Contents

1 Introduction

Welcome to measure theory. This course introduces the modern theory of functions and integration which underpins most advanced analysis topics. In particular the theory of function spaces will be important in PDEs and the notion of measurable functions allows us to rigorously understand random variables.

The key example we will study is *Lebesgue measure* in $\mathbb{R}^d$. The goal of defining Lebesgue measure is to find a way of asigning length/area/volume/whatever its called if $d \geq 4$ to a subset of $\mathbb{R}^d$. It turns out that it is not possible to do this for every possible subset of $\mathbb{R}^d$, but it is possible to do this for every subset you are likely to come accross!

1.1 Integration

One of the most important results of measure theory is the ability to integrate 'against' the measures that we define. We want this new definition of the integral to agree with the Riemann integral on subsets of $\mathbb{R}^d$ and also allow us to integrate over sets that aren't subsets of $\mathbb{R}^d$ or with different weightings of the different parts of $\mathbb{R}^d$. This new theory of integration allows us to rigorously define expectation in probability theory and provides numerous convergence theorems which are some of the results you will use most from this course.

1.2 The most important things you will learn in this course

For your own knowledge of how measure and function 'really' work:

- How Lebesgue measure is constructed.
- How the Lebesgue integral is constructed.
- How product measures/spaces are constructed.
- How L^p spaces are defined.

For use in later courses:

- The fact that Lebesgue measure exists and does what you expect it to do.
- Why you can work with measures just by looking at how they behave on a π-system (find out what that is soon!).
- The different ways in which functions can converge.
- Equivalences between ways things converge.
- Convergence theorems: i.e. when convergence of functions implies convergence of their integrals.
- Important inequalities: Hölder/Cauchy-Schwartz, Minkowski, Jensen.
- When you can switch the order of integration.

2 σ-algebras, definition of a measure

2.1 Collections of subsets

We begin with some dry definitions of collections of sets and functions from collections of sets to $\mathbb{R}$. These will give us the key formal definition of measures.

We begin with the most basic defnition. An algebra is is collection of sets closed under finite set operations.

Definition 2.1 (Algebra)**.** A collection of subsets of a space E, $\mathcal{A}$, is called and algebra if

- $\emptyset \in \mathcal{A}$, $E \in \mathcal{A}$.
- If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.
- If $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$.
- If $A, B \in \mathcal{A}$ then $A \cap B \in \mathcal{A}$.

We next define a σ-algebra. This is the *key* defnition of a collection of sets for measure theory. The letter σ here denotes countability. A σ-algebra is a collection of subsets of a space E, which are closed under countable set operations.

Definition 2.2 (σ-algebra)**.** A collection of subsets of a space E, $\mathcal{A}$ is a σ-algebra if

- $\emptyset \in \mathcal{A}, E \in \mathcal{A}$.
- If $A \in \mathcal{A}$ then $A^c \in \mathcal{A}$.
- If $A_1, A_2, \ldots$ is a countable collection of sets in $\mathcal{A}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.
- If $A_1, A_2, \ldots$ is a countable collection of sets in $\mathcal{A}$ then $\bigcap_{n=1}^{\infty} A_n \in \mathcal{A}$.

Example 2.3. In this course we really only deal with one 'concrete', non-trivial example of a σ-algebra. This is complicated to introduce and we will discuss bellow. However, in order to better understand the definition we give a few examples of things which are, and are not σ-algebras.

- If E is a space the collection of all subsets of E is a σ-algebra.
- If E is an uncountably large space then the collection of all countable subsets of E is not a σ-algebra

We now collect some results and further definitions about σ-algebras.

Lemma 2.4. *Suppose E is a space and $\mathcal{C}$ is a collection of σ algebras possibly uncountable. Then $\bigcap_{\mathcal{A} \in \mathcal{C}} \mathcal{A}$ is also a σ-algebra.*

Proof. It is straightforward to check that every part of the definition of a σ-algebra holds for the intersection. $\square$

Corollary 2.5. *For any collection of subsets of a space E, $\mathcal{F}$ there is a smallest σ-algebra containing $\mathscr{F}$. We call this $\sigma(\mathcal{F})$ or the σ-algebra generated by $\mathcal{F}$.*

Proof. There exists at least one σ-algebra containing $\mathcal{F}$ since the set of all subsets of E is a σ algebra. Then we can consider the non-empty intersection $\bigcap_{\mathcal{A} \in \mathcal{C}} \mathcal{A}$ where $\mathcal{C}$ is the collection of all σ-algebras which contain $\mathcal{F}$. We call this resulting σ- algebra the *σ-algebra generated by $\mathcal{F}$*. $\square$

Example 2.6 (Key example: Borel σ-algebra). If E is a topological space and $\mathcal{O}$ the family of open sets in E, then we write $\mathcal{B}(E)$ to be the σ-algebra generated by $\mathcal{O}$. This is called the Borel σ-algebra.

We are most interested in $\mathcal{B}(\mathbb{R}^d)$. We have the following result

Lemma 2.7. *$\mathcal{B}(\mathbb{R})$ is generated by the following sets.*

- *The collection of closed sets in $\mathbb{R}$.*
- *The collection of intervals of the form $(-\infty, b]$.*
- *The collection of intervals of the form $(a, b]$.*

Proof. Let us call $\mathcal{B}_1, \mathcal{B}_2, \mathcal{B}_3$ to be the σ-algebras generated by the sets above. We then want to show that $\mathcal{B}(\mathbb{R}) \supseteq \mathcal{B}_1 \supseteq \mathcal{B}_2 \supseteq \mathcal{B}_3$.

As $\mathcal{B}(\mathbb{R})$ contains all the open sets, it also contains all the closed sets (whose complements are open). Therefore, it also contains $\mathcal{B}_1$.

As $\mathcal{B}_1$ contains all the closed sets, and all the intervals $(-\infty, b]$ are closed then $\mathcal{B}_1$ contains the σ-algebra generated by these sets, namely $\mathcal{B}_2$.

As $\mathcal{B}_2$ contains $(-\infty, b]$ and $(-\infty, a]$ and is closed under complements it also contains, $(-\infty, b]$ and (a, ∞). As $\mathcal{B}_2$ is closed under intersection, this means it also contains $(a, b]$. This is true for all $a < b$ so $\mathcal{B}_2$ contains all sets of this form. Consequently, it contains $\mathcal{B}_3$.

Now we want to show that $\mathcal{B}(\mathbb{R}) \subseteq \mathcal{B}_3$. This will conclude the proof. First we note, that we can make an open interval $(a, b) = \bigcup_n (a, b - 1/n]$ where the union is taken over all $n < (b-a)^{-1}$. Now we need to show that any open set in $\mathbb{R}$ is a countable union of open intervals. Let U be such an open set then let

$$O = \bigcup_{q \in \mathbb{Q} \cap U} \bigcup_{r \in \mathbb{Q} \text{ s.t } (q-r,q+r) \subseteq U} (q-r, q+r).$$

Then since O is a union of subsets of U then $O \subseteq U$. Suppose that $x \in U$ then there exists some ρ such that $(x-\rho, x_\rho) \subseteq U$. There is some rationals q, r such that $x \in (q-r, q+r) \subseteq (x-\rho, x+\rho)$ therefore $x \in$). Consequently $U = O$. ☐

We have two further defnitions of collections of sets which will be useful. These separate the two parts of the defnition of a σ-algebra.

Definition 2.8 (π-system)**.** A collection of subsets of E, $\mathcal{A}$ is a π-system if

- $\emptyset \in \mathcal{A}$
- If $A, B \in \mathcal{A}$ then $A \cap B \in \mathcal{A}$.

Definition 2.9 (D-system)**.** A collection of subsets of $E, \mathcal{A}$ is a d-system if

- $E \in \mathcal{A}$.
- If $A, B \in \mathcal{A}$ with $A \subset B$ then $B \setminus A \in \mathcal{A}$.
- If $A_1 \subset A_2 \subset A_3 \subset \dots$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

Lemma 2.10 (Dynkin's π-system lemma)**.** *Let $\mathcal{A}$ be a π-system. Then any d-system containing $\mathcal{A}$ also contains the σ-algebra generated by $\mathcal{A}$.*

Proof. This is an exercise. ☐

2.2 Set functions

Definition 2.11 (Set function)**.** A set function ϕ is a function from a family of subsets of a space E, $\mathcal{A}$ to $\mathbb{R}$.

Definition 2.12 (Measure)**.** A measure is a specific type of set function which satisfies certain axioms. A set function μ defined from a σ-algebra $\mathcal{A}$ is a measure if,

- $\mu(A) \geq 0$ for every $A \in \mathcal{A}$.
- $\mu(\emptyset) = 0$
- If $A_1, A_2, A_3, \dots$ are all pairwise disjoint and in $\mathcal{A}$ then

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

We call this last property *countable additivity.*

Example 2.13 (Delta (function)). You've probably seen $\delta_{x_0}(x)$ used before; it is similar to the Kroeneker delta which appears in discrete spaces $\delta_{x,y} = 1$ if and only if $x = y$. This is the 'function' defined by $\int \delta_{x_0}(x) f(x) \mathrm{d}x = f(x_0)$. We can define a measure on $\mathbb{R}^d$ which will have this property by

$$\delta_{x_0}(A) = \left\{ \begin{array}{ll} 1 & x_0 \in A \\ 0 & x_0 \notin A \end{array} \right. .$$

Example 2.14 (Countable space). If $E = \{x_1, x_2, \dots\}$ is a countable space and $F : E \to \mathbb{R}_{\geq 0}$ is a non-negative function then we can define a measure by $\mu(A) = \sum_n F(x_n) 1_{x_n \in A}$. In fact any measure on a countable set can be written this way by choosing $F(x_n) = \mu(\{x_n\})$.

Example 2.15 (Function (informally)). In the course we will define this rigorously later. However we can define a measure on $\mathbb{R}^d$ by integrating a function over subsets of $\mathbb{R}^d$. If f is a non-negative function then we define $\mu_f(A) = \int_A f(x)\mathrm{d}x$.

Definition 2.16 (Measureable space)**.** We call a pair $(E, \mathcal{A})$ of a space and a σ-algebra, a *measureable space.*

Definition 2.17 (Measure space)**.** We call a triple $(E, \mathcal{A}, \mu)$ of a space, a σ-algebra and a measure a *measure space.*

Definition 2.18 (Finite measure space)**.** We call a measure space $(E, \mathcal{A}, \mu)$ *finite* if $\mu(E) < \infty$.

Definition 2.19 (σ-finite measure space)**.** We call a measure space, $(E, \mathcal{A}, \mu)$, *σ-finite* if there exists a countable collection $E_1, E_2, \dots \in \mathcal{A}$ such that

$$E = \bigcup_n E_n,$$

and

$$\mu(E_i) < \infty, \forall i.$$

Definition 2.20 (Borel measures and Radon measures)**.** A measure μ on a subset of a topological space E is called a *Borel measure* if it is a measure with respect to the Borel σ-algebra.

A Borel measure is called a *Radon measure* if for every compact set $K \in \mathcal{B}(E)$ we have that $\mu(K) < \infty$.

Lemma 2.21 (Continuity of measure)**.** *Let $(E, \mathcal{E}, \mu)$ be a measure space. Suppose that $(A_n)_n$ is a sequence of measurable sets with $A_1 \subseteq A_2 \supseteq \dots$ and $(B_n)_n$ is a sequence of measurable sets with $B_1 \supseteq B_2 \supseteq \dots$, then we have*

$$\mu\left(\bigcup_n A_n\right) = \lim_n \mu(A_n)$$

and

$$\mu\left(\bigcap_n B_n\right) = \lim_n \mu(B_n).$$

Proof. Let $\tilde{A}_n = A_n \setminus A_{n-1}$. We have that $\bigcup_n A_n = \bigcup_n \tilde{A}_n$. Furthermore, countable additivity gives us that

$$\mu\left(\bigcup_n \tilde{A}_n\right) = \sum_n \mu(\tilde{A}_n).$$

Therefore, we have $\sum_{n=1}^m \mu(\tilde{A}_n) \to \mu\left(\bigcup_n A_n\right)$. We also have $\sum_{n=1}^m \mu(\tilde{A}_n) = \mu\left(\bigcup_{n=1}^m \tilde{A}_n\right) = \mu(A_m)$.

Now we move onto the B_n, let $C_n = B_1 \setminus B_n$ then the C_n are an increasing sequence of measurable sets with $C_n \uparrow B_1 \setminus \bigcap_n B_n$. So by the first part we have $\mu\left(B_1 \setminus \bigcap_n B_n\right) = \lim_n \mu(C_n)$. Therefore

$$\mu(B_1) - \mu\left(\bigcap_n B_n\right) = \mu(B_1) - \lim_n \mu(B_n).$$

This gives the result as long as $\mu(B_1) < \infty$. If there exist an m such that $\mu(B_m) < \infty$ then we can renumber starting with m and repeat the argument above. $\square$

3 Outer measure, Lebesgue Measure

Definition 3.1 (Outer measure)**.** We write $\mathscr{P}(E)$ to be the power set of E, that is to say the set of all subsets of E. An outer measure is a function, ν, from $\mathscr{P}(E) \to \mathbb{R}_+$ such that

- $\nu(\emptyset) = 0$,
- If $A \subseteq B$ then $\nu(A) \leq \nu(B)$, (this is called *monotonicity*)
- If $A_1, A_2, \ldots$ is a sequence of subsets then $\nu\left(\bigcup_n A_n\right) \leq \sum_n \nu(A_n)$, (this is called *countable subadditivity*).

The key example of an outer measure is *Lebesgue outer measure*, defining this is our first step to defining Lebesgue measure.

Definition 3.2 (Lebesgue measure on unions of intervals)**.** Let us call $\mathcal{I}$ the set of countable unions of disjoint open intervals (a, b). That is to say I is the set of all sets of the form

$$\bigcup_n (a_1, b_1),$$

where the intervals are disjoint. Then we define a set function λ from $\mathcal{I}$ to $\mathbb{R}$ by

$$\lambda\left((a_1, b_1) \cup (a_2, b_2) \cup \cdots \cup (a_n, b_n)\right) = \sum_{i=1}^{n} (b_i - a_i).$$

Using this we can define Lebesgue outer measure.

Definition 3.3 (Lebesgue outer measure)**.** We define Lebesgue outer measure on $\mathscr{P}(\mathbb{R})$ by

$$\lambda^*(A) = \inf\{\lambda(B) \,:\, B \in \mathcal{I}, A \subset B\}.$$

Proposition 3.4. *Lebesgue outer measure is an outer measure and agrees with λ on $\mathcal{I}$.*

Proof. We need to check each part of the definition of outer measure. First the fact that $\lambda^*(\emptyset) = 0$ follows from the fact that $\emptyset \in \mathcal{I}$ and $\lambda(\emptyset) = 0$. Now suppose that $A_1 \subset A_2$, then any set $B \in \mathcal{I}$ with $A_2 \subseteq B$ also has $A_1 \subseteq B$ so

$$\inf\{\lambda(B) \,:\, B \in \mathcal{I}, A_1 \subset B\} \leq \inf\{\lambda(B) \,:\, B \in \mathcal{I}, A_2 \subset B\},$$

as the infimum over a larger set will always be smaller. Now let us turn to the countable subadditivity. Let us take some sequence $A_1, A_2, \ldots$, if $\sum_n \lambda^*(A_n) = \infty$ then we are done. Therefore we can assume that $\sum_n \lambda^*(A_n) < \infty$. Now let us fix an arbitrary $\epsilon > 0$. Now by the definition of λ^* for each n there exists some $I_n \in \mathcal{I}$ such that $A_n \subseteq I_n$ and $\lambda(I_n) \leq \lambda^*(A_n) + \epsilon 2^{-n}$. Then the set $I = \bigcup_n I_n$ is in $\mathcal{I}$ and $\lambda(I) \leq \sum_n \lambda(I_n) \leq \sum_n \lambda(A_n) + \epsilon$. Therefore $\lambda * \left(\bigcup_n A_n\right) \leq \lambda(I) \leq \sum_n \lambda(A_n) + \epsilon$.

Lastly if A is the interval (a, b) then $(a, b) \in \mathcal{I}$ so $\lambda^*(A) \leq b - a$. Also, any element of $\mathcal{I}$ which contains (a, b) must contain an interval (c, d) with $c \leq a$ and $b \leq d$. Therefore $\lambda^*(A) \geq b - a$. $\square$

We want to turn this outer measure into a true measure. In order to do this we need to restrict λ^* to some subset of $\mathscr{P}(\mathbb{R})$.

Definition 3.5 (Lebesgue Measurable sets)**.** We call a set $A \in \mathscr{P}(\mathbb{R})$ is *Lebesgue Measureable* if for any $B \in \mathscr{P}(\mathbb{R})$ we have

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

Proposition 3.6. *The collection of Lebesgue measureable sets, $\mathscr{M}$, is a σ algebra.*

Proof. First let us notice that the definition of a Lebesgue measureable sets is symmetric in A and A^c, so $A \in \mathscr{M}$ implies that $A^c \in \mathscr{M}$.

Secondly we can see that $\emptyset \in \mathscr{M}$ as $\lambda^*(A \cap \emptyset) + \lambda^*(A \cap \emptyset^c) = \lambda^*(\emptyset) + \lambda^*(A \cap E) = 0 + \lambda^*(A)$. This also implies via the first point that $E \in \mathscr{M}$.

We then show that if $A_1, A_2 \in \mathscr{M}$ then $A_1 \cup A_2 \in \mathscr{M}$. Using the fact that $A_1 \in \mathscr{M}$ we have

$$\lambda^*(B \cap (A_1 \cup A_2)) = \lambda^*(B \cap (A_1 \cup A_2) \cap A_1) + \lambda^*(B \cap (A_1 \cup A_2) \cap A_1^c) = \lambda^*(B \cap A_1) + \lambda^*(B \cap A_2 \cap A_1^c).$$

We also have the identity $(A_1 \cup A_2)^c = A_1^c \cap A_2^c$ therefore

$$\lambda^*(B \cap (A_1 \cup A_2)) + \lambda^*(B \cap (A_1 \cup A_2)^c) = \lambda^*(B \cap A_1) + \lambda^*(B \cap A_2 \cap A_1^c) + \lambda^*(B \cap A_1^c \cap A_2^c).$$

Then since $A_2 \in \mathscr{M}$ we have

$$\lambda^*(B \cap A_2 \cap A_1^c) + \lambda^*(B \cap A_1^c \cap A_2^c) = \lambda^*(B \cap A_1^c).$$

Therefore,

$$\lambda^*(B \cap (A_1 \cup A_2)) + \lambda^*(B \cap (A_1 \cup A_2)^c) = \lambda^*(B \cap A_1) + \lambda^*(B \cap A_1^c).$$

Then we use again the fact that $A_1 \in \mathscr{M}$ to get

$$\lambda^*(B \cap (A_1 \cup A_2)) + \lambda^*(B \cap (A_1 \cup A_2)^c) = \lambda^*(B).$$

This shows that $A_1 \cup A_2 \in \mathscr{M}$.

Now let us take an infinite sequence of disjoint sets $A_1, A_2, A_3, \ldots$ then we will show

$$\lambda^*(B) = \sum_{i=1}^{n} \lambda^*(B \cap A_i) + \lambda^*\left(B \cap \left(\bigcap_{i=1}^{n} A_i^c\right)\right).$$

We can show this by induction. For the base case it just follows with $n = 1$ from the fact that $A_1 \in \mathscr{M}$. Then by induction suppose we know that

$$\lambda^*(B) = \sum_{i=1}^{n-1} \lambda^*(B \cap A_i) + \lambda^*\left(B \cap \left(\bigcap_{i=1}^{n-1} A_i^c\right)\right).$$

Now since $A_n \in \mathscr{M}$ we have

$$\lambda^*\left(B \cap \left(\bigcap_{i=1}^{n-1} A_i^c\right)\right) = \lambda^*\left(B \cap A_n \left(\bigcap_{i=1}^{n-1} A_i^c\right)\right) + \lambda^*\left(B \cap A_n^c \left(\bigcap_{i=1}^{n-1} A_i^c\right)\right).$$

Now since A_n is disjoint from $A_1, \ldots, A_{n-1}$ we have that $A_n \cap \left(\bigcap_{i=1}^{n-1} A_i^c\right) = A_n$ so we have

$$\lambda^*\left(B \cap \left(\bigcap_{i=1}^{n-1} A_i^c\right)\right) = \lambda^*(B \cap A_n) + \lambda^*\left(B \cap \left(\bigcap_{i=1}^{n} A_i^c\right)\right).$$

This gives our induction step.

By monotonicity of the outer measure this gives that for any n we have

$$\lambda^*(B) \geq \sum_{i=1}^{n} \lambda^*(B \cap A_i) + \lambda^*\left(B \cap \left(\bigcap_{i=1}^{\infty} A_i^c\right)\right).$$

Consequently we can let n tend to infinity to get

$$\lambda^*(B) \geq \sum_{i=1}^{\infty} \lambda^*(B \cap A_i) + \lambda^* \left(B \cap \left(\bigcap_{i=1}^{\infty} A_i^c \right) \right).$$

Now we can use the countable subadditivity of λ^* to get

$$\lambda^*(B) \geq \lambda^* \left(B \cap \left(\bigcup_{i=1}^{\infty} A_i \right) \right) + \lambda^* \left((B \cap \left(\bigcap_{i=1}^{\infty} A_i^c \right) \right) = \lambda^* \left(B \cap \left(\bigcup_{i=1}^{\infty} A_i \right) \right) + \lambda^* \left(B \cap \left(\bigcup_{i=1}^{\infty} A_i \right)^c \right).$$

Furthermore, the subadditivity of λ^* gives

$$\lambda^*(B) \leq \lambda^* \left(B \cap \left(\bigcup_{i=1}^{\infty} A_i \right) \right) + \lambda^* \left(B \cap \left(\bigcup_{i=1}^{\infty} A_i \right)^c \right).$$

Therefore,

$$\lambda^*(B) = \lambda^* \left(B \cap \left(\bigcup_{i=1}^{\infty} A_i \right) \right) + \lambda^* \left(B \cap \left(\bigcup_{i=1}^{\infty} A_i \right)^c \right).$$

We have now shown that $\mathscr{M}$ is closed under complements and taking countable unions and contains $\emptyset$ which is sufficient to show that $\mathscr{M}$ is a σ-algebra. □

Proposition 3.7. *The restriction of λ^* to $\mathscr{M}$ is a measure.*

Proof. We need to show that λ^* is countably additive on $\mathscr{M}$ so let $A_1, A_2, \ldots$ be a sequence of disjoint subsets in $\mathscr{M}$. In the proof that $\mathscr{M}$ is a σ-algebra we showed that

$$\lambda^*(B) \geq \sum_{i=1}^{\infty} \lambda^*(B \cap A_i) + \lambda^* \left(B \cap \left(\bigcap_{i=1}^{\infty} A_i^c \right) \right).$$

Now let us take the particular case where $B = \bigcup_{i=1}^{\infty} A_i$ this gives

$$\lambda^* \left(\bigcup_{i=1}^{\infty} A_i \right) \geq \sum_{i=1}^{n} \lambda^*(A_i) + \lambda^* \left(\left(\bigcup_{i=1}^{\infty} A_i \right) \cap \left(\bigcup_{i=1}^{\infty} A_i \right)^c \right) = \sum_{i=1}^{\infty} \lambda^*(A_i).$$

Countable subadditivity gives

$$\lambda^* \left(\bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{n} \lambda^*(A_i),$$

so consequently

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

□

Remark 3.8. We now call the restriction of λ^* to $\mathscr{M}$, λ and call is *Lebesgue measure.*

We now want to know that there are some Lebesgue measureable sets. In order to do this we first show that all the intervals of the form $(-\infty, b]$ are Lebesgue measurable.

Lemma 3.9. *The intervals of the form $(-\infty, b]$ are Lebesgue measureable.*

Proof. □

Corollary 3.10. *Every set in $\mathcal{B}(\mathbb{R})$ is Lebesgue measurable.*

Proof. The Borel σ algebra is the σ algebra generated by sets of the form $(-\infty, b]$ as shown in Lemma REF. Therefore, as $\mathscr{M}$ is a σ-algebra and contains all the intervals of the form $(-\infty, b]$ then it contains the Borel σ-algebra. $\square$

The construction of Lebesgue measure via the outer measure can be generalised via Carathéodory's extension theorem. We briefly give the defintion of a ring of subsets.

Definition 3.11 (Ring)**.** A collection of subsets, $\mathcal{A}$, of a space E is called a ring if for every $A, B \in \mathcal{A}$ we have $A \setminus B \in \mathcal{A}$ and $A \cup B \in \mathcal{A}$.

Now we introduce Carathéodory's Extension theorem. We can see that the proof is in many ways very similar to the construction of Lebesgue measure.

Theorem 3.12 (Carathéodory's Extension Theorem)**.** *Let $\mathcal{A}$ be a ring of subsets of E, and let $\mu : \mathcal{A} \to [0, \infty]$ be a countably additive set function. Then μ extends to a measure on $\sigma(\mathcal{A})$.*

Proof. We define the outer measure μ^* on $\mathscr{P}(E)$ by

$$\mu^*(B) = \inf \left\{ \sum_n \mu(A_n) \,:\, A_n \in \mathcal{A} \forall n, B \subset \bigcup_n A_n \right\}.$$

$\mu^*(B) = \infty$ if there is not possible sequence of A_n so that B is contained in their union. We can see immediately that $\mu^*(\emptyset) = 0$ and μ^* is increasing.

As before we define $\mathscr{M}$ to be the set of μ^* measurable sets A that satisfy, for every $B \subseteq E$ that

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

We want to show that $\mathscr{M}$ is a σ-algebra and μ^* restrics to a measure on $\mathscr{M}$.

First we show that μ^* is countably subadditive. Suppose that we have a sequence B_n and want to show that

$$\mu^* \left(\bigcup_n B_n \right) \leq \sum_n \mu^*(B_n).$$

Let us fix some $\epsilon > 0$ then for each n there is a sequence $A_{n,m} \in \mathcal{A}$ such that $B_n \subset \bigcup_m A_{n,m}$ and $\sum_m \mu(A_{n,m}) \leq \mu^*(B_n) + \epsilon 2^{-n}$. Then $\bigcup_n B_n \subset \bigcup_{n,m} A_{n,m}$ and $\sum_{n,m} \mu(A_{n,m}) \leq \sum_n \mu^*(B_n) + \epsilon$. Therefore $\mu^* \left(\bigcup_n B_n \right) \leq \sum_n \mu(B_n) + \epsilon$. Since ϵ is arbitrary this gives the countable subadditivity.

Now we show that μ^* agrees with μ on $\mathcal{A}$. Let us take $A \in \mathcal{A}$ clearly $A \subseteq A$ so $\mu^*(A) \leq \mu(A)$. Now suppose that there is a sequence $A_n \in \mathcal{A}$ such that $A \subseteq \bigcup_n A_n$. Then $A \cap A_n = A \setminus (A \setminus B) \in \mathcal{A}$. Therefore we use the countable subadditivity of μ on $\mathcal{A}$ to get

$$\mu(A) \leq \sum_n \mu(A_n \cap A) \leq \sum_n \mu(A_n).$$

Taking the infinmum over such sequences gives $\mu(A) \leq \mu(A_n)$. Therefore μ and μ^* agree on $\mathcal{A}$.

Now we show that $\mathscr{M}$ contains $\mathcal{A}$. That is to say we want to show that if $A \in \mathcal{A}$ then for every B

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

Using subadditivity of μ^* we have that $\mu^*(B) \leq \mu^*(B \cap A) + \mu^*(B \cap A^c)$. Therefore we want to show $\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c)$. Let A_n be a sequence in $\mathcal{A}$ such that $\mu^*(B) \geq \sum_n \mu(A_n) - \epsilon$, then we already know that $A \cap A_n$ will be in $\mathcal{A}$ we also have that $A^c \cap A_n = A_n \setminus (A \cap A_n) \in \mathcal{A}$. Therefore $\mu(B \cap A) \leq \sum_n \mu(A_n \cap A)$ and $\mu(B \cap A^c) \leq \sum_n \mu(A^c \cap A_n)$ and consequently

$$\mu^*(B \cap A) + \mu^*(B \cap A^c) \leq \sum_n (\mu(A_n \cap A) + \mu(A_n \cap A^c)) = \sum_n \mu(A_n) \leq \mu^*(B) + \epsilon.$$

As ϵ is arbitrary this gives the required result.

The next step is to show that $\mathscr{M}$ is a σ-algebra. We start with the algebra part. E and $\emptyset$ are in $\mathscr{M}$ as

$$\mu^*(B) = \mu^*(B \cap E) + \mu^*(B \cap \emptyset),$$

just because $B \cap E = B$ and $B \cap \emptyset = \emptyset$ and we know $\mu^*(\emptyset) = 0$. We also can see that

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

is symmetric in exchanging A and A^c so if $A \in \mathscr{M}$ then so is A^c. Now suppose $A_1, A_2 \in \mathscr{M}$. We notice that $(A_1 \cap A_2)^c \cap A_1 = (A_1^c \cup A_2^c) \cap A_1 = (A_1^c \cap A_1) \cup (A_2^c \cap A_1) = A_2^c \cap A^1$ and $A_1^c = A_1^c \cap (A_1 \cap A_2)^c$. Using this and the fact that A_1, A_2, A_1^c, A_2^c are in $\mathscr{M}$ we have

$$\begin{aligned}
\text{Using that } A_1 \in \mathscr{M} \quad \mu^*(B) &= \mu^*(B \cap A_1) + \mu^*(B \cap A_1^c) \\
\text{Using that } A_2 \in \mathscr{M} \quad &= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap A_1 \cap A_2^c) + \mu^*(B \cap A_1^c) \\
\text{Using our first identity} \quad &= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap A_1^c) \\
\text{Using our second identiy} \quad &= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1) + \mu^*(B \cap (A_1 \cap A_2)^c \cap A_1^c) \\
\text{Using the fact that } A_1 \in \mathscr{M} \quad &= \mu^*(B \cap (A_1 \cap A_2)) + \mu^*(B \cap (A_1 \cap A_2)^c).
\end{aligned}$$

Now that we have shown that $\mathscr{M}$ contains finite unions we want to show it countains countable unions. Let A_n be a sequence of disjoin sets in $\mathscr{M}$. Let us write $A = \bigcup_n A_n$. Then itterating our previous result we have for any B, n that

$$\mu^*(B) = \sum_{k=1}^{n} \mu^*(B \cap A_k) + \mu^*(B \cap A_1^c \cap \cdots \cap A_n^c).$$

Now as $A^c \subseteq A_1^c \cap A_2^c \cdots \cap A_n^c$ for each n we have $\mu^*(B \cap A^c) \leq \mu^*(B \cap A_1^c \cap \cdots \cap A_n^c)$. Therefore for each n

$$\mu^*(B) \geq \sum_{k=1}^{n} \mu^*(B \cap A_k) + \mu^*(B \cap A^c).$$

Letting $n \to \infty$ we have

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

Now we use the countable subadditivity of μ^* and the fact that $B \cap A = \bigcup_n (B \cap A_n)$ to get

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

As the other inequality holds by subadditivity of μ^* we have that $\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c)$ and hence $A \in \mathscr{M}$.

Lastly, we want to show that μ^* is a measure on $\mathscr{M}$. In order to do this we need to show that μ^* is countably additive on $\mathscr{M}$. In the last step we showed that for any B, and a suquence of disjoint sets A_n in $\mathscr{M}$ with $A = \bigcup_n A_n$, that

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

If we apply this identity with $B = A$ and use the fact that $A_n \cap A = A_n$ we get

$$\mu^*(A) \geq \sum_n \mu^*(A_n).$$

Since we already know that μ^* is countably subadditive this is sufficient to show that μ^* is countably additive and hence a measure on $\mathscr{M}$. □

Theorem 3.13 (Uniqueness of Extension)**.** *Let μ_1 and μ_2 be measures on $(E, \mathcal{E})$ with $\mu_1(E) = \mu_2(E) < \infty$. Suppose that $\mu_1 = \mu_2$ on $\mathcal{A}$ where $\mathcal{A}$ is a π-system generating $\mathcal{E}$, then $\mu_1 = \mu_2$ on $\mathcal{E}$.*

Proof. Let us consider $\mathcal{D} \subseteq \mathcal{E}$ defined as the measurable sets on which $\mu_1(A) = \mu_2(A)$. By hypothesis $E \in \mathcal{D}$ and $\mathcal{A} \subseteq \mathcal{D}$. We want to show that $\mathcal{D}$ is a σ-algebra and therefore $\mathcal{D} = \mathcal{E}$. Suppose that $A, B \in \mathcal{E}$ with $A \subseteq B$ then we have $\mu_i(A) + \mu_i(B \setminus A) = \mu_i(B) < \infty$. This means that if A and B are in $\mathcal{D}$ then so is $A \setminus B$. Therefore, $\mathcal{D}$ is a d-system containing the π-system $\mathcal{A}$ so by Dynkin's lemma is equal to $\mathcal{E}$. □

3.1 Properties of Lebesgue measure

This collection is a section of facts about Lebesgue measure and the set of Lebesgue measurable sets. We start with looking at $\mathscr{M}$ the σ-algebra of Lebesgue measurable sets.

Lemma 3.14 (Null sets are all Lebesgue measurable)**.** *If A in $\mathscr{P}(\mathbb{R})$ and $\lambda^*(A) = 0$ then $A \in \mathscr{M}$.*

Proof. This is on the assignment. □

We can actually characterise all Lebesgue measurable sets in terms of Null sets and Borel sets both of which we have shown are measurable. We have the following propersition which is too hard to prove in this course.

Proposition 3.15. *A set $S \subseteq \mathbb{R}$ is Lebesgue measurable if and only there exists a Borel set B and a null set N such that $S = B \triangle N$.*

The most important thing we want to prove about $\mathscr{M}$ is that there exists a non-Lebesgue measurable set. Before we do this we need to explore a few more properties of Lebesgue measure itself.

Proposition 3.16. *Lebesgue measure is* regular *that is to say*

- $\lambda(A) = \inf\{\lambda(U) : U \textit{ is open}, A \subseteq U\}$,
- $\lambda(A) = \sup\{\lambda(K) : K \textit{ is compact}, K \subseteq A\}$.

Proof. By monotonicity we can see that $\lambda(A) \leq \inf\{\lambda(U) : U \text{ is open}, A \subseteq U\}$. Furthermore we can find a sequence of half open rectangles R_k such that $A \subseteq \bigcup_n R_n$ and $\sum_n \lambda(R_n) \leq \lambda(A) + \epsilon$. By slightly enlarging each of the half open rectangles we can produce another sequence of fully open rectangles $\tilde{R}_n$ such that $A \subseteq \bigcup_n \tilde{R}_n$ and $\lambda(A) \geq \sum_n \lambda(\tilde{R}_n) - 2\epsilon$. The set, $\bigcup_n \tilde{R}_n$ is open and ϵ can be made arbitrarily small so this shows $\lambda(A) \leq \inf\{\lambda(U) : U \text{ is open}, A \subseteq U\}$.

Monotonicity shows that $\lambda(A) \geq \sup\{\lambda(K) : K \text{ is compact}, K \subseteq A\}$. First let us assume that A is contained in some ball, B around 0. Now use the first part to find some open set U such that $B \setminus A \subseteq U$ and $\lambda(U) \leq \lambda(B \setminus A) + \epsilon$. Now let $K = B \setminus U$ then we have $A \subseteq K \subseteq B$ and $\lambda(K) = \lambda(B) - \lambda(U) \geq \lambda(B) - \lambda(B \setminus A) - \epsilon = \lambda(A) - \epsilon$ (here we use the fact that B, A, U, K will all have finite measure as they are inside B). As ϵ is arbitrary this concludes the proof when A is contained in a ball.

Now suppose that A is unbounded. Then let $A_n = A \cap B_n$ where B_n is the closed ball of radius n. We have that $\lambda(A_n) \to \lambda(A)$. If $\lambda(A) = \infty$ then we can find $K_n \subseteq A_n$ with $\lambda(K_n)$ arbitrarily close to

$\lambda(A_n)$ therefore we can find such a sequence with $\lambda(K_n) \to \infty$. If $\lambda(A) \neq \infty$ then, given ϵ, there exists N such that $\lambda(A_n) \geq \lambda(A) - \epsilon$ for $n \geq N$. Then we can fine $K_N \subseteq A_N$ such that $\lambda(K_n) \geq \lambda(A_N) - \epsilon$ therefore $\lambda(K_N) \geq \lambda(A) - 2\epsilon$. This shows we can a compact set which is contained in A, with measure arbitrarily close to that of A. □

We now want to show that Lebesgue measure is the only which assigns each interval the correct measure.

Proposition 3.17. *Lebesgue measure is the only measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which assigns each half open interval its length. This is equally true with half open hyper-rectangles in $\mathbb{R}^d$.*

Proof. The collection of half open intervals is a π-system which generates the Borel σ-algebra. Therefore, we can use Dynkin's uniqueness of extension Lemma to get that any other measure which agrees with Lebesgue measure on the half open intervals must agree with Lebesgue measure on the whole of the Borel σ-algebra. □

Corollary 3.18. *Lebesgue measure is translation invariant. That is to say if we define the set $x + A = \{x + y, y \in A\}$ then $\lambda(x + A) = \lambda(A)$*

Proof. Define a new measure λ_x by $\lambda_x(A) = \lambda(x+A)$ then $\lambda_x((a, b]) = \lambda((a+x, b+x]) = b+x-(a+x) = b - a$. Therefore λ_x agrees with λ on the half open intervals and therefore agrees with λ on the whole of $\mathcal{B}(\mathbb{R})$. Again it is straightforward to extend this to $\mathbb{R}^d$. □

Lastly, in the construction of Lebesgue measure we show that $\mathscr{M}$ is not the whole of $\mathscr{P}(\mathbb{R})$ and that there exist non-Lebesgue measureable sets.

Proposition 3.19. *There exists sets that are in $\mathscr{P}(\mathbb{R})$ which are not in $\mathscr{M}$.*

Proof. This proof involves the use of the axiom of choice. In fact it is known that it is necessary to use some form of the axiom of choice to prove the existence of a non-Lebesgue measurable set in $\mathbb{R}$.

We use an argument by contradiction, we begin by assuming every subset of $\mathbb{R}$ is Lebesgue measurable. We define an equivalence relation on $[0, 1)$ by saying $x \sim y$ exactly when $x - y \in \mathbb{Q}$. Using the axiom of choice we find a subset S of $[0, 1)$ which contains exactly one representative of each equivalence class. Next we define the set $S + q = \{s + q \pmod 1 : s \in S\}$ for each $q \in \mathbb{Q} \cap [0, 1)$. Then by our choice of S we have that

$$[0, 1) = \bigcup_{q \in \mathbb{Q} \cap [0,1)} (S + q),$$

where this union is disjoint. We can also see by translation invariance of λ that if S were Lebesgue measurable then we would have

$$\lambda(S) = \lambda(S + q)$$

for every q. Therefore, by countable additivity we would have

$$\lambda([0, 1)) = \sum_{q \in \mathbb{Q} \cap [0,1)} \lambda(S + q) = \sum_{q \in \mathbb{Q} \cap [0,1)} \lambda(S) = \infty.$$

□

4 Measurable Functions

A big part of measure theory is the study of functions which are compatible with the measure spaces. We begin with a basic definition which will be satisfied by all the functions we are interested in.

Definition 4.1 (Mesasurable functions)**.** If $(E, \mathcal{E})$ and $(F, \mathcal{F})$ are two measurable spaces and f is a function $E \to F$, then we say f is *measurable* if for every $A \in \mathcal{F}$ we have $f^{-1}(A) \in \mathcal{E}$.

Lemma 4.2. *Suppose that $\mathcal{A} \subset \mathcal{F}$ is such that $\sigma(\mathcal{A})$. If f is a function such that for every $A \in \mathcal{A}$ we have $f^{-1}(A) \in \mathcal{E}$ then f is measurable.*

Proof. First we note that

$$f^{-1}\left(\bigcup_i A_i\right) = \bigcup_i f^{-1}(A_i),$$

and

$$f^{-1}(B \setminus A) = f^{-1}(B) \setminus f^{-1}(A).$$

Now if we consider $\{A \in \mathcal{F} : f^{-1}(A) \in \mathcal{E}\}$ then this is a σ-algebra, as $\mathcal{E}$ is a σ-algebra and f^{-1} preserves set operations. Therefore, $\{A \in \mathcal{F} : f^{-1}(A) \in \mathcal{E}\}$ is a σ-algebra containing $\mathcal{A}$ therefore $\mathcal{E} \subseteq \{A \in \mathcal{F} : f^{-1}(A) \in \mathcal{E}\}$ so f is measurable. □

Remark 4.3. In particular note that the above lemma means that whenever we have $f : E \to \mathbb{R}$ and $\mathbb{R}$ is equipped with the Borel σ algebra, we know that f is measurable if $f^{-1}((-\infty, b])$ is a measurable set for every b.

Lemma 4.4. *If E, F are topological spaces, equipped with their Borel σ-algebras, and we have $f : E \to F$ is a continuous map then f is measurable.*

Proof. This is on the exercise sheet. □

Lemma 4.5. *If f_n is a sequence of measurable function then the following functions are all measurable:*

- $-f_1$
- λf_1 *for* $\lambda > 0$ *a fixed contant.*
- $f_1 \wedge f_2$
- $f_1 \vee f_2$
- $f_1 + f_2$,
- $f_1 f_2$,
- $\sup_n f_n$,
- $\inf_n f_n$,
- $\limsup_n f_n$,
- $\liminf_n f_n$.

Proof. We only show the first and third result. The rest are on the assignment.

In order to show that any of these functions are measureable we want to look at $f^{-1}((-\infty, b])$ or a similar set. $(f_1 \wedge f_2)^{-1}((-\infty, b]) = \{x : \max\{f_1(x), f_2(x)\} \leq b\} = \{x : f_1(x) \leq b \,\text{and}\, f_2(x) \leq b\} = \{x : f_1(x) \leq b\} \cap \{x : f_2(x) \leq b\} = f_1^{-1}((-\infty, b]) \cap f_2^{-1}((-\infty, b])$. Now since f_1 and f_2 are both measureable the sets $f_1^{-1}((-\infty, b])$ and $f_2^{-1}((-\infty, b])$ are both measurable. We also know that the intersection of two measurable sets is measurable.

$(f_1+f_2)^{-1}((b, \infty)) = \{x : f_1(x)+f_2(x) > b\}$. Now if $f_1(x) > b - f_2(x)$ then there exists a $q \in \mathbb{Q}$ such that $f_1(x) > q > b - f_2(x)$. Let us define the set $A = \bigcup_{q \in \mathbb{Q}} \{x : f_1(x) > q\} \cap \{x : f_2(x) > b - q\}$. Since f_1, f_2 are both measurable A is a countable union of measurable sets so measurable. We can also see that if $x \in A$ then $f_1(x)+f_2(x) > b$ and our observation shows that in fact $A = \{x : f_1(x)+f_2(x) > b\}$. Therefore, $f_1 + f_2$ is measurable. □

Definition 4.6 (Image measure)**.** We can use a measurable function f to define an image measure. Suppose μ is a measure on $(E, \mathcal{E})$ and f is a measurable function $(E, \mathcal{E}) \to (F, \mathcal{F})$ then we can define a new measure ν by saying that

$$\nu(A) = \mu(f^{-1}(A)),$$

for every $A \in \mathcal{F}$. We write $\nu = \mu \circ f^{-1}$.

We can use the notion of image measure to construct further measures from Lebesgue measure.

Lemma 4.7. *Suppose $g : \mathbb{R} \to \mathbb{R}$ and that g is non-constant, right-continuous and non-decreasing. Let us define $g(-\infty) = \lim_{x \to -\infty} g(x)$ and $g(\infty) = \lim_{x \to \infty} g(x)$ and let us call the interval $I := (g(-\infty), g(\infty))$ (this might be the whole of $\mathbb{R}$. Define a partial inverse to g by $f : I \to \mathbb{R}$ by*

$$f(x) = \inf_y \{x \leq g(y)\}.$$

Then f is left-continuous and non-decreasing and $f(x) \leq y$ if and only if $x \leq g(y)$.

Proof. Fix $x \in I$ and consider the set $J_x = \{y \in \mathbb{R} : x \leq g(y)\}$ by definition of I we know that J_x is non empty and is not the whole of $\mathbb{R}$ (this shows that f is well defined). As g is non-decreasing, if $y \in J_x$ and $y' \geq y$, then $y' \in J_x$. As g is right-continuous, if $y_n \in J_x$ and $y_n \downarrow y$ then $y \in J_x$ (noting the $\leq$ sign in J_x). Now using this we have that if $x \leq x'$ then $J_x \supseteq J_{x'}$ so $f(x) \leq f(x')$. We also have that if $x_n \uparrow x$ then $J_x = \bigcap_n J_{x_n}$, so $f(x_n) \to f(x)$. □

Theorem 4.8. *Let g be a non-constant, right-continuous and non-decreasing function from $\mathbb{R} \to \mathbb{R}$. There exists a unique Radon measure on $\mathbb{R}$ such that for all $a, b \in \mathbb{R}$ with $a < b$*

$$\mathrm{d}g((a, b]) = g(b) - g(a).$$

We call this measure the Lebesgue Steitjles *measure associated with g. Furthermore, every Radon measure on $\mathbb{R}$ can be represented this way.*

Proof. Define I and f as in the Lemma above. Then we can construct $\mathrm{d}g$ as the image measure of Lebesgue measure on I. That is to say we can let $\mathrm{d}g = \lambda \circ f^{-1}$. If this is the case then

$$\mathrm{d}g((a, b]) = \lambda\left(\{x : f(x) > a, f(x) \leq b\}\right) = \lambda((g(a), g(b)]) = g(b) - g(a).$$

The standard argument for uniqueness of measures (as for that of Lebesgue measure) gives uniqueness of this measure.

Finally, if ν is a Radon measure on $\mathbb{R}$ then we can define a function g, by

$$g(y) = \nu((0, y]), \quad y \geq 0, \quad g(y) = -\nu((y, 0]), \quad y < 0.$$

Then $\nu = \mathrm{d}g$ by uniqueness. □

4.1 Convergence of measurable functions

Definition 4.9 (Almost everywhere / Almost surely)**.** We use the short hand *almost everywhere* (or *almost surely* in a probability space) to discuss properties that are true everywhere except a measure zero set.

Definition 4.10 (Convergence almost everywhere)**.** Let $(E, \mathcal{E}, \mu)$ be a measureable space. A sequence of measureable functions, $(f_n)_{n\geq 1} : E \to F$, *converges almost everywhere* to f if

$$\mu\left(\{x \in E \,:\, f_n(x) \not\to f(x)\}\right) = 0$$

Definition 4.11 (Convergence in measure)**.** Let $(E, \mathcal{E}, \mu)$ be a measureable space. A sequence of measureable functions, $(f_n)_{n\geq 1} : E \to F$, *converges in measure* to f if for every $\epsilon > 0$

$$\mu\left(\{x \,:\, |f(x) - f_n(x)| > \epsilon\}\right) \to 0, \quad \text{as}\, n \to \infty.$$

Example 4.12. The sequence of functions $f_n(x) = x^n$ converges to 0 Lebesgue almost everywhere on $[0, 1]$, and in measure, but it doesn't converge pointwise as it doesn't converge at $x = 1$.

Example 4.13. The- sequence of functions $f_n(x) = 1_{[n,n+1]}(x)$ converges to 0 Lebesgue almost everywhere (in fact everywhere) but not in measure.

Example 4.14. Consider the sequence of functions $f_1 = 1_{[0,1/2)}, f_2 = 1_{[1/2,1)}, f_3 = 1_{[0,1/4)}, f_4 = 1_{[1/4,1/2)}, f_5 = 1_{[1/2,3/4)}, f_6 = 1_{[3/4,1)}, f_7 = 1_{[0,1/8)}, f_8 = 1_{[1/8,1/4)} \ldots$ then f_n converges to 0 in measure, but $f_n(x)$ does not converge for any x.

We can prove a quasi-equivalence between these two notions of measure. Before we do this we need to introduce a very useful lemma, the Borel-Cantelli Lemma. We introduce it here as it is used to prove the following theorem but it is a useful tool to have whilst doing measure theory, particularly probability theory. First let us also introduce some more notation

Definition 4.15. Let $(A_n)_n$ be a sequence of measurable sets then we have the following names

$$\limsup_n A_n = \bigcap_n \bigcup_{m\geq n} A_m = \{A_n \,\text{infinitely often}\,\},$$

and

$$\liminf_n A_n = \bigcup_n \bigcap_{m\geq n} A_m = \{A_n \,\text{eventually}\,\}.$$

The last names are more comon when the A_n are events in a probability space.

Lemma 4.16 (First Borel-Cantelli Lemma)**.** *Let $(E, \mathcal{E}, \mu)$ be a measure space. Then if $\sum_n \mu(A_n) < \infty$ it follows that $\mu(\limsup_n A_n) = 0$).*

Proof. For any n we have

$$\mu(\limsup_n A_n) \leq \mu\left(\bigcup_{m\geq n} A_n\right) \leq \sum_{m\geq n} \mu(A_m).$$

Then the right hand side goes to zero as $n \to \infty$, so $\mu(\limsup_n A_n) = 0$. □

Lemma 4.17 (Second Borel-Cantelli Lemma)**.** *Let $(E, \mathcal{E}, \mu)$ be a probability space ($\mu(E) = 1$). Then suppose that $\mu(A_i \cap A_j) = \mu(A_i)\mu(A_j)$ (the events are pairwise independent) for every i, j and that $\sum_n \mu(A_n) = \infty$ then it will follow that $\mu(\limsup_n A_n) = 1$.*

Proof. First we note that $\mu(A_i^c \cap A_j^c) = \mu((A_i \cup A_j)^c) = 1 - \mu(A_i \cup A_j) = 1 - \mu(A_i) - \mu(A_j) + \mu(A_i)\mu(A_j) = (1-\mu(A_i))(1-\mu(A_j))$. We use the inequality $1 - a \leq e^{-a}$. Let $a_n = \mu(A_n)/\mu(E)$ then

$$\mu\left(\bigcap_{m=n}^{N} A_m^c\right)/\mu(E) = \Pi_{m=n}^{N}(1-a_m) \leq \exp\left(-\sum_{m=n}^{N} a_m\right) \to 0, \quad \text{as } N \to 0.$$

Therefore, $\mu\left(\bigcap_{m\geq n} A_m^c\right) = 0$ for every n. So $\mu(\limsup_n A_n) = 1 - \mu(\bigcup_n \bigcap -m \geq n A_m^c) = 1$. ☐

Theorem 4.18. *Let $(E, \mathcal{E}, \mu)$ be a probability space and $(f_n)_n$ be a sequence of measurable functions. Then we have the following:*

- *Suppose that $\mu(E) < \infty$ and that $f_n \to 0$ almost everywhere, then $f_n \to 0$ in measure.*
- *If $f_n \to 0$ in measure then there exists some subsequence $(n_k)_k$ such that $f_{n_k} \to 0$ almost everywhere.*

Proof. Suppose that $f_n \to 0$ almost everywhere. Then

$$\mu(\{x : |f_n(x)| \leq \epsilon\}) \geq \mu\left(\bigcap_{m\geq n}\{x : |f_m(x)| \leq \epsilon\}\right) \uparrow \mu\left(|f_n| \leq \epsilon \,\text{eventually}\right) \geq \mu(\{x : f_n(x) \to 0\}) = \mu(E),$$

therefore,

$$\mu(\{x : |f_n(x)| > \epsilon) = \mu(E) - \mu(\{x : |f_n(x)| \leq \epsilon\}) \to 0.$$

Now suppose that $f_n \to 0$ in measure. We can find a subsequence n_k such that

$$\mu(\{x : |f_{n_k}(x)| > 1/k\}) \leq 2^{-k}.$$

Therefore

$$\sum_k \mu(\{x : |f_{n_k}(x)| > 1/k\}) < \infty.$$

Therefore by the first Borel-Canteli lemma we have that

$$\mu\left(\{x : |f_{n_k}(x)| > 1/k \,\text{infinitely often}\}\right) = 0.$$

Therefore $f_{n_k} \to 0$ almost everywhere. ☐

4.2 Egoroff's Theorem and Lusin's Theorem

Theorem 4.19 (Egoroff's Theorem)**.** *Let $(E, \mathcal{E}, \mu)$ be a finite measure space and $(f_n)_n$ be a sequence of real valued measurable functions on E. If μ is finite and if f_n converges μ-almost everywhere to f then for each positive ϵ there is a set A with $\mu(A^c) < \epsilon$, such that f_n converges uniformly on A to f.*

Proof. For each n let $g_n(x) = \sup_{j\geq n}|f_j(x) - f(x)|$. Then g_n is a positive function which is finite almost everywhere. The sequence $(g_n)_n$ converges to 0 almost everywhere and so in measure. Therefore, for each positive integer k we can find n_k such that

$$\mu\left(\{x : g_{n_k} > 1/k\}\right) < \epsilon 2^{-k}.$$

Define sets $A_k = \{x : g_{n_k}(x) \leq 1/k\}$ and let $A = \bigcap_k A_k$. The set A has

$$\mu(A^c) = \mu\left(\bigcup_k A_k^c\right) \leq \sum_k \mu(A_k^c) \leq \sum_k \epsilon 2^{-k} = \epsilon.$$

We want to show that f_n converges uniformly to f on A. For each δ there exists a k such athat $1/k < \delta$, then as $A \subseteq A_k$, we have that for every $n \geq n_k$,

$$|f_n - f| \leq g_{n_k} \leq 1/k < \delta,$$

uniformly on all $x \in A_k$ and hence in A. $\square$

This proof motivates the following definition of convergence of functions.

Definition 4.20 (Almost uniform convergence)**.** We say a sequence of functions $(f_n)_{n\geq 1}$ converges *almost uniformly* on a measure space $(E, \mathcal{E}, \mu)$ if for every $\epsilon > 0$ there exists a set A with $\mu(A^c) < \epsilon$ with $f_n \to f$ uniformly on A.

We can use Egoroff's theorem to prove a result called Lusin's theorem. First let us recall the definition of regularity

Definition 4.21. Let E be a topological space and μ be a measure on $(E, \mathcal{B}(E))$ then say μ is *regular* if for every $A \in \mathcal{B}(E)$ we have

- $\mu(A) = \inf\{\mu(U) : A \subseteq U, U \text{ is open}\}$,
- $\mu(A) = \sup\{\mu(K) : K \subseteq A, K \text{ is compact}\}$.

Theorem 4.22 (Lusin's Theorem)**.** *Suppose that f is a measurable function and $A \subseteq \mathbb{R}^d$ is a Borel set and $\lambda(A) < \infty$ then for any $\epsilon > 0$ there is a compact subset K of A with $\lambda(A \setminus K) < \epsilon$ such that the restriction of f to K is continuous.*

Remark 4.23. This theorem can be generalised to locally compact Hausdorff spaces, see Cohn's book.

Proof. Suppose first that f only takes countably many values, $a_1, a_2, a_3, \dots$ on A the let $A_k = \{x \in A : f(x) = a_k\}$, by measurablility of f we can see that $A_k = f^{-1}(\{a_k\})$ is measurable. We know that $A = \bigcup_n A_n$ so by continuity of measure $\lambda(\bigcup_{k=1}^n A_k) \uparrow \lambda(A)$. Since $\lambda(A) < \infty$ we have that for any $\epsilon > 0$ there exists n such that $\mu(A \setminus \bigcup_{k=1}^n A_k) < \epsilon/2$. By the regularity of Lebesgue measure we can find compact subsets $K_1, \dots, K_n$ such that $\lambda(A_n \setminus K_n) \leq \epsilon/2n$. Then let $K = \bigcup_{k=1}^n K_k$. This is a compact subset of A and

$$\lambda(A \setminus K) \leq \lambda(A \setminus \bigcup_{k=1}^n A_k) + \lambda(\bigcup_{k=1}^n A_k \setminus \bigcup_{k=1}^n K_k) < \epsilon/2 + \epsilon/2.$$

Now we have proved the special case where f takes countably many values we can use this to prove the theorem for general f. Let $f_n = 2^{-n}\lfloor 2^n f \rfloor$ then $2^{-n} \geq f(x) - f_n(x) \geq 0$ so $f_n(x) \to f(x)$. Now by Egoroff's theorem there exists K with $\lambda(A \setminus K) < \epsilon/2$ such that $f_n \to f$ uniformly on K. Now, f_n can only take countably many values, so by our special case of Lusin's theorem there exists a $K_n \subseteq K$, compact, such that $\lambda(K \setminus K_n) \leq \epsilon 2^{-n-1}$, and f_n is continuous on K_n. Now let $K_\infty = \bigcap_n K_n$, then K_∞ is compact and $\lambda(A \setminus K_\infty) = \lambda(A \setminus K) + \lambda(K \setminus K_\infty) = \lambda(A \setminus K) + \lambda(\bigcup_n (K \setminus K_n)) \leq \epsilon/2 + \sum_n \epsilon 2^{-n-1} = \epsilon$. Now we have that f_n converges uniformly to f on K and f_n is continuous on K for each n. As the uniform limit of continuous functions is continuous this shows that f is continuous on K. $\square$

5 Integration

We now get to the definition of the Lebesgue integral which is the second important object that we construct in this course. There are several different notations for the integral of a function f with respect to a measure μ. We have

$$\mu(f) = \int_E f \mathrm{d}\mu = \int_E f(x)\mu(\mathrm{d}x).$$

When you are integrating with respect to Lebesgue measure the most common notation is

$$\int_E f(x)\mathrm{d}x.$$

Before we start constructing the integral we'll briefly discuss the motivations for how to construct it. Firstly, you've already seen the Riemann integral. We can describe the strategy of Riemann integration - very loosly - as splitting the *domain* of the function into equal sized chunks, estimating the height of the function on each chunk then adding them together. Broadly what happens with Lebesgue integration is that we split the *range* of the function into equal sized chunks, estimate the size of the part of the domain which will end up in that chunk of range then sum everything up. We need the theory of measure in order to do this because the bit of the domain corresponding to chuncks of the range can be quite weird sets whose size it wouldn't be possible to measure. The first motivation for this is that whilst Riemann integration only works for functions from subsets of $\mathbb{R}^d$ to $\mathbb{R}$, Lebesgue integration allows the domain on the function to be quite weird, (as long as it is a measure space). As an example, this is helpful for taking expectations rigorously because expectations are integral of random variables and the domain of a random variable is a probability space which may not be explicit.

The second big motivation for introducing a new theory of integration is the issue of convergence. It is important in many practical applications of integration theory to know when $\lim_n \int f_n(x)\mathrm{d}x = \int \lim_n f_n(x)\mathrm{d}x$ or when $\int_{E_x}\int_{E_y} f(x,y)\mathrm{d}x\mathrm{d}y = \int_{E_y}\int_{E_x} f(x,y)\mathrm{d}y\mathrm{d}x$. Lebesgue integration allows us to rigorously find conditions on f under which these statements will be true. This is often not possible in a satisfactory way with the Riemann theory of integration. We will see some of these convergence theorems next week and then switching the order of integration towards the end of the course (currently planned for week 9). The most important motivation for developing good convergence theorems was the development of Fourier series. We want to know when it is possible to integrate a Fourier series term by term.

The strategy for constructing the integral is to begin by definiting $\mu(f)$ when f belongs to a special class of measurable functions that we call *simple functions*. We then define the integral to progressively larger classes of functions.

Definition 5.1 (Simple functions)**.** Let $(E, \mathcal{E}, \mu)$ be a measure space. The set of simple functions on this space taking values in $\mathbb{R}$ are functions of the form

$$f(x) = \sum_{k=1}^{n} a_k 1_{A_k}(x).$$

Here, the A_k are disjoint sets in $\mathcal{E}$, 1_A represents the indicator function of the set, and the a_k are non-negative real numbers. We note that this representation of f is not unique.

Definition 5.2 (The integral of a simple function)**.** Still working in the setting above, let $f(x) = \sum_{k=1}^{n} a_k 1_{A_k}(x)$, then we can define

$$\mu(f) = \sum_{k=1}^{n} a_k \mu(A_k).$$

Remark 5.3 (A first comparison between Riemann and Lebesgue integration). A simple function on $\mathbb{R}$ will be Riemann integrable proveded that $\lambda(A_k) < \infty$ for each k. As we extend the Lebesgue integral beyond simple functions we are going to approximate other functions by these simple functions. This is similar to the lower sum approximation used in the construction of the Riemann integral.

Lemma 5.4. *The integral of a simple function is well defined (it doesn't depend on the choice of representation of the simple function) and satisfies the following properties.*

- *For $\alpha > 0$ we have $\mu(\alpha f) = \alpha\mu(f)$*
- *$\mu(f + g) = \mu(f) + \mu(g)$.*
- *If $f(x) \leq g(x)$ for every $x \in E$ then $\mu(f) \leq \mu(g)$.*
- *$f = 0$ μ-almost everywhere if and only if $\mu(f) = 0$*

Proof. Let us first look at the well definedness, without loss of generality lets assume the a_k, b_j are strictly positive. Let us suppose that $f = \sum_{k=1}^n a_k 1_{A_k} = \sum_{k=1}^m b_k 1_{B_k}$ which are both simple function representations. Then we can see that $\bigcup_{k=1}^n A_k = \bigcup_{k=1}^m B_k$ and that $a_i = b_j$ if $A_i \cap B_j \neq \emptyset$. Using this we can write

$$\begin{aligned}
\mu(f) &= \sum_{k=1}^n a_k \mu(A_k) \\
&= \sum_{k=1}^n \sum_{j=1}^m a_k \mu(A_k \cap B_j) \\
&= \sum_{k=1}^n \sum_{j=1}^m b_j \mu(A_k \cap B_j) = \sum_{j=1}^m b_j \sum_{k=1}^n \mu(A_k \cap B_j) \\
&= \sum_{j=1}^m b_j \mu(B_j).
\end{aligned}$$

Now we move on to the linearity properties. These come naturally from the defintion,

$$\mu(\alpha f) = \sum_{k=1}^n \alpha a_k \mu(A_k) = \alpha \sum_{k=1}^n a_k \mu(A_k) = \alpha\mu(A_k).$$

If we let $g = \sum_{j=1}^m c_j 1_{C_j}$ then let us write $A_0 = E \setminus \bigcup_{k=1}^n A_k, a_0 = 0$ and define C_0, c_0 similarly then we can write $f + g$ as a simple function via

$$f(x) + g(x) = \sum_{k=0}^n \sum_{j=0}^m (a_k + c_j) 1_{A_k \cap C_j}$$

and we have

$$\mu(f + g) = \sum_{k=0}^n \sum_{j=0}^m (a_k + c_j)\mu(A_k \cap C_j).$$

We note that $\bigcup_{k=0}^n A_k = \bigcup_{j=0}^m C_j = E$ and the A_k are mutually disjont, and the C_j are mutually disjoint. Therefore

$$
\begin{aligned}
\mu(f+g) &= \sum_{k=0}^{n}\sum_{j=0}^{m}(a_k+c_j)\mu(A_k\cap C_j) \\
&= \sum_{k=0}^{n} a_k \sum_{j=0}^{m}\mu(A_k\cap C_j) + \sum_{j=0}^{m} c_j \sum_{k=0}^{n}\mu(A_k\cap C_j) \\
\text{as the unions of the } A_k \text{ or } C_j \text{ fill the space} \quad &= \sum_{k=0}^{n} a_k\mu(A_k) + \sum_{j=0}^{m} c_j\mu(C_j) \\
\text{as } a_0 = c_0 = 0 \quad &= \sum_{k=1}^{n} a_k\mu(A_k) + \sum_{j=1}^{m} c_j\mu(C_j) \\
&= \mu(f)+\mu(g).
\end{aligned}
$$

Now we move onto the monotonicity of the integral. Let us express f and g as before. We can rewrite as

$$f = \sum_{k=1}^{n}\sum_{j=1}^{m} a_k 1_{A_k\cap C_j} = \sum_{k=1}^{n}\sum_{j=1}^{m} a_{k,j} 1_{A_k\cap C_j},$$

where $a_{k,j} = a_k 1_{A_k\cap C_j\neq\emptyset}$. In the same way we can write

$$g = \sum_{k=1}^{n}\sum_{j=1}^{m} c_{k,j} 1_{A_k\cap C_j},$$

where $c_{k,j} = c_j 1_{A_k\cap C_j\neq\emptyset}$. Then if $f(x)\leq g(x)$ for every x we know that this means $a_{k,j}\leq c_{k,j}$ for every k,j. Then by the well definedness of the integral we have

$$\mu(f) = \sum_{k=1}^{n}\sum_{j=1}^{m} a_{k,j}\mu(A_k\cap C_j) \leq \sum_{k=1}^{n}\sum_{j=1}^{m} c_{k,j}\mu(A_k\cap C_j) = \mu(g).$$

Lastly, we look at when $\mu(f)=0$. First if $f=0$ μ-almost everywhere then $a_k=0$ or $\mu(A_k)=0$ for each k, therefore $\mu(f)=0$. If $\mu(f)=0$ then since all the terms $a_k\mu(A_k)\geq 0$ we have that $a_k=0$ or $\mu(A_k)=0$ for each k. □

We are now going to extend the definition of the integral to a larger class of functions.

Definition 5.5 (Lebesgue integral for positive functions)**.** Let f be a positive, measurable function, we define

$$\mu(f) = \sup\{\mu(g) \,:\, g \,\text{is a simple function}, g\leq f\}$$

Lemma 5.6. *The above definition of Lebesgue integral for positive functions is consistent with the defintion for simple functions. That is to say that if $f=\sum_{k=1}^{n} a_k 1_{A_k}$ where $a_k\geq 0$ and the A_k are disjoint measurable sets then*

$$\sum_{k=1}^{n} a_k\mu(A_k) = \sup\{\mu(g) \,:\, g\,\textit{is a simple function}, g\leq f\}.$$

Proof. This is on the exercise sheet. □

Definition 5.7 (Final definition of Lebesgue integral)**.** Suppose that f is a measurable function which is not necessarily positive. Then we call f, *μ-integrable* or *Lebesgue integrable* if $\mu(|f|) < \infty$. In this case we can write $f = f_+ - f_-$ where f_+ and f_- are both positive and measureable ($f_+ = \max\{f, 0\}$). We then define the integral of f by

$$\mu(f) = \mu(f_+) - \mu(f_-).$$

Remark 5.8. Notice that we haven't yet proved that these definitions of the integral behave the way we hope (e.g. are linear, monotone, etc). In order to do this we need to prove some convergence results.

5.1 Convergence theorems for integrals of functions

This is one of the most useful parts of the course. In follow on courses in PDE and probability you will use these theorems again and again.

Theorem 5.9 (Monotone Convergence Theorem)**.** *Let f be a non-negative, measurable function and let f_n be a sequence of such functions. Then if $f_n \uparrow f$ we will have that $\mu(f_n) \uparrow \mu(f)$.*

Proof. We will break this proof down into progressively more complicated cases. First we note that by monotonicity $\lim_n \mu(f_n) \leq \mu(f)$ and therefore it is sufficient to prove $\mu(f) \leq \lim_n \mu(f_n)$.

First let us look at the case where $f_n = 1_{A_n}$ and $f = 1_A$, then the assumptions imply that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \dots$ and $A = \bigcup_n A_n$. Then this result is the same as the continuity of a measure as proved before.

Now let us keep $f = 1_A$ and let f_n be a sequence of simple functions. Pick $\epsilon \in (0, 1)$ arbitrary. Then let $A_n = \{x : f_n(x) > 1 - \epsilon\}$ then we have that $A_n \uparrow A$, and $(1 - \epsilon)1_{A_n} \leq f_n$. Therefore by the first case we have that $\lim_n \mu(f_n) \geq \lim_n \mu((1 - \epsilon)1_{A_n}) = \lim_n(1 - \epsilon)\mu(1_{A_n}) = (1 - \epsilon)\mu(f)$. Since ϵ is arbitrary this gives the result.

Now we look at the case where both f and f_n are simple functions. We can write $f = \sum_{k=1}^n a_k 1_{A_k}$, where wlog each a_k is strictly positive, then we have that $a_k^{-1} f_n 1_{A_k} \uparrow 1_{A_k}$ so we can apply the previous case to each part of f. Specifically

$$\mu(f_n) = \mu(\sum_{k=1}^n 1_{A_k} f_n) = \sum_{k=1}^n a_k \mu(a_k^{-1} 1_{A_k} f_n) \uparrow \sum_{k=1}^n a_k \mu(A_k) = \mu(f).$$

Here the first equality follows from the fact that the support of f_n must be included in the support of f since $0 \leq f_n \leq f$.

Our next case is when f is positive and measurable and f_n are all simple. Let us pick g a simple function with $g \leq f$ then $g_n = f_n \wedge g = \min\{f_n, g\}$ is a sequence of simple functions increasing to g. Therefore, by our previous case $\mu(g_n) \uparrow \mu(g)$. Furthermore $g_n \leq f_n$ so by monotonicity

$$\mu(g) = \lim_n \mu(g_n) \leq \lim_n \mu(f_n).$$

As g is an arbitrary this means that $\sup\{\mu(g) : g \, \text{is a simple function}, g \leq f\} \leq \lim_n \mu(f_n)$.

The last case is the most general where both f_n and f are positive and measureable. In this case we introduce our favorite kind of approximation (which is very similar to what we used in Lusin's theorem)

$$g_n = \left(2^{-n} \lfloor 2^n f \rfloor\right) \wedge n,$$

then g_n is a sequence of simple functions with $g_n \uparrow f$ and $g_n \leq f_n$. Therefore we have

$$\mu(f) = \lim_n \mu(g_n) \leq \lim_n \mu(f_n) \leq \mu(f).$$

Hence we have the required result. $\square$

We can now use this to prove that the integral of positive measurable functions has the desired properties.

Proposition 5.10. *Suppose f and g are non-negative, real valued, measurable function on a space $(E, \mathcal{E}, \mu)$ then*

- *For every $\alpha > 0$ $\mu(\alpha f) = \alpha\mu(f)$,*
- $\mu(f+g) = \mu(f) + \mu(g)$
- *If $f \leq g$ then $\mu(f) \leq \mu(g)$*
- *$\mu(f) = 0$ if and only if f is 0 almost everywhere.*

Proof. Suppose that f_n is a sequence of measurable functions with $f_n \uparrow f$ then αf_n is a sequence of measurable functions with $\alpha f_n \uparrow \alpha f$. Monotone convergence tells us that $\mu(\alpha f) = \lim_n \mu(\alpha f_n)$. We can use our previous results for simple functions to get that $\mu(\alpha f_n) = \alpha\mu(f_n)$. Therefore $\mu(\alpha f) = \lim_n \mu(\alpha f_n) = \lim_n \alpha\mu(f_n) = \alpha\mu(f)$.

For the sum let f_n and g_n be sequences of measurable functions with $f_n \uparrow f$ and $g_n \uparrow g$. Then we have $(f_n + g_n) \uparrow f + g$ and by monotone convergence and the results for simple functions we have $\mu(f+g) = \lim_n \mu(f_n + g_n) = \lim_n \mu(f_n) + \lim_n \mu(g_n) = \mu(f) + \mu(g)$.

If $f \leq g$ then we recall that by definition $\mu(f) = \sup\{\mu(\tilde{f}) : \tilde{f} \text{ is simple}, \tilde{f} \leq f\}$. So if h is a simple function with $h \leq f$ then we also have that $h \leq g$. Therefore, we can write

$$\begin{aligned}
\mu(g) &= \sup\{\mu(h) : h \text{ is simple}, h \leq g\} \\
&= \max\{\sup\{\mu(h) : h \text{ simple}, h \leq f\}, \sup\{\mu(h) : h \text{ simple}, h \leq f, h \text{ is not} \leq \text{g}\}\} \\
&= \max\{\mu(f), \sup\{\mu(h) : h \text{ simple}, h \leq f, h \text{ is not} \leq \text{g}\}\} \geq \mu(f).
\end{aligned}$$

Now $\mu(f) = 0$ if and only if $\sup\{\mu(h) : h \text{ simple}, h \leq f\} = 0$ which is if and only if $\mu(h) = 0$ for every $h \leq f$ and h simple. Now we know that if h is simple $\mu(h) = 0$ if and only if h is zero almost everywhere. This tells us that $\mu(f) = 0$ if and only if $h \leq f$ and h simple, implies $h = 0$ almost everywhere. If we look at the set on which f is positive we have $\{x : f(x) > 0\} = \bigcup_n \{x : f(x) > 1/n\}$ so by continuity of measure if $\mu(f > 0) > 0$ then there exists some n such that $\mu(f > 1/n) > 0$ therefore f is zero almost everywhere if and only if we can't fit any simple function underneath f which is not zero almost everywhere. □

Theorem 5.11 (Fatou's Lemma)**.** *Let f_n be a sequence of non-negative measurable function then we have the following result*

$$\mu\left(\liminf_n f_n\right) \leq \liminf_n \mu(f_n)$$

Remark 5.12. I always have trouble remembering which way around the inequality goes in this lemma. A helpful example is if $f_n = 1_{[n,n+1)}$ and μ is Lebesgue measure. Then $\lambda(f_n) = 1$ for every n and $\liminf_n f_n = 0$. This is also an instructive example for why the limits can fail to be the same. Essentially here the mass we are trying to integrate escapes to infinity.

Proof. This is essentially a consequence of monotone convergence. Let $g_n = \inf_{k \geq n} f_n$, then g_n is a non-decreasing sequence of measurable functions and $g_n \leq f_n$ for each n. By definition of the g_n we also know that $\liminf f_n = \liminf_n g_n = \lim_n g_n$. Using Monotone convergence we then have

$$\mu(\liminf_n f_n) = \mu(\lim_n g_n) = \liminf_n \mu(g_n).$$

Then using monotonicity we have

$$\mu(g_n) \leq \mu(f_n)$$

for each n, so consequently

$$\liminf_n \mu(g_n) \leq \liminf_n \mu(f_n).$$

Putting these all together gives the result. $\square$

Fatou's lemma is key to proving our next important convergence theorem.

Theorem 5.13 (Dominated convergence theorem)**.** *Let f_n be a sequence of functions and f another function such that $f_n \to f$ almost everywhere. Suppose further that there exists a positive function g such that $|f| \leq g, |f_n| \leq g$ for every n and $\mu(g) < \infty$, then $\lim_n \mu(f_n) = \mu(f)$. The function g is called the* dominating function*.*

Proof. Let us first suppose that $f_n \to f$ and the domination conditions hold everywhere. Then we have that $g + f_n$ is a sequence of non-negative measurable functions whose limit is $g + f$. Applying Fatou's lemma gives

$$\mu(g + f) \leq \liminf_n \mu(g + f_n) = \mu(g) + \liminf_n \mu(f_n),$$

subtracting $\mu(g)$ from each side (which we can do as it is finite) gives

$$\mu(f) \leq \liminf_n \mu(f_n).$$

Similarly $g - f_n$ is a sequence of non-negative measurable functions whose limit is $g - f$. Applying Fatou's lemma again gives

$$\mu(g) - \mu(f) \leq \mu(g) + \liminf_n(-\mu(f_n)) = \mu(g) - \limsup_n \mu(f_n).$$

Rearranging this since all the quantities are finite gives

$$\limsup_n \mu(f_n) \leq \mu(f).$$

Putting both parts together gives

$$\mu(f) \leq \liminf_n \mu(f_n) \leq \limsup_n \mu(f_n) \leq \mu(f).$$

Therefore the limit of the sequence $\mu(f_n)$ exists and is equal to $\mu(f)$.

The extension of this result to when the conditions only hold almost everywhere is due to the fact that the integrals of any function is unchanged by modifying that function on a measure zero set. This type of result will be discussed in more detail when we introduce *Lebesgue spaces*. It isn't really the point of this particular theorem, we just give the full version here so we are able to apply it. $\square$

The following is a useful criteria for when we can differentiate under the integral sign which also serves as a good example of how to use the dominated convergence theorem.

Theorem 5.14 (Differentiation under the integral sign)**.** *Let $(E, \mathcal{E}, \mu)$ be a measure space and $f : U \times E \to \mathbb{R}$ be a function such that $x \mapsto f(t, x)$ is integrable for every t, and $t \mapsto f(t, x)$ is differentiable for every x, and suppose further that there exists an integrable function $g(x)$ such that*

$$\left|\frac{\partial f(t,x)}{\partial t}\right| \leq g(x), \quad \forall t \in U$$

then the function $x \mapsto \partial f(t, x)/\partial t$ is integrable and the function $F(t) = \int_E f(t, x)\mu(\mathrm{d}x)$ is differentiable with

$$\frac{\mathrm{d}F}{\mathrm{d}t} = \int_E \frac{\partial f}{\partial t}(t, x)\mu(\mathrm{d}x).$$

Notice here we are using a different notation for the integral with respect to μ. We do this because it is helpful to be able to emphasise that we integrate in x but not t.

Proof. Let ϵ_n be an arbitrary sequence which tends towards 0. Let

$$g_n(t,x) = \frac{f(t+\epsilon_n, x) - f(t,x)}{\epsilon_n} - \frac{\partial f}{\partial t}(t,x).$$

First we notice that $g_n \to 0$ and $g_n + \partial f/\partial t$ is measurable so $\partial f/\partial t$ is the limit of measurable functions so measurable. By the mean value theorem we have $|g_n| \leq 2g$ for each n. Therefore, by dominated convergence we have

$$\int g_n(t,x)\mu(\mathrm{d}x) \to 0.$$

This gives the required result. □

We have a couple of useful facts about integration which don't fit into a big section.

Definition 5.15 (Restriction of a Measure)**.** Suppose that $(E, \mathcal{E}, \mu)$ is a measure space and $A \in \mathcal{E}$ then the set of measurable subsets of A is a σ algebra we call $\mathcal{A}$ and the restriction of μ to $\mathcal{E}_A$ is a measure we call μ_A. Furthermore we have that if f is a measurable function on E then $\mu(f1_A) = \mu_A(f)$.

Remark 5.16. The last part is actually a lemma whose proof is an exercise.

We can use this defintion to make sense of Lebesgue integrals on intervals (for example). If $I = [a, b]$ then $\int_a^b f(x)\mathrm{d}x = \lambda(f1_I)$.

We also notice that we can define a measure using a positive function f.

Proposition 5.17. *Let $(E, \mathcal{E}, \mu)$ be a measure space and let f be a non-negative measurable function. Define $\nu(A) = \mu(f1_A)$ for each $A \in \mathcal{E}$. Then ν is a measure on E and for all non-negative g we have*

$$\nu(g) = \mu(fg)$$

Proof. First we need to show that ν is indeed a measure. $f1_\emptyset = 0$ so we have $\nu(\emptyset) = 0$ as required. We will also have that $\nu(A) \geq 0$ since f is non-negative. We show countable additivity, we note that if A an B are disjoin then $1_{A\cup B} = 1_A + 1_B$ and furthermore if $A_1, A_2, \dots$ is a sequence of disjoint sets then $1_{\bigcup_n A_n} = \sum_n 1_{A_n}$. With this reformulation $\nu(\bigcup_n A_n) = \mu(f1_{\bigcup_n A_n}) = \mu(\sum_n f1_{A_n})$ using the Beppo-Levi reformulation of monotone convergence we have

$$\mu(\sum_n f1_{A_n}) = \sum_n \mu(f1_{A_n}) = \sum_n \nu(A_n),$$

which is our desired result.

Now we want to show that if $g \geq 0$ then $\nu(g) = \mu(fg)$. Let us begin with the case where $g = 1_A$ for some measurable A, then $\nu(g) = \nu(A) = \mu(f1_A) = \mu(fg)$ so the result follows by definition. Then using the linearity of μ we can see that if g is a simple function then $\nu(g) = \mu(fg)$. Now suppose that g is not necessarily simple, we can constuct (in our standard way) a sequence of simple functions, g_n, which increase to g then by monotone convergence we have that $\nu(g) = \lim_n \nu(g_n) = \lim_n \mu(fg_n)$. Now fg_n is a sequence of function which increases to fg so using monotone convergence we have that $\lim_n \mu(fg_n) = \mu(fg)$ so we have that $\nu(g) = \mu(fg)$. □

There are also a few facts about Riemann integration which work in pretty much exactly the same way for Lebesgue integration. For example the fundamental theorem of calculus holds equally well in this case. We will see in general that when something is Riemann integrable it is also Lebesgue integrable which will prove all these in general.

Theorem 5.18 (Fundamental theorem of calculus)**.** *Suppose that $f : [a, b] \to \mathbb{R}$ is a continuous function and set $F(t) = \int_a^t f(x)\mathrm{d}x = \lambda(1_{[a,t]}f)$, then F is differentiable with $F'(t) = f(t)$. Furthermore, let $F : [a, b] \to \mathbb{R}$ have continuous derivative f, then $F(b) - F(a) = \int_a^b f(x)\mathrm{d}x$.*

Proof. Given $\epsilon > 0$ there exists $\delta > 0$ such that $|x-t| \leq \delta$ implies that $|f(x) - f(t)| \leq \epsilon$ therefore if we take $|h| \leq \delta$ then

$$\left|\frac{1}{h}(F(t+h) - F(t)) - f(t)\right| = \frac{1}{|h|}\left|\int_t^{t+h}(f(x) - f(t))\mathrm{d}x\right| \leq \frac{1}{|h|}\int_{t\wedge(t+h)}^{t\vee(t+h)}|f(x) - f(t)|\mathrm{d}x.$$

Now we can use the fact that inside the integral $|x-t| \leq \delta$ so we have

$$[\left|\frac{1}{h}(F(t+h) - F(t)) - f(t)\right| \leq \frac{1}{|h|}\int_{t\wedge(t+h)}^{t\vee(t+h)}\epsilon\mathrm{d}x = \epsilon.$$

Therefore $\lim_{h\to 0}(F(t+h) - F(t))/h = f(t)$.

In the other direction $\mathrm{d}/\mathrm{d}t(F(t) - \int_a^t f(x)\mathrm{d}x) = 0$ so $F(t) - \int_a^t f(x)\mathrm{d}x$ is constant in t (by the mean value theorem), so $F(t) = F(a) + \int_a^t f(x)\mathrm{d}x$. This gives us the result. □

We can use the fundamental theorem of calculus to prove the standard result about change of variables. This time we can exploit our new machinary.

Proposition 5.19. *Let $\phi : [a,b] \to [\phi(a), \phi(b)]$ be continuously differentiable and stricly increasing then for all non-negative g on $[\phi(a), \phi(b)]$ we have*

$$\int_{\phi(a)}^{\phi(b)} g(y)\mathrm{d}y = \int_a^b g(\phi(x))\phi'(x)\mathrm{d}x.$$

Proof. First suppose that g is the indicator function of an interval $(c,d]$ then we want to prove that

$$\int_{\phi(a)}^{\phi(b)} 1_{(c,d]}(x)\mathrm{d}y = \int_a^b 1_{(c,d]}(\phi(x))\phi'(x)\mathrm{d}x.$$

Here the left hand side is equal to $[\phi(a), \phi(b)] \cap (c,d]$ and the right hand side is

$$\int_{a\vee c}^{b\wedge d}\phi'(x)\mathrm{d}x,$$

using the fundamental theorem of calculus this is

$$\phi(b\wedge d) - \phi(a\vee c) = \phi(b)\wedge\phi(d) - \phi(a)\vee\phi(c) = [\phi(a), \phi(b)] \cap (c,d]$$

where here we used the fact that ϕ was increasing to commute it with min and max.

Now we have shown our proposition holds when g is the indicator function of a half open interval. By linearity of the integral it will hold when g is the indicator function of a finite disjoint union of half open intervals. Now let $\mathcal{D}$ be the set of all Borel sets A such that 1_A satisfies our proposition. As the name suggests we want to show that $\mathcal{D}$ is a d-system. If $A \subseteq B$ and $A, B \in \mathcal{D}$ then $1_{B\setminus A} = 1_B - 1_A$ so the proposition will hold for $B \setminus A$ by linearity of the integral. Suppose that $A_1 \subseteq A_2 \subseteq A_3 \ldots$ then let $g_n = 1_{A_n}$ then $g_n \uparrow 1_A =$ g and $g_n \circ \phi \uparrow g \circ \phi$, as ϕ is increasing so if

$$\int_{\phi(a)}^{\phi(b)} g_n(y)\mathrm{d}y = \int_a^b g_n(\phi(x))\phi'(x)\mathrm{d}x$$

for each n applying monotone convergence to each side gives the result for $g = 1_A$. This shows that $\mathcal{D}$ is a d-system. Applying Dynkin's lemma then shows that for every $A \in \mathcal{B}(R)$ we have that the proposition holds with $g = 1_A$.

Linearity of the integral allows us to extend this result to any simple function g. We can then use monotone convergence in exactly the same way as for the last part to extend it to any non-negative measurable g. □

5.2 Agreement with Riemann Integral

We now turn our attention to the case wher μ is Lebesgue measure, λ. We want to show that our new defintion of the integral will agree with the Riemann integral when they are both defined. Let us first recall the definitions associated with the Riemann integral.

Definition 5.20. Let $[a, b]$ be an interval in $\mathbb{R}$ then a finite sequence of real numbers $\{a_k\}_{k=0}^n$ is called a *partition* of the interval if $a = a_0 < a_1 < \cdots < a_n = b$. I usually denote a partition with a lower case p or q. You also often see the notation $\mathscr{P}$.

Definition 5.21. If we have two partitions $p = \{a_k\}_{k=0}^n$ and $q = \{b_j\}_{j=0}^m$ then we say q is a *refinement* of p if every element of p appears in q.

Definition 5.22. We call a sequence of partitions $(p_n)_{n\geq 1}$ *nested* if for every n we have that p_{n+1} is a refinement of p_n.

Definition 5.23. If we have a partition $p = \{a_k\}_{k=0}^n$ and a function f then we can define

$$m_k = \inf\{f(x) : x \in [a_{k-1}, a_k]\} \quad \text{and} \quad M_k = \sup\{f(x) : x \in [a_{k-1}, a_k]\}.$$

Then we have the upper sum and lower sum associated to the partition which are defined as

$$l(f, p) = \sum_{k=1}^n m_k(a_k - a_{k-1}), \quad u(f, p) = \sum_{k=1}^n M_k(a_k - a_{k-1}).$$

Remark 5.24. We can check that if q is a refinement of p then

$$l(f, p) \leq l(f, q) \leq u(f, q) \leq u(f, p).$$

Definition 5.25. We call a function f, *Riemann integrable* on $[a, b]$ if for every sequence of nested partitions $(p_n)_{n\geq 1}$ we have that $\lim_n l(f, p_n) = \lim_n u(f, p_n)$ and we call this limit the Riemann integral of f or $\int_a^b f(x)\mathrm{d}x$.

Lemma 5.26. *A function, f, is Riemann integrable if and only if there exists a partition p such that*

$$u(f, p) - l(f, p) < \epsilon.$$

Proof. First suppose that f satisfies that for every ϵ there exists a p such that $u(f, p) - l(f, p) < \epsilon$ then

$$\inf_q u(f, q) - \sup_q l(f, q) \leq u(f, p) - l(f, p) < \epsilon.$$

Since, ϵ is arbitrary this implies

$$\inf_q u(f, q) = \sup_q l(f, q)$$

and therefore f is Riemann integrable.

Suppose that f is Riemann integrable then we know that $\inf_q u(f, q) = \sup_q l(f, q) = \int f\mathrm{d}x$. Therefore, given ϵ there exists p_1 and p_2 so that $u(f, p_1) \leq \int f\mathrm{d}x + \epsilon/2$ and $l(f, p_2) \geq \int f\mathrm{d}x - \epsilon/2$. Now let $p = p_1 \cup p_2$, that is to say the partition made up of all the point in both p_1 and p_2. In this case p_1 and p_2 are both refinements of p, so we have

$$\int f\mathrm{d}x - \epsilon/2 \leq l(f, p_2) \leq l(f, p) \leq u(f, p) \leq u(f, p_1) \leq \int f\mathrm{d}x + \epsilon/2,$$

so

$$u(f, p) - l(f, p) < \epsilon.$$

□

Theorem 5.27. *Let $[a, b]$ be an interval. Suppose that f is a bounded, function which is Riemann integrable on $[a, b]$, then it is Lebesgue integrable and the Riemann integral agrees witht the Lebesgue integral.*

Proof. As f is bounded we only need to show that it is Lebesgue measurable in order for it to be integrable. Using the lemma above there exists a nested sequence of partitions p_n such that $u(f, p_n) - l(f, p_n) < 1/n$ for each n. Let us define two sequences of functions g_n and h_n. We write $p_n = \{a_k^n\}_{k=0}^{N_n}$, and recall the definition of m_k^n and M_k^n associated to the partition. Then we define

$$g_n := \sum_{k=1}^{N_n} m_k^n 1_{[a_{k-1}^n, a_k^n)}, \quad h_n := \sum_{k=1}^{N_n} M_k^n 1_{[a_{k-1}^n, a_k^n)}.$$

Here we can see that g_n and h_n are both sequences of simple functions. We also have that g_n is a monotonically increasing sequence and h_n is a monotonically decreasing sequence. As f is bounded so are the sequences $g_n(x)$ and $h_n(x)$ for each x so we define $g(x) = \lim_n g_n(x)$ and $h(x) = \lim_n h_n(x)$, these are both bounded, Borel measurable functions. We also have that $g_n(x) \leq f(x) \leq h_n(x) \leq \sup_{[a,b]} f$ so consequently $g(x) \leq f(x) \leq h(x)$. We can see that $\lambda(g_n) = l(f, p_n)$ and $\lambda(h_n) = u(f, p_n)$. We can use $\sup_{[a,b]} f$ as a dominating function, so we have by dominated convergence that

$$\lambda(g) = \lim_n \lambda(g_n) = \lim_n l(f, p_n) = \lim_n u(f, p_n) = \lim_n \lambda(h_n) = \lambda(h).$$

We also have that $h - g \geq 0$ and $\lambda(h - g) = 0$ so we know that $h = g$ Lebesgue almost everywhere and as $h - f \leq h - g$ we know that $f = h = g$ almost everywhere. Therefore, f is almost everywhere equal to a measurable function and it is bounded so is Lebesgue integrable □

We finish this section with some examples of functions which are Lebesgue integrable but are not Riemann integrable. The most classic example of this is

Example 5.28. Let $f(x) = 1_{\mathbb{Q}}$ then f is Lebesgue integrable but not Riemann integrable on $[0, 1]$ (or any other interval). In order to see that this function is not Riemann integrable we can see that for any partition p as the rationals an the irrationals are dense in $[0, 1]$ then $l(f, p) = 0$ and $u(f, p) = 1$ therefore if we take a seqence of nested partitions p_n then we wont have the limits of $l(f, p_n)$ and $u(f, p_n)$ meeting.

6 Norms and inequalities

Let us remember what a normed space is with an emphasis on how this fits in with functions spaces.

Definition 6.1 (Normed space)**.** A normed space is a vector space $\mathcal{V}$ equipped with a norm, $\|\cdot\|$, which should satisfy

- $\|v\| \in \mathbb{R}_+$
- $\|\lambda v\| = |\lambda| \|v\|$
- $\|v + u\| \leq \|v\| + \|u\|$
- $\|v\| = 0$ if and only if $v = 0$.

We are interested in normed spaces of functions, where the norms come from integrating quantites.

Definition 6.2. $L^p(E)$ Suppose that $(E,\mathcal{E},\mu)$ is a measure space, and $p \geq 1$, then we have the associated L^p space, which is the space of bounded measureable functions equipped with the norm

$$\|f\|_p = (\mu(|f|^p))^{1/p}.$$

When we are working on $\Omega \subseteq \mathbb{R}^d$ with Lebesgue measure we often write the space $L^p(\Omega)$ to be the set of functions with $\int_\Omega |f|^p \mathrm{d}x < \infty$. We then often write the norm $\|\cdot\|_{L^p(\Omega)}$ if we have not previously specified which space we are working in. If we are working with the measure $\gamma(x)\mathrm{d}x$ for some positive function γ on $\mathbb{R}^d$ then we write $L^p(\gamma), \|\cdot\|_{L^p(\gamma)}$. We also calle $L^p(E)$ or $L^p(\Omega)$ or $L^p(\gamma)$ the space of measurable functions where the associated norm is finite.

We also have the supremum norm.

Definition 6.3. Suppose $(E,\mathcal{E},\mu)$ is a measure space. We have the following norm on measurable functions

$$\|f\|_\infty = \inf\{\lambda : |f| \leq \lambda \,\text{almost everywhere}\}.$$

We also call the space $L^\infty(E)$ the space of all measurable functions on E with $\|f\|_\infty < \infty$.

Remark 6.4. Strictly speaking the norms defined above are seminorms. This is because all these norms will vanish for a function f, where f is non-zero but is equal to zero almost everywhere. When working in L^p spaces we consider two functions the same if they are equal almost everywhere.

Strictly speaking we no longer consider functions f when we work in L^p spaces we instead consider equivalence classes of functions with the equivalence relation $f \sim g$ if $f = g$ almost everywhere. When working in this setting we write $\mathcal{L}^p$ for the space of measurable function equipped with the p-seminorm and L^p for the space of equivalence classes of functions equipped with the p-norm. Most of the time we wont really think about an element of L^p as an equivalence class and *hopefully* it quickly becomes natural to think about functions as defined up to alteration on a null set.

Theorem 6.5. *For $p \in [1,\infty]$ the space $L^p(E)$ is indeed a vector space.*

Proof. We need to show $f \in L^p(E)$ implies that $\alpha f \in L^p(E)$ for $\alpha \in \mathbb{R}$ (or $\mathbb{C}$) and $f, g \in L^p(E)$ implies that $f + g \in L^p(E)$, specifically we need to show that $\|\alpha f\|_p < \infty$ and $\|f+g\|_p < \infty$.

For $p < \infty$, the first point we can use the linearity of the integral to get

$$\mu\left(|\alpha|^p |f|^p\right) = |\alpha|^p \mu(|f|^p) < \infty.$$

Now, in a slightly more complex way we have

$$\int |f(x)+g(x)|^p \mu(\mathrm{d}x) \leq \int \left(2\max\{|f(x)|,|g(x)|\}\right)^p \mu(\mathrm{d}x) \leq \int 2^p \left(|f(x)|^p + |g(x)|^p\right)\mu(\mathrm{d}x) \leq 2^p\left(\|f\|_p^p + \|g\|_p^p\right) < \infty.$$

For $p = \infty$ for the first point it follows immediately from the definition that $\|\alpha f\|_\infty = |\alpha|\|f\|_\infty < \infty$. For the second point since the union of two null sets is null we have that $f+g$ is equivalent to a function which is uniformly bounded. Therefore it is clear that $\|f+g\|_\infty \leq \|f\|_\infty + \|g\|_\infty$. □

Definition 6.6 (Convergence in L^p)**.** We say that a sequence of measurable functions of function f_n converges to f in L^p if $\|f_n - f\|_p \to 0$.

In order to progress further with normed spaces of functions we need to be able to prove the triangle inequality for the p-norms. This inequality is called *Minkowski's inequality*. In the next section we prove it as well as several other inequalities which are very useful when working with function spaces.

6.1 Inequalities

For our first couple of inequalities let us just look at some useful inequalities between real numbers.

Lemma 6.7 (Young's inequality (watch out there are at least two things with this name))**.** *Let x and y be two positive real numbers and $p \in [1, \infty)$ with $1/p + 1/q = 1$ then we have*

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Proof. We can see the inequality holds when either x or y are zero so we neglect this case and define $u = x^p$ and $v = y^q$. Therefore we want to show that

$$u^{1/p}v^{1/q} \leq \frac{u}{p} + \frac{v}{q}.$$

As everything is strictly positive we can divide both sides by v, and use the relationship between p and q, to get

$$u^{1/p}v^{-1/p} \leq \frac{u/v}{p} + \frac{1}{q}.$$

Now let us define $t = u/v$, so our orriginal inequality will be true if we can show

$$t^{1/p} \leq \frac{t}{p} + \frac{1}{q},$$

or equivalently

$$\frac{t}{p} + \frac{1}{q} - t^{1/p} \geq 0.$$

We can differentiate this function in t and get

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{t}{p} + \frac{1}{q} - t^{1/p}\right) = \frac{1}{p}(1 - t^{-1/q})$$

and differentiate a second time to get

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\left(\frac{t}{p} + \frac{1}{q} - t^{1/p}\right) = \frac{1}{pq}t^{-1-1/q} > 0.$$

So this function achieves a minimum when $t^{-1/q} = 1$, that is when $t = 1$, and it achieves the minimum value 0. Therefore it is always positive and the inequality holds. □

We also have the very simple corollary which is often useful (especially in Analysis of PDE)

Corollary 6.8. *Suppose that x, y are positive then for every $\eta > 0$ we have*

$$xy \leq \frac{x^p\eta^p}{p} + \frac{y^q}{\eta^q q}$$

Proof. Just write $xy = (\eta x)(y/\eta)$. □

Using Young's inequality we can prove an inequality about functions.

Proposition 6.9 (Hölder's Inequality)**.** *Suppose that $(E, \mathcal{E}, \mu)$ is a measure space and $f \in L^p(E), g \in L^q(E)$ with $1/p + 1/q = 1$ then $fg \in L^1(E)$ and we have the following inequality*

$$\|fg\|_1 \leq \|f\|_p\|g\|_q$$

Proof. First let us look at the case where $f \in L^1(E)$ and $g \in L^\infty(E)$ without loss of generality let g be bounded everywhere by $\|g\|_\infty$ then we have

$$|f(x)g(x)| \leq |f(x)| \|g\|_\infty$$

and integrating this inequality (using monotonicity) gives the result.

The more complicated case is where $p \in (1, \infty)$ then we have for each $\eta > 0$ that

$$|f(x)g(x)| \leq \frac{\eta^p |f(x)|^p}{p} + \frac{|g(x)|^q}{\eta^q q}.$$

Integrating this gives

$$\|fg\|_1 \leq \frac{\eta^p}{p} \|f\|_p^p + \frac{1}{\eta^q q} \|g\|_q^q.$$

We can then choose η however we want so we choose it to make the right hand side as small as possible. We can find out how best to choose η by differentiating in η.

$$\frac{\mathrm{d}}{\mathrm{d}\eta} \left(\frac{\eta^p}{p} \|f\|_p^p + \frac{1}{\eta^q q} \|g\|_q^q \right) = \eta^{p-1} \|f\|_p^p - \eta^{-q-1} \|g\|_q^q,$$

and

$$\frac{\mathrm{d}^2}{\mathrm{d}\eta^2} \left(\frac{\eta^p}{p} \|f\|_p^p + \frac{1}{\eta^q q} \|g\|_q^q \right) = (p-1)\eta^{p-2} \|f\|_p^p + (q-1)\eta^{-q-2} \|g\|_q^q > 0.$$

So the right hand side of the inequality is smalles when

$$\eta^{p-1} \|f\|_p^p = \eta^{-q-1} \|g\|_q^q.$$

Which is when

$$\eta = \|g\|_q^{q/(p+q)} \|f\|_p^{-p/(p+q)}.$$

Substituting this value of η in gives

$$\|fg\|_1 \leq \frac{1}{p} \|f\|_p^{p-p^2/(p+q)} \|g\|^{pq/(p+q)} + \frac{1}{q} \|f\|_p^{pq/(p+q)} \|g\|_q^{q-q^2/(p+q)} = (\|f\|_p \|g\|_q)^{pq/(p+q)},$$

and

$$pq/(p+q) = ((p+q)/pq)^{-1} = (1/q + 1/p)^{-1} = 1.$$

□

Second proof of Holder's inequality. This is the more standard proof suppose first that $\|f\|_p = 1, \|g\|_q = 1$ then using Young's inequality $|f(x)g(x)| \leq |f(x)|^p/p + |g(x)|^q/q$. So integrating this gives $\|fg\|_1 = \|f\|_p^p/p + \|g\|_q^p/q = 1/p + 1/q = 1$. Then we have for general f, g that $\|f/\|f\|_p\|_p = 1$ and $\|g/\|g\|_q\|_q = 1$ so

$$\|fg/\|f\|_p \|g\|_q\|_1 \leq 1,$$

and multiplying through gives

$$\|fg\|_1 = \|f\|_p \|g\|_q.$$

□

Remark 6.10 (Cauchy-Schwartz Inequality). The important case of this inequality when $p = q = 2$ is generally known as the Cauchy-Schwartz inequality!!!

We also have Minkowski's inequality which as we discussed is necessary to make sure L^p is a normed space.

Proposition 6.11 (Minkowski's Inequality)**.** *Let $(E, \mathcal{E}, \mu)$ be a measure space and suppose that f, g are in L^p then*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof. We have already shown this when $p = \infty$, the case where $p = 1$ is also straightforward. We have

$$|f(x)+g(x)| \leq |f(x)| + |g(x)|$$

and integrating this gives the required inequality.

We now move on to $p \in (1, \infty)$. We choose q so that $1/p + 1/q = 1$ and observe that $|f+g|^{p-1} \in L^q(E)$ as

$$|f+g|^{q(p-1)} = |f+g|^p.$$

We also have that $|f|, |g| \in L^p(E)$. Therefore we have

$$\begin{aligned} \|f+g\|_p^p = \mu(|f+g|^p) &= \mu(|f+g||f+g|^{p-1}) \\ &\leq \mu(|f||f+g|^{p-1}) + \mu(|g||f+g|^{p-1} \\ \text{using Hölder's ineq} \quad &\leq (\|f\|_p + \|g\|_p) \, \||f+g|^{p-1}\|_q \\ &= (\|f\|_p + \|g\|_p) \, \|f+g\|_p^{p/q}. \end{aligned}$$

Rearranging this gives

$$\|f+g\|^{p(1-1/q)} \leq \|f\|_p + \|g\|_p,$$

and we recall that $p(1-1/q) = 1$. □

Now we move onto some more probabilistically focussed inequalities which do not directly relate to L^p spaces

Proposition 6.12 (Markov's Inequality/ Chebychev's inequality)**.** *Let $(E, \mathcal{E}, \mu)$ be a measure space and f a non-negative measurable function and $\lambda > 0$. Then we have*

$$\mu(\{x : f(x) > \lambda\}) \leq \frac{1}{\lambda}\mu(f).$$

Proof. We have the following inequality

$$\lambda 1_{\{f(x)>\lambda\}} \leq f.$$

We then integrate this and use the monotonicity of the integral to get

$$\lambda\mu(\{f(x) > \lambda\}) \leq \mu(f).$$

□

Remark 6.13 (Tail estimates). On of the powerful consequences of Markov's inequality is that is allows us to estimate how the function will behave at large values. For example suppose that $f \in L^p(\mathbb{R})$ then we know that

$$\lambda(\{x : |f(x)| > t\} = \lambda(\{x : |f(x)|^p > t^p\}) \leq t^{-p}\|f\|_p^p.$$

This is particularly relevant in probability where we are interested in estimating how often extreme events happen and we get inequalities of the form

$$\mathbb{P}(X > x) \leq x^{-p}\mathbb{E}(X^p).$$

Remark 6.14 (Tchernoff bounds). Another common use of Markov's inequality is when we know how $\mu(\exp(\alpha f(x)))$ behaves as we vary α. For example, in a probabilistic setting $\mathbb{E}(e^{\alpha X})$ is the moment generating function which is often known for distributions. We can then use Markov's inequality via

$$\mu(\{f(x) > t\}) = \mu(\{\exp(\alpha f(x)) > e^{\alpha t}\}) \leq \mu(\exp(\alpha f))e^{-\alpha t}.$$

Since the left hand side does not depend on α one can then optimise over α which will often give a superior bound. An example of this is in the probabilistic setting if X is a normal random variable on $\mathbb{R}$ with mean 0 and varaince σ^2 then we have

$$\mathbb{E}\left(e^{\alpha X}\right) = e^{\alpha^2\sigma^2/2}.$$

This leads to

$$\mathbb{P}(X > t) \leq e^{\alpha^2\sigma^2/2-\alpha t}.$$

We can then see that

$$\alpha^2\sigma^2/2 - \alpha t = \frac{1}{2}\left(\alpha\sigma - \frac{t}{\sigma}\right)^2 - \frac{t^2}{2\sigma^2},$$

so we can choose $t = \sigma^2\alpha$ to get

$$\mathbb{P}(X > t) \leq e^{-t^2/2\sigma^2}.$$

Our last big inequality is Jensen's inequality which involves convexity. We briefly recall the definition of convexity and prove a useful lemma before moving onto the inequality.

Definition 6.15 (Convexity)**.** Let I be an interval and let $\phi : I \to \mathbb{R}$ then we call ϕ *convex* if for every $t \in [0, 1]$, and $x, y \in I$, we have

$$\phi(tx + (1-t)y) \leq t\phi(x) + (1-t)\phi(y).$$

Lemma 6.16. *Let $\phi : I \to \mathbb{R}$ be convex and let m be a point in the interior of I then there exists a, b such that $ax + b \leq \phi(x)$ for every $x \in I$ and $am + b = \phi(m)$.*

Proof. Take $x < m < y$ then by convexity

$$\phi(m) \leq \frac{y-m}{y-x}\phi(x) + \frac{m-x}{y-x}\phi(y).$$

We can rearrange this to

$$(y - m + m - x)\phi(m) \leq (y-m)\phi(x) + (m-x)\phi(y),$$

then to

$$(y-m)(\phi(m) - \phi(x)) \leq (m-x)(\phi(y) - \phi(m)),$$

then to

$$\frac{\phi(m) - \phi(x)}{m-x} \leq \frac{\phi(y) - \phi(m)}{y-m}.$$

This is true for any x, y surrounding m so there exists a such that

$$\frac{\phi(m) - \phi(x)}{m-x} \leq a \leq \frac{\phi(y) - \phi(m)}{y-m},$$

for every such x, y. From this we get that $\phi(x) \geq a(x-m) + \phi(m)$. □

Proposition 6.17 (Jensen's inequality). *Suppose that $(E, \mathcal{E}, \mu)$ is a measure space with $\mu(E) = 1$ and let ϕ be a convex function from $\mathbb{R}$ to $\mathbb{R}$ and f is an integrable function then $\phi(f)$ is also integrable and*

$$\mu(\phi(f)) \geq \phi(\mu(f)).$$

Remark 6.18. This is another inequality where I sometimes have trouble remembering which way the inequality sign goes. My key example to check on is

$$\frac{1}{4} = \left(\int_0^1 x\mathrm{d}x\right)^2 \leq \int_0^1 x^2 \mathrm{d}x = \frac{1}{3}.$$

Proof. As $\mu(E) = 1$ we can consider $\mu(f)$ as the average value that f takes over E. Using our lemma we have that there exists a, b such that

$$ax + b \leq \phi(x),$$

and

$$a\mu(f) + b = \phi(\mu(f)).$$

By the monotonicity of the integral

$$\mu(af + b) \leq \mu(\phi(f))$$

and the left hand side is $a\mu(f) + b\mu(E) = a\mu(f) + b$ by linearity which by construction is equal to $\phi(\mu(f))$. $\square$

,

6.2 Back to L^p spaces

Now we are armed with our inequalities, we want to discuss some properties of L^p spaces.

Theorem 6.19 ($L^p(E)$ is complete). *This is for the case $p < \infty$. Suppose that f_n is a sequence in L^p with $\|f_n - f_m\|_p \to 0$ as $n, m \to 0$ then there exists an f in L^p such that $\|f_n - f\|_p \to 0$ as $n \to \infty$*

Proof. Let $n_1 = 1$ and then we can find n_k recursively such that $\|f_{n_k} - f_{n_{k-1}}\|_p \leq 2^{-k}$. Then we have that

$$\sum_k \|f_{n_k} - f_{n_{k-1}}\|_p \leq 1.$$

Choose K arbirtrary, then by Minkowski's inequality we have

$$\|\sum_{k=1}^K |f_{n_k} - f_{n_{k-1}}|\|_p \leq \sum_{k=1}^K \|f_{n_k} - f_{n_{k-1}}\|_p \leq 1.$$

By Monotone convergence we can let $K \to \infty$ to get

$$\|\sum_{k=1}^\infty |f_{n_k} - f_{n_{k-1}}|\|_p \leq 1.$$

Therefore,

$$\sum_{k=1}^\infty |f_{n_k}(x) - f_{n_{k-1}}(x)| < \infty$$

almost everywhere. This implies that $f_{n_k}(x)$ is a Cauchy sequence for almost every x. Since we know that $\mathbb{R}$ is complete, there exists a set E' with $\mu(E \setminus E') = 0$ such that for every $x \in E'$ the sequence $f_{n_k}(x)$. Define

$$f(x) = \begin{cases} \lim_k f_{n_k}(x) & x \in E' \\ 0 & x \notin E' \end{cases}$$

Now we have a candidate for our limit, we want to show $f_n \to f$ in $L^p(E)$. Given $\epsilon > 0$ there exists N such that if $n, m \geq N$ then $\|f_n - f_m\|_p \leq \epsilon$. Therefore, for k sufficiently large and $n \geq N$ we have $\|f_n - f_{n_k}\|_p \leq \epsilon$. Now, using Fatou's lemma

$$\|f_n - f\|_p = \|f_n - \lim_k f_{n_k}\|_p \leq \liminf_k \|f_n - f_{n_k}\|_p \leq \epsilon.$$

Therefore $\|f_n - f\|_p \to 0$. □

Proposition 6.20. *Simple functions, step functions (functions of the form $\phi(x) = \sum_{k=1}^n a_k 1_{(c_k, d_k]}$, and continuous functions are all dense in the space $L^p(\mathbb{R}), p \in [1, \infty)$ that is to say for any $\epsilon > 0$ and any f in $L^p(\mathbb{R})$ there is a function g which is a simple function/step function/ continuous function such that $\|f - g\|_p \leq \epsilon$.*

Proof. The proof for simple functions and step functions is in the fourth assignment. In order to show that it works for continuous functions we notice that the result is true for step functions so for any $f \in L^p(\mathbb{R})$ and any $\epsilon > 0$ there exists a step function ϕ such that $\|f - \phi\|_p \leq \epsilon/2$, if we can find a continuous function g such that $\|\phi - g\|_p \leq \epsilon/2$ then by Minkowski's inquality $\|f - g\|_p \leq \|f - \phi\|_p + \|\phi - g\|_p \leq \epsilon/2 + \epsilon/2$.

Now if we look at the indicator function $1_{(c,d]}(x)$ then let us take

$$g_{\epsilon,c,d}(x) = \begin{cases} 0 & x \notin (c - \epsilon, d + \epsilon) \\ (x - c + \epsilon)/\epsilon & x \in [c - \epsilon, c) \\ 1 & x \in [c, d) \\ -(x - d - \epsilon)/\epsilon & x \in [d, d + \epsilon) \end{cases}$$

Then $\|g_{c,d,\epsilon} - 1_{(c,d]}\|_p \leq 2\epsilon$. Now let $\phi(x) = \sum_{k=1}^n a_k 1_{(c_k,d_k]}$ and let $g = \sum_{k=1}^n a_k g_{c_k,d_k,\epsilon/2|a_k|n}$ then

$$\|\phi - g\|_p \leq \sum_{k=1}^n \|a_k(1_{(c_k,d_k]} - g_{c_k,d_k,\epsilon/2|a_k|n})\|_p \leq \sum_{k=1}^n |a_k|\epsilon/|a_k|n \leq \epsilon.$$

□

7 Product Measures

In this section we look at taking two measure spaces $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ and defining a σ algebra and a measure on the product space $E \times F$. This will give us another way of defining Lebesgue measure on $\mathbb{R}^d$. First we remind ourselves of the definition of cartesian product.

Definition 7.1 (Cartesian product)**.** If E and F are spaces then the cartesian product $E \times F$ is the space of twoples (x, y) where $x \in E$ and $y \in F$.

Example 7.2. $\mathbb{R} \times \mathbb{R} = \mathbb{R}^2$.

Now we want to consider the product σ-algebra.

Definition 7.3. The product σ-algebra $\mathcal{E} \times \mathcal{F}$ is a σ-algebra on $E \times F$ which is generated by the collection

$$\mathcal{A} = \{A \times B \,:\, A \in \mathcal{E}, B \in \mathcal{F}\}.$$

That is to say $\mathcal{E} \times \mathcal{F} = \sigma(\mathcal{A})$.

We now take some time to look at the projection maps π_E and π_F.

Definition 7.4. We define two maps $\pi_E : E \times F \to E$ and $\pi_F : E \times F \to F$ by

$$\pi_E(x, y) = x, \quad \pi_F(x, y) = y.$$

Lemma 7.5. *The maps π_E and π_F are both measurable. Furthermore if $C \in \mathcal{E} \times \mathcal{F}$ then the following sets are measurable*

$$C_x = \{y \in F \,:\, (x, y) \in C\} = \pi_F\left(\pi_E^{-1}(\{x\}) \cap C\right) \in \mathcal{F}, \quad C_y = \{x \in E \,:\, (x, y) \in C\} = \pi_E\left(\pi_F^{-1}(\{y\}) \cap C\right).]$$

Furthermore if $f : E \times F \to G$ is a measurable function then $f_x : F \to G$ defined by $f_x(y) = f(x, y)$ and $f_y : E \to G$ defined by $f_y(x) = f(x, y)$ are both measurable functions.

Proof. First let us show that the projection maps are measurable. Let A be in $\mathcal{E}$ then $\pi_E^{-1}(A) = A \times F$, as $F \in \mathcal{F}$ this is a product set so is in $\mathcal{E} \times \mathcal{F}$.

Now let us look at C_x. Let $\mathcal{C}$ be the collection of sets in $\mathcal{E} \times \mathcal{F}$ such that $C_x \in \mathcal{F}$. Then $\mathcal{C}$ contains all the product sets. We now want to show that $\mathcal{C}$ is a σ-algebra. $(C^c)_x = \{y \in F \,:\, (x, y) \in C^c\} = \{y \in F \,:\, (x, y) \notin C\} = F \setminus \{y \in F \,:\, (x, y) \in C\} = (C_x)^c$. Therefore $C \in \mathcal{C}$ implies that $C^c \in \mathcal{C}$. We also have that $(\bigcup_n C_n)_x = \bigcup_n ((E_n)_x)$. Therefore, $\mathcal{C}$ is closed under complements and countable unions so is a σ-algebra. Therefore $\mathcal{C} \supset \mathcal{E} \times \mathcal{F}$.

Now we move onto f_x. If $A \in \mathcal{F}$ then $f_x^{-1}(A) = \{y \in F \,:\, f(x, y) \in A\} = (f^{-1}(A))_x$. Using the previous part we know that this is a measurable set. Therefore f_x is measurable. $\square$

Theorem 7.6 (Product Measure)**.** *Given two measure spaces $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ there exists a unique measure, $\mu \times \nu$, on $\mathcal{E} \times \mathcal{F}$ such that $(\mu \times \nu)(A \times B) = \mu(A)\nu(B)$ when $A \in \mathcal{E}$ and $B \in \mathcal{F}$. Furthermore*

$$(\mu \times \nu)(C) = \int_E \nu(C_x)\mu(\mathrm{d}x) = \int_F \mu(C_y)\nu(\mathrm{d}y).$$

Proof. As $\mathcal{A} = \{A \times B \,:\, A \in \mathcal{E}, B \in \mathcal{F}\}$ is a π-system generating $\mathcal{E} \times \mathcal{F}$ we can use Carathéodory's extension theorem to prove the first part of this theorem. However we will work directly as defining this measure is straightforward and useful for understanding it.

First we check that $x \mapsto \nu(C_x)$ and $y \mapsto \mu(C_y)$ are both measurable functions so the integrals are well defined. Let us begin in the case that ν is a finite measure. Let $\mathcal{C}$ be the collection of sets for which the function $x \mapsto \nu(C_x)$ is $\mathcal{E}$ measurable. If $C = A \times B$ then $\nu(C_x) = \nu(B)1_{x \in A}$ which is measurable. Now we want to show that $\mathcal{C}$ is a σ-algebra. If $C^1 \subset C^2$ then $\nu((C^2 \setminus C^1)_x) = \nu(C^2_x) - \nu(C^1_x)$ so $C^2 \setminus C^1 \in \mathcal{C}$. Suppose that C^n is an increasing sequence of sets in $\mathcal{C}$ then $\nu\left(\left(\bigcup_n C^n\right)_x\right) = \lim_n \nu\left(C^n_x\right)$ so $\bigcup_n C_n$ is in $\mathcal{C}$. Therefore $\mathcal{C}$ is a σ-algebra and consequently contains $\mathcal{E} \times \mathcal{F}$. Now the only reason that we needed ν to be finite was to ensure that $A \times B \in \mathcal{C}$ as otherwise this function might take the value infinity sometimes. You can solve this problem by putting a σ-algebra on $[0, \infty]$ or we can work in the σ-finite setting and let $\{D_n\}$ be a sequence of disjoint subsets with $\nu(D_n) < \infty$ whose union is the whole of F. By the argument above $x \mapsto \nu((C \cap D_n)_x)$ is always measurable (restricting the space to D_n) and $\nu(C_x) = \lim_n \sum_{k=1}^n \nu((C \cap D_n)_x)$.

Now we move onto the main part of the proof we can define two different candidates for $(\mu \times \nu)$ namely

$$(\mu \times \nu)_1(C) = \int_E \nu(C_x)\mu(\mathrm{d}x), \quad (\mu \times \nu)_2(C) = \int_F \mu(C_y)\nu(\mathrm{d}y).$$

We can see that if C is of the form $A \times B$ then

$$(\mu \times \nu)_1(A \times B) = \int_E \nu(B) 1_{x \in A} \mu(\mathrm{d}x) = \mu(A)\nu(B) = \int_F \mu(A) 1_{y \in B} \nu(\mathrm{d}y) = (\mu \times \nu)_2(A \times B).$$

Now we would like to define a collection of subsets $\tilde{\mathcal{C}}$ which is the sets C for which $(\mu \times \nu)_1(C) = (\mu \times \nu)_2(C)$. We have already shown that this contains all sets of the form $A \times B$. We want to show that $\tilde{\mathcal{C}}$ is a σ-algebra which will then imply that $\tilde{\mathcal{C}} \supseteq \mathcal{E} \times \mathcal{F}$. Suppose $C_1 \subseteq C_2$ are both in $\tilde{\mathcal{C}}$ then as $\nu((C_1 \setminus C_2)_x) = \nu((C_1)_x) - \nu((C_2)_x)$ this shows that $C_1 \setminus C_2 \in \tilde{\mathcal{C}}$. Now let C_n be a sequence of disjoint sets in $\tilde{\mathcal{C}}$ we have that

$$\begin{aligned}
(\mu \times \nu)_1 \left(\bigcup_n C_n \right) &= \int_E \nu \left(\left(\bigcup_n C_n \right)_x \right) \mu(\mathrm{d}x) \\
&= \int_E \sum_n \nu((C_n)_x) \mu(\mathrm{d}x) \\
\text{Using Beppo-Levi} \quad &= \sum_n \int_E \nu((C_n)_x) \mu(\mathrm{d}x) \\
&= \sum_n (\mu \times \nu)_1(C_n) \\
&= \sum_n (\mu \times \nu)_2(C_n) \\
&= \sum_n \int_F \mu((C_n)_y) \nu(\mathrm{d}y) \\
\text{Beppo-Levi} &= \int_F \sum_n \mu((C_n)_y) \nu(\mathrm{d}y) \\
&= \int_F \mu \left(\left(\bigcup_n C_n \right)_y \right) \nu(\mathrm{d}y) = (\mu \times \nu)_2 \left(\bigcup_n C_n \right).
\end{aligned}$$

Therefore $\bigcup_n C_n \in \tilde{\mathcal{M}}$. □

One of the key tools we get when using product measure is Fubini's theorem. There are two theorems one for positive functions, one for integrable functions. The naming gets a bit wooly, but often the theorem for positive functions is called Tonelli's theorem and that for integrable functions is called Fubini's theorem. Sometimes the later is called the Fubini-Tonelli theorem and sometimes both are called Fubini-Tonelli or Fubini. To play it safe I'm going to call both Fubini-Tonelli Theorem.

Theorem 7.7 (Fubini-Tonelli theorem for positive functions)**.** *Suppose that $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ are measure spaces and f is a non-negative $\mathcal{E} \times \mathcal{F}$ measurable function then the functions $x \mapsto \int_F f(x,y)\nu(\mathrm{d}y)$ and $y \mapsto \int_E f(x,y)\mu(\mathrm{d}x)$ are both measurable and*

$$(\mu \times \nu)(f) = \int_E \left(\int_F f(x,y)\nu(\mathrm{d}y) \right) \nu(\mathrm{d}x) = \int_F \left(\int_E f(x,y)\mu(\mathrm{d}x) \right) \nu(\mathrm{d}y).$$

Proof. We build up the proof gradually, beginning with the case where f is the indicator function of a set $C \in \mathcal{E} \times \mathcal{F}$. In this case the measurability of the integrals in x or y and the form for $(\mu \times \nu)(f)$ are given by the construction of the product measure in the previous theorem.

The linearity of the integral then imply that the Fubini-Tonelli theorem holds whenever f is a non-negative simple function. We then note that any non-negative measurable function f, can be approximated from below by non-negative simple functions. Let f_n be a sequence of simple functions approximating f. Then

$$f_n = \sum_{k=1}^{N_n} c_k^n 1_{C_k^n},$$

where $C_k^n \in \mathcal{E} \times \mathcal{F}$. Then we know that

$$(\mu \times \nu)(f_n) = \int_E \left(\int_F c_k^n 1_{C_k^n}(x, y)\nu(\mathrm{d}y) \right) \mu(\mathrm{d}x) = \int_E \left(\int_F c^n 1_{(C_k^n)_x}(y)\nu(\mathrm{d}y) \right) \mu(\mathrm{d}x).$$

By monotone convergence as $n \to \infty$ the left hand side converges to $(\mu \times \nu)(f)$. We can also see that by monotone convergence

$$\int_F c^n 1_{(C_k^n)_x}(y)\nu(\mathrm{d}y) \uparrow \int_F f(x, y)\nu(\mathrm{d}y).$$

Consequently, we use monotone convergence again to get that the right hand side converges to

$$\int_E \left(\int_F f(x, y)\nu(\mathrm{d}y) \right) \mu(\mathrm{d}x).$$

This gives the desired conclusion for positive f. □

Theorem 7.8 (Fubini-Tonelli theorem for integrable functions). *Suppose that $(E, \mathcal{E}, \mu)$ and $(F, \mathcal{F}, \nu)$ are measure spaces and f is a $\mathcal{E} \times \mathcal{F}$ measurable function which is integrable with respect to $(\mu \times \nu)$ then the functions*

$$g(x) = \begin{cases} \int_F f(x, y)\nu(\mathrm{d}y) & \int_F |f(x, y)|\nu(\mathrm{d}y) < \infty \\ 0 & \int_F |f(x, y)|\nu(\mathrm{d}y) = \infty \end{cases}$$

and

$$h(y) = \begin{cases} \int_E f(x, y)\mu(\mathrm{d}x) & \int_E |f(x, y)|\mu(\mathrm{d}x) < \infty \\ 0 & \int_E |f(x, y)|\mu(\mathrm{d}x) = \infty \end{cases}$$

are both measurable and integrable. Furthermore,

$$(\mu \times \nu)(f) = \int_E \left(\int_F f(x, y)\nu(\mathrm{d}y) \right) \nu(\mathrm{d}x) = \int_F \left(\int_E f(x, y)\mu(\mathrm{d}x) \right) \nu(\mathrm{d}y).$$

Proof. Now we turn to the case where f is not necessarily non-negative but is $(\mu \times \nu)$ integrable. By our result for non-negative functions we know that

$$(\mu \times \nu)(|f|) = \int_E \left(\inf_F |f(x, y)|\nu(\mathrm{d}y) \right) \mu(\mathrm{d}x),$$

which proves that the function $x \mapsto \int_F |f(x, y)|\nu(\mathrm{d}y)$ is μ-integrable, and is consequently finite almost everywhere, therefore restricting the functions g, h to where they would be finite is not a problem. Let A be the set on which $x \mapsto \int_F |f(x, y)|\nu(\mathrm{d}y)$ is finite. Now we write $f = f_+ - f_-$ in our usual way. Then by definition

$$\int f(x, y)\nu(\mathrm{d}y) 1_{x \in A} = \left(\int f_+(x, y)\nu(\mathrm{d}y) - \int f_-(x, y)\nu(\mathrm{d}y) \right) 1_{x \in A}.$$

Then using the fact that $\mu(A^c) = 0$, and our result for non-negative functions we have

$$\begin{aligned}(\mu\times\nu)(f) = (\mu\times\nu)(f_+) - (\mu\times\nu)(f_-) &= \int_E\int_F f_+(x,y)\nu(\mathrm{d}y)\mu(\mathrm{d}x) - \int_E\int_F f_-(x,y)\nu(\mathrm{d}y)\mu(\mathrm{d}x)\\ &= \int_E\left(\int_F f_+(x,y)\nu(\mathrm{d}y) - \int_F f_-(x,y)\nu(\mathrm{d}y)\right)1_{x\in A}\mu(\mathrm{d}x)\\ &= \int_E\left(\int_F f(x,y)\nu(\mathrm{d}y)1_A\right)\mu(\mathrm{d}x)\\ &= \int_E\int_F f(x,y)\nu(\mathrm{d}y)\mu(\mathrm{d}x).\end{aligned}$$

□

7.1 Applications of product measure and Fubini's theorem

This section is a collection of examples and applications of product measure and Fubini's theorem

Example 7.9. Suppose $(E,\mathcal{E},\mu)$ is a measure space we look at its product with $(\mathbb{R},\mathcal{B}(\mathbb{R}),\lambda)$ and suppose that $f : E \to \mathbb{R}$ is a non-negative measurable, then the set

$$A = \{(x,y) \,:\, 0 \le y \le f(x)\}$$

is measurable with the product σ-algebra and its measure is the *area under the graph* of f. We have that

$$(\mu\times\lambda)(A) = \mu(\lambda(A_x)) = \mu(f),$$

and

$$(\mu\times\lambda)(A) = \lambda(\mu(A_y)) = \lambda(\{x \,:\, f(x)\ge y\}) = \int_0^\infty \mu(\{x \,:\, f(x)\ge y\})\mathrm{d}y.$$

Before the next example we need a result I forgot about in measurable functions (I will move it there in the final version of the notes).

Lemma 7.10. *If $(E,\mathcal{E}),(F,\mathcal{F})$ and $(G,\mathcal{G})$ are all measurable spaces and $f : E\to F$ and $g : F\to G$ are measurable then so is $g\circ f$.*

Proof. Take any set $A\in\mathcal{G}$ then $(g\circ f)^{-1}(A) = \{x\in E \,:\, g(f(x))\in A\}$. Let us call $B = \{y\in F \,:\, g(y)\in A\} = g^{-1}(A)$ then $(g\circ f)^{-1}(A) = \{x\in E \,:\, f(x)\in B\} = f^{-1}(B)$. Then as g is measurable and $A\in\mathcal{G}$ then $B\in\mathcal{F}$. In the same way as f is measurable and $B\in\mathcal{F}$ then $f^{-1}(B)\in\mathcal{E}$. As $f^{-1}(B) = (g\circ f)^{-1}(A)$ this shows that $(g\circ f)^{-1}(A)\in\mathcal{E}$ for every $A\in\mathcal{G}$ and hence $g\circ f$ is measurable. □

Example 7.11 (Convolutions). Suppose that both f and g are in $L^1(\mathbb{R})$ then for almost every x the function $t\mapsto f(x-t)g(t)$ is also in $L^1(\mathbb{R})$. We have that the function $f*g$ defined by

$$x\mapsto\begin{cases}\int_{\mathbb{R}} f(x-t)g(t)\mathrm{d}t & \text{if } t\mapsto f(x-t)g(t) \text{ is Lebesgue integrable}\\ 0 & \text{Otherwise}\end{cases}$$

is in L^1 and satisfies $\|f*g\|_1 \le \|f\|_1\|g\|_1$.

We can prove this using Fubini-Tonelli. First we want to check that $t\mapsto f(x-t)g(t)$ is measurable. Write $h(t) = x-t$ this continuous function (for fixed x) and $t\mapsto f(x-t) = f(h(t))$ so it is the composition of two measurable functions so measurable. We also know that the product of two measurable functions is measurable to $f(x-t)g(t)$ is a measurable function of t. Now we want to check that it is integrable

$$\int \left| \int f(x-t)g(t)\mathrm{d}t \right| \mathrm{d}x \leq \int \int |f(x-t)g(t)|\mathrm{d}t\mathrm{d}x$$

as $f(x-t)g(t) \leq |f(x-t)g(t)|$ and $-f(x-t)g(t) \leq |f(x-t)g(t)|$. Now we apply Fubini-Tonelli and get

$$\int \int |f(x-t)g(t)|\mathrm{d}t\mathrm{d}x = \int \left(\int |f(x-t)|\mathrm{d}x \right) |g(t)|\mathrm{d}t = \int \|f\|_1 |g(t)|\mathrm{d}t = \|f\|_1 \|g\|_1.$$

We can also show that convolutions of functions are continuous functions using the tools from measure theory. For this we need to show that shifts are continuous in L^1.

Lemma 7.12. *Define the map $T_\tau : L^p(\mathbb{R}) \to L^p(\mathbb{R})$ by $(T_\tau f)(x) = f(x+\tau)$ then*

$$\lim_{\tau \to 0} \|T_\tau f - f\|_p = 0.$$

Proof. We want to show that for any ϵ there exists τ_* such that if $\tau \leq \tau_*$ then $\|T_\tau f - f\|_p \leq \epsilon$. First let us show the result for step functions, that is to say functions of the form

$$\phi(x) = \sum_{k=1}^{n} a_k 1_{[c_k, d_k)}.$$

First by Minkowski's inequality we have

$$\|T_\tau \phi - \phi\|_p \leq \sum_{k=1}^{n} |a_k| \|T_\tau 1_{[c_k,d_k)} - 1_{[c_k,d_k)}\|_p = \sum_{k=1}^{n} |a_k| \lambda([c_k+\tau, d_k+\tau) \Delta [c_k, d_k) \leq \sum_{k=1}^{n} |a_k| 2|\tau|.$$

Therefore we can make τ small enough so that this is less than ϵ.

Now let us look at a general f we know (from Assignment 4) that there is a step function ϕ such that $\|f-\phi\|_p \leq \epsilon/3$. We also can change variables $x \leftrightarrow x+\tau$ so $\|T_\tau f - T_\tau \phi\|_p = \left(\int |f(x+\tau) - \phi(x+\tau)|^p \mathrm{d}x \right)^{1/p} = \|f-\phi\|_p$ for any τ. For this ϕ we can find τ sufficiently small such that $\|T_\tau \phi - \phi\|_p \leq \epsilon/3$. Hence

$$\|T_\tau f - f\|_p \leq \|T_\tau f - T_\tau \phi\|_p + \|T_\tau \phi - \phi\|_p + \|\phi - f\|_p \leq \epsilon.$$

□

Now we go back to convolutions, we can show that if $f, \in L^p(\mathbb{R})$ and $g \in L^q(\mathbb{R})$ then $f * g$ is continous.

$$|f*g(y) - f*g(x)| = |\int_{\mathbb{R}} (f(x-t) - f(y-t))g(t)\mathrm{d}t| \leq \int_{\mathbb{R}} |f(x-t) - f(y-t)||g(t)|\mathrm{d}t$$

we can bound this using Hölder's inequailty by

$$\|g\|_q \left(\int |f(x-t) - f(y-t)|^p \mathrm{d}t \right)^{1/p} = \|g\|_q \left(\int |f(t-x+y) - f(t)|\mathrm{d}t \right)^{1/p} = \|g\|_q \|T_{-x+y} f - f\|_p.$$

So if $|x-y|$ is small enough then $|f*g(x) - f*g(y)|$ will also be small.

8 Radon-Nikodym Theorem

8.1 Signed measures

We introduce the notion of signed measures which will be useful in the proof of the Radon-Nikodym theorem.

Definition 8.1 (Finite signed measure)**.** A function μ from a σ-algebra $\mathcal{E}$ to $\mathbb{R}$ is a *finite signed measure* if

- $\mu(\emptyset) = 0$,
- If $(A_n)_{n\geq 1}$ is a sequence of disjoint sets then $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$

Example 8.2. If $(E, \mathcal{E}, \mu)$ is a measure space and $f \in L^1(E)$ then ν defined by $\nu(A) = \mu(f 1_A)$ is a signed measure.

We want to show two decomposition theorems which basically allow us to reduce the situation back to measures. First we need some more definitions and a useful Lemma.

Definition 8.3. If $(E, \mathcal{E})$ is a measurable space and ν is a finite signed measure then we call A a *positive set* if for every $B \in \mathcal{E}$ with $B \subseteq A$ then $\nu(B) \geq 0$. The *negative sets* are defined analagously.

Lemma 8.4. *Suppose that ν is a finite signed measure on $(E, \mathcal{E})$ and suppose $A \in \mathcal{E}$ with $\nu(A) < 0$ then there exists a negative set B with $B \subseteq A$ and $\nu(B) \leq \nu(A)$. We will produce this set A by an itterative process, define*

$$\delta_1 = \sup\{\nu(C) \,:\, C \subseteq A\},$$

then since $\emptyset \subseteq A$ we have that $\delta_1 \geq 0$. If $\delta_1 = 0$ then we have a negative set so are done. If not we can find a set $C_1 \subseteq A$ with

$$\nu(C_1) \geq \min\{\delta_1/2, 1\}.$$

Now we will define a sequence of δ_n and C_n by setting

$$\delta_n = \sum\{\nu(C) \,:\, C \subseteq (A \setminus \bigcup_{i=1}^{n-1} C_i)\}$$

and C_n a set so that

$$\nu(C_n) \geq \min\{\delta_n/2, 1\}.$$

Now let $C_\infty = \bigcup_n C_n$ and $B = A \setminus C_\infty$. We now need to check that B has the required properties,

$$\nu(A) = \nu(C_\infty) + \nu(B) \geq \nu(B),$$

as $\nu(C_\infty) \geq 0$ by construction. As ν is a finite measure we must have $\nu(C_\infty) < \infty$ and as the C_n are constructed to be disjoint this means we must have $\lim_n \nu(C_n) = 0$*. Therefore* $\lim_n \delta_n = 0$*. If $D \subseteq B$ then we must have that $\nu(D) \leq \delta_n$ for every n, therefore $\nu(D) \leq 0$.*

Now we are able to state and prove our two decomposition theorems.

Theorem 8.5 (Hahn Decomposition theorem)**.** *Let $(E, \mathcal{E})$ be a measure space and ν a finite signed measure. Then there exists a positive set P and a negative set N for ν such that $E = P \cup N$.*

Proof. Let $L = \inf\{\nu(A) : A \text{ is a negative set for } \nu\}$ then L is finite as otherwise we could construct a set with measure $-\infty$. Then let A_n be a negative set with $\nu(A_n) \leq L + 1/n$ then let $N = \bigcup_n A_n$. It is easy to check that N is a negative set and that $\nu(N) = L$. Let $P = N^c$ we want to check that P is a positive set. Suppose there exists a set $A \subseteq P$ with $\nu(A) < 0$, then by our lemma there exists a negative set $B \subseteq A$ with $\nu(B) \leq \nu(A) < 0$. Then $N \cup B$ is a negative set and N and B are disjoint so $\nu(N \cup B) = \nu(N) + \nu(B) < \nu(N)$ which contradicts the fact that $\nu(N) = L = \inf\{\nu(A) : A \text{ is a negative set for } \nu\}$ so we are done. □

Theorem 8.6 (Jordan decompostion theorem)**.** *Every signed measure is the difference of two positive measures. Precisely, if $(E, \mathcal{E})$ is a measure space and ν is a signed measure then there exist measures ν_+ and ν_- such that for every $A \in \mathcal{E}$ we have $\nu(A) = \nu_+(A) - \nu_-(A)$.*

Proof. Take some Hahn decomposition (P, N) then let $\nu_+(A) = \nu(A \cap P)$, as $A \cap P \subseteq P$ then $\nu(A \cap P) \geq 0$. Similarly let $\nu_-(A) = -\nu(A \cap N)$. By additivity of ν we have that $\nu(A) = \nu_+(A) - \nu_-(A)$. Countable additivity of ν_+ and ν_- follow immediately from countable additivity of ν.

Now we notice that if $B \subseteq A$ then

$$\nu(B) = \nu_+(B) - \nu_-(B) \leq \nu_+(B) \leq \nu_+(A)$$

and $\nu_+(A) = \nu(A \cap P)$ therefore we have that

$$\nu_+(A) = \sup\{\nu(B) : B \subseteq A, B \in \mathcal{E}\}$$

in the same way

$$\nu_-(A) = \sup\{-\nu(B) : B \subseteq A, B \in \mathcal{E}\}.$$

This shows that the values of ν_+, ν_- do not depend on the particular choice of Hahn decomposition. □

8.2 Absolute Continuity

We now move on to the main focus of this section, the Radon-Nikodym theorem. In order to understand the theorem we need a definition.

Definition 8.7. Let $(E, \mathcal{E})$ be a measurable space and μ and ν be two measures then we say that ν is absolutely continuous with respect to μ of $\nu \ll \mu$ if for every $A \in \mathcal{E}$ with $\mu(A) = 0$ we also have $\nu(A) = 0$.

We can characterise absolute continuity

Definition 8.8. Suppose that $(E, \mathcal{E})$ is a measurable space and μ a measure, ν a finite measure then $\nu \ll \mu$ if and only if for earch $\epsilon > 0$ there exists a $\delta > 0$ such that $\mu(A) < \delta$ implies that $\nu(A) < \epsilon$.

Proof. First let us suppose there exists such at δ for each ϵ, then if $\mu(A) = 0$ we have that $\mu(A) < \delta$ for every δ so we must have $\nu(A) < \epsilon$ for every ϵ so $\nu(A) = 0$.

Now let us suppose that $\nu \ll \mu$. We prove the result by contradiction. Suppose there exists an ϵ such that for every δ there exists a set A with $\mu(A) < \delta$ but $\nu(A) > \epsilon$. Then we can find a sequence of sets A_k such that $\mu(A_k) < 2^{-k}$ but $\nu(A_k) \geq \epsilon$. By the first Borel-Cantelli lemma we have that

$$\mu\left(\bigcap_n \bigcup_{m \geq n} A_m\right) = 0.$$

We also have that $\nu(\bigcup_{m\geq n} A_m) \geq \nu(A_n) \geq \epsilon$ and

$$\nu\left(\bigcap_n \bigcup_{m\geq n} A_m\right) = \lim_n \nu\left(\bigcup_{m\geq n} A_m\right) \geq \epsilon.$$

This show gives us a set with $\mu(B) = 0$ but $\nu(B) \geq 0$ which contradicts $\nu \ll \mu$. $\square$

Now we can prove the main theorem for this section.

Theorem 8.9 (Radon-Nikodym Theorem)**.** *Let $(E, \mathcal{E})$ be a measure space and let μ, ν be two finite measures with $\nu \ll \mu$. Then there exists a measurable function $g : E \to [0, \infty)$ such that $\nu(A) = \mu(g1_A)$. The function g is unique up to identifying almost everywhere equal functions. We write $g = \mathrm{d}\nu/\mathrm{d}\mu$ and call it the Radon-Nikodym derivative of ν with respect to μ.*

Proof. Let us define the set $\mathcal{F}$ which is the set of all measurable functions, f, with $\mu(f1_A) \leq \nu(A)$ for every $A \in \mathcal{E}$. The idea is that $\mathcal{F}$ contains a function g which achieves $\mu(g) = \sup_{f\in\mathcal{F}} \mu(f)$.

As a first step we show that $f_1 \vee f_2 = \max\{f_1, f_2\} \in \mathcal{F}$ when $f_1, f_2 \in \mathcal{F}$. Let us take any $A \in \mathcal{E}$ then let $A_1 = A \cap \{f_1 \geq f_2\}$ and $A_2 = A \cap \{f_1 < f_2\}$. Then

$$\mu(f_1 \vee f_2 1_A) = \mu(f_1 \vee f_2 1_{A_1}) + \mu(f_1 \vee f_2 1_{A_2}) = \mu(f_1 1_{A_1}) + \mu(f_2 1_{A_2}) \leq \nu(A_1) + \nu(A_2) = \nu(A).$$

Therefore $f_1 \vee f_2 \in \mathcal{F}$.

Now take a sequence f_n such that $\mu(f_n) \geq \sup_{f\in\mathcal{F}} \mu(f) - 1/n$. Then let $g_n = f_1 \vee f_2 \vee \cdots \vee f_n$, so that the sequence of function g_n is increasing and $\mu(g_n) \geq \sup_{f\in\mathcal{F}} \mu(f) - 1/n$. Then as g_n is increasing it has a limit g and the monotone convergence theorem shows that

$$\mu(g1_A) = \lim_n \mu(g_n 1_A) \leq \nu(A).$$

So $g \in \mathcal{F}$.

Now we can define another positive measure $\nu_0(A) = \nu(A) - \mu(g1_A)$. We want to show that $\nu_0 = 0$ and will do this by contradiction. Suppose that there exists $A \in \mathcal{E}$ such that $\nu_0(A) > 0$ then by monotonicity we will have $\nu_0(E) > 0$ and since μ is a finite measure there exists a number $\epsilon > 0$ such that $\nu_0(E) > \epsilon\mu(E)$. Now $\nu_0 - \epsilon\mu$ is a finite signed measure. Let (P, N) be a Hahn decomposition for this signed measure. Then $(\nu_0 - \epsilon\mu)(A \cap P) \geq 0$ so $\nu_0(A \cap P) \geq \epsilon\mu(A \cap P)$. Hence we have

$$\begin{aligned}\nu(A) &= \mu(g1_A) + \nu_0(A) \geq \mu(g1_A) + \nu_0(A \cap P)\\ &\geq \mu(g1_A) + \epsilon\mu(A \cap P) = \mu(1_A(g + \epsilon 1_P)).\end{aligned}$$

We also have that $\mu(P) > 0$ as if $\mu(P) = 0$ then we would have $\nu_0(P) = 0$ as $\nu_0 \ll \nu \ll \mu$, and this would mean

$$(\nu_0 - \epsilon\mu)(E) = (\nu_0 - \epsilon\mu)(N) \leq 0,$$

which would contradict $\nu_0(E) > \epsilon\mu(E)$. Therefore, $g + 1_P$ belongs to $\mathcal{F}$ but $\mu(g + 1_P) > \mu(g)$ which contradicts the fact that g achieves $\mu(g) = \sup_{f\in\mathcal{F}} \mu(f)$. Hence $\nu(A) = \mu(g1_A)$.

Now we turn to uniqueness suppose that we have two positive functions g, h such that $\nu(A) = \mu(g1_A) = \mu(h1_A)$ for every A, then as ν is finite g and h are integrable so $g - h$ is integrable and $\mu((g - h)1_A) = 0$ for every A. As $g - h$ is measurable then $\{x \in E : g - h \geq 0\}$ is a measurable set so $\mu((g - h)1_{\{x\in E\,:\, g-h\geq 0\}}) = 0$. This shows that $(g - h)1_{\{x\in E\,:\, g-h\geq 0\}} = 0$ almost everywhere. In the same way $(g - h)1_{\{x\in E\,:\, g-h\leq 0\}} = 0$ almost everywhere. Therefore $g = h$ μ-almost everywhere. $\square$

8.3 Duality in L^p spaces

The goal of this section is to prove that if $1/p + 1/q = 1$ then the dual space of $L^p(E)$ is isomorphic to the space $L^q(E)$. First let us define a dual space.

Definition 8.10. Let $\mathcal{V}$ be a Banach space (a complete, normed vector space) then the dual space of $\mathcal{V}$ is written $\mathcal{V}'$ and is the space of all bounded linear operators from $\mathcal{V}$ to $\mathbb{R}$. We recall that we call an operator K on $\mathcal{V}$ bounded if $|K(v)| \leq C\|v\|$ for all $v \in \mathcal{V}$. We can define a norm on $\mathcal{V}'$ by $\|K\| = sup_{\|v\|=1}|K(v)|$.

The first thing to note is that if $g \in L^q(E)$ then we can define a bounded linear operator on $L^p(E)$ by $K_g(f) = \mu(fg)$. This is bounded by Hölder's inequality $|\mu(fg)| \leq \mu(|fg|) = \|fg\|_1 \leq \|f\|_p\|g\|_p$. It is also linear thanks to the linearity of the integral. Therfore we can produce a map from $L^q(E) \to (L^p(E))'$ by $g \mapsto K_g$.

Theorem 8.11. *Let $(E, \mathcal{E}, \mu)$ be a measure space and $p \in (1, \infty)$. The dual space of $L^p(E)$ is $L^q(E)$ where $1/p + 1/q = 1$. Furthermore the map defined by $g \mapsto K_g$ is an isometry.*

Proof. Remark: This result is similar in spirit to the Riesz representation result that was a non-examinable topic in week 6.

First we note that the map $g \mapsto K_g$ is linear and $\|K_g\|_{(L^p)'} \leq \|g\|_q$. Therefore the map is injective we want to show that $\|K_g\| = \|g\|$ and that it is surjective.

First for the fact that $\|K_g\| = \|g\|$ we look at the function $f(x) = \text{sgn}(g)|g(x)|^{q-1}$ then $\mu(|f|^p) = \mu(|g|^q) < \infty$. Therfore we can look at the action of K_g on f and we have $K_g(f) = \mu(|g|^q)$ so we know that $\|K_g\| \geq K_g(f)/\|f\|_p = \mu(|g|^q)/\mu(|g|^q)^{1/p} = \mu(|g|^q)^{1-1/p} = \|g\|_q$. Therefore $g \mapsto K_g$ preserves norms.

Now we want to show that this map is surgective, let us begin with the case where $\mu(E) < \infty$. Let us take K an arbirary element of $(L^p(E))'$. In this case $1_A \in L^p(E)$ for every $A \in \mathcal{E}$ so we can define a function on $\square$