsilon$$

which completes the proof. The proof for limit superiors is analagous (or you can use Proposition D.1.1). □

PROPOSITION D.1.4. *The sequence $\{a_n\}$ converges if and only if $\limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n$ in which case*

$$\lim_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n.$$

PROOF. If $\lim_{n\to\infty} a_n = L$, then all subsequential limits are also equal to L, which implies (by Proposition D.1.3) that $\lim_{n\to\infty} a_n = \limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n$. Conversely, assuming that $L := \limsup_{n\to\infty} a_n = \liminf_{n\to\infty} a_n$, we can show that for any $\epsilon > 0$,there's some large enough $n_0 \in \mathbb{N}$, |such that

$$\sup_{i\geq n_0} x_i - \inf_{i\geq n_0} x_i \leq |\inf_{i\geq n_0} x_i - L| + |\sup_{i\geq n_0} x_i - L| < \epsilon.$$

Then for any $m, n \geq n_0$

$$|x_m - x_n| \leq \sup_{i \geq n_0} x_i - \inf_{i \geq n_0} x_i \leq \epsilon$$

and thus the sequence is Cauchy and converges to L by Proposition D.1.3. □

PROPOSITION D.1.5. $\liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \leq \liminf_{n\to\infty} (a_n + b_n) \leq \limsup_{n\to\infty} (a_n + b_n) \leq \limsup_{n\to\infty} a_n + \limsup_{n\to\infty} b_n$.

PROOF. Note that $\inf_{i\geq n} a_i + \inf_{i\geq n} b_i \leq \inf_{i\geq n} (a_i + b_i)$ and take limits. The other inequalities follow by Propositions D.1.2 and D.1.1. □

REMARK. We are implicitly excluding the $\infty - \infty$ situations in the limits. This should be assumed throughout this section.

PROPOSITION D.1.6. *For* $a_n, b_n \geq 0$ *we have that* $\liminf a_n \liminf b_n \leq \liminf a_n b_n \leq \limsup a_n b_n \leq \limsup a_n \limsup b_n$.

PROOF. Note that we have the inequalities

$$0 \leq \inf_{i\geq n} a_i \leq a_i$$
$$0 \leq \inf_{i\geq n} b_i \leq b_i$$

for all $i \geq n$ and so by multiplying the inequalities we get

$$0 \leq \inf_{i\geq n} a_i \inf_{i\geq n} b_i \leq a_i b_i \implies 0 \leq \inf_{i\geq n} a_i \inf_{i\geq n} b_i \leq \inf_{i\geq n} a_i b_i$$

and taking limits finishes the proof. □

REMARK D.1.7. Note that the condition that $a_n, b_n \geq 0$ is necessary: see $a_n = (-1)^n, b_n = (-1)^{n+1}$.

PROPOSITION D.1.8. $\liminf_{n\to\infty} (a_n + b_n) \leq \liminf_{n\to\infty} a_n + \limsup_{n\to\infty} b_n \leq \limsup_{n\to\infty} (a_n + b_n)$

PROOF. Note that

$$a_n + b_n \leq a_n + \sup_{i\geq n} b_i \implies \inf_{i\geq n} (a_n + b_n) \leq \inf_{i\geq n} \left(a_n + \sup_{i\geq n} b_n \right) = \inf_{i\geq n} a_i + \sup_{i\geq n} b_i$$

where the last equality follows because $\sup b_i$ is a constant for a fixed n. Taking limits then yields the result. □

PROPOSITION D.1.9. *If* $\lim_{n\to\infty} a_n = L$ *then*

$$\liminf_{n\to\infty} (a_n + b_n) = L + \liminf_{n\to\infty} b_n$$
$$\limsup_{n\to\infty} (a_n + b_n) = L + \limsup b_n.$$

PROOF. Note first that

$$\begin{aligned} L + \liminf_{n\to\infty} b_n &= \liminf_{n\to\infty} a_n + \liminf_{n\to\infty} b_N \\ &\leq \liminf_{n\to\infty} (a_n + b_n) \\ &\leq \limsup_{n\to\infty} a_n + \liminf_{n\to\infty} b_n \\ &= L + \liminf_{n\to\infty} b_n \end{aligned}$$

by PropositionsD.1.4, D.1.5, and D.1.8. □

PROPOSITION D.1.10. *If* $a_n, b_n \geq 0$ *then*

$$\liminf_{n\to\infty} a_n b_n \leq \liminf_{n\to\infty} a_n \limsup_{n\to\infty} b_n \leq \limsup_{n\to\infty} a_n b_n.$$

PROOF. The proof proceeds in the exact same way as in Proposition D.1.8. □

PROPOSITION D.1.11. *If* $a_n, b_n \geq 0$, $\lim a_n = L$ *and* $\limsup_{n\to\infty} b_n < \infty$ *then*

$$\liminf (a_n b_n) = L \liminf b_n$$
$$\limsup (a_n b_n) = L \limsup b_n.$$

PROOF. This is akin to Proposition D.1.11; the following chain of inequalities establish the claim

$$\begin{aligned} L \liminf b_n &= \liminf a_n \liminf b_n \\ &\leq \liminf a_n b_n \\ &\leq \limsup a_n \liminf b_n \\ &= L \liminf b_n. \end{aligned}$$

□

PROPOSITION D.1.12. *Let* a_n *be a real sequence and let* $f : \mathbb{R} \to \mathbb{R}$ *be a continuous and increasing function, then*

$$f\left(\liminf_{n\to\infty} a_n\right) = \liminf_{n\to\infty} f(a_n)$$
$$f\left(\limsup_{n\to\infty} a_n\right) = \limsup_{n\to\infty} f(a_n).$$

PROOF. First observe that since f is increasing, we have that $\inf_{i\geq n} a_i \leq a_i \implies f(\inf_{i\geq n} a_i) \leq f(a_i)$ for all $i \geq n$ and so

$$f\left(\inf_{i\geq n} a_i\right) \leq \inf_{i\geq n} f(a_i).$$

Conversely, fix n and suppose that $f(\inf_{i\geq n} a_i) < \inf_{i\geq n} f(a_i)$. Fixing $\epsilon = \frac{\inf_{i\geq n} f(a_i) - f\left(\inf_{i\geq n} a_i\right)}{2}$, note that by continuity there exists a $\delta > 0$ such that $|\inf_{i\geq n} a_i - x| < \delta \implies |f(\inf_{i\geq n} a) - f(x)| < \epsilon$. By the definition of infimum, there is some $i_0 \geq n$ where $a_{i_0} - \inf_{i\geq n} a_i < \delta$ and so

$$f(a_{i_0}) - f\left(\inf_{i\geq n} a_i\right) < \epsilon$$

which is a contradiction and hence

$$f\left(\inf_{i\geq n} a_i\right) \geq \inf_{i\geq n} f(a_i).$$

Taking limits and applying the continuity of f once again yields the result. □

PROPOSITION D.1.13.

$$\limsup_{n\to\infty} \max\{a_n, b_n\} = \max\left\{\limsup_{n\to\infty} a_n, \limsup_{n\to\infty} b_n\right\}$$

and

$$\liminf_{n\to\infty} \min\{a_n, b_n\} = \min\left\{\liminf_{n\to\infty} a_n, \liminf_{n\to\infty} b_n\right\}.$$

EXERCISE D.1.14. Let $\{a_n\}$ and $\{b_n\}$ be positive real sequences with such that $\limsup_{n\to\infty} \frac{\log(b_n)}{n} = -\infty$. Prove that

$$\liminf_{n\to\infty} \frac{\log(a_n + b_n)}{n} = \liminf_{n\to\infty} \frac{\log(a_n)}{n}.$$

SOLUTION. Note that since exponential function is increasing and continuous, by Proposition D.1.12

$$\exp\left(\limsup_{n\to\infty} \log\left(b_n^{\frac{1}{n}}\right)\right) = \limsup_{n\to\infty} b_n^{\frac{1}{n}} = 0$$

which implies that $\lim_{n\to\infty} b_n^{\frac{1}{n}} = 0$ since $b_n > 0$ for all $n \in \mathbb{N}$. Since eventually, $b_n < b_n^{\frac{1}{n}}$ we have that $\lim_{n\to\infty} b_n = 0$.

$$\frac{\log\left(a_n+b_n\right)}{n} = \frac{\log\left(a_n\right)}{n} + \frac{\log\left(1+\frac{b_n}{a_n}\right)}{n} \leq \frac{\log\left(a_n\right)}{n} +$$

D.1.1.2. *Base-b representations.*

THEOREM D.1.15. *For any real number $x \in [0,1)$ and any positive integer $b \geq 2$, there exists a sequence of non-negative integers $\{a_i\}_{i=1}^{\infty} \in \{0,1,2,\ldots,b-1\}$ such that*

$$\sum_{i=1}^{\infty} \frac{a_i}{b^i} = x.$$

Moreover, if there is no $n_0 \in \mathbb{N}$ such that for all $n \geq n_0 : a_n = b-1$, then the sequence $\{a_i\}$ is unique.

PROOF. Note that we can partitionTODO □

EXAMPLE D.1.16. For $n \geq 1$ let $x_n = \frac{1}{n\alpha_n}$ where α_n is such that $2^{\alpha_n-2} < n \leq 2^{\alpha_n-1}$. Is the series $\sum_{n=1}^{\infty} x_n$ convergent or divergent? Justify your answer.

D.1.2. Limits of functions.

D.2. Metric spaces

D.2.1. Topology of metric spaces.

D.2.2. Complete metric spaces.

D.2.3. Compact metric spaces.

D.2.4. Continuity of functions between metric spaces.

DEFINITION D.2.1. Let (X,d_X) and (Y,d_Y) be a metric space. A function $f : X \to Y$ is said to be *continuous at a point* $c \in X$ if for any sequence $x_n \to c$ we have that $f\left(x_n\right) \to f\left(c\right)$. The function f is called *continuous* if it is continuous at every point $c \in X$.

PROPOSITION D.2.2. *Let (X,d_X) and (Y,d_Y) be a metric space. A function $f : X \to Y$ is said to be* continuous at a point $c \in X$ *if and only if for every $\epsilon > 0$ there exists a δ such that for any $x \in X$ $d_X\left(x,c\right) < \delta \implies d_Y\left(f\left(x\right), f\left(c\right)\right) < \epsilon$.*

PROOF. Suppose that the $\epsilon - \delta$ definition doesnt hold i.e. there exists some $\epsilon_0 > 0$ such that for all $\delta > 0$ there exists some x such that $d_X\left(x,c\right) < \delta$ but $d_Y\left(f\left(x\right), f\left(c\right)\right) \geq \epsilon_0$. Letting $\delta = 1/n$, we can find a corresponding x_n such that $d_X\left(x_n,c\right) < \frac{1}{n}$ but $d_Y\left(f\left(x_n\right), f\left(c\right)\right) \geq \epsilon_0$. Since this can be done for any $n \in \mathbb{N}$, we have a seqeunce $x_n \to c$ but $f\left(x_n\right) \not\to c$.

Conversely, suppose the $\epsilon - \delta$ definition holds and let $x_n \to c$ by arbitrary. For a large enough N, $d_X\left(x_n,c\right) < \delta$ for all $n \geq N$ and so $d_Y\left(f\left(x_n\right), f\left(c\right)\right) < \epsilon$ which completes the proof. □

PROPOSITION D.2.3. *Let (X,d_X) and (Y,d_Y) be a metric space. A function $f : X \to Y$ is said to be* continuous if and only if for every open set $O_Y \in \tau_Y$ (here τ_y denotes the topology on Y) the preimage $f^{-1}\left[o_Y\right] \in \tau_X$.

PROOF. First assume that f is continuous and let O_Y be an open set in Y. We wish to show that for any $x \in f^{-1}\left[O_Y\right]$,there is some open ball $B\left(x,\delta\right) \subseteq f^{-1}\left[O_y\right]$. Note that since O_Y is open, there exists some ball $B\left(f\left(x\right),\epsilon\right) \subseteq O_Y$. Note that by continuity, there exists some $\delta > 0$ such that $f\left[B\left(x,\delta\right)\right] \subseteq B\left(f\left(x\right),\epsilon\right)$. Then by the property of preimages

$$B\left(x,\delta\right) \subseteq f^{-1}f\left[B\left(x,\delta\right)\right] \subseteq f^{-1}\left[o_Y\right]$$

which proves the claim.

Conversely, suppose that preimages of open sets are open. Fix $\epsilon > 0$ and note that for any $x \in X$, the preimage of the ball $f^{-1}[B(f(x),\epsilon)]$ is open and thus there exists some $\delta > 0$ such that $B(x,\delta) \subseteq f^{-1}[B(f(x),\epsilon)]$. Then

$$f[B(x,\delta)] \subseteq ff^{-1}[B(f(x),\epsilon)] \subseteq B(f(x),\epsilon)$$

which completes the proof. □

Proposition D.2.4. *Let $(X,d_X),(Y,d_Y)$, and (Z,d_Z) be metric spaces. Then for any continuous functions $f : X \to Y$ and $g : Y \to Z$, the composition $g \circ f : X \to Z$ given by*

$$g \circ f(x) = g(f(x))$$

is continuous.

Proof. Let O_Z be an open set in Z. Note that $(g \circ f)^{-1}[O_Z] = f^{-1}\left[g^{-1}[O_Z]\right]$ and $g^{-1}[O_Z]$ is open since g is continuous and so $f^{-1}\left[g^{-1}[O_Z]\right]$ is open. □

Remark. This proof applies *mutis mutandis* to continuous functions between general topological spaces, of which metric spaces are a special case.

Definition D.2.5. Let (X,d_X) and (Y,d_Y) be a metric spaces. A function $f : X \to Y$ is called *Lipschitz continuous* if there exists some $K > 0$ such that

$$d_Y(f(x),f(y)) \leq K d_X(x,y).$$

Proposition D.2.6. *Let (X,d_X) and (Y,d_Y) be a metric spaces. Every Lipschitz continuous function $f : X \to Y$ is uniformly continuous.*

Proof. Fix $\epsilon > 0$ and note that if $\delta = \frac{\epsilon}{K}$ then $d_X(x,y) < \delta \implies d_Y(f(x),f(y)) < \epsilon$. □

Proposition D.2.7. *Let$f : D \subseteq \mathbb{R} \to \mathbb{R}$ be a differentiable function. Then f is Lipschitz if and only if it has a bounded derivative.*

Proof. First suppose that the function has a bounded derivative. Then there exists some $M > 0$ such that for any $x \in D$, $|f'(x)| \leq M$. Then for any $x,y \in D$, the mean value theorem implies that there exists some $c \in (\min\{x,y\},\max\{x,y\})$ such that

$$\left|\frac{f(x)-f(y)}{x-y}\right| = |f'(c)| \leq M$$

Rearranging yields the result.

Conversely, suppose that the function is Lipschitz with constant K. Then

$$\left|\frac{f(x)-f(y)}{x-y}\right| \leq K$$

and taking limits $x \to y$ gives the result. □

Definition D.2.8. Let (X,d_X) and (Y,d_Y) be metric spaces. A function $f : X \to Y$ is called *Cauchy-continuous* if for any Cauchy sequence $\{x_n\} \in X$, the image $\{f(x_n)\}$ is Cauchy.

Proposition D.2.9. *Let (X,d_X) and (Y,d_Y) be metric spaces. If $f : X \to Y$ is Cauchy continuous then it is continuous. Moreover, if X is complete then every continuous function $f : X \to Y$ is also Cauchy-continuous.*

Proof. Fix $\epsilon > 0$ and let $c \in X$ be arbitrary. Then there exists at least one (possibly eventually constant) Cauchy sequence $x_m \to c.$. Now consider a new sequence y_n such that $y_n = x_n$ for even $n$and $y_n = c$ for odd n. Then $y_n \to c$ (and is Cauchy) and so $f(y_n)$ is Cauchy. This implies that for large n

$$d(f(x_n),f(c)) < \epsilon$$

which implies $f(y_n) \to f(c)$. Extracting the even indexed subsequence shows that $f(x_n) \to c$.

Next, suppose that X is complete and f is continuous. Let $\{x_n\}$ be a Cauchy sequence. Then by completenes $x_n \to c \in X$. By continuity, $f(x_n) \to f(c)$ and so $\{f(x_n)\}$ is Cauchy. □

PROPOSITION D.2.10. *Let (X, d_X) and (Y, d_Y) be a metric spaces. If a function $f : X \to Y$ is uniformly continuous, it is Cauchy continuous.*

PROOF. Fix $\epsilon > 0$ and let $\{x_n\} \in X$ be a Cauchy sequence. By uniform continuity, there exists some $\delta > 0$ such that for any $m, n \in \mathbb{N}$ $d_X(x_n, x_m) < \delta \implies d_Y(f(x_n), f(x_m)) < \epsilon$. Since $\{x_n\}$ is Cauchy, there exists some N such that for any $m, n \geq N$ $d_X(x_n, x_m) < \delta$ which completes the proof. □

THEOREM D.2.11. *Let (X, d_X) be a compact metric space and let (Y, d_Y) be an arbitrary metric space. The following are equivalent for a function $f : X \to Y$*

(i) f is continuous
(ii) f is uniformly continuous
(iii) f is Cauchy continuous.

PROOF. We prove $(i) \implies (ii)$; $(ii) \implies (iii)$ is Proposition D.2.10 and $(iii) \implies (i)$ is Proposition D.2.9.

Let $\epsilon > 0$ be fixed. By continuity, we know that for every $c \in X$ there exists some $\delta_c > 0$ such that when $d_X(x, c) < \delta_c \implies d_Y(f(x), f(c)) < \frac{\epsilon}{2}$. Note that the balls $\left\{B_{d_X}\left(c, \frac{\delta_c}{2}\right)\right\}_{c \in X}$ form a cover of X and by compactness we can extract a finite subcover $\left\{B_{d_X}\left(c_i, \frac{\delta_{c_i}}{2}\right)\right\}_{i=1}^{n}$. Let $\delta := \min_i \frac{\delta_{c_i}}{2}$ and note that for any $x, y \in X$, there's some $1 \leq i_0 \leq n$ such that $x \in B_{d_X}\left(c_{i_0} \frac{d_{c_{i_0}}}{2}\right)$. Then, if $d_X(x, y) < \delta$

$$\begin{aligned} d_X(c_{i_0}, y) &\leq d_X(x, y) + d_X(x, c_{i_0}) \\ &< \delta + \frac{\delta_{c_{i_0}}}{2} \\ &\leq \delta_{c_{i_0}} \end{aligned}$$

Then,

$$\begin{aligned} d_Y(f(x), f(y)) &\leq d_Y(f(x), f(c_{i_0})) + \delta_Y(f(y), f(c_{i_0})) \\ &< \frac{\epsilon}{2} + \frac{\epsilon}{2} \\ &= \epsilon \end{aligned}$$

which completes the proof. □

DEFINITION D.2.12. Let (X, d_X) and (Y, d_Y) be a metric spaces. A function $f : X \to Y$ is called a contraction if it is Lipschitz with a Lipschitz constant $0 < K < 1$. In other words, for any $x, y \in X$

$$d_Y(f(x), f(y)) \leq K d_X(x, y).$$

THEOREM D.2.13 (Banach Fixed Point Theorem). *Let (X, d) be a complete metric space and let a function $f : X \to X$ be a contraction. Then there exists a unique $x \in X$ such that $f(x) = x$.*

PROOF. Let $x_0 \in X$ be arbitrary, define a sequence $x_{n+1} = f(x_n)$ and fix $\epsilon > 0$. For any $n \geq m \geq 1$

$$\begin{aligned} d(x_n, x_m) &\leq d(x_n, x_{n-1}) + d(x_{n-1}, x_{n-2}) + \ldots + d(x_{m+1}, x_m) \\ &\leq K^{n-1} d(x_1, x_0) + K^{n-2} d(x_1, x_0) + \ldots + K^m d(x_1, x_0) \\ &= d(x_1, x_0) \sum_{i=m}^{n-1} K^i \end{aligned} \tag{26}$$

where the first inequalitity is simply the triangle inequality and the second is induction on the contraction property. Note that since $\left\{\sum_{i=1}^{n} K^i\right\}_{i=1}^{\infty}$ is a convergent geometric series, it's Cauchy as in we can find an $N \in \mathbb{N}$ such that for all $n, m \geq N$ we have that $\left|\sum_{i=1}^{n} K_i - \sum_{i=1}^{m} K_i\right| = \sum_{i=\min\{n.m\}-1}^{\max\{n.m\}} K^i < \epsilon$. Therefore, for $m, n \geq N$, the inequality (26) implies

$$d(x_m, x_n) \leq d(x_1, x_0)\, \epsilon.$$

Since ϵ can be arbitrarily small, our sequence x_n is Cauchy and so by the completeness of X it converges to some limit x. Then, by the continuity of f

$$\begin{aligned} x &= \lim_{n\to\infty} x_n \\ &= \lim_{n\to\infty} f(x_{n-1}) \\ &= f(x). \end{aligned}$$

which yields our fixed point.

Now suppose that there were two fixed points i.e. $x \neq y$ such that $f(x) = x$ and $f(y) = y$. By the contraction property,

$$d(x,y) \leq Kd(x,y)$$

which is a contradiction. □

D.2.4.1. *Continuity of real valued function on* $\mathbb{R}$.

Proposition D.2.14. *Let $f : [a,b] \to \mathbb{R}$ be a strictly increasing, continuous function. Then f is bijective on its range, and its inverse is strictly increasing and continuous.*

Proof. Note that injectivity is straightforward since if $x \neq y$ then either $x > y$ or $y < x$ in which case $f(x) > f(y)$ or $f(x) < f(y)$, respectively. To see that the inverse is strictly increasing, let $c, d \in R$ where $R \subseteq \mathbb{R}$ is the range of f. If $c > d$, then since f is strictly increasing it must be that $f^{-1}(c) > f^{-1}(d)$. Finally, notice that f^{-1} is a strictly increasing function that is onto $[a,b]$. Suppose that $y_0 \in R$ is a point of discontinuity for f^{-1}, then $f^{-1}\left(y_0^-\right) < f^{-1}\left(y_0^+\right)$ where both limits are in $[a,b]$. Then of course there's some $f^{-1}\left(y_0^-\right) < c < f^{-1}\left(y_0^+\right)$ such that there's no $y \in R$ such that $f^{-1}(y) = c$ which is a contradiction since f^{-1} is onto $[a,b]$. □

Example D.2.15. Let g be a continuous function with $g(1) = 1$ such that

$$g(x+y) = 5g(x)g(y)$$

for all x, y. Find $g(x)$. [Hint: You may use the following result. If f is a continuous function that satisfies $f(x+y) = f(x) + f(y)$ for all x, y, then $f(x) = xf(1)$.TODO

D.2.5. Separable metric spaces.

APPENDIX E

Common probability distributions

E.1. General families of distributions

E.1.1. Location-scale families.

E.1.2. General exponential families.

E.1.3. Stable distributions.

E.1.4. Infinitely divisible distributions.

E.1.5. Power series distributions.

E.2. Special parametric families of distributions

E.2.1. Normal distributions and their associates.

E.2.1.1. *The univariate normal distribution.*

E.2.1.2. *The multivariate normal distribution.*

EXAMPLE E.2.1. Let $\underline{Y} = (Y_1, Y_2)'$ have the bivariate normal distribution $N_2(\underline{0}, \Sigma)$, where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}.$$

Obtain the mean and variance of $U = \underline{Y}'\Sigma^{-1}\underline{Y} - Y_1^2/\sigma_1^2$. TODO

E.2.1.3. *The lognormal distribution.*

E.2.1.4. *The folded normal distribution.*

E.2.1.5. *The Rayleigh distribution.*

E.2.1.6. *The Maxwell distribution.*

E.2.1.7. *The Levy distribution.*

E.2.2. Distributions useful for basic statistical inference.

E.2.2.1. *The Gamma distribution.*

E.2.2.2. *The Chi-squared distributtion.*

E.2.2.3. *Student's t distribution.*

E.2.2.4. *The F distribution.*

E.2.3. Continuous distributions with bounded support.

E.2.3.1. *The uniform distribution.* The uniform distribution on the interval $[a, b]$ is the simplest example of a distribution that is absolutely continuous with respect to the Lebesgue measure. In fact, its density is $\frac{1}{b-a}\mathbb{1}_{[a,b]}$ which means the distribution is simply the restriction of the Lebesgue measure to the interval $[a, b]$, with an appropriate normalization to ensure that it is a probability measure. A random variable X distributed uniformly on $[a, b]$ is often denoted $X \sim U[a, b]$. The CDF of such an X is $F_X(x) = \bar{\lambda}\left(\frac{1}{b-a}\mathbb{1}_{[a,x]}\right) = \frac{x-a}{b-a}\mathbb{1}\{a \le x \le b\} + \mathbb{1}\{x > b\}$. The moments can be computed easily.

PROPOSITION E.2.2. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X \sim U[a,b]$. Then,*

$$\mathbb{E}\left[X^k\right] = \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}.$$

PROOF. Note that by change of variables and the fundamental theorem of calculus

$$\begin{aligned}\mathbb{E}\left[X^k\right] &= \bar{\lambda}_x\left(x^k \frac{1}{b-a}\mathbb{1}[a,b]\right)\\ &= \frac{1}{b-a}\left.\frac{x^{k+1}}{k+1}\right|_a^b\\ &= \frac{b^{k+1} - a^{k+1}}{(k+1)(b-a)}.\end{aligned}$$

□

PROPOSITION E.2.3. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $X \sim U[a,b]$. Then,*

$$M_X(t) = \frac{e^{bt} - e^{at}}{t(b-a)}\mathbb{1}\{t \neq 0\} + \mathbb{1}\{t = 0\}.$$

PROOF. Suppose $t \neq 0$, in which case

$$\begin{aligned}M_X(t) &= \mathbb{E}\left[e^{tX}\right]\\ &= \frac{1}{b-a}\bar{\lambda}_x\left(e^{tx}\mathbb{1}[a,b]\right)\\ &= \frac{1}{b-a}\frac{1}{t}\left.e^{tx}\right|_a^b\\ &= \frac{e^{tb} - e^{ta}}{t(b-a)}.\end{aligned}$$

The other case is trivial. □

Other interesting moments of the distribution are the variance, skewness and kurtosis.The variance is found easily by the identity

$$\begin{aligned}\mathbb{V}\mathrm{ar}[X] &= \mathbb{E}\left[X^2\right] - (\mathbb{E}[X])^2\\ &= \frac{b^3 - a^3}{3(b-a)} - \left(\frac{a+b}{2}\right)^2\\ &= \frac{(a+b)^2 - ab}{3} - \frac{(a+b)^2}{4}\\ &= \frac{(b-a)^2}{12}.\end{aligned}$$

Similarly, the skewness is given

$$\begin{aligned}\mathrm{skew}(X) &= \mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right]\\ &=\end{aligned}$$

E.2.3.2. *The Beta distribution.*
E.2.3.3. *The Beta Prime distribution.*
E.2.3.4. *The arcsine distribution.*
E.2.3.5. *The semicircle distribution.*
E.2.3.6. *The triangle distribution.*
E.2.3.7. *The Irwin-Hall distribution.*

E.2.4. Continuous distributions with positve support.

E.2.4.1. *Exponential-logarithmic distribution.*

E.2.4.2. *The Gompertz distribution.*

E.2.4.3. *The Log-logistic distribution.*

E.2.4.4. *The Pareto distribution.*

E.2.4.5. *The Wald distribution.*

E.2.4.6. *The Weibull distribution.*

E.2.5. Continuous distributions supported on the real line.

E.2.5.1. *The Laplace distribution.*

E.2.5.2. *The logistic distribution.*

E.2.5.3. *The exreme value distribution.*

E.2.5.4. *The hyperbolic secant distribution.*

E.2.5.5. *The Cauchy distribution.*

E.2.6. Distributions associated with modeling Bernoulli trials.

E.2.6.1. *The Bernoulli distribution.* The Bernoulli distribution is the simplest of all discrete probability distributions. It represents the mathematical abstraction of coin tossing with not-necessarily fair coins.

Definition E.2.4. A random variable X on probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is said to have a Bernoulli distribution with parameter $p \in [0,1]$ if $\mathbb{P}(X=1) = p$ and $\mathbb{P}(X=0) = 1-p$.

Formally, any indicator variable of an event $A \in \mathcal{F}$ is a Bernoulli random variable. However, we typically reserve this description for random variables used to model the outcome of a binary experiment trial. Usually we are interested in the *sequence* of such trials.

Proposition E.2.5. *The raw moments of a Bernoulli random variable X with parameter p is given*

$$\mathbb{E}\left[X^k\right] = p$$

for any $k \in \mathbb{N}$. The central moments are

$$\mathbb{E}\left[(X-p)^k\right] = (1-p)^k p + (1-p)(-p)^k$$

Proof. Note that

$$\begin{aligned}\mathbb{E}\left[(X-p)^k\right] &= (1-p)^k \mathbb{P}(X=1) + (0-p)^k \mathbb{P}(X=0) \\ &= (1-p)^k p + (1-p)(-p)^k .\end{aligned}$$

The raw moment case is simpler and follows exactly in the same way. □

Proposition E.2.6. *The moment generating function of a Bernoulli random variable X with parameter p is given*

$$M_X(t) = 1 - p + pe^t.$$

Proof. Again follow the same approach as in E.2.5. □

E.2.6.2. *The Binomial distribution.* The Binomial distribution is perhaps the most well known of all discrete distributions. It models the number of successes in a fixed number of trials (say coin tosses) and as such is a defined as a sum of Bernoulli trials.

Definition E.2.7. A random variable Y has a Binomial distribution with parameters n and p if

$$Y = \sum_{i=1}^{n} X_i$$

where X_i are independent and identically distributed Bernoulli random variables with parameter p. In this case we write $Y \sim \mathrm{Bin}(n,p)$

PROPOSITION E.2.8. *The mass function of a random variable Y with a Binomial distribution with parameters n and p is given by*

$$\mathbb{P}(Y=k)=\binom{n}{k}p^k(1-p)^{n-k}\mathbb{1}\{0\leq k\leq n\}$$

PROOF. For $n=2$, we use the result about convolutions. More specifically, by Corollary 9.4.25

$$\begin{aligned}\mathbb{P}(Y=k)&=\sum_{x\in\{0,1\}}\mathbb{P}(X_1=x)\,\mathbb{P}(X_2=k-x)\\&=(1-p)\left[p\mathbb{1}\{k=1\}+(1-p)\mathbb{1}\{k=0\}\right]\\&\quad+p\left[p\mathbb{1}\{k=2\}+(1-p)\mathbb{1}\{k=1\}\right]\\&=p^2\mathbb{1}\{k=2\}+2p(1-p)\mathbb{1}\{k=1\}+(1-p)^2\mathbb{1}\{k=0\}\\&=\binom{2}{k}p^k(1-p)^{2-k}\mathbb{1}\{0\leq k\leq 2\}.\end{aligned}$$

Now for the induction step, assume that the result holds for n and then we write $Y=\sum_{i=1}^{n+1}X_i=Z+X_{n+1}$ where $Z\sim\mathrm{Bin}(n-1,p)$. Then, another convolution argument shows

$$\begin{aligned}\mathbb{P}(Y=k)&=\sum_{z=0}^{n}\mathbb{P}(Z=z)\,\mathbb{P}(X_{n+1}=k-z)\\&=\sum_{z=0}^{n}\binom{n}{z}p^z(1-p)^{n-z}\mathbb{1}\{0\leq z\leq n\}\left(p\mathbb{1}\{z=k-1\}+(1-p)\mathbb{1}\{z=k\}\right)\\&=\sum_{z=0}^{n}\binom{n}{z}p^{z+1}(1-p)^{n-z}\mathbb{1}\{0\leq z\leq n,z=k-1\}\\&\quad+\sum_{z=0}^{n}\binom{n}{z}p^{z}(1-p)^{n-z+1}\mathbb{1}\{0\leq z\leq n,z=k\}\\&=\binom{n}{k-1}p^k(1-p)^{n-k+1}+\binom{n}{k}p^k(1-p)^{n-k+1}\\&=\binom{n+1}{k}p^k(1-p)^{n+1-k}\end{aligned}$$

where we used Pascal's rule in the last equality. □

The moments of the Binomial distribution are easily characterized using the Bernoulli moments and the multinomial theorem

PROPOSITION E.2.9. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and let $Y\sim\mathrm{Bin}(n,p)$ for some $n\in\mathbb{N}$ and $p\in[0,1]$. Then,*

$$\mathbb{E}\left[Y^k\right]=\sum_{i_1+i_2+\ldots+i_n=k,i_j\geq 0}\binom{k}{i_1,\ i_2,\ \ldots\ i_n}p^{\sum_{j=1}^{n}\mathbb{1}\{i_j\geq 1\}}.$$

PROOF. Note that $Y = \sum_{i=1}^{n} X_i$ where X_i are i.i.d Bernoulli random variables with parameter p. Then

$$\begin{aligned}
\mathbb{E}\left[Y^k\right] &= \mathbb{E}\left[\left(\sum_{i=1}^{n} X_i\right)^k\right] \\
&= \mathbb{E}\left[\sum_{i_1+i_2+\ldots+i_n=k, i_j\geq 0} \binom{k}{i_1, \quad i_2, \quad \ldots \quad i_n} \prod_{j=1}^{n} X_j^{i_j}\right] \\
&= \sum_{i_1+i_2+\ldots+i_n=k, i_j\geq 0} \binom{k}{i_1, \quad i_2, \quad \ldots \quad i_n} \mathbb{E}\left[\prod_{j=1}^{n} X_j^{i_j}\right] \\
&= \sum_{i_1+i_2+\ldots+i_n=k, i_j\geq 0} \binom{k}{i_1, \quad i_2, \quad \ldots \quad i_n} \prod_{j=1}^{n} \mathbb{E}\left[X_j^{i_j}\right] \\
&= \sum_{i_1+i_2+\ldots+i_n=k, i_j\geq 0} \binom{k}{i_1, \quad i_2, \quad \ldots \quad i_n} p^{\sum_{j=1}^{n} \mathbb{1}\{i_j\geq 1\}}
\end{aligned}$$

where in the second equality we have used the multinomial theorem, the third is linearity of integration, the fourth is Proposition 9.4.21, and the last equality uses Proposition E.2.5. □

The first moment is $\mathbb{E}[Y] = np$ which can be seen just by linearity and the definition. The second moment is

$$\begin{aligned}
\mathbb{E}\left[Y^2\right] &= \mathbb{E}\left[\sum_{i=1}^{n} X_i^2 + \sum_{i=1}^{n} \sum_{j=1, i\neq j}^{n} X_i X_j\right] \\
&= np + n(n-1)p^2.
\end{aligned}$$

We are more interested in higher *central* moments like the variance. Of course, $\mathbb{V}\mathrm{ar}[Y] = \mathbb{E}\left[Y^2\right] - (\mathbb{E}[Y])^2 = np + n(n-1)p^2 - n^2p^2 = np - np^2 = np(1-p)$. The moment generating function is can be computed in a similar way.

PROPOSITION E.2.10. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space and let* $Y \sim \mathrm{Bin}(n, p)$ *for some* $n \in \mathbb{N}$ *and* $p \in [0,1]$. *Then,*

$$M_Y(t) = \left(1 - p + pe^t\right)^n.$$

PROOF. Note that $Y = \sum_{i=1}^{n} X_i$ where as before X_i are iid Bernoulli with parameter p and so

$$\begin{aligned}
M_Y(t) &= \mathbb{E}\left[e^{tY}\right] \\
&= \mathbb{E}\left[e^{t\sum X_i}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} e^{tX_i}\right] \\
&= \prod_{i=1}^{n} \mathbb{E}\left[e^{tX_i}\right] \\
&= \left(1 - p + pe^t\right)^n
\end{aligned}$$

wheree the fourth equality is again due to Proposition 9.4.21 and the last is Proposition E.2.6. □

E.2.6.3. *The geometric distribution.*
E.2.6.4. *The negative binomial distribution.*
E.2.6.5. *The multinomial distribution.*
E.2.6.6. *The discrete arcine distribution.*

E.2.6.7. *The Beta-binomial distribution.*

E.2.6.8. *The Beta-negative binomial distribution.*

E.2.6.9. *The discrete uniform distribution.* The discrete uniform distribution is the generalization of the distribution we encountered in Example 9.4.3. The idea is that it is a distribution supported on any finite set, where each element of the set has equal probability. Thus for a finite set A and a discrete uniform random variable X supported on A we have that $\mathbb{P}(X=k)=\frac{1}{|A|}\mathbb{1}\{k\in A\}$. Typically, the set $A=\{a,a+1,a+2,\ldots,b\}$ in which case the $|A|=b-a+1$. In the discussion below we will focus on the special case where $a=0$ and $b=n-1$ for some $n\in\mathbb{N}$.

Proposition E.2.11. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and suppose X is a discrete uniform distribution supported on $A=\{0,1,\ldots,n-1\}$. Then, the CDF of X is given*

$$F_X(x)=\frac{\lfloor x\rfloor+1}{n}\mathbb{1}\{x\in[0,n-1]\}+\mathbb{1}\{x>n-1\}.$$

Proof. Observe

$$\begin{aligned}F_X(x)&=\mathbb{E}[\mathbb{1}\{X\le x\}]\\&=\frac{1}{n}\sum_{a=0}^{n-1}\mathbb{1}\{a\le x\}\\&=\frac{\lfloor x\rfloor+1}{n}\mathbb{1}\{x\in[0,n-1]\}+\mathbb{1}\{x>n-1\}.\end{aligned}$$

□

Proposition E.2.12. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and suppose X is a discrete uniform distribution supported on $A=\{0,1,\ldots,n-1\}$. Then for any $k\in\mathbb{N}$*

$$\mathbb{E}\left[X^k\right]=\frac{1}{n}\sum_{a=0}^{n-1}a^k.$$

In particular

$$\mathbb{E}[X]=\frac{1}{2}(n-1)$$

and

$$\mathbb{E}\left[X^2\right]=\frac{(n-1)(2n-1)}{6}$$

and so

$$\mathbb{V}\mathrm{ar}[X]=\frac{n^2-1}{12}.$$

Proof. This is a standard application of Corollary 3.2.14. The specific formulas for the first and second moments follow from partial sum formulas for consecutive integers and squares, respectively. The variance then is given by $\mathbb{V}\mathrm{ar}[X]=\mathbb{E}\left[X^2\right]-(\mathbb{E}[X])^2$. □

Proposition E.2.13. *Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and let X be a discrete uniform random variable supported on $A=\{0,1,\ldots,n-1\}$. Then, the MGF of X is given*

$$M_X(t)=\frac{1-e^{nt}}{n(1-e^t)}\mathbb{1}\{t\neq 0\}+\mathbb{1}\{t=0\}.$$

PROOF. Observe that

$$\begin{aligned} M_X(t) &= \mathbb{E}\left[e^{tX}\right] \\ &= \frac{1}{n}\sum_{a=0}^{n-1} e^{ta} \\ &= \frac{1}{n}\sum_{a=0}^{n-1} \left(e^t\right)^a \\ &= \frac{1-e^{nt}}{n\left(1-e^t\right)}\mathbb{1}\left\{t \neq 0\right\} + \mathbb{1}\left\{t = 0\right\} \end{aligned}$$

by the geometric partial sum formula. □

E.2.7. Distributions associated with finite sampling models.

E.2.7.1. *The hypergeometric distribution.*
E.2.7.2. *The multivariate hypergeometric distribution.*
E.2.7.3. *The matching distribution.*
E.2.7.4. *The birthday distribution.*
E.2.7.5. *The coupon collector distribution.*
E.2.7.6. *The Polya distribution.*

E.2.8. Distributions associated with the Poisson process.

E.2.8.1. *The exponential distribution.*
E.2.8.2. *The Erlang distribution.*
E.2.8.3. *The Poisson distribution.*