Lebesgue Measure & Integration

Kexing Ying

January 11, 2021

Contents

1 Motivation

We recall from **Analysis I** the definition of the Darboux integral. While this notion of integration was sufficient for our use case last year, as we shall see, there are some limitations with this notion of integration. These limitations will be addressed by the means of measure theory.

Definition 1.1 (Darboux Integrable)**.** A function $f : [a, b] \to \mathbb{R}$ is called Darboux integrable if for any partition $\mathcal{P} = \{a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b\}$ for some $n \geq 1$ if $[a, b]$, by defining the lower and upper Darboux sums,

$$L(f, \mathcal{P}) = \sum_{i=1}^{n} (t_i - t_{i-1}) \inf_{t \in [t_{i-1}, t_i]} f(t),$$

and

$$U(f, \mathcal{P}) = \sum_{i=1}^{n} (t_i - t_{i-1}) \sup_{t \in [t_{i-1}, t_i]} f(t),$$

one has

$$\sup_{\mathcal{P}} L(f, \mathcal{P}) = \inf_{\mathcal{P}} U(f, \mathcal{P}).$$

If this is the case we define the integral of f over $[a, b]$ to be this value, i.e.

$$\int_a^b f := \sup_{\mathcal{P}} L(f, \mathcal{P}) = \inf_{\mathcal{P}} U(f, \mathcal{P}).$$

Many functions are Darboux integrable and in fact, as demonstrated last year, all functions in $C^{\circ}_{pw}([a, b])$, that is piecewise continuous functions on $[a, b]$ are Darboux integrable. Nonetheless, however, the class of Darboux integrable functions is also rather limited.

Consider the Dirichlet function

$$\mathbf{1}_{\mathbb{Q}}(x) := \begin{cases} 1, & x \in \mathbb{Q}; \\ 0, & x \in \mathbb{R} \setminus \mathbb{Q}. \end{cases}$$

That is, the indicator function for $\mathbb{Q}$. We see that $\mathbf{1}_{\mathbb{Q}}$ is not Darboux integrable since both $\mathbb{Q}$ and $\mathbb{R} \setminus \mathbb{Q}$ are dense in $\mathbb{R}$ and so, for any partition $\mathcal{P}$ of $[a, b]$, $L(\mathbf{1}_{\mathbb{Q}}, \mathcal{P}) = 0$ while $U(\mathbf{1}_{\mathbb{Q}}, \mathcal{P}) = 1$. This is not ideal, since, as $\mathbb{Q}$ is countable while $\mathbb{R} \setminus \mathbb{Q}$ is not, we intuitively expect that a satisfactory theory of integration would assign $\int_a^b \mathbf{1}_{\mathbb{Q}} = 0$.

Moreover, by defining $P = (q_n)_{n \in \mathbb{N}} \subseteq \mathbb{Q}$ be some enumeration of $\mathbb{Q} \cap [a, b]$, we can define the following sequence of functions,

$$f_n(x) := \begin{cases} 1, & x \in \{q_0, \cdots, q_n\}; \\ 0, & \text{otherwise.} \end{cases}$$

It is not difficult to see that $\int_a^b f_n = 0$ for all n and $f_n \to \mathbf{1}_{\mathbb{Q}}$ pointwise. However, this implies

$$0 = \lim_{n \to \infty} \int_a^b f_n \neq \int_a^b \lim_{n \to \infty} f_n = \int_a^b \mathbf{1}_{\mathbb{Q}},$$

and in fact, the right hand side is not even defined (as $\mathbf{1}_{\mathbb{Q}}$ is not Darboux integrable)!

To solve this issue we will introduce the notion of the Lebesgue measure and furthermore, its associated Lebesgue integral which extends our Darboux integral such that it has the "nice" properties we desire.

We will in this course also look at L^p spaces. From the perspective of analysis, it is often convenient to work in Banach spaces (complete normed vector spaces) such that we can utilise many existing theorems we have proved in **Analysis II**, e.g. Banach's fixed point theorem. For instance, one can endow $C^{\circ}_{pw}([a,b])$ with the (semi-)norm

$$\|f\|_{L^1} := \int_a^b |f| \, .$$

Then, by considering the aforementioned sequence $(f_n) \subseteq C^{\circ}_{pw}([a,b])$, one can easily show that (f_n) is a Cauchy sequence with respect to $\| \cdot \|_{L^1}$. However, $f_n \to \mathbf{1}_{\mathbb{Q}}$ pointwise. This motivates us to introduce the Banach space $L^1([a,b])$ of integrable functions, and more generally, L^p-spaces later in the course.

Lastly, as we have seen within last term's probability module, measure theory lays below as the foundations for probability theory. As a quick reminder, we recall that a probability space is a special type of measure space and random variables defined on these probability spaces are simply measurable functions to $\mathbb{R}$ (or more exotic fields). This can be interpreted with connotations to real world situations in several ways.

2 Measures and Measure Spaces

2.1 Algebra, σ-Algebra and Measures

As we would like an adequate theory to assign a notion of "size" on sets, we need to construct a function from a set of sets to $\overline{\mathbb{R}_0^+}$. To achieve this, the natural idea is to construct a function with domain being the power set, however, this is not necessarily always possible or meaningful. Thus, instead of assigning every subset of some set a size, we only look at some collection of "nice" sets.

Definition 2.1 (Algebra)**.** Let X be some set and suppose we denote $\mathcal{X}$ for the power set of X, then a family $\mathcal{A} \subseteq \mathcal{P}(X)$ is called an algebra over X if

- $X \in \mathcal{A}$;
- for all $A \in \mathcal{A}$, $A^c = X \setminus A \in \mathcal{A}$;
- if $(A_k)_{i=1}^n$ is a finite sequence of sets in $\mathcal{A}$, then $\bigcup_{k=1}^n A_k \in \mathcal{A}$.

Definition 2.2 (σ-algebra)**.** Let X be a set, then a σ-algebra $\mathcal{A}$ on X is an algebra on X such that $\mathcal{A}$ is closed under countable unions, i.e. if $(A_k)_{k=1}^\infty$ is a sequence of sets in $\mathcal{A}$, then $\bigcup_{k=1}^\infty A_k \in \mathcal{A}$.

As σ-algebras (and algebras) are simply sets of sets, there is an induced order on σ-algebras by $\subseteq$. If there are two σ-algebras $\mathcal{A}, \mathcal{B}$ such that $\mathcal{A} \subseteq \mathcal{B}$, then we say $\mathcal{A}$ is coarser than $\mathcal{B}$.

Trivially, we find $\{\varnothing, X\}$ is a σ-algebra. Indeed, this is the coarsest σ-algebra. Furthermore, given a set X and a subset $A \subseteq X$, we have $\{\varnothing, A, A^c, X\}$ is also a σ-algebra.

While every σ-algebra is also an algebra, the converse is not true. An counter-example of this is by consider the algebra

$$\mathcal{A} := \{\varnothing\} \cup \left\{ U \mid \exists \bigcup_{k=1}^m (a_k, b_k], m \geq 1, 0 \leq a_k < b_k \leq 1 \right\},$$

on $X = (0, 1]$. We see that $\mathcal{A}$ is an algebra (since $(a, b]^c = (0, a] \cup (b, 0]$) however $\mathcal{A}$ is not a σ-algebra on X since we can define the sequence $A_k = (0, 1 - 1/k] \in \mathcal{A}$ but $\bigcup_k A_k = (0, 1) \notin \mathcal{A}$.

Proposition 1. Let $\mathcal{F}$ be an arbitrary collection of σ-algebra (or algebras) over X. Then the intersections

$$\bigcap \mathcal{F} := \bigcap_{\mathcal{A} \in \mathcal{F}} \mathcal{A},$$

is a σ-algebra (or algebra).

Proof. Straight forward by definition. $\square$

With this, we have a notion of infimum on σ-algebras and hence, we can also define a notion closure.

Definition 2.3 (σ-algebra Generated by a set)**.** Let $\mathcal{C} \subseteq \mathcal{P}(X)$, then

$$\sigma(\mathcal{C}) := \bigcap \{\mathcal{A} \mid \mathcal{C} \subseteq \mathcal{A} \wedge \mathcal{A} \in \mathcal{F}\},$$

where $\mathcal{F}$ is the set of all σ-algebras on X.

As previously shown, $\sigma(\mathcal{C})$ is a intersection of σ-algebras, and so the name suggests, the σ-algebra generated by $\mathcal{C}$ is the smallest σ-algebra containing $\mathcal{C}$. Indeed, we find $\sigma(\varnothing) = \{\varnothing, X\}$ and $\sigma(A) = \{\varnothing, A, A^c, X\}$. Moreover, we see that $\mathcal{C}$ is a σ-algebra if and only if $\sigma(\mathcal{C}) = \mathcal{C}$.

Definition 2.4 (Borel σ-algebra)**.** If $(X, \mathcal{T})$ is a topological space, then the Borel σ-algebra over X is

$$B(X) := \sigma(\mathcal{T}).$$

Unlike topologies, the unions and intersections in a σ-algebra is treated symmetrically.

Proposition 2. If $\mathcal{A}$ is a σ-algebra, then if $(A_k)_{k=1}^{\infty}$ is a sequence of sets in $\mathcal{A}$, then

$$\bigcap_{k=1}^{\infty} A_k \in \mathcal{A}.$$

Proof. By considering de Morgen's identity, we have $\bigcap_{k=1}^{\infty} A_k = (\bigcup_{k=1}^{\infty} A_k^c)^c$. So, since $\bigcup_{k=1}^{\infty} A_k^c \in \mathcal{A}$ as each component is, $\bigcup_{k=1}^{\infty} A_k^c)^c$ and hence $\bigcap_{k=1}^{\infty} A_k$ is also in $\mathcal{A}$. $\square$

Definition 2.5 (Measurable Space)**.** A set X equipped with a σ-algebra $\mathcal{A}$ is called a measurable space and is written as a tuple $(X, \mathcal{A})$. Furthermore, if $A \subseteq X$ is in $\mathcal{A}$, then we say A is a measurable set.

Definition 2.6 (Measure)**.** Let $(X, \mathcal{A})$ is a measurable space. Then a measure on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ such that

- $\mu(\varnothing) = 0$;
- if $(A_k)_{k=1}^{\infty} \subseteq \mathcal{A}$ is a sequence of pairwise disjoint sets, then

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

.

We call the second property σ-additivity.

Definition 2.7 (Measure Space)**.** A measurable space $(X, \mathcal{A})$ equipped with the measure μ is called a measure space and is written as a triplet $(X, \mathcal{A}, \mu)$.

A commonly used measure on any arbitrary measurable space $(X, \mathcal{A})$ is the counting measure μ. As the name suggests, for all $A \in \mathcal{A}$, $\mu(A) = |A|$ if A is finite and ∞ otherwise. Another example of a measure is the Dirac measure $\delta_x : \mathcal{A} \to [0, \infty]$ for some $x \in X$ where for all $A \in \mathcal{A}$, $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. For the last example, let X be uncountable and let $\mathcal{A} := \{A \subseteq X \mid A \text{ or } A^c \text{ is uncountable}\}$. Then, one can show that $\{X, \mathcal{A}\}$ forms a measurable space and we find that the function $\mu : \mathcal{A} \to [0, \infty]$ defined as $\mu(A) = 0$ if A is countable and $\mu(A) = 1$ if A^c is countable is a measure on $(X, \mathcal{A})$.

Proposition 3. Let $(X, \mathcal{A}, \mu)$ be a measure space. Then,

- if $A, B \in \mathcal{A}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$;
- if $n \geq 1$, $(A_k)_{k=1}^n$ is a sequence of pairwise disjoint sets in $\mathcal{A}$, then

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

- if $(A_k)_{k=1}^\infty$ is a sequence of monotonically increasing sets in $\mathcal{A}$, then

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

We note that the limit in part 3 exists since the limit in monotonically increasing on the extended reals (so bounded by ∞).

- if $(A_k)_{k=1}^\infty$ is a sequence of monotonically decreasing sets in $\mathcal{A}$, if $\mu(A_1) < \infty$, then

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

- if $A \in \mathcal{A}$ and $(A_k)_{k=1}^\infty$ is a sequence in $\mathcal{A}$, then $\mu(A) \leq \sum_{k=1}^\infty \mu(A_k)$.

The last property is refereed to as σ-sub-additivity.

Proof. Part 2 is trivial.

(Part 1) As $B = A \sqcup B \setminus A = A \sqcup (A^c \cap B)$ where $A^c \cap B$ is measurable since both A^c and B are. So, by σ-additivity,

$$\mu(B) = \mu(A \sqcup (A^c \cap B)) = \mu(A) + \mu(A^c \cap B) \geq \mu(A).$$

(Part 3) Define $B_1 = A_1$ and $B_{k+1} = A_{k+1} \setminus A_k$. Then, $(B_k)_{k=1}^\infty$ is a sequence of disjoint subset in $\mathcal{A}$. So, by σ-additivity,

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

(Part 4) Define $B_k = A_1 \setminus A_k$, then $(B_k)_{k=1}^\infty$ is a sequence of monotonically increasing sets in $\mathcal{A}$. So, by part 3,

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

Furthermore, as $A \subseteq B$ implies $\mu(B) - \mu(A) = \mu(B \setminus A)$, we have

$$\mu(A_1) - \mu\left(\bigcap_{k=1}^\infty A_k\right) = \mu\left(A_1 \setminus \bigcap_{k=1}^\infty A_k\right) = \mu\left(\bigcup_{k=1}^\infty B_k\right) = \lim_{k\to\infty} \mu(B_k) = \mu(A_1) - \lim_{k\to\infty} \mu(A_k),$$

hence the result.

(Part 5) Exercise. $\square$

2.2 Hahn-Carathéodory Extension Theorem

While it is fun to work with abstract measures in general, it is also useful to consider specific measures. In this section we will look at how to construct measures on arbitrary spaces. The general idea for constructing measures is that, given a *premeasure* $\tilde{\mu}$, we may extend it to an *outer measure* μ^*, and finally with the outer measure, we may restrict it to a measure μ.

Definition 2.8 (Premeasure)**.** Let $\mathcal{A}$ be an algebra on X. A function $\mu : \mathcal{A} \to [0, \infty]$ is a premeasure if it satisfies $\mu(\varnothing) = 0$ and countable additivity; that is, given $(A_n)_{n=1}^{\infty} \subseteq \mathcal{A}$ that is pairwise disjoint,

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

provided $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Definition 2.9 (Outer Measure)**.** A function $\mu : \mathcal{P}(X) \to [0, \infty]$ is an outer measure if it satisfies $\mu(\varnothing) = 0$ and sub-additivity, that is, if $(A_k)_{k=1}^{\infty}$ is a sequence of subsets of X, and $A \subseteq \bigcup A_k$ then

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

Lemma 2.1 (Monoticity of Outer Measures)**.** Let $\mu : \mathcal{P}(X) \to [0, \infty]$ be an outer measure on X. Then for all $A \subseteq B \subseteq X$,

$$\mu(A) \leq \mu(B).$$

Proof. By defining the sequence $A_1 = B$ and $A_n = \varnothing$ for all $n > 1$, by the sub-additivity property of outer measure, we have $A \subseteq B = \bigcup A_n$, and so

$$\mu(A) \leq \sum \mu(A_n) = \mu(B) + \mu(\varnothing) + \cdots = \mu(B).$$

$\square$

In order to prove the aforementioned argument, we will require some proposition. However, as will be demonstrated below, the notion of premeasure is in fact, too strict. Thus, for that reason, let us introduce the following definition.

Definition 2.10 (Cover)**.** A family of sets $\mathcal{K} \subseteq \mathcal{P}(X)$ is called a cover of X if $\varnothing \in \mathcal{K}$ and there exists a sequence $(K_n)_{n=1}^{\infty}$ in $\mathcal{K}$ such that

$$X \subseteq \bigcup_{n=1}^{\infty} K_n.$$

Remark. We note that this is a different definition of covers than that we had learnt for the definition compactness of topological spaces.

Straight away, we see that every algebra $\mathcal{A}$ of X is a cover of X as $X \in \mathcal{A}$ and so we can simply let the the countable sequence be $K_n = X$. Thus, every proposition we prove about covers apply also to algebras.

Proposition 4. Let $\mathcal{K}$ be a cover of X, and $\tilde{\mu} : \mathcal{K} \to [0, \infty]$ be a map such that $\tilde{\mu}(\varnothing) = 0$. Then

$$\mu^* : \mathcal{P}(X) \to [0, \infty] := A \mapsto \inf \left\{ \sum_{j=1}^{\infty} \tilde{\mu}(K_j) \mid (K_j)_{j=1}^{\infty} \subseteq \mathcal{K} \wedge A \subseteq \bigcup_{j=1}^{\infty} K_j \right\}$$

is an outer measure on X.

Proof. μ^* is well defined since for all $A \subseteq X$, by the definition of covers, there exists some $(K_n)_{n=1}^{\infty}$, $\bigcup K_n \supseteq X \supseteq A$, and so, we are taking the infimum of a non-empty set. Clearly $\mu^*(\varnothing) = 0$ since we can choose $K_n = \varnothing$ and so, it remains to show sub-additivity.

Let $(A_n)_{n=1}^{\infty}$ be a sequence of subsets of X, and suppose $A \subseteq \bigcup A_n$. For each $n \in \mathbb{N}$ and $\epsilon > 0$, by the definition of infimum, there exists some $(K_{n,j})_{j=1}^{\infty} \subseteq \mathcal{K}$ such that $A_n \subseteq \bigcup K_{n,j}$ and

$$\sum_{j=1}^{\infty} \tilde{\mu}(K_{n,j}) < \mu^*(A_n) + 2^{-n}\epsilon.$$

Now, by the fact that $A \subseteq \bigcup A_n$, we have $A \subseteq \bigcup_n \bigcup_j K_{n,j}$, and so, again by the definition of μ^*, we have

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

As $\epsilon > 0$ was arbitrary,

$$\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n),$$

so μ^* is a outer measure. ☐

With that, we can naturally construct a outer measure from a premeasure (and in fact, weaker notions suffices). This, outer measure however, in some sense overshoots measures as it is defined on all subsets of X, and does not attain σ-additivity but only σ-sub-additivity.

Lemma 2.2. Let μ^* be an outer measure on X, then

$$\Sigma := \{A \subseteq X \mid \mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)\}$$

is a σ-algebra on X.

Remark. By sub-additivity, we see that Σ is equivalently defined by requiring $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$.

Proof. $\varnothing \in \Sigma$ since for all $B \subseteq X$, $B \cap \varnothing = \varnothing$ and $B \cap \varnothing^c = B$ and so

$$\mu^*(B \cap \varnothing) + \mu^*(B \cap \varnothing^c) = \mu^*(\varnothing) + \mu^*(B) = \mu^*(B).$$

Let $A \in \Sigma$, then for all $B \subseteq X$,

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c) = \mu^*(B \cap (A^c)^c) + \mu^*(B \cap A^c),$$

and so, $A^c \in \Sigma$.

To show that Σ is closed under union, let us first show that Σ is closed under finite intersections. Let $A_1, A_2 \in \Sigma$, then for all $B \subseteq X$,

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\ &= \mu^*(B \cap A_1 \cap A_2) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) \\ &= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\ &= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c),\end{aligned}$$

and so $A_1 \cap A_2 \in \Sigma$. Now, suppose (A_n) is a sequence of disjoint sets in Σ, it suffices to show $\bigcup_{n=1}^\infty A_n \in \Sigma$ (we can assume this since given a sequence of sets, we may define another disjoint sequence such that their union are equal). Since the sequence is pairwise disjoint, we have

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\ &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c).\end{aligned}$$

Indeed, as $A_1 \cap A_2 = \varnothing$, $A_1^c \cap A_2 = A_2$ and so,

$$\begin{aligned}\mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_2) + \mu^*(B \cap A_1^c \cap A_2^c) \\ &= \cdots \\ &= \sum_{j=1}^n \mu^*(B \cap A_j) + \mu^*\left(B \cap \bigcap_{j=1}^n A_j^c\right) \\ &= \sum_{j=1}^n \mu^*(B \cap A_j) + \mu^*\left(B \cap \overline{\bigcup_{j=1}^n A_j}\right)\end{aligned}$$

for all $n \in \mathbb{N}$. By sub-additivity, we have

$$\mu^*\left(B \cap \bigcup_{j=1}^n A_j\right) \leq \sum_{j=1}^n \mu^*(B \cap A_j),$$

and so, by taking $n \to \infty$, we have

$$\mu^*(B) \geq \mu^*\left(B \cap \bigcup_{j=1}^\infty A_j\right) + \mu^*\left(B \cap \overline{\bigcup_{j=1}^\infty A_j}\right),$$

implying $\bigcup A_n \in \Sigma$. □

Theorem 1 (Hahn-Carathéodory Extension Theorem)**.** Let X be an arbitrary set, $\mathcal{A}$ an algebra over X and $\tilde{\mu} : \mathcal{A} \to [0, \infty]$ a premeasure on X. Then by defining μ^* as in proposition 4, Σ as in lemma 2.2, and by defining $\mu = \mu^* \mid_\Sigma$, then,

- (X, Σ, μ) is a measure space;
- $\mathcal{A} \subseteq \Sigma$;

- for all $A \in \mathcal{A}$, $\mu(A) := \mu^*(A) = \tilde{\mu}(A)$.

Proof. By construction, we have μ^* is an outer measure on X and Σ is a σ-algebra on X, and so, for the first part of the proof, it remains to show that $\mu = \mu^* \mid_\Sigma$ is a measure on the measurable space (X, Σ).

Indeed, $\mu(\varnothing) = 0$ as $\mu^*(\varnothing) = 0$ since it is an outer measure, so let us consider σ-additivity. Let $A, B \in \Sigma$, $A \cap B = \varnothing$, then

$$\begin{aligned} \mu(A \cup B) = \mu^*(A \cup B) &= \mu^*((A \cup B) \cap A) + \mu^*((A \cup B) \cap A^C) \\ &= \mu^*(A) + \mu^*(B), \end{aligned}$$

and so, by induction μ is finitely additive. So, if $(A_n)_{n=1}^\infty$ is a sequence of pairwise disjoint measurable sets, then, for all $m \in \mathbb{N}$,

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

where the inequality is due to the monotonicity of outer measures. So, by taking $m \to \infty$ we have

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

However, by the sub-additivity of outer measures,

$$\sum_{n=1}^\infty \mu(A_n) \geq \mu\left(\bigcup_{n=1}^\infty A_n\right),$$

and so $\sum \mu(A_n) \geq \mu\left(\bigcup A_n\right)$ resulting in μ being a measure.

For the second part of the theorem, it suffices to show that for all $A \in \mathcal{A}$, for all $B \subseteq X$,

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c).$$

By recalling the definition of μ^*, for all $\epsilon > 0$, there exists some sequence $(K_j)_{j=1}^\infty \subseteq \mathcal{A}$, $B \subseteq \bigcup K_j$,

$$\sum_{j=1}^\infty \tilde{\mu}(K_j) \leq \mu^*(B) + \epsilon.$$

Now, as $B \subseteq \bigcup K_j$, $B \cap A \subseteq \bigcup (K_j \cap A)$ and $B \cap A^c \subseteq \bigcup (K_j \cap A^c)$, and so, by sub-additivity

$$\begin{aligned} \mu^*(B \cap A) + \mu^*(B \cap A^c) &\leq \sum \mu^*(K_j \cap A) + \sum \mu^*(K_j \cap A^c) \\ &= \sum (\tilde{\mu}(K_j \cap A) + \tilde{\mu}(K_j \cap A^c)) \\ &= \sum \tilde{\mu}(K_j) \leq \mu^*(B) + \epsilon, \end{aligned}$$

where we used the last part of the theorem to exchange μ^* for $\tilde{\mu}$. Thus, as $\epsilon > 0$ was arbitrary, the inequality follows.

Lastly, to show that μ^* agrees with $\tilde{\mu}$ on $\mathcal{A}$, let $A \in \mathcal{A}$. By the monotonicity of outer measures, we have $\mu^*(A) \geq \tilde{\mu}(A)$ and so, it suffices to show there exists some $(K_j) \subseteq \mathcal{A}$ such that $A \subseteq \bigcup K_j$ and $\sum \tilde{\mu}(K_j) = \tilde{\mu}(A)$. But, this is trivial by simply choosing $A_1 = A$ and $A_n = \varnothing$ for all $n > 1$ and so, we are done! $\square$

With the Hahn-Carathéodory extension theorem, we can obtain a measure (and also a measurable space), just from a premeasure on some algebra. This is a very useful theorem as algebras and premeasures are rather weak notions and can be easily constructed through many means. However, this theorem does not yet guarantee the uniqueness, so one might ask whether or not there are other extensions. We shall take a look at that now.

Definition 2.11 (σ-Finite)**.** A premeasure $\tilde{\mu}$ on the algebra $\mathcal{A}$ on X is σ-finite if there exists disjoint sets $(S_n)_{n=1}^{\infty} \subseteq \mathcal{A}$ such that $X = \bigcup_{n=1}^{\infty} S_n$ and $\tilde{\mu}(S_n) < \infty$ for all $n \geq 1$.

Theorem 2 (Uniqueness of the Hahn-Carathéodory Extension)**.** Under the assumption of the Hahn-Carathéodory extension theorem, and if $\tilde{\mu}$ is σ-finite. Then, for all outer measures $\nu : \mathcal{P}(X) \to [0, \infty]$,

$$\nu \mid_{\Sigma} = \mu,$$

where μ is the outer measure obtained from the Hahn-Carathéodory extension.

Proof. Let $A \in \Sigma$, and we will first show $\nu(A) \leq \mu(A)$. Let $(A_n)_{n=1}^{\infty} \subseteq \mathcal{A}$ be some sequence such that $A \subseteq \bigcup A_n$. Then, by the sub-additivity of ν,

$$\nu(A) \leq \sum \nu(A_n) = \sum \tilde{\mu}(A_n).$$

So, by taking the infimum over all possible (A_n), the inequality is achieved.

To show the reverse inequality, let us suppose $(*)$ that there exists some $S \in \mathcal{A}$ such that $A \subseteq S$ and $\mu(S) < \infty$. Then,

$$\nu(A) + \nu(S \cap A^c) \leq \mu(A) + \mu(S \cap A^c) = \mu(S) = \tilde{\mu}(S) = \nu(S).$$

Now, by sub-additivity,

$$\nu(S) \leq \nu(A) + \nu(S \cap A^c),$$

and so, $\nu(A) + \nu(S \cap A^c) \geq \mu(A) + \mu(S \cap A^c)$, hence[1],

$$\nu(A) - \mu(A) \geq \mu(S \cap A^c) - \nu(S \cap A^c) \geq 0,$$

as $\mu \geq \nu$ as shown previously. Thus, $\nu(A) \geq \mu(A)$ and so, $\nu \mid_{\Sigma} = \mu$.

Now, to relax the assumption that such S exists, we shall use the σ-finiteness of $\tilde{\mu}$. Let $(S_n) \subseteq \mathcal{A}$ be the sequence as described by the σ-finiteness of $\tilde{\mu}$, and define $A_n = A \cap S_n$. Then, by construction, the sets A_n are pairwise disjoint and $A = \bigcup A_n$. Since $\mu = \nu$ on sets, which $(*)$ is satisfied, we have

$$\mu\left(\bigcup_{n=1}^{m} A_n\right) = \nu\left(\bigcup_{n=1}^{m} A_n\right),$$

[1]Wlog. $\mu(A) < \infty$ since if otherwise, $\infty < \nu(A) + \nu(S \cap A^c)$. Since $\nu(S \cap A^c) \leq \nu(S) \leq \mu(S) < \infty$, $\nu(A) = \infty$ and hence $\nu(A) \geq \mu(A)$.

for all $m \geq 1$ (since we can choose $S = \bigcup_{n=1}^m S_n$). Now, by monotonicity,

$$\nu(A) = \nu\left(\bigcup_{n=1}^{\infty} A_n\right) \geq \nu\left(\bigcup_{n=1}^{m} A_n\right) = \mu\left(\bigcup_{n=1}^{m} A_n\right),$$

for all m, the inequality is achieved by taking $m \to \infty$. □

2.3 The Lebesgue Measure

With the Hahn-Carathéodory extension theorem in mind, we may construct the Lebesgue measure on Euclidean spaces. As with the proof of the extension theorem, we shall construct a premeasure assigning $b - a$ to every interval (a, b) and then, extend that to a general measure.

Definition 2.12. For $a = (a_1, \cdots, a_n) \in \mathbb{R}^n$, $b = (b_1, \cdots, b_n) \in \mathbb{R}^n$, we define the interval between a and b to be the set

$$(a, b) := \prod_{k=1}^{n} (a_k, b_k) = \{(x_1, \cdots, x_n) \mid a_k < x_k < b_k \;\; 1 \leq k \leq n\},$$

if $a_k < b_k$ for all k and $(a, b) = \varnothing$ otherwise.

Similarly, we define the half open and closed intervals in $\mathbb{R}^n$ with $(a, b] := \prod_{k=1}^n (a_k, b_k]$, $[a, b) := \prod_{k=1}^n [a_k, b_k)$ and $[a, b] := \prod_{k=1}^n [a_k, b_k]$. Indeed, we also allow $\pm\infty$ as an endpoint and we shall from this point forward refer to such sets as intervals in $\mathbb{R}^n$.

Definition 2.13 (Elementary Figure)**.** A set $I \subseteq \mathbb{R}^n$ is an elementary figure if its the union of finitely many disjoint intervals; and in this section, we shall denote $\mathcal{A} \subseteq \mathcal{P}(\mathbb{R}^n)$ for the set of elementary figures in $\mathbb{R}^n$.

We see that $\mathcal{A}$ is an algebra on $X = \mathbb{R}^n$ and so, we may define a premeasure on $\mathcal{A}$. Let $\tilde{\lambda} : \mathcal{A} \to [0, \infty]$ be the function such that for all $a, b \in \mathbb{R}^n$,

$$\tilde{\lambda}((a, b)) = \tilde{\lambda}((a, b]) = \tilde{\lambda}([a, b)) = \tilde{\lambda}([a, b]) := \prod_{k=1}^{n} (b_k - a_k),$$

if $a_k < b_k$ and 0 otherwise. Furthermore, for $\bigcup_{k=1}^m I_k \in \mathcal{A}$,

$$\tilde{\lambda}\left(\bigcup_{k=1}^{m} I_k\right) = \sum_{k=1}^{m} \tilde{\lambda}(I_k),$$

where $I_1, \cdots, I_k$ are disjoint intervals.

Lemma 2.3. The map $\tilde{\lambda}$ is a premeasure on $\mathcal{A}$.

Proof. As it is trivially true that $\tilde{\lambda}(\varnothing) = 0$ to show that $\tilde{\lambda}$ is a premeasure, it suffices to show σ-additivity within $\mathcal{A}$.

We note that $\tilde{\lambda}$ is finitely additive by definition, hence monotone on $\mathcal{A}$, and therefore if $I = \bigcup_{k=1}^{\infty}$ where I_k are disjoint intervals,

$$\tilde{\lambda}(I) \geq \tilde{\lambda}\left(\bigcup_{k=1}^{m} I_k\right) = \sum_{k=1}^{m} \tilde{\lambda}(I_k),$$

for all $m \in \mathbb{N}$. Thus, by taking $m \to \infty$,

$$\tilde{\lambda}(I) \geq \sum_{k=1}^{\infty} \tilde{\lambda}(I_k).$$

For the reverse inequality, we use the what is called a *compactness argument* to reduce to finite additivity. Wlog. we may assume $\sum \tilde{\lambda}(I_k) < \infty$ (since if otherwise the reverse inequality is trivial) and that I is a single interval with end points a, b (since if the inequality is true for a single interval, it is also true for the sum of finitely many intervals).

Let $\bar{I}$ be the closure of I and for all $L > 0$ we define $\bar{I}_L := \bar{I} \cap [-L, L]^n$. By Heine-Borel, $\bar{I}_L$ is compact and moreover, by taking $L \to \infty$,

$$\tilde{\lambda}(\bar{I}_L) \to \tilde{\lambda}(\bar{I}) = \tilde{\lambda}(I).$$

Now, for all intervals J with end points α, β, we define the interval $J^\epsilon \supseteq J$ with endpoints $\alpha^\epsilon \neq \alpha$, $\beta^\epsilon \neq \beta$ such that

$$\tilde{\lambda}(J^\epsilon) \leq (1+\epsilon)^n \tilde{\lambda}(J).$$

Lastly, for all $k \in \mathbb{N}$, we define the open intervals $\tilde{I}_k$ with $\tilde{I}_k \supseteq I_k^\epsilon$ satisfying

$$\tilde{\lambda}(\tilde{I}_k) < (1+\epsilon)^n \lambda(I_k) + \epsilon 2^{-k}.$$

So $\bar{I}_L \subseteq \bar{I} \subseteq \bigcup_{k=1}^{\infty} I_k^\epsilon \subseteq \bigcup_{k=1}^{\infty} \tilde{I}_k$, and $\{\tilde{I}_k\}$ forms an open cover of $\bar{I}_L$, and hence, by compactness, there exists a finite subcover for $\bar{I}_L$, that is, there exists some m such that (by reordering), $\bar{I}_L \subseteq \bigcup_{k=1}^{m} I_k$. It follows that

$$\tilde{\lambda}(\bar{I}_L) \leq \tilde{\lambda}\left(\bigcup_{k=1}^{m} \tilde{I}_k\right) \leq \sum_{k=1}^{m} \tilde{\lambda}(\tilde{I}_k) \leq (1+\epsilon)^n \sum_{k=1}^{\infty} \lambda(I_k) + \epsilon.$$

Thus, by taking $\epsilon \to 0$ and then $L \to \infty$ we have the reverse inequality and so $\tilde{\lambda}$ is a premeasure on $\mathcal{A}$. ☐

With this lemma, one can immediately apply the Hahn-Carathéodory extension theorem resulting in the Lebesgue measure on Euclidean spaces, and furthermore, by considering $\tilde{\lambda}((z, z+1]) = 1 < \infty$, and $\bigcup_{z \in \mathbb{Z}^n} (z, z+1] = \mathbb{R}^n$, we have $\tilde{\lambda}$ is σ-finite, and hence, the Lebesgue measure is unique.

Lemma 2.4. Let $\mathcal{B}(\mathbb{R}^n)$ be the Borel σ-algebra on $\mathbb{R}^n$, then $\mathcal{B}(\mathbb{R}^n) \subseteq \Sigma$ where Σ is the σ-algebra induced by the Lebesgue measure.

Proof. Since $\mathcal{B}(\mathbb{R}^n)$ is the σ-algebra generated by the set of open sets of $\mathbb{R}^n$, it suffices to show that for all open sets $O \subseteq \mathbb{R}^n$, $O \in \sigma(\mathcal{A}) \subseteq \Sigma$.

For all $m \in \mathbb{N}$, define
$$C_m := \{[z, z + 2^{-m}) \mid z \in 2^{-m}\mathbb{Z}^n\} \subseteq \mathcal{A},$$
that is the grid of half open cubes covering $\mathbb{R}^n$ with individual cubes having length 2^{-m}. Then, by letting $C'_m := \{U \in C_m \mid U \subseteq O\}$, we have $C = \bigcup_{m\in\mathbb{N}} C'_m$ which is a countable set of half open cubes. Now, we see that $C = O$ since $C \subseteq O$ trivially and for all $o \in O$, as O is open, there exists some $\epsilon > 0$ such that $B_\epsilon(o) \subseteq O$ and so, o is contained in one of the cubes with length $< 2^{-m}$ where $m > 2/\epsilon$ and so $O \in \sigma(\mathcal{A}) \subseteq \Sigma$. □

With that, we see that all Borel sets in $\mathbb{R}^n$ are Lebesgue measurable and restricting $\tilde{\lambda}$ onto $\mathcal{B}(\mathbb{R}^n)$, we have the following measure space on $\mathbb{R}^n$.

Definition 2.14. The Lebesgue measure $\lambda : \mathcal{B}(\mathbb{R}^n) \to [0, \infty]$ is the Hahn-Carathéodory extension of $\tilde{\lambda}$ restricted onto $\mathcal{B}(\mathbb{R}^n)$.

We note that, if $(X, \mathcal{F}, \mu)$ is a measure space and $A \in \mathcal{F}$, then by defining $\mathcal{F} \mid_A := \{A \cap B \mid B \in \mathcal{F}\}$ and $\mu \mid_A (B) := \mu(B)$ for all $B \in \mathcal{F}$, $B \subseteq A$, it is easy to see that $\mathcal{F} \mid_A$ is a σ-algebra on A and $\mu \mid_A$ is a measure on $(A, \mathcal{F} \mid_A)$ and is called the restriction of μ to A. Indeed, with this in mind, we see that the Lebesgue measure can be restricted on small sets such as intervals; in particular, by restricting the Lebesgue measure on $[0, 1]$, the resulting measure $\lambda \mid_{[0,1]}$ is a probability measure.

2.4 Lebesgue Measurable Sets

We investigate which real sets are Lebesgue measurable.

Proposition 5. For all $A \in \mathcal{B}(\mathbb{R}^n)$,
$$\lambda(A) = \inf_{G \supseteq A; G \text{open}} \lambda(G).$$

Proof. Since measures are monotone, $(\leq)$ is established. By recalling the construction of the Lebesgue measure, and by the properties of inf, for all $\epsilon > 0$, there exists some $(K_n)_{n=1}^\infty \subseteq \mathcal{A}$, such that $A \subseteq \bigcup K_n$ and
$$\lambda(A) + \epsilon = \lambda^*(A) + \epsilon \geq \sum_{n=1}^{\infty} \tilde{\lambda}(K_n) \geq \tilde{\lambda}\left(\bigcup K_n\right) = \lambda\left(\bigcup K_n\right) \geq \inf_{G \supseteq A; G \text{open}} \lambda(G).$$
since the union of open sets is open, $\bigcup K_n$ is open and contains A. So, as $\epsilon > 0$ was arbitrary, $(\geq)$ is established. □

By inspection of the proof, we find the statement to be true for any real sets provided with change λ to the outer measure λ^* on the left hand side. This *regularity* of measurable sets is expressed in the following.

Proposition 6. If $A \in \mathcal{B}(\mathbb{R}^n)$, then for all $\epsilon > 0$, there exists some $G \supseteq A$, G open such that $\lambda(G \setminus A) < \epsilon$.

Proof. One can of course prove it using the construction of the Lebesgue measure, however, a stronger proposition holds. $\square$

Proposition 7. Let $\mathcal{A}$ be an algebra and μ an measure on $\sigma(\mathcal{A})$ which is σ-finite on $\mathcal{A}$. Then, for all $A \in \sigma(\mathcal{A})$, and $\epsilon > 0$, there exist disjoint sets $(A_n)_{n=1}^{\infty} \in \mathcal{A}$ such that $A \subseteq \bigcup A_n$ and $\mu\left(\bigcup A_n \setminus A\right) < \epsilon$.

Proof. Since μ is σ-finite on $\mathcal{A}$, there exists a sequence of sets $(S_n) \subseteq \mathcal{A}$ such that $X \subseteq \bigcup S_n$ and $\mu(S_n) < \infty$ for all n.

By restricting μ on $\mathcal{A}$, we have a pre-measure on $\mathcal{A}$ whose extended outer-measure μ^* agrees with μ on $\sigma(\mathcal{A})$ by the uniqueness of the Hahn-Carathéodory extension. Thus,

$$\mu(A) = \mu^*(A) = \inf\left\{\sum \mu\mid_{\mathcal{A}}(K_i) \mid (K_i) \subseteq \mathcal{A} \wedge A \subseteq \bigcup K_j\right\}.$$

Now, since $\mu(A)$ is the least lower bound of $B := \{\sum \mu\mid_{\mathcal{A}}(K_i) \mid (K_i) \subseteq \mathcal{A} \wedge A \subseteq \bigcup K_j\}$, there exists some $(K_{n,m}) \subseteq A_i$ such that for all $n \in \mathbb{N}$, $A \cap S_n \subseteq \bigcup_{m=1}^{\infty} K_{n,m}$ and,

$$\mu(A \cap S_n) \geq \sum_{m=1}^{\infty} \mu(K_{n,m}) - 2^{-n}\epsilon.$$

Then, by defining $\bigcup_{k=1}^{\infty} A_k = \bigcup_{n=1}^{\infty}\bigcup_{m=1}^{\infty} K_{n,m}$ where $A_k = K_{i,j}$ for some $(i,j) \in \mathbb{N}$, we have a disjoint sequence of sets and,

$$\begin{aligned}\mu\left(\bigcup_{n=1}^{\infty}(A_n \setminus A)\right) &= \mu\left(\bigcup_{n=1}^{\infty}\bigcup_{m=1}^{\infty}(K_{n,m} \setminus (A \cap S_n))\right) \\ &\leq \sum_{n=1}^{\infty}\left(\sum_{m=1}^{\infty}\mu(K_{n,m}) - \mu(A \cap S_n)\right) \leq \epsilon.\end{aligned}$$

$\square$

Furthermore, by applying proposition 6 to A^c for some $A \in \mathcal{B}(\mathbb{R}^n)$, there exists some $\tilde{G}$ open, containing A^c such that $\lambda(\tilde{G} \setminus A^c) < \epsilon$ for all $\epsilon > 0$. Then, by defining $F := \tilde{G}^c$, we have $F \subseteq A$ closed, such that

$$\lambda(A \setminus F) = \lambda(A \cap F^c) = \lambda(A \cap \tilde{G}) = \lambda(\tilde{G} \setminus A^c) < \epsilon.$$

Thus,

$$\lambda(G \setminus F) = (\lambda(G) - \lambda(A)) + (\lambda(A) - \lambda(F)) = \lambda(G \setminus A) + \lambda(A \setminus F) < 2\epsilon.$$

So, for all $A \in \mathcal{B}(\mathbb{R}^n)$, there exists some open $G \supseteq A$ and some closed $F \subseteq A$ such that $\lambda(G \setminus F) < \epsilon$ for all $\epsilon > 0$.

Proposition 8 (Transitional Invariance of λ)**.** Let $\Phi_{x_0} : \mathbb{R}^n \to \mathbb{R}^n : x \mapsto x + x_0$ for some $x_0 \in \mathbb{R}^n$. Then,

$$\lambda(\Phi_{x_0}(A)) = \lambda(A),$$

for all $A \in \mathcal{B}(\mathbb{R}^n)$.

Proof. Since, for all $A \in \mathcal{B}(\mathbb{R}^n)$, there exists some open $G \supseteq A$ such that $\lambda(G) - \lambda(A) < \epsilon$, it suffices to show transitional invariance for open sets. Now, as shown previously, any open sets in $\mathbb{R}^n$ can be written as a disjoint union of countable intervals, the result follows since Lebesgue measures are transitional invariant on intervals by definition. $\square$

Proposition 9. $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$.

Proof. As we have seen last term, the Vitali set is a classical example of a non-measurable real set.

Define the equivalence relation $\sim$ such that, for all $x, y \in (0,1]$, $x \sim y \iff x - y \in \mathbb{Q}$. Then, with axiom of choice, for each equivalence class $[x] \in (0,1] \setminus \sim$, we choose exactly one $v \in [x]$ and we define the Vitali set V to be the set of these choices.

Now, let $A := \mathbb{Q} \cap (-1,1]$ and then,

$$(0,1] \subseteq \bigcup_{q \in A} (q + V) \subseteq [-1,2],$$

where the first inclusion is true since, for all $x \in (0,1]$, x belongs to an equivalence class $[x']$ where $x' \in V$. Now, as $x \in [x']$ implies $x - x' \in \mathbb{Q}$, we can simply choose $q = x - x'$ and so $x = q + x' \in \bigcup q + V$.

Thus, if V is measurable, then $1 \leq \lambda(\bigcup(q+V)) \leq 3$. Now, by observing that for all $p, q \in \mathbb{Q}$, if $p \neq q$ then $p + V \cap q + V = \varnothing$,

$$\lambda\left(\bigcup_{q \in A}(q+V)\right) = \sum_{q \in A} \lambda(q+V) = \sum_{q \in A} \lambda(V),$$

since λ is transitional invariant. However, as $\lambda(V)$ is bounded above by 3, $\lambda(V) = 0$ # as $\lambda(V) \geq 1$.

So $V \notin \mathcal{B}(\mathbb{R})$ and hence $\mathcal{B}(\mathbb{R}) \neq \mathcal{P}(\mathbb{R})$. $\square$

3 Functions and Integrals on Measure Spaces

3.1 Measurable Functions

Definition 3.1. Let $(X, \mathcal{A})$ and $(Y, \mathcal{A}')$ be measurable spaces and $f : X \to Y$ be a function. Then f is $\mathcal{A} - \mathcal{A}'$ measurable if for all $A \in \mathcal{A}'$,

$$f^{-1}(A) := \{x \in X \mid f(x) \in A\} \in \mathcal{A}.$$

In the special case that $Y = \mathbb{R}$ and $\mathcal{A}' = \mathcal{B}(\mathbb{R})$ we say f is Borel-measurable.

In applications to probability theory, we recall that one work with the measure space $(\Omega, \mathcal{A}, \mathcal{P})$, and in particular, if $X : \Omega \to \mathbb{R}$ is $\mathcal{A} - \mathcal{B}(\mathbb{R})$ measurable, then X is refereed to as a real-valued random variable.

While we may check the measurability of a function directly from definition, the following lemma provides us with a easier method.

Lemma 3.1. Let X, Y, $\mathcal{E} \subseteq \mathcal{P}(Y)$ be sets and $\mathcal{A}$ a σ-algebra on X. Then a function $f : X \to Y$ is $\mathcal{A} - \sigma(\mathcal{E})$ measurable if and only if for all $A \in E$, $f^{-1}(A) \in \mathcal{A}$.

Proof. The forward direction is trivial so let us consider the reverse.

Consider the set $\mathcal{Q} := \{S \in \sigma(\mathcal{E}) \mid f^{-1}(S) \in \mathcal{A}\}$. It suffices to show $\mathcal{Q}$ is a σ-algebra since, since if that is the case, as $\mathcal{E} \subseteq \mathcal{Q}$, we have $\sigma(\mathcal{E}) \subseteq \mathcal{Q}$.

Straight away, we have $\varnothing \in \mathcal{Q}$ since $f^{-1}(\varnothing) = \varnothing \in \mathcal{A}$. Furthermore, if $A \in \mathcal{Q}$, then $f^{-1}(A) \in \mathcal{A}$ and so $f^{-1}(A^c) = f^{-1}(A)^c \in \mathcal{A}$; and hence, $A^c \in \mathcal{Q}$. Lastly, if $(A_n)_{n=1}^{\infty} \subseteq \mathcal{Q}$, then $f^{-1}\left(\bigcup A_n\right) = \bigcup f^{-1}(A_n) \in \mathcal{A}$, and so, $\bigcup A_n \in \mathcal{Q}$. Thus, $\mathcal{Q}$ is a σ-algebra. ☐

With this lemma in mind, on can check the measurability of a function by simply insuring the measurable of the preimage of sets belonging to the generating set. As an example, we may show a function f to be Borel-measurable by simply showing that for all $U \subseteq \mathbb{R}$, if U is open then $f^{-1}(U) \in \mathcal{B}(\mathbb{R})$. Indeed, we may further relax this constraint by recalling that $\mathcal{B}(\mathbb{R}) = \sigma(\{(-\infty, y] \mid y \in \mathbb{R}\})$, and so, it suffices to ensure the measurability of $(-\infty, y]$ ofr all $y \in \mathbb{R}$.

Definition 3.2. If $f : X \to \overline{\mathbb{R}} = [-\infty, \infty]$ be a function where $(X, \mathcal{A})$ is a measurable space. Then f is measurable if $f^{-1}(\mathcal{B}(\mathbb{R})) \subseteq \mathcal{A}$ and $f^{-1}(\{\infty\}), f^{-1}(\{-\infty\}) \in \mathcal{A}$.

With this definition in mind, let us consider some functions that are measurable.

Proposition 10. Any continuous function $f : X \to Y$ where $(X, \mathcal{T}_X), (Y, \mathcal{T}_Y)$ are topological spaces is $\mathcal{B}(X) - \mathcal{B}(Y)$ measurable.

Proof. Follows directly by the definition of continuity and lemma 3.1. ☐

Corollary 2.1. Any continuous function $f : \mathbb{R}^n \to \mathbb{R}^m$ is Borel-measurable.

Proposition 11. Let $(X, \mathcal{A})$ be a measurable space and $A \subseteq X$. Then the indicator function of A, defined as

$$\mathbf{1}_A : X \to \mathbb{R} : x \mapsto \begin{cases} 1, & x \in A \\ 0, & x \in A^c \end{cases}$$

is measurable if and only if $A \in \mathcal{A}$.

Proof. Indeed, if $A \notin \mathcal{A}$, then, while $\{1\}$ is Borel-measurable, $f^{-1}(\{1\}) = A \notin \mathcal{A}$, and so $\mathbf{1}_A$ is not measurable.

On the other hand, if $A \in \mathcal{A}$, then, for all $S \subseteq \mathbb{R}$, either $0 \in S$, $1 \in S$, both are in S, or neither are in S resulting in $f^{-1}(S) = A, A^c, X$ or $\varnothing$, all of which are members of $\mathcal{A}$. So f is measurable since $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{P}(\mathbb{R})$. □

With this proposition, we have $\mathbf{1}_{\mathbb{Q}}$ is measurable since $\mathbb{Q}$ is countable and so, is a countable union of singletons and hence is measurable.

Proposition 12. Let $(X, \mathcal{A}_X), (Y, \mathcal{A}_Y), (Z, \mathcal{A}_Z)$ be measurable spaces. Then, if $f : X \to Y$ and $g : Y \to Z$ are both measurable, then $g \circ f : X \to Z$ is also measurable.

Proof. This follows directly since for all $U \in \mathcal{A}_Z$, $(g \circ f)^{-1}(U) = f^{-1}(g^{-1}(U))$, and as $g^{-1}(U)$ is measurable as g is, $f^{-1}(g^{-1}(U)) \in \mathcal{A}_X$ and so $g \circ f : X \to Z$ is measurable. □

Theorem 3. Let $(X, \mathcal{A})$ be a measurable space and $f, g : X \to \mathbb{R}$ are $\mathcal{A} - \mathcal{B}(\mathbb{R})$ measurable functions. Then,

$$f + g,\ fg,\ |f|,\ f \wedge g := \min\{f, g\},\ f \vee g := \max\{f, g\},\ \frac{1}{f},$$

are all measurable (where the last function is measurable provided $f(x) \neq 0$ for all x). Furthermore, if $(f_n) : \mathbb{N} \to X \to \mathbb{R}$ is a sequence of measurable functions, then

$$\inf_{n \in \mathbb{N}} f_n,\ \sup_{n \in \mathbb{N}} f_n,\ \liminf_{n \to \infty} f_n,\ \limsup_{n \to \infty} f_n$$

are also measurable.

As we shall see later, the construction of the Lebesgue integral allows us to integral measurable function. Thus, with this theorem, we see the limit of integrable functions are also integrable. This contrasts with the Darboux integral where this is not necessarily the case (see motivation).

Proof. $(f + g)$. It is not hard to see that

$$(f + g)^{-1}(-\infty, a) = \bigcup_{r \in \mathbb{Q}_{\leq a}} \bigcup_{s \in \mathbb{Q}_{\leq a - r}} f^{-1}(-\infty, r) \cap g^{-1}(-\infty, s),$$

where the right hand side is measurable, and so $f + g$ is measurable.

(fg). We use the identity that $fg = \frac{1}{2}[(f + g)^2 - f^2 - g^2]$. Since $f^2 = (x \mapsto x^2 : \mathbb{R} \to \mathbb{R}) \circ f$ where $x \mapsto x^2 : \mathbb{R} \to \mathbb{R}$ is continuous and hence, measurable, we have fg as a linear combination of measurable functions, and so, by our previous result, fg is measurable.

($|f|$) ($f \wedge g$) ($f \vee g$). We define $(\cdot)^+ : \mathbb{R} \to \mathbb{R} := s \mapsto \max\{s, 0\}$ and $(\cdot)^- : \mathbb{R} \to \mathbb{R} := s \mapsto \max\{-s, 0\} = (-s)^+$. Now, as both $(\cdot)^+$ and $(\cdot)^-$ are continuous, they are measurable. So, since $|f| = ((\cdot)^+ \circ f) + ((\cdot)^- \circ f))$, we have $|f|$ is measurable. Similarly we see that

$$f \wedge g = f - (g - f)^-,$$

and

$$f \vee g = f + (g - f)^+,$$

and so, both $(f \wedge g)$ and $(f \vee g)$ are also measurable.

($1/f$). Follows as $(1/f)^{-1}(\infty, a)$ is $f^{-1}(1/a, 0)$ for $a < 0$, $f^{-1}(-\infty, 0)$ for $a = 0$ and $f^{-1}((-\infty, 0) \cup (1/a, \infty))$ for $a > 0$.

($\inf_{n \in \mathbb{N}} f_n$) ($\sup_{n \in \mathbb{N}} f_n$). Follows straight way as,

$$(\inf_{n \in \mathbb{N}} f_n)^{-1}(-\infty, a) = \bigcup_{n \in \mathbb{N}} f_n^{-1}(-\infty, a),$$

and $\sup f_n = -\inf(-f_n)$.

($\liminf_{n \to \infty} f_n$) ($\limsup_{n \to \infty} f_n$). Follows as, by definition,

$$\liminf_{n \to \infty} f_n = \sup_{l \geq 1} \left(\inf_{k \geq l} f_k \right),$$

and

$$\limsup_{n \to \infty} f_n = \inf_{l \geq 1} \left(\sup_{k \geq l} f_k \right).$$

□

For the sake of completeness, we shall quickly introduce a method of generating σ-algebras from functions. However, the proofs and many properties are left on the problem sheet (sheet 3) as exercises.

Definition 3.3 (σ-algebra Generated by Functions)**.** Let X be a space, $(Y, \mathcal{A}_Y)$ a measurable space and $f : X \to Y$ a function. Define

$$\sigma(f) = \{f^{-1}(A) \mid A \in \mathcal{A}_Y\}.$$

This is a σ-algebra on X, and one can show that this is indeed the smallest σ-algebra on X on which f is measurable. Furthermore, for notation sake, if $\{f_i \mid i \in I\}$ is a family of functions, then

$$\sigma(f_i \mid i \in I) := \sigma\left(\bigcup_{i \in I} \sigma(f_i) \right).$$

3.2 Sequences of Measurable Functions

A nice method to determine whether or not a function is measurable is to show that it is a limit of measurable simple functions. Simple functions are very useful in many instances and we will later also see their applications in Lebesgue integration.

Definition 3.4 (Simple Function)**.** Let X be a set. A function $s : X \to \mathbb{R}$ is called a simple function (or a step function) if it takes on finitely many values. That is, the set $\{y \in \mathbb{R} \mid \exists\, x \in X, y = s(x)\}$ is finite.

Writing the image of a simple function s as $\{\alpha_1, \cdots, \alpha_l\}$, and by writing $A_i = s^{-1}(\{\alpha_i\})$ for all $1 \leq i \leq l$, we see straight way that $\{A_i \mid i\}$ forms a partition of X and for all $x \in X$,

$$s(x) = \sum_{i=1}^{l} \alpha_i \mathbf{1}_{A_i}(x).$$

In particular, if $\mathcal{A}$ is a σ-algebra over X, by recalling that a indicator function of the set A is measurable if and only if A is measurable, we see that s is measurable if and only if A_i is measurable for all i.

Theorem 4 (Approximation by Simple Functions)**.** Let $(X, \mathcal{A})$ be a measurable space and let $f : X \to [0, \infty]$ be a function. Then f is measurable if and only if there exists a sequence of measurable simple functions $s_m : X \to [0, \infty)$, such that for all $x \in X$,

$$0 \leq s_1(x) \leq s_2(x) \leq \cdots \leq f(x),$$

and furthermore, $f(x) = \lim_{n\to\infty} s_n(x)$.

Proof. If such a sequence exits, then, by the fact that the $\liminf$ of measurable functions is measurable, we have f is measurable.

Conversely, suppose f is measurable. Then, for $n \geq 1$, we define

$$\phi_n : [0, \infty] \to \mathbb{R} : t \mapsto \begin{cases} k2^{-n}, & k2^{-n} \leq t < (k+1)2^{-n}, k = 0, 1, \cdots, n2^{-n}; \\ n, & t \geq n. \end{cases}$$

The functions (ϕ_n) are trivially measurable, increasing and for all $t < n$, we have

$$t - 2^{-n} \leq \phi_n(t) \leq t.$$

Hence, we have $\lim_{n\to\infty} \phi_n(t) = t$ for all $t \in [0, \infty]$. Then, by defining $s_n = \phi_n \circ f$, we find that this satisfy our condition. ☐

Definition 3.5 (Almost Everywhere)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space. A property $P : X \to \text{Prop}$ of X is said to hold μ-almost everywhere (abbreviated as μ-a.e.) if the set $\{x \in X \mid \neg P(x)\}$ is measurable and has μ-measure 0.

A commonly application of the above definition is the following notion of convergence.

Definition 3.6 (Convergence Almost Everywhere)**.** Let $(f_n : X \to \mathbb{R})_{n=1}^{\infty}$ be a sequence of measurable functions. The sequence (f_n) converges μ-almost everywhere if

$$\mu\left(\{x \in X \mid \lim_{n\to\infty} f_n(x)\ \nexists\}\right) = 0.$$

We note that we assumed that the set $\{x \in X \mid \lim_{n\to\infty} f_n(x) \nexists\}$ is measurable. We shall prove this claim here.

Proof. It suffices to show the measurable of the complement, so let us consider $S := \{x \in X \mid \lim_{n\to\infty} f_n(x)\exists\}$. We see that S is equivalent to the set of x such that $(f_n(x))$ is Cauchy, so we have

$$S = \bigcap_{K=1}^{\infty} \bigcup_{N=1}^{\infty} \{x \in X \mid \forall n, m \geq N, |f_n(x) - f_m(x)| < 1/K\}.$$

Indeed, taking the for all outside, we have,

$$S = \bigcap_{K=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n \geq N} \bigcap_{m \geq N} \{x \in X \mid |f_n(x) - f_m(x)| < 1/K\},$$

and so, it suffices to show that $\{x \in X \mid |f_n(x) - f_m(x)| < 1/K\}$ is measurable for appropriate K, N, n, m. Thus, as

$$\{x \in X \mid |f_n(x) - f_m(x)| < 1/K\} = (f_n - f_m)^{-1}(-1/K, 1/K),$$

since f_n and f_m are measurable, $(f_n - f_m)^{-1}(-1/K, 1/K)$ is a measurable set, and we are done! $\square$

Throughout the remainder of this section, we consider $\Omega \subseteq \mathbb{R}^n$ measurable with respect to the Lebesgue measure on the Borel σ-algebra with $\lambda(\Omega) < \infty$.

Theorem 5 (Ergorov)**.** Let $(f_n : \Omega \to \mathbb{R})_{n=1}^{\infty}$ be a sequence of measurable functions and $f : \Omega \to \mathbb{R}$ be a measurable function. Suppose that $f_k \to f$ as $k \to \infty$ almost everywhere on Ω. Then, for all $\delta > 0$, there exists a compact set $F \subseteq \Omega$ such that

$$\lambda(\Omega \setminus F) < \delta, \text{ and } \sup_{x \in F} |f_k(x) - f(x)| \to 0$$

as $k \to \infty$.

Conceptually, the Ergorov theorem tells us that for any sequence of almost convergent measurable functions, there exists a "large" set such that the convergence is uniform.

One cannot choose $\delta = 0$ in general. Indeed, by considering the sequence $f_k(x) = x^k$ on $[0, 1]$, we see that $f_k \to \mathbf{1}_{\{1\}}$ pointwise. However, as we have see from first year's analysis, this function does not uniformly converge on $[0, 1]$ and so, we may only choose $F = [0, 1 - \delta]$ for any $\delta \in (0, 1)$.

Proof. (Ergorov). Let $\delta > 0$ and for $i, j \geq 1$, define

$$C_{ij} := \bigcup_{k=j}^{\infty} \{x \in \Omega \mid |f_k(x) - f(x)| > 2^{-i}\}.$$

We see that C_{ij} is measurable and $C_{i,j+1} \subseteq C_{i,j}$. So, by the continuity of measures from above, we have

$$\lim_{j\to\infty} \lambda(C_{ij}) = \lambda\left(\bigcap_{j=1}^{\infty} C_{ij}\right) = 0$$

since f_k converges to f almost everywhere. Hence, by the definition of the limit, there exists some $J(i) \geq 1$ such that $\lambda(C_{i,J(i)}) < \delta 2^{-i-1}$. Then, by defining

$$A := \Omega \setminus \bigcup_{i=1}^{\infty} C_{i,J(i)},$$

we have

$$\lambda(\Omega \setminus A) \leq \sum_{i=1}^{\infty} \lambda(C_{i,J(i)}) < \sum_{i=1}^{\infty} \delta 2^{-i-1} = \frac{\delta}{2}.$$

Thus, by recalling that for all $\epsilon > 0$, there exists some $F \subset A$ closed such that $\lambda(A \setminus F) < \epsilon$, the result follows by taking $\epsilon = \delta/2$. ☐

The Ergorov theorem allows us to conclude a powerful result about measurable functions. Indeed, by removing the compactness requirement for F, this theorem can be easily generalised to arbitrary measure spaces by the same proof.

Theorem 6. Let $\Omega \subseteq \mathbb{R}^n$ with $\lambda(\Omega) < \infty$. If the function $f : \Omega \to \mathbb{R}$ is measurable, then, for all $\delta > 0$, there exists $F \subseteq \Omega$ compact such that $\lambda(\Omega \setminus F) < \delta$ and $f \mid_F : F \to \mathbb{R}$ is continuous.

Proof. Suppose first that $f = s$ for some s a simple function. Then, we can write

$$s = \sum_{i=1}^{n} \alpha_i \mathbf{1}_{A_i}$$

for some partition of $\Omega - \{A_i\}$. Now, since for all A_i, we may find $F_i \subseteq A_i$ such that F_i is closed (and hence compact) and $\lambda(A_i \setminus F_i) < \delta 2^{-i}$, we have $F = \bigcup_{i=1}^{n} F_i$ which suffices our condition.

Now, for the general case for f, we apply the approximation theorem for measurable functions. However, as the approximation theorem only works for non-negative functions, let us write $f = f^+ - f^-$ where f^+ and f^- are both non-negative. Now, by the approximation theorem, there exists $(s_n^{\pm})$ a sequence of simple functions approximating $f^{\pm}$, the sequence of function $s_n := s_n^+ - s_n^-$ is a simply function such that

$$\lim_{n \to \infty} s_n = \lim_{n \to \infty} s_n^+ - s_n^- = f^+ - f^- = f.$$

As we have already proved the theorem for simple functions, there exists compact sets F_n such that $\lambda(\Omega \setminus F_n) < \delta 2^{-n-1}$ and $s_n \mid_{F_n} : F_n \to \mathbb{R}$ is continuous. Furthermore, by Ergorov's theorem, there exists some compact $F_0 \subseteq \Omega$ such that $\lambda(\Omega \setminus F_0) < \delta/2$ and

$$\sup_{x \in F_0} |s_n(x) - f(x)| \to 0,$$

as $n \to \infty$. Then, by defining

$$F := \bigcap_{i=0}^{\infty} F_i,$$

we have a compact set (as the intersection of closed sets is closed) on which $s_n \mid_F$ is continuous and uniformly converge to $f \mid_F$, and so $f \mid_F$ is continuous. Furthermore, since

$$\lambda(\Omega \setminus F) = \lambda\left(\Omega \setminus \bigcap F_i\right) = \lambda\left(\bigcup \Omega \setminus F_i\right) \leq \sum \lambda(\Omega \setminus F_i) < \delta/2 + \delta/2 = \delta,$$

we have the result. $\square$

This is a powerful theorem that provides us with a necessary condition on measurable functions however, it does not mean that measurable function must be continuous at some point. Indeed, by recalling that $\mathbf{1}_{\mathbb{Q}\cap[0,1]}$ on $\Omega = [0,1]$ is measurable and nowhere continuous, we have a counterexample. This not contradict our theorem however, since we can simply choose $F = [0,1] \setminus \mathbb{Q}$ and so $\mathbf{1}_{\mathbb{Q}\cap[0,1]} \mid_F = 0$. Thus, by any sufficiently large compact subset of F, the result follows.

3.3 The Integral

We will finally construct the integral on some arbitrary measure space $(X, \mathcal{A}, \mu)$, in particular the case $X = \mathbb{R}$, $\mathcal{A} = \mathcal{B}(\mathbb{R})$ and $\mu = \lambda$. This results in the Lebesgue integral and it turns out the Lebesgue integral extend the Darboux integral, i.e. the Lebesgue integral and the Darboux integral agrees on Darboux integrable functions.

We will introduct the integral of *suitable* measurable functions in three steps:

1. simple functions;
2. non-negative functions (by approximation theorem);
3. *integrable* real-valued functions.

As we shall see, while the first two definitions are applicable for all appropriate measurable functions, the third definition applies only to certain real-valued measurable functions in order to prevent cases such as $\infty - \infty$.

The integral of the simple functions is a one might expect.

Definition 3.7 (Integral of Simple Functions)**.** Let

$$S^+ := \{s : X \to [0, \infty) \mid s \text{ is simple and measurable}\},$$

and for $s \in S^+$, Wlog. we write $s = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ with $\alpha_i \in (0, \infty)$ as a representation of s (not unique). Then, for $s \in S^+$ with representation $s = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$, the integral of s with respect to the measure μ is

$$\mu(s) \equiv \int_X s \mathrm{d}\mu \equiv \int s \mathrm{d}\mu := \sum_{i=1}^n \alpha_i \mu(A_i) \in [0, \infty].$$

While this definition seems reasonable, as the representation of s is not unique, we need to check whether or not it is well-defined. Indeed, if $s = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i} = \sum_{i=1}^k \beta_i \mathbf{1}_{B_i}$, Wlog. we may assume $\bigcup A_i = \bigcup B_i = X$ since if otherwise, we may add a term (without changing the value of the integral) $\alpha_{n+1} A_{n+1}$ with $\alpha_{n+1} = 0$ and $A_{n+1} = X \setminus \bigcup A_i$. Then, by additivity,

$$\sum_{i=1}^n \alpha_i \mu(A_i) = \sum_{i=1}^n \alpha_i \sum_{j=1}^k \mu(A_i \cap B_j) = \sum_{i=1}^n \sum_{j=1}^k \beta_j \mu(A_i \cap B_j) = \sum_{j=1}^k \beta_j \mu(B_j),$$

where the second equality is true since on $A_i \cap B_j$, $\alpha_i = \beta_j$, and so, the integral is well-defined.

Lemma 3.2. Let $f, g \in S^+$, then, for all $\alpha, \beta \in [0, \infty)$,

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

Furthermore, if $f \leq g$, then

$$\int f \mathrm{d}\mu \leq \int g \mathrm{d}\mu.$$

Proof. $\int \alpha f \mathrm{d}\mu = \alpha \int f \mathrm{d}\mu$ immediately, so let us show Linearity of addition. Write $f = \sum_{i=1}^n \alpha_i \mathbf{1}_{A_i}$ and $g = \sum_{i=1}^k \beta_i \mathbf{1}_{B_i}$; then, again, Wlog. we may assume $\bigcup A_i = \bigcup B_i = X$. Then,

$$f + g = \sum_{i=1}^n \sum_{j=1}^k (\alpha_i + \beta_j) \mathbf{1}_{A_i \cap B_j},$$

and hence,

$$\begin{aligned}
\int (f+g) \mathrm{d}\mu &= \sum_{i=1}^n \sum_{j=1}^k (\alpha_i + \beta_j) \mu(A_i \cap B_j) \\
&= \sum \alpha_i \sum \mu(A_i \cap B_j) + \sum \beta_j \sum \mu(A_i \cap B_j) \\
&= \sum \alpha_i \mu(A_i) + \sum \beta_j \mu(B_j) \\
&= \int f \mathrm{d}\mu + \int g \mathrm{d}\mu.
\end{aligned}$$

(Monotonicity is left as an exercise.) □

Definition 3.8 (Integral of Non-negative Functions)**.** For $f : X \to [0, \infty]$ a measurable function. Then the integral of f is

$$\int f \mathrm{d}\mu := \sup \left\{ \int g \mathrm{d}\mu \mid g \in S^+, g \leq f \right\}.$$

Proposition 13. Given $f, g : X \to [0, \infty]$ measurable and $\alpha, \beta \in [0, \infty)$,

$$\int \alpha f + \beta g \mathrm{d}\mu = \alpha \int f \mathrm{d}\mu + \beta \int f \mathrm{d}\mu.$$

Furthermore, if $f \leq g$ for all $x \in X$ then,

$$\int f \mathrm{d}\mu \leq \int g \mathrm{d}\mu.$$

Proof. Follows by the combination of the next proposition with the approximation of measurable functions by simple functions. □

Proposition 14 (Monotone Convergence)**.** Let $f : X \to [0, \infty]$ be a measurable function and $(f_n : X \to [0, \infty])_{n=1}^{\infty}$ be a sequence of measurable functions such that $f_n \uparrow f$. Then,

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

Proof. First, we note that the limit $\lim_{n\to\infty} \int f_n \mathrm{d}\mu$ exists since it is monotone and bounded above by $\int f \mathrm{d}\mu$. To show the reverse inequality, it suffices to show that

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu \geq \int g \mathrm{d}\mu,$$

for all simple functions $g \leq f$. Let $g = \sum_{i=1}^{l} \alpha_i \mathbf{1}_i$, and fix $\epsilon > 0$ with

$$G_n^\epsilon := \{x \in X \mid f_n(x) \geq (1-\epsilon)g(x)\} \in \mathcal{A}.$$

Then, since $f_n \uparrow f$, (G_n^ϵ) is a increasing sequence of sets with $\bigcup G_n^\epsilon = X$. Thus, for all $n \in \mathbb{N}$,

$$\int f_n \mathrm{d}\mu \geq \int f_n \mathbf{1}_{G_n^\epsilon} \mathrm{d}\mu \geq \int (1-\epsilon) g \mathbf{1}_{G_n^\epsilon} \mathrm{d}\mu = (1-\epsilon) \int \sum_{i=1}^{l} \alpha_i \mathbf{1}_{A_i \cap G_n^\epsilon} \mathrm{d}\mu,$$

so, by the definition of the integral over simple functions, we have

$$\int f_n \mathrm{d}\mu \geq (1-\epsilon) \sum_{i=1}^{l} \alpha_i \mu(A_i \cap G_n^\epsilon).$$

Thus, by taking $n \to \infty$, by the continuity of measures, we have

$$\lim_{n\to\infty} \int f_n \mathrm{d}\mu \geq \lim_{n\to\infty} (1-\epsilon) \sum_{i=1}^{l} \alpha_i \mu(A_i \cap G_n^\epsilon) = (1-\epsilon) \sum_{i=1}^{l} \alpha_i \mu(A_i) = (1-\epsilon) \int g \mathrm{d}\mu.$$

Hence, as ϵ is arbitrary, the reverse inequality is achieved. $\square$

We remark that in the definition of G_n^ϵ, the introduction of ϵ is necessary since, if $g(x) = f(x)$ at some $x \in X$ while $f_n(x) < f(x)$ for all n, $x \notin \bigcup G_n^\epsilon$, breaking our argument.

Corollary 6.1. For $f : X \to [0, \infty]$ measurable,

- $f = 0$ μ-a.e. $\iff \int f \mathrm{d}\mu = 0$;
- if $\int f \mathrm{d}\mu < \infty$, then $f < \infty$, μ-a.e.

Proof.

(1) ($\implies$) By the approximation theorem, there exists a sequence of increasing measurable simple functions $(s_n : X \to [0, \infty])$ such that $s_n \uparrow f$. Now, since $s_n \leq f$, $s_n = 0$, μ-a.e. since

$$\{s_n \neq 0\} \subseteq \{f \neq 0\} \implies \mu(\{s_n \neq 0\}) \leq \mu(\{f \neq 0\}) = 0.$$

and so, $\int s_n \mathrm{d}\mu = \sum \alpha_i \mu(A_i) = 0$ where (A_n) forms a partition of $\{s_n \neq 0\}$. Thus, by the above proposition,

$$0 = \lim_{n\to\infty} \int s_n \mathrm{d}\mu = \int \lim_{n\to\infty} s_n \mathrm{d}\mu = \int f \mathrm{d}\mu.$$

($\Longleftarrow$) Suppose otherwise, i.e. there exists $A \in \mathcal{A}$, such that $\mu(A) \neq 0$ and for all $a \in A$, $f(a) \neq 0$. Then, I claim that there exists $B \subseteq A$, $B \in \mathcal{A}$ and there exists some $\epsilon > 0$ such that $\mu(B) > \epsilon$. Suppose otherwise, then for all $\epsilon > 0$, $\mu(f^{-1}[\epsilon, \infty]) = 0$. Thus, as $f^{-1}[\epsilon, \infty] \uparrow f^{-1}(0, \infty]$, by the continuity of measures, $\mu(f^{-1}(0, \infty]) = 0$ #. So, choose $\epsilon > 0$ and B such that $\mu(B) > 0$ and $f(B) > \epsilon$. Then $f \geq \epsilon \mathbf{1}_B$ on X and hence by monotonicity,

$$\int f \mathrm{d}\mu \geq \int \epsilon \mathbf{1}_B \mathrm{d}\mu = \epsilon \mu(B) > 0,$$

contradicting $\int f \mathrm{d}\mu = 0$.

(2) If otherwise, then there exists $A \in \mathcal{A}$, $f(A) = \{\infty\}$ and $\mu(A) > 0$. So, $f \geq \infty \mathbf{1}_A$ (where $\infty \times 0 := 0$) and

$$\int f \mathrm{d}\mu \geq \int \infty \mathbf{1}_A \mathrm{d}\mu = \infty \mu(A) = \infty.$$

$\square$

This result allows us to strengthen the previous results regarding monotonicity. For instance, monotonicity holds if $f \leq g$ μ-a.e. Indeed, if that is the case, then by defining $\hat{f} = f \times \mathbf{1}_{\{f \leq g\}}$, we have

$$\int f \mathrm{d}\mu - \int \hat{f} \mathrm{d}\mu = \int f - \hat{f} \mathrm{d}\mu = 0$$

since $f - \hat{f} = 0$ μ-a.e. Now, since $\hat{f} \leq g$ everywhere on X, the result follows.

Lastly, we may finally define the integral for real-valued functions.

Definition 3.9 (μ-integrable Functions)**.** A measurable function $f : X \to [-\infty, \infty]$ is called μ-integrable if $\int |f| \mathrm{d}\mu < \infty$. We write

$$\mathcal{L}^1(\mu) \equiv \mathcal{L}^1(X, \mathcal{A}, \mu) := \{f : X \to [-\infty, \infty] \mid f \text{ measurable and } \mu\text{-integrable}\}$$

for the space of measurable and integrable functions.

Definition 3.10 (Integral of Real-valued Functions)**.** Let $f \in \mathcal{L}^1(\mu)$. Then, the integral of f is

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

where $f = f^+ - f^-$. Furthermore, for $A \in \mathcal{A}$, we define

$$\int_A f \mathrm{d}\mu := \int f \mathbf{1}_A \mathrm{d}\mu.$$

(Hence $\int f \mathrm{d}\mu = \int_X f \mathrm{d}\mu$).

Theorem 7 (Properties of the Integral)**.** Let $f, g \in \mathcal{L}^1(\mu)$, then

- (Monotonicity) $f \leq g$ μ-a.e. $\implies \int f \mathrm{d}\mu \leq \int g \mathrm{d}\mu$. In particular, if $f = g$ μ-a.e. then $\int f \mathrm{d}\mu = \int g \mathrm{d}\mu$;

- (Triangle inequality) $\left|\int f \mathrm{d}\mu\right| \leq \int |f| \mathrm{d}\mu$;
- (Linearity) for all $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g \in \mathcal{L}^1(\mu)$ and $\int \alpha f + \beta g \mathrm{d}\mu = \alpha \int f \mathrm{d}\mu + \beta \int g \mathrm{d}\mu$.

Proof. Follows easily from the definition and the previous properties for non-negative functions. ☐

In the case that $(X, \mathcal{A}, \mu) = (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \lambda)$ with λ denoting the Lebesgue measure, the corresponding integral $\int f \mathrm{d}\lambda$ is called the **Lebesgue integral**. In fact, one commonly "complete the space" by adding all subsets of λ-null sets, i.e. one considers λ on the "completed" σ-algebra

$$\mathcal{B}^*(\mathbb{R}^n) := \sigma(\mathcal{B}(\mathbb{R}^n) \cup \mathcal{N}),$$

where $\mathcal{N} := \{A \subseteq \mathbb{R}^n \mid \exists N \in \mathcal{B}(\mathbb{R}^n), A \subseteq N \wedge \lambda(N) = 0\}$. Furthermore, the Lebesgue integral extends the Darboux integral in the following sense:

Proposition 15 (Extension of Darboux Integral)**.** Let $f : I \to \mathbb{R}$ be Darboux-integrable on $I = [a, b]$ with

$$\left|\int_a^b f(x) \mathrm{d}x\right| < \infty.$$

Then, $f\mathbf{1}_I$ is Lebesgue integrable and

$$\int_I f \mathrm{d}\lambda = \int_a^b f(x) \mathrm{d}x.$$

Proof. See problem sheet. ☐

Furthermore, this extension is a strict extension, as in the Lebesgue integral is strictly stronger than the Darboux integral. Indeed, by recalling the motivational example, we saw that $\mathbf{1}_{\mathbb{Q}\cap[0,1]}$ we not Darboux integrable by, we see now that this is easily Lebesgue integrable as it is simply

$$\int \mathbf{1}_{\mathbb{Q}\cap[0,1]} \mathrm{d}\mu = \mu(\mathbb{Q} \cap [0, 1]) = 0.$$

Now consider when $X = \mathbb{N}$, $\mathcal{A} = \mathcal{P}(X)$ and $\mu =$ the counting measure, then, we define

$$l^1 := \mathcal{L}^1(\mu) = \left\{a : \mathbb{N} \to \mathbb{R} \mid \sum |a_n| < \infty\right\},$$

and so, l^1 is simply the space of absolutely convergent sequences with $\int a \mathrm{d}\mu = \sum a_n$.

Lastly, returning to our example of applications of measure theory, if $\mu = \mathbb{P}$ – the probability measure and $f \in \mathcal{L}^1(\mathcal{P})$, then one usually writes

$$\mathbb{E}[f] := \int_\Omega f \mathrm{d}\mathbb{P},$$

the expectation of f with respect to $\mathbb{P}$.

3.4 Convergence Theorems

As before, let $(X, \mathcal{A}, \mu)$ be a measure space. The convergence theorems regard the interchange of limits and integrals. In fact, we have already seen on of such theorems – the monotone convergence theorem, and as an application, we obtain the following lemma.

Proposition 16 (Fatou's Lemma)**.** Let $f_n : X \to [0, \infty]$ for all $n \in \mathbb{N}$ be a sequence of measurable functions. Then

$$\int \liminf_{n\to\infty} f_n \mathrm{d}\mu \leq \liminf_{n\to\infty} \int f_n \mathrm{d}\mu.$$

Proof. For $n \in \mathbb{N}$, define $g_n : X \to [0, \infty]$ by

$$g_n(x) := \inf_{k \geq n} f_k(x).$$

We see that g_n is measurable and $g_n \leq f_k$ for all $k \geq n$ and hence, by monotonicity,

$$\int g_n \mathrm{d}\mu \leq \inf_{k\geq n} \int f_k \mathrm{d}\mu.$$

Now, since g_n is monotonically increasing, by the monotone convergence theorem, we have

$$\int \liminf_{n\to\infty} f_n \mathrm{d}\mu = \int \lim_{n\to\infty} g_n \mathrm{d}\mu = \lim_{n\to\infty} \int g_n \mathrm{d}\mu \leq \lim_{n\to\infty} \inf_{k\geq n} \int f_k \mathrm{d}\mu.$$

$\square$

We note that we have not assumed that f_n is convergent, and so, Fatou's lemma is useful in that it can apply to a wide range of sequences of functions.

Theorem 8 (Lebesgue Dominated Convergence Theorem)**.** Let $g : X \to [0, \infty]$ be integrable, i.e. $g \in \mathcal{L}^1(\mu)$, and $f, f_n : X \to [-\infty, \infty]$ for $n \geq 1$ be measurable functions such that $f_n \to f$ μ-almost everywhere and

$$|f_n(x)| < g(x)$$

for all $x \in X$. Then,

$$\left| \int f_n \mathrm{d}\mu - \int f \mathrm{d}\mu \right| \leq \int |f_n - f| \mathrm{d}\mu \to 0,$$

as $n \to \infty$.

Proof. Straight, we have f_n and f are integrable since both are bounded by g, μ-a.e. and so, their integral are finite (so the theorem is well-defined). The inequality is simply the triangle-inequality for integrals, and so, it suffices to show $\int |f_n - f| \mathrm{d}\mu \to 0$ as $n \to \infty$.

Thus, by applying Fatou's lemma on $2g - |f_n - f|$ (which is a increasing sequence of non-negative measurable functions), we have

$$\int 2g \mathrm{d}\mu = \int \liminf_{n\to\infty} \{2g - |f_n - f|\} \mathrm{d}\mu \leq \liminf_{n\to\infty} \int 2g - |f_n - f| \mathrm{d}\mu,$$

since $2g - |f_n - f| \to 2g$, μ-a.e. Thus, by applying linearity, we have

$$\int 2g \mathrm{d}\mu \leq \int 2g \mathrm{d}\mu - \liminf_{n\to\infty} \int |f_n - f| \mathrm{d}\mu.$$

So,

$$0 \leq \limsup_{n\to\infty} \int |f_n - f| \mathrm{d}\mu \leq 0$$

and the result follows. □

Straight away we see that the dominated convergence theorem shows that, for appropriate sequences of functions $f_n \to f$,

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

Definition 3.11 (Convergence in Measure)**.** Let $f : X \to \mathbb{R}$, $f_n : X \to [-\infty, \infty]$ for all $n \in \mathbb{N}$ be measurable. Then f_n converges to f is measure if for all $\epsilon > 0$,

$$\mu(\{x \in X \mid |f_n(x) - f(x)| > \epsilon\}) \to 0$$

as $n \to \infty$.

Proposition 17. Let $f, f_n : X \to \mathbb{R}$ for $n \in \mathbb{N}$ be measurable with $\mu(X) < \infty$. Then,

- if $\int |f_n - f| \mathrm{d}\mu \to 0$, then $f_n \to f$ in measure;
- if $f_n \to f$ μ-a.e. then $f_n \to f$ in measure;
- if $f_n \to f$ in measure, then there exists a subsequence $\Lambda \subseteq \mathbb{N}$ such that $f_n \to f$ μ-a.e. as $n \in \Lambda \to \infty$.

Proof. See problem sheet 4 and 6. □

Definition 3.12 (Density)**.** Let $f : X \to [0, \infty)$ be measurable and μ is a measure on $(X, \mathcal{A})$. Then, we can construct a new measure ν where

$$\nu(A) := \int_A f \mathrm{d}\mu,$$

for all $A \in \mathcal{A}$ (easy proof). This measure ν is said to have a density with respect to μ and f is a density. For shorthand, we often write $\nu = f\mu$ and $f = \frac{\mathrm{d}\nu}{\mathrm{d}\mu}$.

Definition 3.13 (Absolutely Continuous)**.** Let μ, ν be two measures on $(X, \mathcal{A})$. Then ν is absolutely continuous with respect to μ (denoted $\nu \ll \mu$) if $\nu(A) = 0$ for all $A \in \mathcal{A}$ with $\mu(A) = 0$.

The reason for this definition is the following.

Proposition 18. Let $f \in \mathcal{L}^1(\mu)$, then for all $\epsilon > 0$, there exists some $\delta > 0$ such that for all $A \in \mathcal{A}$, $\mu(A) < \delta$, we have $\int_A |f| \mathrm{d}\mu < \epsilon$.

Proof. See problem sheet 5. □

Clearly $\nu = f\mu$ implies $\nu \ll \mu$ since for all $A \in \mathcal{A}$, $\mu(A) = 0$ implies that $\nu(A) = \int_A |f| \mathrm{d}\mu < \epsilon$ for all $\epsilon > 0$ and so, a measure is absolutely continuous if it has a measure. Indeed, as we shall see later (in the Radon - Nikodym theorem), the reverse is also true – every absolutely continuous measure has a density. Thus, in light of this, we see that a measure ν is absolutely continuous is equivalent to saying that there exists some μ such that for all $\epsilon > 0$ there exists a $\delta > 0$ such that for all $A \in \mathcal{A}$, $\mu(A) < \delta$, we have $\nu(A) < \epsilon$ (provided X has finite measure).

Definition 3.14 (Uniformly Absolutely Continuous Integrals)**.** Let $\mathcal{F} \subseteq \mathcal{L}^1(\mu)$. The family $\mathcal{F}$ is said to have uniformly absolutely continuous integrals if for all $\epsilon > 0$, there exists some $\delta > 0$, such that for all $f \in \mathcal{F}$, $A \in \mathcal{A}$ such that $\mu(A) < \delta$, we have

$$\int_A |f| \mathrm{d}\mu < \epsilon.$$

Straight away, we see that any singleton $\mathcal{F} = \{f\}$ has uniformly absolutely continuous integrals. Indeed, by induction, this is also true for any finite families $\mathcal{F}$ by induction.

Theorem 9 (Vitali)**.** Let $\mu(X) < \infty$, $f, f_n : X \to \mathbb{R}$ be measurable for all $n \in \mathbb{N}$. Then, the following are equivalent:

- $f_n \to f$ in measure and $\mathcal{F} := \{f_n \mid n \in \mathbb{N}\}$ has uniformly absolutely continuous integrals;
- $\int |f_n - f| \mathrm{d}\mu \to 0$ as $n \to \infty$.

We see that the requirements for the dominated convergence theorem implies Vitali's theorem since if there exists such a dominating g, for all $\epsilon > 0$, we can choose $\delta > 0$ to be that of g's. So, for all $A \in \mathcal{A}$, if $\mu(A) < \delta$,

$$\int_A |f_n| \mathrm{d}\mu \le \int_A g \mathrm{d}\mu < \epsilon,$$

for all $n \in \mathbb{N}$.

Proof. (Vitali).

($\Longleftarrow$) $f_n \to f$ in measure by proposition above and so, it suffices to show that $\mathcal{F}$ has uniformly absolutely continuous integrals.

For all $\epsilon > 0$, choose $n_0 \in \mathbb{N}$ such that

$$\int |f_n - f| \mathrm{d}\mu < \frac{\epsilon}{2}$$

for all $n \ge n_0$. Then, by defining $\mathcal{F}' := \{f, f_1, \cdots, f_{n_0}\}$, since $\mathcal{F}'$ is finite, it has uniformly absolutely continuous integrals. Thus, there exists some $\delta > 0$ such that for all $A \in \mathcal{A}$ with $\mu(A) < \delta$,

$$\int_A |f| \mathrm{d}\mu < \frac{\epsilon}{2}, \text{ and } \max_{n \le n_0} \int_A |f_n| \mathrm{d}\mu < \frac{\epsilon}{2}.$$

Hence, by choosing this to be our δ, for $n > n_0$,

$$\int_A |f_n| \mathrm{d}\mu = \int_A |f_n - f + f| \mathrm{d}\mu \leq \int_A |f_n - f| \mathrm{d}\mu + \int_A |f| \mathrm{d}\mu < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

and we have $\mathcal{F}$ has uniformly absolutely continuous integrals.

($\implies$) By contradiction, assume $\limsup_{n\to\infty} \int |f_n - f| \mathrm{d}\mu > 0$. Then, there exists some subsequence $\Lambda \subseteq \mathbb{N}$, such that

$$\lim_{n \in \Lambda \to \infty} \int |f_n - f| \mathrm{d}\mu > 0,$$

and $f_n \to f$ μ-a.e. as $n \in \Lambda \to \infty$. Now, since $\mathcal{F} \cup \{f\}$ has uniformly absolutely continuous integrals, for all $\epsilon > 0$, there exists some $\delta > 0$ such that for all $A \in \mathcal{A}$, $\mu(A) < \delta$, we have

$$\int_A |f| \mathrm{d}\mu < \frac{\epsilon}{3}, \text{ and } \int_A |f_n| \mathrm{d}\mu < \frac{\epsilon}{3}$$

for all $n \in \mathbb{N}$. Applying Ergorov's theorem to $(f_n)_{n \in \Lambda}$, one finds a measurable set F such that $\mu(X \setminus F) < \delta$ and

$$\sup_{x \in F} |f_n(x) - f(x)| \to 0,$$

as $n \in \Lambda \to \infty$. Thus, choosing n_0 such that

$$\sup_{x \in F} |f_n(x) - f(x)| < \frac{\epsilon}{3\mu(X)},$$

for all $n \in \lambda_{\geq n_0}$, it follows that for all $n \in \lambda_{\geq n_0}$,

$$\int |f_n - f| \mathrm{d}\mu = \int |f_n - f| \mathbf{1}_F \mathrm{d}\mu + \int |f_n - f| \mathbf{1}_{X \setminus F} \mathrm{d}\mu \leq \frac{\epsilon}{3\mu(X)} \mu(F) + \frac{\epsilon}{3} + \frac{\epsilon}{3} \leq \epsilon,$$

which contradicts our assumption. $\square$

3.5 L^p Spaces

L^p spaces are amongst the most important vector spaces of functions in analysis and we shall in this section look at some of their properties.

Definition 3.15. Let $(X, \mathcal{A}, \mu)$ be a measure space and $1 \leq p \leq \infty$. Then, for $f : X \to [-\infty, \infty]$ measurable, define

$$\|f\|_{L^p(\mu)} := \left(\int |f|^p \mathrm{d}\mu \right)^{1/p}$$

if $p < \infty$, and

$$\|f\|_{L^\infty(\mu)} := \inf\{C \in \mathbb{R} \mid |f| \leq C, \mu - \text{a.e.}\},$$

if $p = \infty$. With that, we define $\mathcal{L}^p$ to be the space

$$\mathcal{L}^p := \{f : X \to [-\infty, \infty] \mid f \text{ measurable}, \|f\|_{L^p(\mu)} < \infty\}.$$

For shorthand, whenever the measure is clear from the context, we denote the L^p norm as $\| \cdot \|_p$.

Sometimes, the norm $\| \cdot \|_{L^\infty(\mu)}$ is refereed to as the essential supremum norm.

Clearly, for all $f \in \mathcal{L}^\infty(\mu)$, we have

$$|f(x)| \leq \|f\|_{L^\infty(\mu)}$$

μ-a.e. Indeed, by definition of the infimum, we can construct the sequence $(C_k)_{k=0}^\infty \subseteq \{C \in \mathbb{R} \mid |f| \leq C, \mu - \text{a.e.}\}$ such that $C_k \downarrow \|f\|_{L^\infty(\mu)}$. Hence, we have $\{x \mid |f(x)| > C_k\} \uparrow \{x \mid |f(x)| > \|f\|_{L^\infty(\mu)}\}$, and so, by continuity of measures,

$$\mu(\{x \mid |f(x)| > \|f\|_{L^\infty(\mu)}\}) = \lim_{k \to \infty} \mu(\{x \mid |f(x)| > C_k\}) = 0.$$

We would like to turn $\mathcal{L}^p(\mu)$ into a normed vector space but the obvious choice for the norm – $\| \cdot \|_{L^p(\mu)}$ is problematic since $\| \cdot \|_{L^p(\mu)}$ is not positive-definite[2] as if $f = g$ μ-a.e. then $\|f - g\|_p = 0$ but $f - g \neq 0$. To fix this, we quotient $\mathcal{L}^p(\mu)$ by the equivalence relation by $f \sim g \iff f = g$ μ-a.e. With that, we may define the L^p space.

Definition 3.16 ($L^p(\mu)$ Space)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space and $1 \leq p \leq \infty$. Then, the L^p space if defined to be the vector space

$$L^p(\mu) := \{[f] \mid f \in \mathcal{L}^p(\mu)\}$$

equipped with the norm $\|[f]\|_{L^p(\mu)} := \inf_{g \in [f]} \|g\|_{L^p(\mu)} = \|f\|_{L^p(\mu)}$.

As a convention, we often identify the equivalent class $[f]$ by some choice $f \in [f]$, thus writing $f \in L^p(\mu)$. This is justified as long as any manipulations we apply on this equivalence class traverses across equality μ-a.e.

Lemma 3.3 (Young's Inequality)**.** Let $1 < p, q < \infty$ be conjugate, i.e. $1/p + 1/q = 1$. Then

$$ab \leq \frac{a^p}{p} + \frac{b^q}{q},$$

for all $a, b \geq 0$.

Proof. See exercises. $\square$

Corollary 9.1 (Hölder's Inequality)**.** Let $1 \leq p, q \leq \infty$ be conjugate and $f \in L^p(\mu)$ and $g \in L^q(\mu)$. Then $fg \in L^1(\mu)$ and

$$\|fg\|_{L^1(\mu)} \leq \|f\|_{L^p(\mu)} \|g\|_{L^q(\mu)}.$$

Proof. Wlog. assume $p \leq q$. If $p = 1$ and $q = \infty$, then the inequality follows as $|fg| = |f||g| \leq |f| \|g\|_\infty$ μ-a.e. so let us suppose $p, q < \infty$.

[2] $\|f\|_{L^p(\mu)} > 0$ for all $f \neq 0$

In the case that $\|f\|_p = \|g\|_p = 1$, we have

$$\int |fg| \mathrm{d}\mu \leq \int \frac{|f|^p}{p} + \frac{|g|^q}{q} \mathrm{d}\mu = \frac{1}{p}\|f\|_p^p + \frac{1}{q}\|g\|_q^q = \frac{1}{p} + \frac{1}{q} = 1.$$

For the general case, we can simply reduce the inequality to the previous case by normalising the functions, i.e. consider the inequality on $\tilde{f} = f/\|f\|_p$ and $\tilde{g} = g/\|g\|_q$. $\square$

Corollary 9.2 (Minkowski's Inequality)**.** Let $1 \leq p \leq \infty$ and $f, g \in L^p(\mu)$. Then

$$f + g \in L^p(\mu) \text{ and } \|f + g\|_{L^p(\mu)} \leq \|f\|_{L^p(\mu)} + \|g\|_{L^p(\mu)}.$$

Proof. See exercises. $\square$

Theorem 10. The space $L^p(\mu)$ is a complete, normed vector space ($\equiv$ a Banach space) for all $1 \leq p \leq \infty$.

Proof. Clearly, we have positive definiteness since $\|f\|_p = 0 \iff f \in [0] \iff [f] = [0]$. Furthermore, linearity of the norm is also clear by linearity of the integrals. Thus, as the triangle inequality is simple Minkowski's inequality, we have $L^p(\mu)$ forms a normed vector space.

Let $(f_n)_{n=1}^\infty$ be a Cauchy sequence in L^p where $p < \infty$, then, it suffices to show that (f_n) converges in L^p. Let $1 \leq p < \infty$ and choose a subsequence (f_{n_k}) such that $\|f_n - f_m\|_p < 2^{-k}$ for all $k \geq 1$, $m, n \geq n_k$. Define $g_l = \sum_{k=1}^l |f_{n_{k+1}} - f_{n_k}|$ and $g = g_\infty$, then, since $0 \geq g_l \uparrow g$, by the monotone convergence theorem,

$$\|g\|_p = \left(\int |g|^p \mathrm{d}\mu\right)^{1/p} = \lim_{l \to \infty} \left(\int |g_l|^p \mathrm{d}\mu\right)^{1/p} \leq \lim_{l \to \infty} \sum_{k=1}^l \|f_{n_{k+1}} - f_{n_k}\| < \infty,$$

and hence, $g \in L^p(\mu)$. It follows that

$$f(x) := f_{n_1}(x) + \sum_{k=1}^\infty (f_{n_{k+1}}(x) - f_{n_k}(x))$$

whenever $g(x) < \infty$ and 0 otherwise is well-defined. By observing that the definition of $f(x)$ is a telescoping sum for finitely many terms, we have

$$f(x) = f_{n_j}(x) + \sum_{k=j}^\infty (f_{n_{k+1}}(x) - f_{n_k}(x))$$

for all $j \in \mathbb{N}$. Thus, for all $k \in \mathbb{N}$, we have

$$|f(x) - f_{n_k}(x)| = \left|\sum_{l=k}^\infty f_{n_{k+1}}(x) - f_{n_k}(x)\right| \leq g(x).$$

μ-a.e. Thus, by the (convergence in measure version) dominated convergence theorem, we have $f \in L^p$ and

$$\int |f - f_{n_k}|^p \mathrm{d}\mu \to 0.$$

Then $f_n \to f$ in general by Minkowski's inequality as

$$\|f - f_n\|_p \leq \|f - f_{n_k}\|_p + \|f_{n_k} - f_n\| \leq 2^{-k} + \|f - f_{n_k}\|_p,$$

for some appropriate $n_k > n$. Thus, by taking $n \to \infty$, the result follows.

For the case that $p = \infty$, the result follows easily from the definition (exercise). $\square$

The L^p space might appear "vast" but for instance, we have (proof omitted) that the space of continuous functions with compact support $C_c^\circ(\mathbb{R}^n)$ is dense in $L^p(\mathbb{R}^n)$ and so, we may approximate every function in L^p by a sequence of continuous functions with compact support, providing us with a easier mental image to work with.