

MATH3969: Measure Theory and Fourier Analysis

Charles Christopher Hyland

Semester 2 2018

Abstract

Thank you for stopping by to read this. These are notes collated from lectures and tutorials as I took this course.

Contents

1.1	Introduction	1
1.1.1	Basic Set and Function Definition	1
1.1.2	Introduction to σ -algebra and Measures.	1
2.1.3	Set Theory	1
2.1.4	Constructing σ – algebra from intersections	2
2.1.5	Construction of σ – algebra From Functions.	2
2.1.6	Measures	3
2.1.7	Constructing Measure Spaces	4
2.1.8	σ – algebra with disjoint sets	4
3.1.9	Properties of Measures	1
3.1.10	Metric Spaces	1
3.1.11	The Construction of measures from outer measures	3
4.1.12	Exploring Outer Measures and their properties.	1
4.1.13	Constructing σ – algebra For Measures.	2
5.1.14	μ^* -measurable set properties.	1
5.1.15	Lebesgue Measure	1
6.1.16	Regularity of the Lebesgue Measure.	1
6.1.17	Uniqueness of the Lebesgue Measure.	2
7.1.18	Uniqueness of Lebesgue Measure.	1
7.1.19	Linear Transformation of Lebesgue Measure.	1
7.1.20	The Lebesgue-Stieltjes Measure	1
8.2	Measurable Functions	1
8.2.1	Measurable Functions.	1
9.2.2	Convergence of sequences of functions.	1
9.2.3	Sequences of measurable functions	1
10.2.4	Simple Measurable Functions.	1
10.2.5	Non-negative Measurable Functions approximation by simple functions.	2
11.3	Integration and properties	1
11.3.1	Integration	1
11.3.2	Integration of non-negative simple functions.	1
11.3.3	Examples of integrals with respect to measures.	3
12.3.4	Monotone Convergence Theorem.	1
13.3.5	Integrable Functions.	1
14.3.6	Measures with a density.	1
14.3.7	Special Transformation Formula	2
15.3.8	Limit Theorems	1
16.4	The Lebesgue Spaces	1
16.4.1	The Lebesgue Spaces	1
17.4.2	Further properties of Lebesgue Spaces	1
18.4.3	The L^p spaces	1
19.4.4	Density Theorems	1
19.4.5	The space L^∞	3
20.4.6	Tonelli’s Theorem	1
21.5	Convolution and Approximations	1
21.5.1	Translation of a function	1
22.5.2	Convex Functions	1

22.5.3	Convolution	3
23.5.4	Approximate Identities	1
24.5.5	Approximate Theorems	1
25.6	The Fourier Transform	1
25.6.1	The Fourier Transform	1
26.6.2	Properties of the Fourier transform	1
27.6.3	Further Properties of the Fourier transform	1
28.6.4	Fourier Transform in L^2	1
29.7	Absolute Continuity and the Radon-Nikodym Theorem	1
29.7.1	Hilbert Spaces	1
30.7.2	Absolute Continuity	1
30.7.3	Radon-Nikodym Theorem	2
31.8	Probability Theory	1
31.8.1	Introduction to Probability Theory	1
31.8.2	Definitions	1
32.8.3	Distribution of a random variable	1
33.8.4	Conditional Expectation	1

1. Introduction.

1.1 Introduction

1.1.1 Basic Set and Function Definition

Definition 1.1 (*Domain*). If we are given a function $f: X \rightarrow Y$, then X is the **domain** of f .

Definition 1.2 (*Codomain*). If we are given a function $f: X \rightarrow Y$, then Y is the **codomain** of f . It is the set in which the output of the function is constrained to fall into.

Definition 1.3 (*Range/Image*). The set of all actual output values.

Definition 1.4 (*Preimage*). The subset of the domain which gets mapped to the image.

If $A, B \subseteq X$, then $A = B$ if $A \subseteq B$ and $B \subseteq A$.

Recall that in sets, you cannot simply do subtractions. Instead, you go A/B as meaning "A minus B".

Definition 1.5 (*De Morgan's Law*). De Morgan's law are simply

1. $(A \cup B)^c = A^c \cap B^c$;
2. $(A \cap B)^c = A^c \cup B^c$.

Intersection of sets are **associative**, that is, $(A \cap B) \cap C = A \cap (B \cap C)$.

1.1.2 Introduction to σ -algebra and Measures.

Definition 1.6 (σ -algebra). Let \mathcal{A} be a collection of subsets of the set X . \mathcal{A} is a σ -algebra if it has the 3 properties:

1. $\emptyset \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.
3. If $A_k \in \mathcal{A}$ for all $k \in \mathbb{N}$, then $\bigcup_{k \in \mathbb{N}} A_k \in \mathcal{A}$.

Remark 1.7 If we replace property 3 with $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$, then \mathcal{A} is known as an **algebra**.

Definition 1.8 (*Power Set*). The class of all subsets of the set X is known as the **power set** denoted by $\mathcal{P}(X)$.

Proposition 1.9 *Let \mathcal{A} be a σ -algebra of the set X . The following properties are true.*

1. $\emptyset, X \in \mathcal{A}$.
2. If $A_k \in \mathcal{A}$ for all $k \in \mathbb{N}$, then $\bigcap_{k \in \mathbb{N}} A_k \in \mathcal{A}$.
3. If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$ and $A \cap B \in \mathcal{A}$.

Remark 1.10 *To prove property 3, recall that the 3rd property of a σ -algebra is for finite countable unions. Hence, if we set A and B to union with a countable number of empty sets, this can then prove property 3.*

Remark 1.11 *De Morgan's Law of*

1. $(A \cup B)^c = A^c \cap B^c$.
2. $(A \cap B)^c = A^c \cup B^c$.

can help us prove property 2.

Definition 1.12 *We define the distribution laws of set theory*

$$\bigcup (A_i) \cap B = \bigcup (A_i \cap B);$$

$$\bigcap (A_i) \cup B = \bigcap (A_i \cup B).$$

Additionally, we note that $A \cap B \subseteq A$ and hence

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

Proposition 1.13 *The countable intersection of σ -algebra is itself a σ -algebra.*

MATH3969: Measure Theory and Fourier Analysis

2. Constructing σ -algebra.

2.1.3 Set Theory

Definition 2.14 (*Bounded Set*). A set K is bounded if the set $K \subset \text{Ball}(x, r)$ for some $x \in \mathbb{R}^N$ and $r \in (0, \infty)$.

Definition 2.15 (*Open Set*). A set X is open if for all $x \in X$, the ball of x of radius ϵ for all $\epsilon > 0$ is in the set X .

Definition 2.16 (*Closed Set*). A closed set is the compliment to the open set.

We note some interesting properties.

Proposition 2.17 Let $\{E_1, E_2, \dots\}$ be arbitrary collection of open sets and $\{F_1, F_2, \dots\}$ be arbitrary collection of closed sets. We have that

- (i) $\bigcup_{j=1}^{\infty} E_j$ is open,
- (ii) $\bigcap_{j=1}^{\infty} E_j$ does not need to be open,
- (iii) $\bigcup_{j=1}^{\infty} F_j$ does not need to be open,
- (iv) $\bigcap_{j=1}^{\infty} F_j$ is closed.

We also note infinite intervals of the form

- (i) $(-\infty, \infty)$ is open and closed;
- (ii) (a, ∞) and $(-\infty, b)$ are open;
- (iii) $[a, \infty)$ and $(-\infty, b]$ are closed.

Definition 2.18 (*Compact Set*). A set $K \subset \mathbb{R}^N$ is compact if every open cover of K has a finite subcover.

Remark 2.19 To elaborate, if had a set K that can be covered by a union of open sets, there exists a finite subset of that open cover which stills covers the set K . However, it is difficult to check this hence we can use the Heine-Borel theorem.

Theorem 2.20 (*Heine-Borel Theorem*). A set $X \in \mathbb{R}^N$ is compact if and only if it is closed and bounded.

2.1.4 Constructing σ – algebra from intersections

Let us define a set X and we define \mathbb{C} to be a collection of subsets of X . We define $\mathbb{A}(\mathbb{C})$ to be the intersection of all σ -algebra of subsets of X that contains \mathbb{C} . $\mathbb{A}(\mathbb{C})$ is the smallest σ -algebra *generated* by \mathbb{C} . It can be characterised as the intersection of all σ – algebra containing \mathbb{C} .

Proposition 2.21 (*σ – algebra generated by a system of sets*). Let \mathcal{C} be a collection of subsets of X . We call $\mathcal{A}(\mathcal{C})$ the σ – algebra generated by \mathcal{C} whereby $\mathcal{A}(\mathcal{C})$ is the intersection of σ – algebra that contains \mathcal{C} .

Definition 2.22 (*Borel σ -algebra*). A σ -algebra generated by the class of **open sets** in \mathbb{R}^n . We write $\mathcal{B} := \mathcal{A}(\mathcal{C})$ where \mathcal{C} as the collection of all open subsets of X . A set $A \in \mathcal{B}(\mathbb{R}^N)$ is called a Borel set.

We can't simply just take the collection of open sets to be the Borel σ – algebra, as it does not contain the complement of these sets, the closed set in it. Hence we need to generate the Borel σ – algebra.

To generate a Borel σ -algebra, take the intersection of all σ -algebras containing open sets. Examples of Borel sets include the open sets, closed, and all countable unions and intersection of open/close sets. We can also generate \mathcal{B} through half open intervals $[a, b)$ with $a, b \in \mathbb{R}$ as we can express $[a, b) = \cap_{n=1}^{\infty} (a - \frac{1}{n}, b)$ which is a countable intersection of open sets, which we can then generate \mathcal{B} from. Additionally, we can rewrite a semi-infinite set as $(-\infty, b) = \cup_{n=1}^{\infty} (-n, b)$ and an infinite set $\mathbb{R} = \cup_{n=1}^{\infty} (-n, n)$.

2.1.5 Construction of σ – algebra From Functions.

We can use *functions* to generate σ – algebra from another σ – algebra. Let X and Y be sets with the function $f: X \rightarrow Y$.

We have 2 scenarios here, we can generate σ – algebra of subsets of X from a σ – algebra of subsets of Y or vice versa.

For $A \subseteq Y$, we define the *inverse image (pre-image)* of A by:

$$f^{-1}[A] = \{x \in X : f(x) \in A\} \subseteq X$$

Note the $[\cdot]$ symbol, as the function may not necessarily be invertible (therefore you may have multiple things in the domain map to the same thing).

Proposition 2.23 (*Constructing σ – algebra From functions*). Suppose that X, Y are sets and $f: X \rightarrow Y$ is a function.

(i) If \mathcal{A} is a σ – algebra in Y , then:

$$\mathcal{A}_0 = \{f^{-1}[A] : A \in \mathcal{A}\} \subseteq \mathbb{P}(X)$$

is a σ – algebra in X .

(ii) If \mathcal{A} is a σ – algebra in X , then:

$$\mathcal{A}_1 = \{A \subseteq Y : f^{-1}[A] \in \mathcal{A}\} \subseteq \mathbb{P}(Y)$$

is a σ – algebra in Y .

To prove this, we note some properties of inverse images.

$$(i) \quad (f^{-1}[A])^c = f^{-1}[A^c]$$

(ii) If $A_i \subseteq Y$ for all $i \in I$, where I is an index set, then we have that:

$$\bigcup_{i \in I} f^{-1}[A_i] = f^{-1}\left[\bigcup_{i \in I} A_i\right];$$

(iii) If $A_i \subseteq Y$ for all $i \in I$, where I is an index set, then we have that:

$$\bigcap_{i \in I} f^{-1}[A_i] = f^{-1}\left[\bigcap_{i \in I} A_i\right];$$

2.1.6 Measures

Definition 2.24 (Measure). Let \mathcal{A} be a σ -algebra of subsets of X . A measure is a function $\mu: \mathcal{A} \rightarrow [0, \infty]$ such that:

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

$$(ii) \quad \mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) \leq \sum \mu(A_i) \text{ with equality if } A_i \cap A_j = \emptyset \text{ for } i \neq j. \text{ (Countable Additivity)}$$

Remark 2.25 If $A \in \mathcal{A}$, we call the set A **measurable** with respect to μ . Notice that that property (ii) requires a countable union of sets.

Definition 2.26 (Measurable Function). A function is **measurable** if it is a function between two **measurable spaces** (X, \mathcal{A}) such that the preimage of any measurable set is measurable.

Remark 2.27 A **probability measure** is a measure where $P(X) = 1$.

Definition 2.28 (Measurable Space). Given a nonempty set X and a σ -algebra \mathcal{A} on X , then the tuple (X, \mathcal{A}) is called a **measurable space**.

Definition 2.29 (Measure Space). Given a nonempty set X , a σ -algebra \mathcal{A} on X and a measure μ on the measurable space (X, \mathcal{A}) then the triple (X, \mathcal{A}, μ) is called a **measure space**.

To show something is a measure space, we first show that it contains a σ -algebra and a measure.

We define some examples of measures.

Example 2.30 (Dirac Measure). Let X be a set and fix $x \in X$. For $A \subseteq X$, we set

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

Here, we take every subset $A \subseteq X$ and see whether x is in that subset A . This is known as the Dirac Measure **concentrated at x** . $\delta_x: \mathbb{P}(X) \rightarrow [0, \infty)$ is a measure. The associated measure space is $(X, \mathbb{P}(X), \delta_x)$.

Note to show countable additivity for this measure, note that if A_j are disjoint, then only one of them can contain the element x .

Example 2.31 (Count Measure). Let X be a set and **fix** $x \in X$. For $A \subseteq X$, we set

$$\delta_x(A) = \begin{cases} \text{card}(A) & A \text{ has finite cardinality,} \\ \infty & \text{Otherwise.} \end{cases}$$

$\text{card}(A)$ is the cardinality of the set A . This is known as the counting measure.

Remark 2.32 Note that a simple probability measure is simply the counting measure scaled by the cardinality of possible outcomes. Here, the set X is denoted by Ω as the **sample space** whilst events are denoted as $A \subseteq \Omega$. The σ – algebra is the powerset $\mathbb{P}(\Omega)$.

2.1.7 Constructing Measure Spaces

We can use functions to construct measures on the *codomain* given that the *domain* is a measure space. The steps involved are first constructing a σ – algebra and then constructing a measure.

Proposition 2.33 Let (X, \mathcal{A}, μ) be a measure space, Y a set, and $f: X \rightarrow Y$ a function. We define that:

$$\mathcal{A}_1 = \{A \subseteq Y : f^{-1}[A] \in \mathcal{A}\}$$

and

$$\mu_f(A) = \mu(f^{-1}[A])$$

for all $A \in \mathcal{A}_1$. Then $(Y, \mathcal{A}_1, \mu_f)$ is a measure space.

Remark 2.34 So what we have done is that we first constructed a σ – algebra from another σ – algebra by using a function. From that, we then constructed a measure space by constructing a measure from another measure.

Remark 2.35 μ_f is called the **distribution of f with respect to μ** .

Remark 2.36 This notion of distribution is also in probability theory. A random variable is a **measurable** function between sample space $\Omega \rightarrow \mathbb{R}$.

2.1.8 σ – algebra with disjoint sets

Here, we want to set up σ – algebras with disjoint sets as the 3rd property does not require disjoint sets yet working with arbitrary sets and disjoint sets aren't the same thing. Hence, we need to add more properties for the σ – algebra. The reason we want to do this is that when we look at **Lebesgue measures**, it is easier to work with disjoint sets.

We come up with new ways of constructing σ – algebra.

Proposition 2.37 Suppose that \mathcal{A} is a collection of subsets of X with the following properties.

1. $\emptyset \in \mathcal{A}$.
2. If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.
3. If $A_k \in \mathcal{A}$, $k \in \mathbb{N}$ are disjoint, then $\bigcup_{k \in \mathbb{N}} A_k \in \mathcal{A}$.
4. If $A, B \in \mathcal{A}$, then $A \cap B \in \mathcal{A}$.

Then \mathcal{A} is a σ - algebra.

Proposition 2.38 If $A_k \in \mathcal{A}$ for all $k \in \mathbb{N}$, then

$$\bigcup_{j=0}^{\infty} \left(\bigcap_{k=j}^{\infty} A_k \right) \in \mathcal{A}.$$

Furthermore, we note that if $\sum_{k=0}^{\infty} \mu^*(A_k) < \infty$, then $\sum_{k=j}^{\infty} \mu^*(A_k) \rightarrow 0$ as $j \rightarrow \infty$ or else we will not have had a finite series.

MATH3969: Measure Theory and Fourier Analysis

3. Elementary properties of Measures

3.1.9 Properties of Measures

Proposition 3.39 *Let (X, \mathcal{A}, μ) be a **measure space**.*

- (i) *If $A, B \in \mathcal{A}$ and $A \cap B = \emptyset$, then $\mu(A + B) = \mu(A) + \mu(B)$;*
- (ii) *(Monotonicity of measures) If $A, B \in \mathcal{A}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$;*
- (iii) *(Countable Sub-additivity) If $A, A_k \in \mathcal{A}$, $k \in \mathbb{N}$, and $A = \bigcup_{k=0}^{\infty} A_k$, then*

$$\mu(A) \leq \sum_{k=0}^{\infty} \mu(A_k)$$

Remark 3.40 *The last property refers to the fact that this allows for overlaps of sets.*

Proposition 3.41 *(Monotone Continuity of Measures). Let (X, \mathcal{A}, μ) be a measurable space.*

- (i) *(Continuity From Below). If $A_i \in \mathcal{A}$, $k \in \mathbb{N}$ and $A_0 \subseteq A_1 \subseteq A_2 \subseteq \dots$, then:*

$$\lim_{k \rightarrow \infty} \mu(A_k) = \mu\left(\bigcup_{k=0}^{\infty} A_j\right)$$

- (ii) *(Continuity From Above). If $A_i \in \mathcal{A}$, $k \in \mathbb{N}$ and $A_0 \supseteq A_1 \supseteq A_2 \supseteq \dots$, **and** $\mu(A_k) < \infty$ for some $k \in \mathbb{N}$ then:*

$$\lim_{k \rightarrow \infty} \mu(A_k) = \mu\left(\bigcap_{k=0}^{\infty} A_j\right)$$

Remark 3.42 *We note that for (ii), we now have an additional condition of a finite measure. If not, condition (ii) does not necessarily hold, as we can find cases where this causes things to break. Additionally, we can just also have $\mu(A_0) < \infty$ as an equivalent condition. Note that this is a case of **nested sets**.*

3.1.10 Metric Spaces

We now define new objects called metric spaces.

Definition 3.43 *(Topology). Let X be a set and τ a collection of subsets of X , satisfying*

- (i) $\emptyset, X \in \tau$;
- (ii) *Any union of members of τ belongs to τ ;*
- (iii) *The intersection of any finite number of members of τ still belongs to τ .*

The elements of τ are called **open sets** and the collection τ is called a **topology** on X .

Remark 3.44 An ordered pair (X, τ) is called a **topological space**.

Definition 3.45 (Metric). A metric d on a set X is a function

$$d : X \times X \rightarrow [0, \infty),$$

where $[0, \infty)$ is the set of non-negative real numbers and for all $x, y, z \in X$, the following conditions are satisfied:

1. (Non-negativity) $d(x, y) \geq 0$;
2. (Identity of indiscernibles) $d(x, y) = 0 \leftrightarrow x = y$;
3. (Symmetry) $d(x, y) = d(y, x)$;
4. (Triangle inequality) $d(x, z) \leq d(x, y) + d(y, z)$

Remark 3.46 Every metric space is a topological space. A metric space has a notion of distance, while a topological space only has a notion of closeness. If we have a notion of distance then we can say when things are close to each other. However, distance is not necessary to determine when things are close to each other.

Definition 3.47 (Metric Space). A metric space is an ordered pair (M, d) where M is a set and d is a metric on M .

Definition 3.48 (Cover). If the set X is a topological space, then a cover C of X is a collection of sets G_α of X whose union contains the set X as a the whole space X .

Remark 3.49 In this case, if we had a cover C , we say that C covers X .

Definition 3.50 (Open Cover). Let E be a set and X be a metric space. An **open cover** of E in X , is a **collection of open sets** $\{G_\alpha\}$ whose union **contains** E

Definition 3.51 (Subcover). A **subcover** of $\{G_\alpha\}$ is a **subcollection** $\{G_{\alpha_\gamma}\}$ that stills cover E .

Definition 3.52 (Compact Set). A subset S of a topological space X is **compact** if for every open cover of S , there exists a finite subcover of S . That is, X is compact if for every collection C of open subsets of X such that

$$X = \bigcup_{x \in C} x,$$

there exists a finite subset F of C such that

$$X = \bigcup_{x \in F} x.$$

3.1.11 The Construction of measures from outer measures

Now we are interested in being able to get N -dimensional measures for arbitrary subsets of \mathbb{R}^N , not just σ -algebra. In particular, we can relax the condition of countable additivity and hence we can define a "outer measure" that is defined **on all** subsets of X , not just a σ -algebra which could be smaller. We first define a function:

$$\mu^* : \mathcal{P}(X) \rightarrow [0, \infty]$$

We first define this function and then find a large enough σ -algebra such that μ^* becomes a measure, in order to "maximize" the number of sets in which we can assign a measure too.

We define an **open** rectangular box to be

$$R = (a_1, b_1) \times (a_2, b_2) \times \dots \times (a_n, b_n) \subseteq \mathbb{R}^N$$

where $a_j \leq b_j$ for all j .

We then define the volume of an open rectangular box by

$$\text{vol}(R) = (b_1 - a_1) \times (b_2 - a_2) \times \dots \times (b_n - a_n).$$

Definition 3.53 (*Lebesgue Outer Measure*). For every subset $A \subseteq \mathbb{R}^N$, we set

$$m_N^*(A) := \inf \left\{ \sum_{k=0}^{\infty} \text{vol}(R_k) : R_k, k \in \mathbb{N}, \text{ open rectangles with } A \subseteq \bigcup_{k=0}^{\infty} R_k \right\}.$$

We call $m_N^*(A)$ the **Lebesgue outer measure** of the set A .

Remark 3.54 We look at the sum of the volume of rectangles, where when we union them, it covers the subset A . Here, the Lebesgue outer measure is simply the infimum of the sum of volume of rectangles. The Lebesgue outer measure is not actually a measure. It is the infimum of the volume of all possible coverings that contains A . It is not a measure as it does not satisfy the **countable additivity** property.

Remark 3.55 It is actually possible to replace the open rectangles with closed rectangles! It is equivalent.

Proposition 3.56 The Lebesgue outer measure $m_n^* : \mathcal{P}(\mathbb{R}^N) \rightarrow [0, \infty]$ has the following properties.

- (i) $m^*(\emptyset) = 0$;
- (ii) (*Countable Subadditivity*). For every **countable collection** $A, A_k \subseteq \mathbb{R}^N, k \in \mathbb{N}$ with $A \subseteq \bigcup_{k=0}^{\infty} A_k$

$$m^*(A) \leq \sum_{k=0}^{\infty} m^*(A_k);$$

- (iii) $\{x\}$ is Lebesgue outer-measurable;
- (iv) $m^*(\{x\}) = 0$ for all $x \in \mathbb{R}^N$;
- (v) $m^*(C) = 0$ for any countable collection c .

Proof: (Sketch).

- (i) Take the trivial rectangles to show this.
- (ii) Use the definition of the infimum over countably many open rectangles. Use the ϵ property of infimum to then prove the inequality. Note that we need to choose $\frac{\epsilon}{2^{k+1}}$ specifically to get things to work out nicely.

■

Remark 3.57 *The name "outer" refers to the fact that we are approximating the volume from the outside with countably many rectangles. Here, m_N^* is defined for all elements of the powerset.*

Note that we say that $A_k = \cup_{j=0}^{\infty} R_{k,j}$. So then we have that $\cup_{k=0}^{\infty} A_k = \cup_{k=0}^{\infty} \cup_{j=0}^{\infty} R_{k,j}$.

MATH3969: Measure Theory and Fourier Analysis

4. Outer Measures.

4.1.12 Exploring Outer Measures and their properties.

Currently, our goal is to be able to define measures for any arbitrary subset. We first looked at the famous Lebesgue outer measure in the previous section. Now we generalise to other outer measures. We eventually work our way to constructing σ – algebra such that if we restrict the outer measure to this σ – algebra, we get a measure. Sets that belong to this σ – algebra are known as **measurable**.

Definition 4.58 (*Outer Measure*). A function $\mu^*: \mathcal{P}(X) \rightarrow [0, \infty]$ is called an **outer measure** if it satisfies the properties

(i) $\mu^*(\emptyset) = 0$;

(ii) (*Countable Subadditivity*) For every countable collection $A, A_k \subseteq X, k \in \mathbb{N}$ with $A \subseteq \bigcup_{k=0}^{\infty} A_k$ we have that

$$\mu^*(A) \leq \sum_{k=0}^{\infty} \mu^*(A_k).$$

Note that the properties are similar to that of a measure except that the countable subadditivity is seen as a relaxed version of countable additivity.

Remark 4.59 (*Properties of outer measures*).

(i) In the special case that $A = \bigcup_{k=0}^{\infty} A_k$, we have

$$\mu^*\left(\bigcup_{k=0}^{\infty} A_k\right) \leq \sum_{k=0}^{\infty} \mu^*(A_k)$$

even if the sets A_k are disjoint.

(ii) Outer measures does not need to satisfy countable additivity even if the sets A_k are disjoint. Hence, it is a weaker condition.

(iii) If $A \subseteq B$, then

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

Use the fact that $A \subseteq B \cup \emptyset \cup \dots \cup \emptyset$ to prove this.

(iv) For $A, B \subseteq X$, we have that

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

Proposition 4.60 (*Monotonicity*). The Lebesgue outer measure has the property of monotonicity. That is, if $A \subseteq B$, then

$$m^*(A) \leq m^*(B).$$

Proof: We look at the set of rectangles that covers A and B. The set of rectangles that covers A is **bigger** than the set of rectangles covering B as A is smaller so more intervals can cover it. Hence, recall from analysis that if $A \subseteq B \implies \inf A \leq \inf B$. Resultantly, the infimum of the volume of rectangles covering B will be greater than or equal to the volume of rectangles covering A, this is the definition of the Lebesgue Outer measure. Hence,

$$m^*(A) \leq m^*(B).$$

■

4.1.13 Constructing σ – algebra For Measures.

We now construct condition for measures. Just like in the Riemann integral, you want the upper and lower sums to coincide, here, we can approximate a set from the outside and from the inside. If the measures coincide, then we have the actual measure. However, an issue is that having a finite outer measure in this scenario will imply an inner measure of ∞ as we are taking the measure over the entire space. So what we will do is "restrict the domain" to \mathcal{D} , which contains our set of interest A. We then approximate A inside this new domain. Let $A, S \subseteq X$ and furthermore, $A \subseteq S$ and S has finite measure.

We can now express the set S as a disjoint union with A.

$$S = (S \cap A) \dot{\cup} (S \cap A^c).$$

From this, we can apply the measure function to get

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

The inner measure of our set of interest A is the difference of $m^*(S) - \mu^*(S \cap A^c)$. The outer measure of our set of interest A is then just $\mu^*(S \cap A)$. The issue is that sometimes we may get weird stuff with the 2 terms on the right for a random set S. Hence, we look at every set $S \subseteq X$ where the above statement holds. This is useful as we now have an additive function. $\mu^*((S \cap A))$ is known as the **outer approximation** and $\mu^*(S) - \mu^*((S \cap A^c))$ is known as the **inner approximation** for S.

With the above properties, we can now construct σ – algebra on which we can then also define a measure. First we define one more term.

Definition 4.61 (μ^* -measurable Set). Let $\mu^* : \mathbb{P}(X) \rightarrow [0, \infty]$ be an **outer measure** on the set X. We call a set $A \subseteq X$ a μ^* -measurable set if

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

for all $S \subseteq X$. We denote the class of all measurable sets by \mathcal{A} .

With this, we come to one of the most important theorems.

Definition 4.62 (Carathéodory). Let $\mu^* : \mathbb{P}(X) \rightarrow [0, \infty]$ be an outer measure and let the family of sets \mathcal{A} be

$$\mathcal{A} := \{A \subseteq X : \mu^*(S) = \mu^*(S \cap A) + \mu^*(S \cap A^c) \text{ for all } S \subseteq X\}.$$

Then \mathcal{A} is a σ – algebra. Furthermore, $\mu := \mu^*|_{\mathcal{A}} : \mathcal{A} \rightarrow [0, \infty]$ is a **measure**.

Recall that a set A is a μ^* -measurable set if for all $S \subseteq X$, we have that $\mu^*(S) = \mu^*(S \cap A) + \mu^*(S \cap A^c)$. From that, we note as $S = (S \cap A) \cup (S \cap A^c)$ for all sets $A, S \subseteq X$. Recall if $A \subseteq B \implies \mu^*(A) \leq \mu^*(B)$. It follows that $\mu^*(S) \leq \mu^*(S \cap A) + \mu^*(S \cap A^c)$. Therefore, to show a set A is μ^* -measurable then it just requires us to show that $\mu^*(S) \geq \mu^*(S \cap A) + \mu^*(S \cap A^c)$ for all $S \subseteq X$.

Hence, we have just constructed a σ -algebra \mathcal{A} associated with μ^* or of μ^* -measurable sets. We have just constructed a σ -algebra and a measure from an outer measure!

MATH3969: Measure Theory and Fourier Analysis

5. μ^* -measurable sets and their properties.

5.1.14 μ^* -measurable set properties.

We now look at the properties of μ^* -measurable sets.

Proposition 5.63 (*Properties of μ^* -measurable sets and outer measures*). Let μ^* be an outer measure and \mathcal{A} be the set of μ^* -measurable sets. Then the following assertions are true.

- (i) If $\mu^*(A) = 0$, then $A \in \mathcal{A}$;
- (ii) If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$.
- (iii) If $A \in \mathcal{A}$, $B, S \subseteq X$ **and** $A \cap B = \emptyset$, then

$$\mu^*(S \cap (A \cup B)) = \mu^*(S \cap A) + \mu^*(S \cap B);$$

- (iv) If $A, B \in \mathcal{A}$, then $A \cup B \in \mathcal{A}$ **and** $A \cap B \in \mathcal{A}$.

Remark 5.64 *Some comments on each of the above property in the proposition.*

- (i) A is **measurable** if the outer-measure of A is 0;
- (ii) If A is measurable, then its compliment is also measurable;
- (iii) If **only** A is measurable and A is disjoint from B , then additivity holds for the **disjoint union**.
- (iv) If A and B are both measurable, then their union and interesection are measurable.

Refer to page 18-19 of the course notes for the proof.

With the above properties, we can show that **Carathéodory's** theorem constructs a σ - *algebra* and a measure after restricting the outer measure to that σ - *algebra*. We only need to show that the measure is actually a measure by showing the **countable additivity property** and that the σ - *algebra* is closed under **countable union**. You can read the proof in page 19 and 20 of the course notes.

5.1.15 Lebesgue Measure

From this, we can now define the **Lebesgue measure** which is the main result of the Carathéodory theorem.

Definition 5.65 (*Lebesgue Measure*). Let $m_N^*(A)$ be the **Lebesgue outer measure** for all subsets $A \subset \mathbb{R}^N$. We call

$$\mathcal{M}_N := \{A \subseteq \mathbb{R}^N : m_N^*(S) = m_N^*(S \cap A) + m_N^*(S \cap A^c) \text{ for all } S \subseteq \mathbb{R}^N\}$$

the **Lebesgue σ - algebra** and $m_n := m_n^*|_{\mathcal{M}_N}$ the N -dimensional **Lebesgue measure**. Sets in \mathcal{M}_N are called **Lebesgue measurable**.

Remark 5.66 *Alot of the properties we showed earlier also holds here. Examples include*

- (i) A set with Lebesgue measure 0 is Lebesgue measurable.*
- (ii) A set is Lebesgue measurable iff it's complement is Lebesgue measurable.*
- (iii) The empty set and the real line is Lebesgue measurable.*
- (iv) If the set is Lebesgue measurable, any translations of it is Lebesgue measurable.*
- (v) All Borel sets (open and closed sets) are Lebesgue measurable.*
- (vi) The union of a finite collection of Lebesgue measurable sets is Lebesgue measurable.*
- (vii) The union of a countable collection of Lebesgue measurable sets is Lebesgue measurable.*
- (viii) Singleton sets are Lebesgue measurable as every singleton set is a Borel set, whereby Borel sets are Lebesgue measurable.*

Proposition 5.67 *(Sets with outer measure 0 is measurable). A set $N \subset X$ with $\mu^*(N) = 0$ is measurable.*

Proof:

$$\mu^*(S) \leq \mu^*(S \cap N) + \mu^*(S \cap N^c)$$

is trivial. For the other inequality, note that

$$\mu^*(S \cap N) \leq \mu^*(N) = 0.$$

Hence, we have the other inequality of

$$\mu^*(S) \geq \mu^*(S \cap N) + \mu^*(S \cap N^c).$$

■

MATH3969: Measure Theory and Fourier Analysis

6. Properties of Lebesgue Measure.

6.1.16 Regularity of the Lebesgue Measure.

Theorem 6.68 *All Borel sets, in particular all open and closed sets, are Lebesgue Measurable.*

Recall that the Borel σ – algebra in \mathbb{R}^N is the smallest σ – algebra containing all open sets.

Theorem 6.69 *The Lebesgue Measure is a Borel Measure, that is, $\mathcal{B}_N \subseteq \mathcal{M}_N$.*

We now look at the regularity properties of the Lebesgue measure. The reason we do so is that we can move away from using the infimum of volume of rectangle coverings to using open and closed sets to define our Lebesgue measure. This is much easier to work with.

Regularity properties of measures refers to the possibility of approximating in measure one class of sets by another class of sets. So in the Lebesgue measure, for the outer regularity, we look at approximating arbitrary sets from the outside by open sets whilst for the inner regularity, we look at approximating Lebesgue Measurable sets by compact sets.

A **regular measure** on a topological space is a measure for which **every** measurable set can be approximated from above by **open** sets and from below by **compact and measurable** sets. To recall the definition of a compact set, we say that a set S is a **compact subset** of X if for every open cover of S , there exists a finite subcover of S . A cover \mathbb{C} of X is a collection of sets of X , where X is contained in the union of the collection of sets in the cover. A subcover is a subset of \mathbb{C} that covers X . Finally, an open cover is when every set in the cover is an open set, or in other words, belongs to the topology on X .

Proposition 6.70 (*Translation invariance of the Lebesgue Measure*). *For every $t \in \mathbb{R}^N$ and every $A \subseteq \mathbb{R}^N$, we have that $m_N^*(t + A) = m_N^*(A)$.*

Proof: The proof follows directly by the fact that the Lebesgue outer measure is simply the volume of the open rectangle whose union covers A and the fact that the volume of a rectangle is translation invariant for every triangle. In other words, the Lebesgue outer measure is the infimum of the sum of volume of rectangles. We note that $t + A \subseteq \bigcup_{k=1}^{\infty} (t + R_k)$ if and only if $A \subseteq \bigcup_{k=1}^{\infty} (R_k)$. There is a 1-1 correspondence of rectangles covering in both cases and hence, $vol(R_k) = vol(t + R_k)$ and therefore $m^*(A) = m^*(A + t)$. ■

We next want to show that the Lebesgue measure is unique. We look at the **regularity** of the Lebesgue measure to help us prove more properties.

Proposition 6.71 (*Outer Regularity of Lebesgue Measure*). *Let $A \subseteq \mathbb{R}^N$ be an **arbitrary set**. The outer regularity is defined to be*

$$m_N^*(A) = \inf\{m_N(U) : A \subseteq U, U \text{ is open}\}.$$

Proof: (Sketch). We assume $m_N^*(A) < \infty$. By monotonicity of outer measures, $m_N^*(A) < m_N^*(U)$ for all open sets U as $A \subseteq U$. We need to show that for every $\epsilon > 0$, there exists an open set U with $A \subseteq U$ such that $m_N^*(A) + \epsilon \geq m_N(U)$, as this is a definition of an infimum. ■

Proposition 6.72 (*Inner Regularity of Lebesgue Measure*). Let $A \subseteq \mathbb{R}^N$ be an **Lebesgue Measurable set**. The **inner regularity** is defined to be

$$m_N(A) = \sup\{m_N(K) : K \subseteq A, K \text{ is compact}\}.$$

Remark 6.73 Recall that compact sets in \mathbb{R}^N implies closed and bounded. Compact sets have the additional advantage that the measure is always finite even if that of A is not finite.

Notice that we have m_N in the outer regularity. This is because U is open and therefore contained in \mathcal{M}_N . Hence the Lebesgue outer measure and Lebesgue measure coincide.

Definition 6.74 (*Regular Borel Measure*). A Borel measure on a metric space X is called outer regular if for every Borel set $A \subseteq X$

$$\mu(A) = \inf\{\mu(U) : A \subset U, U \text{ is open}\}.$$

The measure is called inner regular if

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

Notice again we don't have an outer measure since the Borel outer measure coincide with the Borel measure on open and compact sets. We also have that every Lebesgue measurable set is between two Borel set of the same measure.

Remark 6.75 If a set A is Lebesgue measurable, there exists a Borel set $C \subseteq A$ and $m(A) = m(C)$.

6.1.17 Uniqueness of the Lebesgue Measure.

We now work our way towards proving uniqueness of the Lebesgue Measure, that is, any regular Borel measure that is translation invariant measure on unit cubes is THE Lebesgue Measure.

We use the fact that every open set can be written as a countable union of disjoint cubes.

Definition 6.76 (*Cube in \mathbb{R}^N*). A cube in \mathbb{R}^N is simply a set of the form

$$Q = I \times I \times \dots \times I = I^N.$$

In particular, it is the cross product of N intervals.

In terms of intuition, imagine we want to decompose the plane into disjoint cubes. We can do this by taking unit squares where the bottom and left corner of the squares are open and the top and right are closed. This will cover each boundary exactly **once**. In particular, we denote the set of collection of cubes as

$$\mathbb{G}_n = \left\{ \frac{k}{2^n} + (0, \frac{1}{2^n}] : k \in \mathbb{Z}^N \right\};$$

where $(0, \frac{1}{2^n}]$ is the interval whilst $\frac{k}{2^n}$ is the translation of the interval. We note the fact that every cube in \mathbb{G}_n has the form

$$Q_{n,k} = \left(\frac{k_1}{2^n}, \frac{k_1+1}{2^n} \right] \times \left(\frac{k_2}{2^n}, \frac{k_2+1}{2^n} \right] \times \dots \times \left(\frac{k_N}{2^n}, \frac{k_N+1}{2^n} \right]$$

where every cube is simply the product of multiple unit cubes. These cubes $Q_{n,k}$ are called **Dyadic cubes**. Hence, each cube in one scale can be written as an union of cubes of a smaller scale.

It is clear that *every open set can be written as a countable union of disjoint cubes*, that is

$$\mathbb{R}^N = \dot{\bigcup}_{k \in \mathbb{Z}^N} Q_{n,k}.$$

Proposition 6.77 (*Dyadic Decomposition*). *Let $U \subseteq \mathbb{R}^N$ be an open set. Then there exists countably many disjoint dyadic cubes $Q_j \in \bigcup_{n \in \mathbb{N}} \mathbb{G}_n$ such that $U = \bigcup_{j=0}^{\infty} Q_j$.*

Proof: (Sketch). We can prove this through induction by constructing cubes and at each value of n , we collect all cubes that are inside the set but have not been in any of the $n-1$ collections. Due to the fact that it is an open set, for each $x \in U$, there exists a ball $B(x, \epsilon)$ which is inside the set U and hence, we can fit a cube into that ball to contain x . Hence, each $x \in U$ is contained in a cube and therefore the union $\mathbb{C} = \bigcup_{n \in \mathbb{N}} C_n = U$, which each cube being at most countable and hence the union is countable as well. ■

Using dyadic decompositions, we can compute the measure of any open set of any translation invariant Borel measure in terms of the Lebesgue measure.

Lemma 6.78 *Let μ be a translation invariant Borel measure on \mathbb{R}^N . If $Q_{n,k}$ is the dyadic cube and we define $\alpha := \mu((0, 1]^N) < \infty$ to be the measure on unit cube, then*

$$\mu(Q_{n,k}) = \alpha m_N(Q_{n,k}) = \alpha 2^{-nN}$$

for all $k \in \mathbb{Z}^N$ and all $n \in \mathbb{N}$. This is the measure of any dyadic cube.

Remark 6.79 α is defined to be the measure of the unit cube, which is the first cube we try to fit the shape with. Then, to find the measure of the dyadic cube, we divide the **Lebesgue measure** of the unit cube by the number of dyadic cubes there are which is 2^{-nN} .

Recall that the dyadic cube is

$$Q_{n,k} = \left(\frac{k_1}{2^n}, \frac{k_1+1}{2^n}\right] \times \left(\frac{k_2}{2^n}, \frac{k_2+1}{2^n}\right] \times \dots \times \left(\frac{k_N}{2^n}, \frac{k_N+1}{2^n}\right]$$

Proof: First, we showed that the cube at the origin has a measure. Then, all other unit cubes are translations of that cube and due to the fact of translation invariance, they all have the same measure. Hence, to know the total measure, we just need to know how many cubes are inside. The number of cubes in $(0, 1]^n$ of side length $\frac{1}{2^n}$ is 2^n for each dimension. Therefore, for N dimensions, we have $(2^n)^N$ cubes in total. ■

MATH3969: Measure Theory and Fourier Analysis

7. Lebesgue Measure and Linear Transformations

7.1.18 Uniqueness of Lebesgue Measure.

Here, we seek to prove that there is a unique outer regular translation invariant Borel measure on \mathbb{R}^N such that the unit cube has measure one.

Theorem 7.80 (*Uniqueness Theorem for Lebesgue Measure*). *Let μ be an outer regular Borel measure on \mathbb{R}^N . If μ is translation invariant and $\alpha := \mu((0, 1]^N) < \infty$, then*

$$\mu(A) = \alpha m_N(A)$$

for all Borel sets $A \subseteq \mathbb{R}^N$. In particular, there is a unique outer regular translation invariant Borel measure on \mathbb{R}^N such that the unit cube has measure one.

This states that the Lebesgue measure is the unique (up to multiplication by a constant) translation invariant Borel measure.

7.1.19 Linear Transformation of Lebesgue Measure.

We want to show how the Lebesgue measure behaves under any linear transformations. Additionally, the Lebesgue measure is not only translation invariant but invariant to any rigid or orthogonal transformation.

Theorem 7.81 *Let $T : \mathbb{R}^N \rightarrow \mathbb{R}^N$ be a linear transformation. Then*

$$m_N^*(T(A)) = |\det T| m_N^*(A)$$

for every (arbitrary) $A \subseteq \mathbb{R}^N$. Moreover, if A is Lebesgue measurable, then $T(A)$ is Lebesgue measurable as well.

In otherwords, linear transformations maps measurable sets to measurable sets.

7.1.20 The Lebesgue-Stieltjes Measure

We are interested in showing the connection between the class of right continuous increasing functions on \mathbb{R} and the class of outer regular Borel measures on \mathbb{R} that are finite on bounded sets. There is a 1-1 correspondence between such functions and measures. To show that every regular Borel measure has a right continuous increasing function, we look at the **distribution function**. The distribution function of μ is defined to be $F(t) := \mu((-\infty, t])$, where μ is the finite Borel measure. It satisfies all the properties of distribution functions in probability theory.

Now we want to see whether every right continuous increasing function on \mathbb{R} is associated with a Borel measure on \mathbb{R} .

Theorem 7.82 (*Lebesgue-Stieltjes Measure*). *There is 1-1 correspondence between right continuous increasing functions and regular Borel measures on \mathbb{R} .*

For every right continuous increasing function on \mathbb{R} , there exists a regular Borel measure such that $\mu((a, b]) = F(b) - F(a)$ for $a \leq b$.

The measure μ_F that is induced by the function F , is called the **Lebesgue-Stieltjes measure**. We define $v_F((a, b]) := F(b) - F(a)$ on half open intervals. From that, we define the outer measure

$$\mu_F^*(A) := \inf \left\{ \sum_{k=0}^{\infty} v_F((a_k, b_k)) : A \subseteq \bigcup_{k=0}^{\infty} (a_k, b_k) \right\}.$$

Then we can construct a σ - algebra from Carathéodory theorem. Finally, we can then restrict our outer measure μ_F^* on this new σ - algebra to get the Lebesgue-Stieltjes measure μ_F . We can show this is a outer regular Borel measure.

MATH3969: Measure Theory and Fourier Analysis

8. Measurable Functions.

8.2 Measurable Functions

8.2.1 Measurable Functions.

Here, we introduce the idea of describing properties of functions by classes of sets. In particular, we use open sets to describe the properties of continuous functions. A function is a **continuous function** iff the preimage of every open set is an open set.

Definition 8.83 (*Measurable Function*). Let (X, \mathcal{A}, μ) be a measure space and Y a metric space. We call a function $f: X \rightarrow Y$ a μ -measurable function if $f^{-1}[U] \in \mathcal{A}$ for **all** open sets $U \subseteq Y$.

Remark 8.84 Notice that we use open sets rather than measurable sets in the definition. The reason for doing so is that when we seek to prove things, open sets are a much smaller class and hence we have less cases to consider in the proof.

If X is a metric space and μ a Borel measure, then every continuous function $f: X \rightarrow Y$ is measurable. \mathcal{A} is a Borel σ -algebra and hence contains every open set. Therefore, every open set, which contains the preimage of the function, is in the σ -algebra and hence it is a measurable function.

Hence, if we had a Borel σ -algebra on the target space, then it follows that we only need to check every open set in Y is measurable, hence, any continuous function is automatically measurable.

Proposition 8.85 Let (X, \mathcal{A}, μ) be a measure space, Y a topological space and $f: X \rightarrow Y$ is measurable. Then $f^{-1}[A] \in \mathcal{A}$ for all Borel sets $A \subseteq Y$.

Proof: The set $\{A \subseteq Y : f^{-1}[A] \in \mathcal{A}\}$ is a σ -algebra in Y . As f is a measurable function, the preimage of every open set in Y is in \mathcal{A} and hence, the σ -algebra constructed on Y contains all open sets. This σ -algebra is a superset of the Borel σ -algebra and hence, all Borel sets are measurable as it belongs to the σ -algebra of Y , which contains all measurable open sets whose preimage belongs to \mathcal{A} . This is because the Borel σ -algebra \mathcal{B} is the smallest σ -algebra containing all open sets. ■

Definition 8.86 (*Indicator Function*). Let X be a set and $A \subseteq X$ a subset. The function $1_A: X \rightarrow \mathbb{R}$ is given by

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \in X - A \end{cases}$$

This function is called the **indicator function** of A .

Proposition 8.87 Let (X, \mathcal{A}, μ) be a measure space and $A \subseteq X$. Then 1_A is a measurable function iff A is a measurable set.

Proof: Let $U \subseteq \mathbb{R}$ be open. Then from the definition of 1_A

$$1_A^{-1}(U) = \begin{cases} \emptyset & \text{if } 0, 1 \notin U \\ A & \text{if } 1 \in U \text{ and } 0 \notin U \\ A^c & \text{if } 0 \in U \text{ and } 1 \notin U \\ X & \text{if } 0, 1 \in U. \end{cases}$$

Hence 1_A is measurable iff A is measurable. ■

When we check measurability of a function, we want to test whether is $f^{-1}[A]$ measurable for as few sets as possible. To save time, when testing real valued functions, we can use countably many half lines to check measurability. This is often taken for definition for measurability in other books but for us, this means that we do not require functions to be real valued.

Theorem 8.88 (*Measurability of Real Valued Functions*). Let (X, \mathcal{A}, μ) be a measure space and $f : X \rightarrow [-\infty, \infty]$ be a function. We define the Borel σ -algebra on \mathbb{R} by $\mathcal{B}(\mathbb{R})$. Then the following assertions are equivalent:

- (i) f is measurable, $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}(\mathbb{R})$;
- (ii) $f^{-1}[(\alpha, \infty]]$ is measurable for all $\alpha \in \mathbb{Q}$;
- (iii) $f^{-1}[[\alpha, \infty]]$ is measurable for all $\alpha \in \mathbb{Q}$;
- (iv) $f^{-1}[[-\infty, \alpha]]$ is measurable for all $\alpha \in \mathbb{Q}$;
- (v) $f^{-1}[[-\infty, \alpha]]$ is measurable for all $\alpha \in \mathbb{Q}$.

We can also replace \mathbb{Q} with \mathbb{R} .

Proof: (i) \Rightarrow (ii) – (v).

All these half intervals are **Borel sets** as closed sets, open sets, and half open sets (which can be expressed as countable union of open sets) are Borel sets. As stated earlier, all preimages of Borel sets $A \subseteq Y$ are in \mathcal{A} for a measurable function f from the Borel σ -algebra induced on Y by the measurable function f .

Now we note that (ii) \Rightarrow (iii), (iv), and (v).

(ii) \Rightarrow (iii).

$$[\alpha, \infty] = \bigcap_{n=1}^{\infty} (\alpha - \frac{1}{n}, \infty),$$

$$f^{-1}([\alpha, \infty]) = \bigcap_{n=1}^{\infty} f^{-1}[(\alpha - \frac{1}{n}, \infty)],$$

where by assumption (ii), this half open interval is measurable.

(ii) \Rightarrow (iv). Use $f^{-1}[(\alpha, \infty)]^c = f^{-1}[(\alpha, \infty]^c$. This then means that (iii) \Rightarrow (v).

(ii) – (v) \Rightarrow (i). Hence, if we assume (ii) is true, then (iii)–(v) are also true. First, we note that

$$f^{-1}[(a, b]] = f^{-1}[(a, \infty]] \cap f^{-1}[[-\infty, b]]$$

and by (ii), (v), this is in \mathcal{A} .

Now, we choose an open set $U \in \mathbb{R}$ which we can be expressed as disjoint union of dyadic intervals $(a_k, b_k]$. Hence,

$$\begin{aligned} f^{-1}[U] &= f^{-1}\left[\bigcup_{k \in \mathbb{N}} (a_k, b_k]\right] \\ &= \bigcup_{k \in \mathbb{N}} f^{-1}[(a_k, b_k]] \end{aligned}$$

where the preimage of each dyadic interval is in \mathcal{A} . Hence, the countable union of measurable sets are measurable through induction. ■

Furthermore, for \mathbb{R} valued function, recall that Borel σ - algebra can be generated by intervals. Hence, an alternative proposition can be formulated as

Theorem 8.89 (Real Valued Functions). *Let (X, \mathcal{A}, μ) be a measure space and $f : X \rightarrow [-\infty, \infty]$ be a function. We define the Borel σ - algebra on \mathbb{R} by $\mathcal{B}(\mathbb{R})$. Then the following assertions are equivalent:*

- (i) f is measurable, $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}(\mathbb{R})$;
- (ii) $\{x \in X : f(x) < b\} \in \mathcal{A}$ is measurable for all $b \in \mathbb{Q}$;
- (iii) $\{x \in X : f(x) \leq b\} \in \mathcal{A}$ is measurable for all $b \in \mathbb{Q}$;
- (iv) $\{x \in X : f(x) > b\} \in \mathcal{A}$ is measurable for all $b \in \mathbb{Q}$;
- (v) $\{x \in X : f(x) \geq b\} \in \mathcal{A}$ is measurable for all $b \in \mathbb{Q}$;

We can also replace \mathbb{Q} with \mathbb{R} .

Proof: (Sketch). Note that $\{x \in X : f(x) < b\} \in \mathcal{A} = f^{-1}((-\infty, b))$. We know that this interval is in the Borel σ - algebra of the target space and then we need to check if these intervals are in the $\mathcal{B}(\mathbb{R})$, which we know to be the case. ■

Proposition 8.90 *Let $\mu : \mathcal{A} \rightarrow [0, \infty]$ be a measure on X . Suppose $f = (f_1, f_2, \dots, f_n) \rightarrow \mathbb{R}^N$, then $f^{-1}[A_1 \times A_2 \times \dots \times A_N] = \bigcap_{k=1}^N f_k^{-1}[A_k]$.*

Proof: Any preimage x of the product of sets means that $f_k(x)$ is in each set A_k . Hence, that means x is in the intersection of component preimages. ■

Proposition 8.91 *Let (X, \mathcal{A}, μ) be a measure space and $f = (f_1, \dots, f_N) : X \rightarrow \mathbb{K}^N$ a function. Then f is measurable iff every component function $f_k : X \rightarrow \mathbb{K}$ is measurable.*

Proof: \Rightarrow Let every set A_k be \mathbb{R} except for one being an open set. Use the previous proposition to then show that f_k is measurable.

\Leftarrow Express an open set U as a disjoint countable union of dyadic cubes. Use the previous proposition to arrive at $U_{j=1}^\infty f^{-1}[Q_j]$. We can then use the previous proposition to express the preimage of each dyadic cube as the finite intersection of pre-images of intervals for each component function.

$$U_{j=1}^\infty f^{-1}[\bigcap_{k=1}^n f_k^{-1}[I_{j,k}]]$$

The intersection of measurable functions is measurable, hence f is measurable. ■

Proposition 8.92 (*Composition of a continuous and measurable function*). Let (X, \mathcal{A}, μ) be a measure space and Y, Z metric spaces. If $f : X \rightarrow Y$ is **measurable** and $\phi : Y \rightarrow Z$ **continuous**, then $\phi \circ f : X \rightarrow Z$ is measurable.

Proof: If U is an open set in Z , then $\phi^{-1}[U]$ is an open set in Y since ϕ is continuous. Then, as f is measurable, then $f^{-1}[\phi^{-1}[U]]$ will be in \mathcal{A} . Hence, the composition of $\phi \circ f$ is measurable. ■

Compositions of measurable functions are not necessarily measurable, which means that the preimage of an open set is not necessarily an open set, but the composition of continuous functions are continuous.

Theorem 8.93 (*Real Valued Function Operations*). Let (X, \mathcal{A}, μ) be a measure space and $f, g : X \rightarrow [\infty, \infty]$ measurable. Then the following functions are measurable as well

(i) $f+g$ is measurable;

(ii) fg is measurable;

(iii) $\frac{f}{g}$ is measurable;

(iv) $|f|$, $\max\{f, g\}$ and $\min\{f, g\}$ are measurable.

Proof: We look at the composition of a continuous and a measurable function. First we map from $X \rightarrow \mathbb{R}^2$ and then map from $\mathbb{R}^2 \rightarrow \mathbb{R}$. The first mapping is measurable as the two component functions are measurable whilst the second mapping is continuous, hence the composition of functions is measurable. For the case of addition, we have

$$X \rightarrow (f(x), g(x)) \rightarrow f(x) + g(x)$$

The first mapping, recall that a vector valued function is measurable iff each of its component is measurable, which is the case here. The second mapping is a continuous mapping. Hence, by the composition of a measurable and continuous mapping, we have a measurable composition. We repeat this argument for argument for multiplication division, and modulus too. For the max/min, we define

$$\max/\min\{f, g\} = \frac{1}{2}(f + g \pm |f - g|).$$
■

MATH3969: Measure Theory and Fourier Analysis

9. Sequences of Measurable Functions.

9.2.2 Convergence of sequences of functions.

We revisit concepts from analysis.

Definition 9.94 (*Pointwise Convergence*). The sequence of functions $\{f_n\}$ converges pointwise on X to a function f if for each $x \in X$ **and** $\epsilon > 0$, there is a $N \in \mathbb{N}$ such that

$$|f_n - f| < \epsilon$$

Here, N depends on x **and** ϵ , where we fix both of them. N changes if we change x .

Definition 9.95 (*Uniform Convergence*). The sequence of functions $\{f_n\}$ converges uniformly on X to a function f if for each $\epsilon > 0$, there is a $N \in \mathbb{N}$ such that

$$|f_n - f| < \epsilon \text{ for all } x \in X.$$

Here, N depends only on ϵ , where we now only fix ϵ . Hence, we can plug in any x into our function and this statement holds. N does not change if we change x .

Note that **uniform convergence implies pointwise convergence** but the converse is not true.

9.2.3 Sequences of measurable functions

Pointwise limits of measurable functions are measurable, unlike the case for continuous functions. Measurability is a "weaker" condition than continuity.

Proposition 9.96 Let (X, \mathcal{A}, μ) be a measure space and $f_n : X \rightarrow [-\infty, \infty]$ be measurable for all $n \in \mathbb{N}$. For $x \in X$ define 2 new functions $g(x)$ and $h(x)$ where

$$g(x) := \inf_{n \in \mathbb{N}} f_n(x) = \inf\{f_0(x), f_1(x), f_2(x), \dots\}$$

$$h(x) := \sup_{n \in \mathbb{N}} f_n(x) = \sup\{f_0(x), f_1(x), f_2(x), \dots\}$$

Then g and h are **measurable**.

Proof: Recall that to show a function is measurable, first we fix $\alpha \in \mathbb{R}$ and check if $g(x) \geq \alpha$. As $g(x)$ is the infimum (the smallest function), this implies that $f_n(x) \geq \alpha$ for all $n \in \mathbb{N}$. Hence, all the f_n are measurable.

To prove the converse statement, if $f_n(x) \geq \alpha$ for all $n \in \mathbb{N}$, then $g(x) \leq \alpha$ by definition of infimum. Therefore,

$$g^{-1}[[\alpha, \infty]] = \bigcap_{n \in \mathbb{N}} f_n^{-1}[[\alpha, \infty]]$$

which is measurable for all $n \in \mathbb{N}$ as the intersection of f_n is measurable.

We repeat the argument again as $h(x) = -\inf_{n \in \mathbb{N}}(-f_n(x)) = \sup_{n \in \mathbb{N}}(f_n(x))$. Hence, h is measurable too. ■

Here, we constructed a new function which is the infimum or supremum of the sequence of measurable function. This function is itself measurable.

Recall that the limit inferior takes the infimum of an **increasing** sequence whilst the limit superior takes the supremum of an **decreasing** sequence. The limit only exists if the limit superior = limit inferior.

However, as $u_n(x)$ is increasing and $v_n(x)$ is decreasing, we have

$$u(x) = \lim_{n \rightarrow \infty} \inf f_n(x) = \lim_{n \rightarrow \infty} (\inf_{k \geq n} f_n) = \sup_{n \in \mathbb{N}} (u_n(x)).$$

$$v(x) = \lim_{n \rightarrow \infty} \sup f_n(x) = \lim_{n \rightarrow \infty} (\sup_{k \geq n} f_n) = \inf_{n \in \mathbb{N}} (v_n(x)).$$

Hence, since we argued that the supremum/infimum of a sequence of measurable function is itself measurable, this shows that clearly measurability is preserved under pointwise convergence when the limit superior = limit inferior = limit.

Theorem 9.97 *Let (X, \mathcal{A}, μ) be a measure space and let $f_n : X \rightarrow [-\infty, \infty]$ be measurable functions. Then the functions u and v are given by*

$$u(x) = \lim_{n \rightarrow \infty} \inf f_n(x).$$

$$v(x) = \lim_{n \rightarrow \infty} \sup f_n(x).$$

*u and v are **measurable functions**. This includes $\lim_{n \rightarrow \infty} f_n(x)$ if it exists, that is, if $f_n \rightarrow f$ pointwise, then f is measurable..*

Proof: Recall that

$$u(x) = \lim_{n \rightarrow \infty} \inf f_n(x) = \sup_{n \in \mathbb{N}} (u_n(x)).$$

$$v(x) = \lim_{n \rightarrow \infty} \sup f_n(x) = \inf_{n \in \mathbb{N}} (v_n(x)).$$

Recall u_n and v_n are the infimum/supremum of the sequence of measurable functions. By the previous proposition, the *infimum/supremum* of a sequence of measurable functions is measurable itself, hence each u_n and v_n in our sequence is measurable. Using the proposition again as u/v is the limit superior/limit inferior, then u and v are respectively measurable.

Finally, if $f_n \rightarrow f$ pointwise, then we know that $u = v = f$, hence f is measurable. ■

MATH3969: Measure Theory and Fourier Analysis

10. Simple Measurable Functions.

10.2.4 Simple Measurable Functions.

We look at functions that attain only finite many values.

Definition 10.98 (*Simple Function*). Let X, Y be sets and $f : X \rightarrow Y$ be finite. We call f a **simple function** if its range is a finite set $\{a_0, a_1, \dots, a_n\}$.

A simple function that is real or complex valued can be written in the canonical representation of being a linear combination of indicator functions.

Definition 10.99 (*Canonical Representation of simple functions*). Let $f : X \rightarrow \mathbb{K}$. f is simple if and only if there exists $\alpha_k \in \mathbb{K}$ and **disjoint** subsets $A_k \subseteq X$, $k=0, \dots, n$ such that

$$f = \sum_{k=0}^n \alpha_k 1_{A_k}$$

where 1_{A_k} is the indicator function of A_k and α_k is the range of the function. So each set A_i uniquely maps to a value α_i and if the value $x \in X$ belongs to the set A_i , then it gets mapped to α_i .

Remark 10.100 We define the sets $A_k := f^{-1}[\{a_k\}]$. Note that we generate disjoint sets A_k as $f(A_j) = \alpha$ and $f(A_j) = \beta$ is impossible for a function if $\alpha \neq \beta$. Hence, A_k are disjoint sets.

We note that a linear combination of simple functions is also a simple function.

Proposition 10.101 For a measure space (X, \mathcal{A}, μ) and metric space Y , the function ϕ is simple **and** measurable if and only if ϕ has a canonical representation with the sets $A_k \in \mathcal{A}$ for all $k = 0, \dots, n$.

By convention, we tend to set $A_0 := f^{-1}[\{0\}]$ and hence write from the index 1

$$\sum_{k=1}^n \alpha_k 1_{A_k}.$$

Even if A_k are not disjoint, we can construct disjoint unions A_k by intersecting them with nested disjoint sets B_k (which are measurable if A_k is measurable).

Construct the disjoint sets $B_k = A_k \cap (A_1 \cup \dots \cup A_{k-1})^c$. Now, we can construct disjoint A_k 's by

$$A_k = \bigcup_{j=0}^n A_k \cap B_j$$

which is now a disjoint union. From this, we get the new indicator function

$$1_{A_k} = \sum_{j=0}^n 1_{A_k \cap B_j}$$

where it is a disjoint union of A_k with all B_j .

Hence, even if A_k is not disjoint, we can still express simple functions in their canonical form as

$$\sum_{k=0}^n \alpha_k 1_{A_k} = \sum_{k=0}^n \sum_{j=0}^n \alpha_k 1_{A_k \cap B_j}.$$

for disjoint unions of B_j .

Proposition 10.102 *Suppose that $f, g : X \rightarrow \mathbb{C}$ are simple measurable functions. Then the following functions are also simple measurable functions*

- (i) $f+g$;
- (ii) αf for all $\alpha \in \mathbb{K}$;
- (iii) fg ;
- (iv) $\frac{f}{g}$ if $g(x) \neq 0$ for all $x \in X$.

Proof: We can express the following

- (i) $f+g = \sum_{k=0}^n \sum_{j=0}^m (\alpha_k + \beta_j) 1_{A_k \cap B_j}$,
- (ii) $fg = \sum_{k=0}^n \sum_{j=0}^m (\alpha_k \beta_j) 1_{A_k \cap B_j}$,
- (iii) $\frac{f}{g} = \sum_{k=0}^n \sum_{j=0}^m (\frac{\alpha_k}{\beta_j}) 1_{A_k \cap B_j}$,

■

10.2.5 Non-negative Measurable Functions approximation by simple functions.

This section shows that every measurable function can be approximated from below by simple measurable functions. Looking at the big picture here to build up to the Lebesgue integral, we have a function f , which we can approximate by a sequence of simple functions. We then define the Lebesgue integral on each of these simple functions and in doing so, this gives us the Lebesgue integral on the original function f .

Theorem 10.103 *Let (X, \mathcal{A}, μ) be a measure space and $f : X \rightarrow [0, \infty]$ be a measurable function. Then there exists a sequence of simple measurable function $\phi : X \rightarrow [0, \infty)$ such that*

$$0 \leq \phi_n(x) \leq \phi_{n+1}(x) \leq f$$

for all $n \in \mathbb{N}$ and all $x \in X$. Moreover, $\phi_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$ for all $x \in X$, that is, $\phi_n \rightarrow f$ pointwise.

Remark 10.104 *This theorem tells us that for any non-negative measurable function on a measure space, we can approximate it by a sequence of finite simple measurable functions that converges pointwise to the function f , i.e. $\lim_{n \rightarrow \infty} \phi_n(x) = f(x)$ for all $x \in X$.*

We can generalise from non-negative functions to functions mapping to \mathbb{K} .

Corollary 10.105 (*Simple Approximation Lemma*). Let (X, \mathcal{A}, μ) be a measure space and $f : X \rightarrow \mathbb{K}$ be measurable. Then there exists simple measurable functions $\phi_n : X \rightarrow \mathbb{K}$ such that

$$0 \leq |\phi_1(x)| \leq |\phi_2(x)| \leq \dots \leq |f(x)|$$

for all $n \in \mathbb{N}$ and all $x \in X$. Moreover, $\phi_n(x) \rightarrow f(x)$ pointwise.

Note that we don't have monotone property for complex numbers, hence why we take the modulus.

Proof: We look at the first case of: $f : X \rightarrow \mathbb{R}$

We split X into X^+ and X^- where the former is the set of x that causes $f(\cdot)$ to be positive and the other negative. We can then use the fact that there exists a sequence of simple functions that converges pointwise to f^+ and $-f^-$ for any non-negative function. From that, we have

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

which now can either be

$$\begin{cases} f^+ & \text{if } x \in X^+ \\ f^- & \text{if } x \in X^- \end{cases}$$

where in each case, it converges to f as $n \rightarrow \infty$.

Now we look at the case where: $f : X \rightarrow \mathbb{C}$

We split it up into the real and complex components. In particular, we let $f = u + iv$, where $u, v : X \rightarrow \mathbb{R}$ and we can use the argument again that there exists sequences of simple measurable functions that approximates them. We get $\Phi_n(x) = \phi_n(x) + i\psi(x)$. Here, $\Phi_n(x)$ is a complex-valued simple function. Finally, $\Phi_n(x) \rightarrow f(x)$ for each fixed $x \in X$. ■

MATH3969: Measure Theory and Fourier Analysis

11. Integration and properties

11.3 Integration and properties

11.3.1 Integration

Recall that for a function $f : [a, b] \rightarrow [0, \infty)$, the Riemann integral is defined to be the limit of Riemann sums. We choose a partition $a = x_0 < x_1 < \dots < x_n = b$ and intermediate points $x_i^* \in [x_{i-1}, x_i]$. The Riemann sums are defined to be

$$\sum_{k=1}^n f(x_k^*) \Delta x_k$$

where $\Delta x_k := x_k - x_{k-1}$. These can be expressed as a simple function of the form

$$\psi = \sum_{k=1}^n f(x_k^*) 1_{(x_{k-1}, x_k]}$$

which takes the value $f(x_k^*)$ on the interval $(x_{k-1}, x_k]$.

To construct the Lebesgue integral, we do the same except for 2 things.

1) We define the general simple function

$$\psi = \sum_{k=1}^n \alpha_k 1_{A_k}$$

with A_k being disjoint arbitrary measurable sets. We can think of α as being simply the height whilst the measure of the base A will be given by $\mu(A)$.

2) We approximate the function from below and take the supremum, as every non-negative measurable function can be approximated from below. This is similar to taking the supremum of the lower Riemann sums.

11.3.2 Integration of non-negative simple functions.

Definition 11.106 (*Integral of simple measurable function*). Let (X, \mathcal{A}, μ) be a measure space and $\psi = \sum_{k=0}^{\infty} \alpha_k 1_{A_k}$ be a simple measurable function. We let

$$\int_X \psi d\mu := \sum_{k=0}^{\infty} \alpha_k \mu(A_k).$$

If $\mu(A_k) = \infty$, we set $\alpha_k \mu(A_k) = \infty$ if $\alpha_k > 0$ or 0 if $\alpha_k = 0$.

Proposition 11.107 Let (X, \mathcal{A}, μ) be a measure space and ϕ, ψ be simple measurable functions. Then the following assertions are valid.

$$(i) \int_X \psi + \phi d\mu = \int_X \psi d\mu + \int_X \phi d\mu;$$

(ii) $0 \leq \psi \leq \phi$, then $\int_X \psi d\mu + \int_X \phi d\mu$;

(iii) $\alpha \int_X \phi d\mu = \int_X \alpha \phi d\mu$ for all $\alpha \geq 0$ (since only non-negative functions means we can only use non-negative scalars);

(iv) If $N \in \mathcal{A}$ and $\mu(N) = 0$, then $\int_X \psi d\mu = \int_{X/N} \psi d\mu = \int_X 1_{X/N} \psi d\mu$.

Proof: We only prove (iv) as others are relatively trivial.

We can express $\phi = 1_{X/N} \phi + 1_N \phi$. From (ii), we have

$$\int_X \phi d\mu = \int_X 1_{X/N} \phi d\mu + \int_X 1_N \phi d\mu.$$

We look at the last term. $\phi = \sum_{k=0}^n \alpha_k 1_{A_k}$ and hence we can express $1_N \phi = \sum_{k=0}^n \alpha_k 1_N 1_{A_k} = \sum_{k=0}^n \alpha_k 1_{N \cap A_k}$ as A_k is disjoint and we can express it as a simple function with N by taking these disjoint intersections. Furthermore, from monotonicity of measures $\mu(A \cap N) \leq \mu(N) = 0$. From this, we get that the last term integrates to

$$\int_X 1_N \phi d\mu = \sum_{k=0}^n \alpha_k \mu(N \cap A_k) = \sum_{k=0}^n \alpha_k (0) = 0.$$

Hence, we have that

$$\int_X \phi d\mu = \int_X 1_{X/N} \phi d\mu.$$

■

Now we look at the integration of **non-negative** measurable functions.

Definition 11.108 (*Abstract Lebesgue Integral*). Let (X, \mathcal{A}, μ) be a measure space and $f : X \rightarrow [0, \infty]$ a measurable function. We then set

$$\int_X f d\mu = \sup \left\{ \int_X \phi d\mu : \phi \text{ is simple, measurable and } 0 \leq \phi \leq f \right\}.$$

Here, ϕ is a finite non-negative function.

With the Lebesgue integral, we are approximation it only from below with simple measurable functions. Note that from now on, when we talk about integrals \int , we are referring to the Abstract Lebesgue integral unless specified otherwise.

We note the special property that for non-negative measurable functions f, g where $0 \leq g \leq f$, we have that

$$\int_X g d\mu \leq \int_X f d\mu$$

Integrals "do not see" sets of measure 0, which means modifying functions on sets of measure zero does not change the value of the integral. Hence, we want to look at functions that hold almost everywhere.

Definition 11.109 (*Almost Everywhere*). Let (X, \mathcal{A}, μ) be a measure space and (P) some property. We say that (P) holds **almost everywhere** on X if there exists a measurable set N such that $\mu(N) = 0$ and (P) holds for all $x \in X - N$.

This is also known as pointwise property. So a property holds almost everywhere if the property holds after we remove all sets of measure 0. This is actually where the idea of **almost surely** from probability theory comes from!

We introduce a tool to help us prove properties of non-negative functions given information regarding their integrals.

Lemma 11.110 (*Markov's Inequality*). Suppose that $f : X \rightarrow [0, \infty]$ is measurable. Then for every $\alpha > 0$, we have that

$$\mu(f^{-1}[\alpha, \infty]) = \mu(\{x \in X : f(x) \geq \alpha\}) \leq \frac{1}{\alpha} \int_X f d\mu.$$

Proof: (Sketch). Construct the set of $x \in X : f(x) \geq \alpha$. This is measurable and we know the integral of w.r.t to this new set is less than integral with respect to the whole function. ■

Proposition 11.111 Suppose that $f : X \rightarrow [0, \infty]$ is measurable

(i) If $\int_X f d\mu < \infty$ then $f(x) < \infty$ almost everywhere, that is

$$\mu(\{x \in X : f(x) = \infty\}) = 0.$$

(ii) If $\int_X f d\mu = 0$ then $f(x) = 0$ almost everywhere, that is

$$\mu(\{x \in X : f(x) > 0\}) = 0.$$

Hence, if the integral of a non-negative function is zero, we cannot conclude that the function f is zero. We only know that the measure on those sets where f is positive is simply zero. However, if X is a metric space, μ a Borel measure, and f is continuous, then we can conclude that $f = 0$.

11.3.3 Examples of integrals with respect to measures.

We now look at the integral of functions with respect to the **Dirac measure**. Recall that the Dirac measure is:

Let X be a set and fix $x \in X$. For $A \subseteq X$, we set

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

This is the Dirac measure on a set X concentrated on $x \in X$.

We define this on the power set $\mathcal{P}(X)$ so that every set and every function is measurable. We let ϕ be a non-negative simple function and find its integral with respect to the Dirac measure.

$$\int_X \phi d\delta_a = \sum_{k=0}^n \alpha_k \delta_a(A_k) = \sum_{k=0}^n \alpha_k 1_{A_k}.$$

Now recall that the sets A_k are a disjoint union that covers X . Hence, x can only be in exactly one of these sets, denote this as set A_m .

$$= \alpha_m = \phi(x).$$

Hence, the integral with respect to the Dirac measure of a simple non-negative function is equivalent to evaluating the point x on the simple non-negative function.

We now look at the integral of non-negative functions,

$$\int_X f d\delta_a$$

which is the supremum of all Lebesgue integrals of simple non-negative functions with respect to δ_a . We have that

$$\int_X \phi d\delta_a = \phi(x) \leq f(x).$$

Hence

$$\int_X f d\delta_a \leq f(x).$$

To get the other inequality, define $\phi(x) = f(x)$ when $x = a$ otherwise it's 0. We have that $0 \leq \phi(x) \leq f(x)$ and by monotonicity

$$\begin{aligned} \int_X f d\delta_a &\geq \int_X \phi d\delta_a = f(x) \\ &\Rightarrow f(x) \leq \int_X f d\delta_a. \end{aligned}$$

Hence, we get that

$$\int_X f d\delta_a = f(x).$$

MATH3969: Measure Theory and Fourier Analysis

12. Monotone Convergence Theorem.

12.3.4 Monotone Convergence Theorem.

We now look at one of the most important theorems in this course.

Theorem 12.112 (*Monotone Convergence Theorem*). Let (X, \mathcal{A}, μ) be a measure space. For every $n \in \mathbb{N}$, let

$$f_n : X \rightarrow [0, \infty]$$

be a measurable function and suppose that

$$0 \leq f_n(x) \leq f_{n+1}(x)$$

for **almost every** $x \in X$. Then there exists a measurable function $f : X \rightarrow [0, \infty]$ such that $f_n(x) \rightarrow f(x)$ as $n \rightarrow \infty$ for almost every $x \in X$ and more importantly

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu.$$

Remark 12.113 This differs to the simple approximation theorem which states that for any non-negative measurable function f , there exists a sequence of monotone simple measurable functions ϕ_n that converges to f pointwise. The MCT refers to the property of the limit of the integral.

So if we had a sequence of non-negative measurable functions, they converge to $f(x)$ for almost every x and that the limit of the Lebesgue integral of the sequence of functions is equal to the Lebesgue integral of a function. This is one of the ways that the Lebesgue integral differs to the Riemann integral.

Theorem 12.114 (*Fatou's Lemma*). Let (X, \mathcal{A}, μ) be a measure space. For every $n \in \mathbb{N}$, let

$$f_n : X \rightarrow [0, \infty]$$

be measurable functions. Then

$$\int_X \lim_{n \rightarrow \infty} \inf f_n d\mu \leq \lim_{n \rightarrow \infty} \inf \int_X f_n d\mu$$

Proof: We first define

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

We see that $u_n(x) \leq f_n(x)$ and u_n is monotonically increasing. We define the limit inferior to be

$$u(x) := \lim_{n \rightarrow \infty} \inf f_n(x) = \lim_{n \rightarrow \infty} u_n(x).$$

for all $x \in X$. u_n and u are measurable as infimum/limit inferiors of a sequence of measurable functions are measurable. As $u_n \leq f_n$, by monotonicity

$$\int_X u_n d\mu \leq \int_X f_n d\mu.$$

Using the monotone convergence theorem, we have that

$$\int_X \lim_{n \rightarrow \infty} \inf f_n d\mu = \int_X u d\mu = \lim_{n \rightarrow \infty} \int_X u_n d\mu \leq \lim_{n \rightarrow \infty} \inf \int_X f_n d\mu.$$

■

We just interchanged the limit and the Lebesgue integral. Fatou's lemma is an extension of the MCT. Fatou's lemma looks at the limit inferior, which exists for every sequence whilst the limit may not necessarily exist.

Proposition 12.115 *Suppose that $f, g : X \rightarrow [0, \infty]$ are measurable functions and $\alpha \geq 0$ is a constant. Then*

- (i) $\int_X f + g d\mu = \int_X f d\mu + \int_X g d\mu;$
- (ii) $\alpha \int_X f d\mu = \int_X \alpha f d\mu.$

Proof: (Sketch). Break it up into simple measurable non-negative functions. Apply the properties there and use the MCT to get back to original measurable functions. ■

Recall that working with series and sequences are equivalent as series by definition are sequences of partial sums. We now apply the MCT to series with non-negative terms.

Theorem 12.116 *(Monotone Convergent Theorem for Non-negative Series). For every $k \in \mathbb{K}^N$, let $g_k : X \rightarrow [0, \infty]$ be measurable. Then*

$$\sum_{k=0}^{\infty} \int_X g_k dx = \int_X \sum_{k=0}^{\infty} g_k dx.$$

Proof: (Sketch). Transform into series of partial sums, apply MCT, take limits, and interchange integral and limits. ■

MATH3969: Measure Theory and Fourier Analysis

13. Integrable Functions.

13.3.5 Integrable Functions.

Now we look at functions that are not necessarily non-negative. In particular, $f : X \rightarrow \mathbb{K}^N$.

Definition 13.117 (*Integrable Function*). Let (X, \mathcal{A}, μ) be a measure space and $f : X \rightarrow \mathbb{K}$ be measurable. We call f a μ -integrable function if

$$\int_X |f| d\mu < \infty.$$

We let

$$\mathcal{L}^1(X, \mathcal{A}, \mu; \mathbb{K}^N) = \{f : X \rightarrow \mathbb{K} \mid f \text{ is } \mu\text{-integrable}\}$$

\mathcal{L}^1 is known as the space of integrable functions.

Recall that if f is measurable, then $|f|$ is also measurable.

The space \mathcal{L}^1 is a vector space. Here, $f \in \mathcal{L}^1(X, \mathcal{A}, \mu; \mathbb{K}^N)$ iff f^+ and f^- is integrable.

Definition 13.118 (*Lebesgue Integral for real or complex valued functions*).

(i) For $f \in \mathcal{L}^1(X, \mathbb{R})$, we define

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

(ii) For $f \in \mathcal{L}^1(X, \mathbb{C})$, we define

$$\int_X f d\mu = \int_X \operatorname{Re} f d\mu + i \int_X \operatorname{Im} f d\mu.$$

(iii) For any μ -measurable set A and $f \in \mathcal{L}^1(X, \mathbb{K})$, we define

$$\int_A f d\mu = \int_X 1_A f d\mu.$$

Integration is a linear operation as seen in next theorem.

Theorem 13.119 Let $f, g \in \mathcal{L}^1(X, \mathcal{A}, \mu; \mathbb{K})$, then the following assertions are true.

(i) For all $\alpha, \beta \in \mathbb{K}$

$$\int_X \alpha f + \beta g d\mu = \alpha \int_X f d\mu + \beta \int_X g d\mu$$

(ii) If $A, B \in \mathcal{A}$ are disjoint, then

$$\int_{A \cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

(iii) The triangle inequality holds

$$|\int_X f d\mu| \leq \int_X |f| d\mu.$$

Finally, we look at examples of measures and what their corresponding integrals are.

The Lebesgue Measure

If $\mu = m_N$ is the Lebesgue measure and $A \subseteq \mathbb{R}^N$ is Lebesgue measurable, we define

$$\int_A f(x) dx := \int_A f dm_N.$$

Note that any Riemann integrable function is also Lebesgue integrable but not vice versa. We can construct pathological examples involving indicator functions whereby the upper and lower Riemann sums differ whilst Lebesgue integration will work. A good one is $f(x) = 1$ if $x \in \mathbb{R} - \mathbb{Q} \cap [0, 1]$ and 0 otherwise. The Lebesgue integral works since the function is just an indicator function but not for Riemann sums as the lower sum is zero whilst the other one is one.

The Lebesgue Stieltjes Integral

Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a **right continuous** increasing function. That is, $F(s) \leq F(t)$ for $s \leq t$ and

$$\lim_{s \rightarrow t^+} F(s) = F(t)$$

for all $t \in \mathbb{R}$. For $A \subseteq \mathbb{R}$, we define the **Lebesgue-Stieltjes** measure induced by F as

$$\mu_F^*(A) := \inf \left\{ \sum_{k=0}^{\infty} (F(b_k) - F(a_k)) : A \subseteq \bigcup_{k=0}^{\infty} (a_k, b_k] \right\}.$$

This μ_F is an outer measure on \mathbb{R} that induces a regular Borel measure μ_F on \mathbb{R} . If $F(x) = x$, we get the Lebesgue measure. We define the **Lebesgue-Stieltjes integral** as

$$\int_{\mathbb{R}} f dF := \int_{\mathbb{R}} f d\mu_F.$$

MATH3969: Measure Theory and Fourier Analysis

14. Measures with a density.

14.3.6 Measures with a density.

We now seek to express a measure in terms of an integral of a measurable function with respect to another measure. The intuition is that if the sample space is $X = \mathbb{R}$, then μ is the Lebesgue measure, $g(\cdot)$ will be the normal density, and v will be the normal distribution (normal probability measure).

Definition 14.120 (*Density of a measure with respect to another measure*).

Let (X, \mathcal{A}, μ) be a measure space and $g : X \rightarrow [0, \infty]$ be a measurable function. For $A \in \mathcal{A}$, we define

$$v(A) := \int_A g d\mu = \int_X 1_A g d\mu.$$

Here, v is a measure. Hence, the measure of A by v is the Lebesgue integral of g over the set A . It is easy to check this. Hence, we call g the density of v measure with respect to the measure μ .

Proof: 1)

$$v(\emptyset) = \int_{\emptyset} g d\mu = \int_X 1_{\emptyset} g d\mu = 0.$$

2)

$$\begin{aligned} v(\cup_{n \in \mathbb{N}} A_k) &= \int_{\cup_{n \in \mathbb{N}} A_k} g d\mu = \int_X 1_{\cup_{n \in \mathbb{N}} A_k} g d\mu \\ &= \int_X \sum_{k=1}^{\infty} 1_{A_k} g d\mu \end{aligned}$$

Since g is non-negative, the monotone convergence theorem applies so

$$= \sum_{k=1}^{\infty} \int_X 1_{A_k} g d\mu = \sum_{k=1}^{\infty} \int_{1_{A_k}} g d\mu = \sum_{k=1}^{\infty} v(A_k).$$

■

We can show that

$$\int_X f dv = \int_X f g d\mu,$$

for all $f \in \mathcal{L}^1(X, \mathcal{A}, v, \mathbb{K})$. Here, we can integrate a function f with respect to measure v in terms of measure μ . For a simple non-negative function $f = \sum_{k=1}^n \alpha_k 1_{A_k}$, we have

$$\int_X f dx = \sum_{k=0}^{\infty} \alpha_k v(A_k) = \sum_{k=0}^{\infty} \alpha_k \int_X 1_{A_k} g d\mu = \int_X \sum_{k=0}^{\infty} \alpha_k 1_{A_k} g d\mu = \int_X f g d\mu.$$

If f is non-negative, then there exists a sequence $\{f_n\}$ of simple functions with $0 \leq f_n \leq f_{n+1}$ and $f_n \rightarrow f$ pointwise. Using MCT, we get the same result.

14.3.7 Special Transformation Formula

We show a transformation formula for linear transformations. This is similar to the idea of integration by substitution. We first show a substitution for linear transformations only.

Theorem 14.121 (*Special Transformation Formula*).

Let $f \in \mathcal{L}^1(\mathbb{R}^N)$ and $T : \mathbb{R}^N \rightarrow \mathbb{R}^N$ be linear. Then, we have that

$$\int_{imT} f dx = |\det T| \int_{\mathbb{R}^N} f \circ T dx$$

where imT is the image of T .

It is also possible to show the full transformation formula for functions g with continuous partial derivatives.

$$\int_{img} f dx = \int_{\mathbb{R}^N} f \circ g |\det J_g| dx$$

where J_g is the Jacobian matrix of g .

MATH3969: Measure Theory and Fourier Analysis

15. Limit Theorems

15.3.8 Limit Theorems

Recall that we are interested in that if we had a sequence of functions $\{f_n\}$ which converges pointwise to some limit function f , does

$$\int \lim_{n \rightarrow \infty} f_n = \lim_{n \rightarrow \infty} \int f_n,$$

hold or in other words, does the limit and integral commute. We have already seen 2 limit theorems so far. The monotone convergence theorem required a monotone sequence of non-negative functions and if this is satisfied, then they do commute. Fatou's lemma was a case where we did not make any assumptions but instead, we can make a guess about the relationship of interchanging them. We can now construct a new limit theorem that requires that we just need f_n to be dominated in order to interchange the limit and integral.

We now construct a case for arbitrary sequences of non-negative functions by using the limit inferior.

Theorem 15.122 (*Dominated Convergence Theorem*).

Let $f_n : X \rightarrow \mathbb{K}$ be a measurable function and $f : X \rightarrow \mathbb{K}$ be such that $f_n(x) \rightarrow f(x)$ pointwise almost everywhere. Furthermore, assume that there exists a dominating function $g(x) \in \mathcal{L}^1(X)$ (so it is integrable, measurable, and independent of n) such that

$$|f_n(x)| \leq g(x)$$

for almost all $x \in X$. Then

- (i) $f_n, f \in \mathcal{L}^1(X, \mathbb{K})$ for all $n \in \mathbb{N}$;
- (ii) $\int_X |f_n - f| d\mu \rightarrow 0$ as $n \rightarrow \infty$;
- (iii) $\int_X f_n d\mu \rightarrow \int_X f d\mu$ as $n \rightarrow \infty$.

We now have an analogue to improper Riemann integrals for the Lebesgue integral.

Corollary 15.123 (*Improper Lebesgue Integral*).

Let $f \in \mathcal{L}^1((a, b), \mathbb{K})$. Then

$$\lim_{x \rightarrow b^-} \int_a^x f(t) dt = \int_a^b f(t) dt.$$

We now look at integral with parameters.

Theorem 15.124 (*Continuity of parameter integrals*).

Let (X, \mathcal{A}, μ) be a measure space and Y a metric space (which is usually a subset of \mathbb{R} or \mathbb{C}). Suppose that $f : X \times Y \rightarrow \mathbb{K}$ such that

- (i) $x \rightarrow f(x, y)$ is μ -measurable for all $y \in Y$;
- (ii) $y \rightarrow f(x, y)$ is continuous at y_0 for almost all $x \in X$;
- (iii) there exists $g \in \mathcal{L}^1(X, \mathbb{R})$ such that

$$|f(x, y)| \leq g(x)$$

for almost all $x \in X$ and all $y \in Y$. For $y \in Y$ define

$$F(y) := \int_X f(x, y) d\mu(x).$$

Then F is continuous at y_0 .

For the first and third point, if the third point holds, then initially, instead of showing $x \rightarrow f(x, y)$ is integrable, we can weaken the condition to showing it is measurable. To show measurability, find the inverse image for every value the function can take on and show that is measurable. Furthermore, note that for the dominating function $g(x)$, this function has to be independent of y and integrable for all $y \in Y$.

Theorem 15.125 (*Differentiation of Parameter Integrals*). Let (X, \mathcal{A}, μ) be a measure space and $J \subseteq \mathbb{R}$ an interval. Suppose that $f : X \times J \rightarrow \mathbb{K}$ be a function such that

- 1. $x \rightarrow f(x, t)$ is μ -integrable for all $t \in J$;
- 2. For almost all $x \in X$, $\frac{\partial f}{\partial t}(x, t)$ exists and is continuous on J ;
- 3. There exists $g \in \mathcal{L}^1(X, \mathbb{R})$ such that

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

for almost all $x \in X$ and all $t \in J$.

For $t \in J$, define

$$F(t) := \int_X f(x, t) d\mu(x).$$

Then, $F : J \rightarrow \mathbb{K}$ is differentiable and

$$F'(t) := \int_X \frac{\partial f}{\partial t}(x, t) d\mu(x)$$

for all $t \in J$.

Definition 15.126 (*Fundamental Theorem of Calculus*).

- (i) If $f : [a, b] \rightarrow \mathbb{R}$ is continuous, then

$$\frac{d}{dt} \int_a^t f(s) ds = f(t)$$

for all $t \in [a, b]$.

- (ii) If $f : [a, b] \rightarrow \mathbb{R}$ is continuous and $f : [a, b] \rightarrow \mathbb{R}$ is continuously differentiable, then

$$\int_a^b f'(s) ds = f(b) - f(a).$$

Continuity and differentiability are **local properties**, that is, it is sufficient to check on some neighbourhood of every point. An example would be if you wanted to check continuity on $(0, \infty)$, it is sufficient to check on every closed interval $[a, b]$ where $0 < a < b < \infty$.

MATH3969: Measure Theory and Fourier Analysis

16. The Lebesgue Spaces

16.4 The Lebesgue Spaces

16.4.1 The Lebesgue Spaces

We now look at a class of function spaces that is quite important.

Definition 16.127 (\mathcal{L}^p -norm of f). We call

$$\|f\|_p$$

the \mathcal{L}^p -norm of f .

Definition 16.128 (\mathcal{L}^p -spaces). Let $1 \leq p < \infty$ and $f : X \rightarrow \mathbb{K}$. We call

$$\|f\|_p := \left(\int_X |f|^p d\mu \right)^{1/p}$$

the \mathcal{L}^p -norm of f . We set

$$\mathcal{L}^p(X, \mathcal{A}, \mu, \mathbb{K}) : \{f : X \rightarrow \mathbb{K} : f \text{ is measurable and } \|f\|_p < \infty\}.$$

If no confusion about the measure, σ -algebra or codomain seems likely we write $\mathcal{L}^p(X, \mathbb{K})$ or $\mathcal{L}^p(X)$. The spaces are called the **Lebesgue spaces**.

Remark 16.129 The \mathcal{L}^p space is a space of functions for which the p -th power of the absolute value is Lebesgue integrable almost everywhere.

To think about what it means for $(\int_X |x| d\mu)^{1/p}$, you can first think of $(\sum_{k=0}^{\infty} |x_k|)^{1/p}$ and then generalise it. Now recall that Lebesgue integrals are fundamentally $\int f d\mu = \sum_{k=1}^n \alpha_k \mu(A_k)$ and hence this is why we can just take the norm of such a vector.

This is similar to our definition of \mathcal{L}^1 earlier if we let $p=1$.

The \mathcal{L}^p -spaces are vector spaces. We can show that from the convexity of $t \rightarrow t^p$, we can see that $f + g \in \mathcal{L}^p(X)$.

We now aim to show that it is actually a norm but first we require some useful inequalities.

Lemma 16.130 (Young's Inequality for Products). Let $p, q > 1$ such that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Then,

$$st \leq \frac{1}{p} s^p + \frac{1}{q} t^q$$

for all $s, t \geq 0$.

Proof: If $s = 0$ or $t = 0$, then the inequality is obvious.

If $s, t > 0$, then by concavity of the logarithm function and since $1/p + 1/q = 1$, we have that

$$\log(st) = \log(s) + \log(t) = \frac{1}{p}\log(s)^p + \frac{1}{q}\log(t)^q \leq \log\left(\frac{1}{p}s^p + \frac{1}{q}t^q\right)$$

where the last inequality arises due to the concave nature of the logarithm function.

We then take exponentials on both sides to get

$$st \leq \frac{1}{p}s^p + \frac{1}{q}t^q.$$

■

Remark 16.131 *This inequality is just a mathematical inequality about the product of two numbers. This will then be used to help prove the more important Hölder's inequality, which is a fundamental inequality between integrals.*

Theorem 16.132 (Hölder's Inequality). *Let $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Then*

$$\left| \int_x fg d\mu \right| \leq \int_x |fg| d\mu \leq \|f\|_p \|g\|_q$$

for all $f \in \mathcal{L}^p(X)$ and $g \in \mathcal{L}^q(X)$. Recall we defined $\|f\|_p := (\int_x |f|^p d\mu)^{1/p}$.

Proof: If $\|f\|_p = 0$ or $\|g\|_q = 0$, then $|f| = 0$ or $|g| = 0$ almost everywhere as the integral is 0, so then $f(x) = 0$ or $g(x) = 0$ almost everywhere. Furthermore, $fg = 0$ almost everywhere.

Hence, assume that $\|f\|_p, \|g\|_q > 0$. By Young's inequality and by the definition of the norms

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

by Young's inequality. Then taking out the constant from integral and using the definition of \mathcal{L} -norm

$$= \frac{1}{p\|f\|_p^p} \int_x |f|^p d\mu + \frac{1}{q\|g\|_q^q} \int_x |g|^q d\mu$$

Here, $\int_x |f|^p d\mu = \|f\|_p^p$, so then

$$\begin{aligned} &= \frac{\|f\|_p^p}{p\|f\|_p^p} + \frac{\|g\|_q^q}{q\|g\|_q^q} \\ &= \frac{1}{p} + \frac{1}{q} = 1. \end{aligned}$$

Hence, multiplying through,

$$\int_x |fg| d\mu \leq \|f\|_p \|g\|_q.$$

■

Remark 16.133 *Hölder's inequality will be used to establish the **Minkowski inequality**, which is the triangle inequality in the space L^p and also to establish that L^q is the **dual space** of L^p for $p \in [1, \infty)$.*

When $p=q=2$ holds, then we get the **Cauch-Schwarz inequality**.

Corollary 16.134 (*Cauchy-Schwarz Inequality*). For $f, g \in \mathcal{L}^2$, where $p=q=2$, then

$$\int |fg| d\mu \leq \|f\|_2 \|g\|_2.$$

We can look at relationship between \mathcal{L}^p -spaces as we vary p .

Proposition 16.135 Let $1 \leq p < q < \infty$. If $\mu(X) < \infty$, then $\mathcal{L}^q(X) \subseteq \mathcal{L}^p(X)$. More precisely, if $f \in \mathcal{L}^q(X)$, then

$$\|f\|_p \leq \mu(X)^{\frac{1}{p} - \frac{1}{q}} \|f\|_q.$$

Generally, $\mathcal{L}^p(X) \not\subseteq \mathcal{L}^q(X)$. Moreover, if $\mu(X) = \infty$, then neither sets are subsets to each other.

Remark 16.136 What this means is that if we had $f \in \mathcal{L}^q(X)$, where $p < q$, then we can show, from Holder's inequality, that $f \in \mathcal{L}^p(X)$. Hence, $\mathcal{L}^q(X) \subseteq \mathcal{L}^p(X)$.

For applications in probability theory, that means the vector space L^2 is a vector subspace of L^1 , but only if $\mu(\Omega) < \infty$, which always happens by construction.

MATH3969: Measure Theory and Fourier Analysis

17. Further properties of Lebesgue Spaces

17.4.2 Further properties of Lebesgue Spaces

Now we derive the **triangle inequality** for the \mathcal{L}^p -norm.

Proposition 17.137 (*Minkowski's Inequality*). *For $1 \leq p < \infty$, we have that*

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

for all $f, g \in \mathcal{L}^p(X)$.

Proof: We assume that $\|f + g\|_p > 0$ or else it is trivial.

If $p=1$, then we have

$$\|f + g\|_1 = \int_X |f + g| d\mu \leq \int_X |f| + |g| d\mu = \int_X |f| d\mu + \int_X |g| d\mu = \|f\|_1 + \|g\|_1$$

as claimed.

Now assume that $p \in (1, \infty)$. Note that

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

In order to invoke Hölder's inequality, we set

$$q = \frac{p}{p-1}$$

so that

$$\frac{1}{p} + \frac{1}{q} = 1.$$

From Hölder's inequality,

$$\begin{aligned} \int_X |f| |f + g|^{p-1} d\mu &\leq \|f\|_p \left(\int_X |f + g|^{(p-1)q} d\mu \right)^{1/q} = \|f\|_p \left(\int_X |f + g|^p d\mu \right)^{(p-1)/p} \\ &= \|f\|_p \|f + g\|_p^{p-1}. \end{aligned}$$

By similar argument, we have

$$\int_X |g| |f + g|^{p-1} d\mu \leq \|g\|_p \|f + g\|_p^{p-1}.$$

Combining everything,

$$\begin{aligned} \|f + g\|_p^p &\leq \|f\|_p \|f + g\|_p^{p-1} + \|g\|_p \|f + g\|_p^{p-1} \\ &= (\|f\|_p + \|g\|_p) \|f + g\|_p^{p-1}. \end{aligned}$$

Then, divide by $\|f + g\|^{p-1}$, we get

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

■

We now show some properties of L^p -norm, to show that it is indeed a norm.

Theorem 17.138 (*Properties of L^p -norm*). Let $1 \leq p < \infty$. Then for $f, g \in \mathcal{L}^p(X)$ and $\alpha \in \mathbb{K}$, we have

- (i) $\|f\|_p \geq 0$ with equality if and only if $f = 0$ almost everywhere;
- (ii) $\|\alpha f\|_p = |\alpha| \|f\|_p$;
- (iii) $\|f + g\|_p \leq \|f\|_p + \|g\|_p$ (triangle or Minkowski's inequality).

Proof: (i) $\|f\|_p \geq 0$ is evident from definition. Clearly, $\|f\|_p = 0$ if $f = 0$ almost everywhere.

(ii) From definition of \mathcal{L}^p -norm

$$\|\alpha f\|_p^p = \left(\int_X |\alpha f|^p \right)^{1/p} = (|\alpha|^p \int_X |f|^p)^{1/p} = |\alpha| \|f\|_p.$$

(iii) From the earlier proof of the Minkowski inequality. ■

However, due to the fact that the first property has the "almost everywhere" condition, this is not a proper norm. It is in fact, a **pseudo-norm** as $\|f\|_p = 0$ does not imply that f is the zero function $f = 0$. Therefore, we can call $(\mathcal{L}^p, \|\cdot\|_p)$ a **pseudo-normed vector space**.

The p -norm gives us a notion of distance between functions in the space $\mathcal{L}^p(X)$. This allows us to start considering ideas of convergence. We look at sequences of functions f_n so that $\|f_n - f\|_p \rightarrow 0$. We look at such sequences in $\mathcal{L}^p(X)$ where we use the metric $d(f, g) := \|f - g\|_p$. However, we will formulate to make this metric a proper metric later on.

Theorem 17.139 (*Properties of the pseudo metric*). The pseudo metric has similar properties to a metric. Defined on a vector space V , the properties of the pseudo metric are

- (i) $d(x, y) \in [0, \infty)$ for $x, y \in V$;
- (ii) $d(x, y) + d(y, z) \geq d(x, z)$ for $x, y, z \in V$;
- (iii) $d(x, y) = d(y, x)$ for $x, y \in V$;
- (iv) $x = y$ implies that $d(x, y) = 0$ for $x, y \in V$.

To turn this pseudo metric into a proper metric, we need to strengthen condition (iv) into $x = y$ if and only if $d(x, y) = 0$ for $x, y \in V$.

Fortunately, a pseudo metric is sufficient for convergence.

Definition 17.140 (*Convergence Sequences in \mathcal{L}^p*). Let $f_n, f \in \mathcal{L}^p(X)$.

We say that $f_n \rightarrow f$ in $\mathcal{L}^p(X)$ if $\|f_n - f\|_p \rightarrow 0$ as $n \rightarrow \infty$;

Definition 17.141 (*Cauchy Sequences in \mathcal{L}^p*). Let $f_n, f \in \mathcal{L}^p(X)$.

We say that $\{f_n\}$ is a **Cauchy sequence** in $\mathcal{L}^p(X)$ if for every $\epsilon > 0$, there exists n_0 such that

$$\|f_n - f_m\|_p < \epsilon$$

for all $m > n > n_0$.

Proposition 17.142 *Every convergent sequence in $\mathcal{L}^p(X)$ is a Cauchy sequence.*

Proof: As $\|f_n - f\|_p \rightarrow 0$, there exists a $n_0 \in \mathbb{N}$ such that $\|f_n - f\|_p < \frac{\epsilon}{2}$ for all $n \geq n_0$. Using triangle inequality

$$\|f_n - f_m\|_p = \|(f_n - f) - (f_m - f)\|_p \leq \|f_n - f\|_p + \|f_m - f\|_p < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for all $m > n > n_0$. ■

We would now like to show that \mathcal{L}^p is **complete**, that is, every Cauchy sequence converges. So we want to upgrade our earlier statement to "Every sequence converges if and only if it is a Cauchy sequence".

First, we require that a Cauchy sequence is bounded, which then means it has a convergent subsequence, and hence the Cauchy sequence converges.

Lemma 17.143 *Suppose that $\{f_n\}$ is a Cauchy sequence in $\mathcal{L}^p(X)$. If $\{f_{n_k}\}$ is a convergent subsequence with $f_{n_k} \rightarrow f$ in $\mathcal{L}^p(X)$, then $f_n \rightarrow f$ in $\mathcal{L}^p(X)$.*

Proof: For a Cauchy sequence, there exists $n_0 \in \mathbb{N}$ such that

$$\|f_n - f_m\|_p < \frac{\epsilon}{2}$$

for $m > n > n_0$. Since $f_{n_k} \rightarrow f$, there exists $\ell \in \mathbb{N}$, where we then define the rank $m := k_\ell$ such that

$$\|f_m - f\|_p < \frac{\epsilon}{2}.$$

Hence, by the triangle inequality and using the rank m for which the subsequence converges

$$\|f_n - f\|_p = \|(f_n - f_m) + (f_m - f)\|_p \leq \|f_n - f_m\|_p + \|f_m - f\|_p < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

for $m > n > n_0$. As the argument works for any $\epsilon > 0$, we conclude that $\|f_n - f\|_p \rightarrow 0$ as claimed. ■

Remark 17.144 *Recall that the sequence converges to the same thing that the subsequence converges to.*

Now we look at the convergence of a sequence such that a certain series converges.

Proposition 17.145 *Let $\{g_k\}$ be a sequence in $\mathcal{L}^p(X)$ such that the series*

$$\sum_{k=1}^{\infty} \|g_k\|_p < \infty.$$

Then there exists a function $f \in \mathcal{L}^p(X)$ such that $f = \sum_{k=0}^{\infty} g_k$ converges pointwise almost everywhere and in $\mathcal{L}^p(X)$.

With all this, we can now finally show that every Cauchy sequence in $\mathcal{L}^p(X)$ converges.

Theorem 17.146 (*Completeness of \mathcal{L}^p*). *Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{L}^p(X)$. Then there exists a function $f \in \mathcal{L}^p(X)$ such that $f_n \rightarrow f$ in $\mathcal{L}^p(X)$. Moreover, $\{f_n\}$ has a subsequence that converges almost everywhere.*

Proof: Since $\{f_n\}$ is a Cauchy sequence, for every $k \in \mathbb{N}$, there exists $m_k \in \mathbb{N}$ such that

$$\|f_n - f_m\| < \frac{1}{2^k}$$

for all $m > n > m_k$.

We set $n_0 := m_0$ and $m_1 := 1 + \max\{1, n_0\}$. Inductively, we set $n_{k+1} := 1 + \max\{k, n_k\}$. Then, we have that n_k is strictly increasing and

$$\|f_{n_k} - f_{n_{k-1}}\|_p < \frac{1}{2^{k-1}}.$$

Here, we have chosen a specific subsequence, as we will soon see has useful properties.

Hence, this series

$$\sum_{k=1}^{\infty} \|f_{n_k} - f_{n_{k-1}}\|_p < \sum_{k=1}^{\infty} \frac{1}{2^{k-1}} = 2$$

converges.

We set $g_k := f_{n_k} - f_{n_{k-1}}$. We note that using a telescoping sum,

$$f_{n_m} = f_{n_0} + \sum_{k=1}^m (f_{n_k} - f_{n_{k-1}}) = f_{n_0} + \sum_{k=1}^m g_k.$$

By the earlier proposition, there exists $f \in \mathcal{L}^p(X)$ such that $f_{n_m} \rightarrow f$ in $\mathcal{L}^p(X)$ as $m \rightarrow \infty$. Therefore, $\{f_n\}$ has a convergent subsequence.

By the earlier lemma, as the subsequence converges to f in $\mathcal{L}^p(X)$, then that means $f_n \rightarrow f$ in $\mathcal{L}^p(X)$. ■

Remark 17.147 *Note that f_n does not converge to f pointwise almost everywhere. We can only state the $\{f_n\}$ has a subsequence that is pointwise convergent.*

Hence, we have the important property of $\mathcal{L}^p(X)$ that

Proposition 17.148 *A sequence in $\mathcal{L}^p(X)$ is convergent if and only if it is a Cauchy sequence.*

MATH3969: Measure Theory and Fourier Analysis

18. The L^p spaces

18.4.3 The L^p spaces

The issue is that the \mathcal{L}^p -norm is not a proper norm since $\|f\|_p = 0$ does **not** imply that $f = 0$ is the zero function. We can only conclude that $f = 0$ almost everywhere.

So what we do is that if 2 functions are equal almost everywhere, we consider them equal.

We define an **equivalence relation** on (X, \mathcal{A}, μ) . For two measurable functions $f, g : X \rightarrow \mathbb{K}$, we say that

$$f \sim g \text{ if } f = g \text{ almost everywhere.}$$

In other words, we could also say $f \sim g$ if $f - g = 0$ almost everywhere. This is an equivalence relation as

- (i) (Reflexivity) $f \sim f$ as $f = f$ everywhere;
- (ii) (Symmetry) $f \sim g$ implies $g \sim f$;
- (iii) (Transitivity) $f \sim g$ and $g \sim h$ implies that $f \sim h$.

We define the **equivalence class** of f by

$$[f] = \{g : X \rightarrow \mathbb{K} : g \text{ is measurable and } f \sim g \text{ almost everywhere}\}.$$

We then introduce a family of vector spaces of these equivalence classes.

Definition 18.149 (L^p -space). For $1 \leq p < \infty$, we define

$$L^p(X, \mathcal{A}, \mu, \mathbb{K}) := \{[f] : f \in \mathcal{L}^p(X, \mathcal{A}, \mu, \mathbb{K})\}.$$

We set

$$\|[f]\|_p := \|f\|_p.$$

Similarly, we set

$$[f] + [g] := [f + g] \text{ and } \alpha[f] := [\alpha f]$$

for all $[f], [g] \in L^p(X)$ and $\alpha \in \mathbb{K}$.

Remark 18.150 Therefore, the $L^p(X)$ space is just the set of equivalence classes for $\int_X |f|^p < \infty$ functions! Furthermore, the norm of the equivalence class is just the p -norm of the function f itself.

It can be shown that the norm, addition, and multiplication by scalars are independent of the choice of function from the equivalence class. So whatever function f_1 we pick, the equivalence class generated would be the same if we picked another function f_2 from the same equivalence class. In particular, $[f_1 + g_1] = [f_2 + g_2]$, $[\alpha f_1] = [\alpha f_2]$, and $\|f_1\|_p = \|f_2\|_p$.

We define the null space as

$$\mathcal{N}(X) = \{g : X \rightarrow \mathbb{K} : g \text{ is measurable and } g = 0 \text{ almost everywhere.}\}$$

which is a vector subspace of $\mathcal{L}_p(X)$.

We then define the quotient space

$$L_p(X) = \mathcal{L}_p(X)/\mathcal{N}(X).$$

Remark 18.151 *In practice, we write $f \in L^p(X)$ and call it an L^p -function rather than $[f]$. Generally, we just choose a representative from $\mathcal{L}^p(X)$.*

Now, if we had $f = 0 \in L^p(X)$, f is the equivalence class of functions which are zero almost everywhere. $\|\cdot\|_p$ is a proper norm now as we have redefined it on a vector space where function that are equal almost everywhere are just considered to be the same function. Hence, the first property of the norm is now properly satisfied.

We can also talk about Cauchy sequences in $L^p(X)$ now, where we call $([f_n])$ a Cauchy sequence in $L^p(X)$ if there exists $f_n \in \mathcal{L}^p(X)$ such that $\{f_n\}$ is a Cauchy sequence in $\mathcal{L}^p(X)$. We know that such a sequence converges in $\mathcal{L}^p(X)$. From that, the sequence $\{f_n\}$ converges in $L^p(X)$. Therefore, we get the following theorem.

We finally have a **complete norm**, that is, a p -norm that satisfies all the properties of a norm.

Theorem 18.152 *For $1 \leq p < \infty$, the space $L^p(X)$ is a complete normed space with respect to the norm $\|\cdot\|_p$.*

Remark 18.153 (a) *A complete normed vector space is called a **Banach space**, which the $L^p(X)$ space is a good example of. The \mathcal{L}^p space is a **pseudo-Banach space**.*

(b) *If we have that $p=2$, then the L^2 -norm is induced by the inner product of equivalence class. A complete inner product space is called a **Hilbert space**.*

Definition 18.154 (Hilbert Space). *A Hilbert space H is a real or complex inner product space that is also a complete metric space with respect to the distance function induced by the inner product.*

\mathcal{L}^p is a Hilbert space if and only if $p = 2$.

Remark 18.155 *This section looks at the differences between a normed vector space and an inner product space. An inner product space is a subspace of the normed vector space. We have norms that do not come from inner products, but all inner products have norms. All inner products give rise to a norm as $\langle x, x \rangle = \|x\|^2$, but there are some norms that do not arise from inner products. A norm is induced by an inner product if and only if the parallelogram law holds, which in this case, only holds for $\|\cdot\|_2$. In other words, inner product spaces satisfies the parallelogram law. Hence, the normed vector space $L^p(X)$ only has an inner product if $p = 2$. So the $L^2(X)$ space is an example of a Hilbert space.*

MATH3969: Measure Theory and Fourier Analysis

19. Density Theorems

19.4.4 Density Theorems

So far, we've defined integrals by approximation of simple function and used the monotone convergence theorem to help us. There is an alternative approach by using **density theorems**. We can use the fact that an arbitrary function in $L^p(X)$ can be approximated by a functions from some smaller class of functions such as continuous or smooth functions.

We revisit some stuff from analysis that we will need frequently.

Definition 19.156 (*Closure of a set*). The closure of a set A , denoted by \bar{A} is defined as the **smallest closed set** in \mathbb{K}^N that **contains** A . More precisely,

$$\bar{A} = \cap \{F \subseteq \mathbb{K}^N : A \subseteq F\}.$$

Hence, if $\bar{A} = A \leftrightarrow A$ is closed.

Definition 19.157 (*Density*). Given a metric space (X, ρ) , we say that a subset A of X is dense in X if for each $\epsilon > 0$ and $x \in X$, we have that

$$B(x, \epsilon) \cap A \neq \emptyset.$$

More succinctly, the closure of $A = X$.

For this definition of dense, it says that for each x , a ball of **any size** around x will intersect with the set A .

So now, we will prove properties on **dense subsets** of L^p and then show it is true for all of L^p . (Recall that the closure of dense is the whole space).

Definition 19.158 (*Dense*). For every $f \in L^p(X)$, there exists a sequence $\{\phi_n\}$ of simple functions in $L^p(X)$ such that $\phi_n \rightarrow f$ in $L^p(X)$.

We now approximate by simple functions.

Proposition 19.159 (*Density of simple functions*). If $1 \leq p < \infty$, then the set of subspace of simple functions

$$\{\phi \in L^p(X) : \phi \text{ is simple}\} = \text{span}\{1_A : A \in \mathcal{A}, \mu(A) < \infty\}$$

is dense in $L^p(X)$.

Proof: We take an arbitrary function $f \in L^p(X)$, by the simple approximation lemma, there exists simple measurable functions ϕ_n such that

$$\phi_n \rightarrow f \text{ pointwise}$$

$$|\phi_n| \leq |\phi_{n+1}| \leq \dots$$

If we take $f \in L^p(X)$, it does not make sense to talk about pointwise properties but here we did. What we did is that we took a representative from $L^p(X)$ and then discuss pointwise convergence.

In particular, this increasing property means that $|\phi_n| \leq |f|$ for all $n \in \mathbb{N}$. Hence we have,

$$|\phi_n - f|^p \rightarrow 0 \text{ pointwise}$$

$$|\phi_n - f|^p \leq (|\phi_n| + |f|)^p \leq 2^{p-1}(|f|^p + |\phi_n|^p) \leq 2^p |f|^p$$

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

By the dominated convergence theorem

$$\|\phi_n - f\|_p = \left(\int_X |\phi_n - f|^p d\mu \right)^{1/p} \rightarrow 0$$

as $n \rightarrow \infty$.

Also, $|\phi_n|^p \leq |f|^p$ so then $\|\phi_n\|_p \leq \|f\|_p$, so $\phi_n \in L^p(X)$. ■

What we mean here is that the equivalence class of a simple function is dense in $L^p(X)$. By dense, we mean that the closure is the whole space. So now matter which function f we pick, there is a simple function near it. In particular, by dense, we mean, for every $f \in L^p(X)$, there exists a sequence (ϕ_n) of simple functions in $L^p(X)$ so that $\phi_n \rightarrow f$ in $L^p(X)$,

Hence, we can approximate every $f \in L^p(X)$ arbitrarily closely by a sequence of simple functions in $L^p(X)$.

We now restrict ourselves to the case of $X = \mathbb{R}^N$ and $\mu = m_N$ to be the Lebesgue measure.

Definition 19.160 (*Support*). Let $f : \mathbb{R}^N \rightarrow \mathbb{K}$ be a function. We call the set

$$\text{supp}(f) := \overline{\{x \in \mathbb{R}^N : f(x) \neq 0\}}$$

the support of f . If $\text{supp}(f)$ is compact, we say that f has a compact support.

Remark 19.161 The support of f is the closure of the set where f is non-zero and hence it is always closed. If the set is also bounded, it is compact because in \mathbb{R}^N , being compact is equivalent to closed and bounded.

We now extend on our earlier proposition of the density of simple functions to simple functions with a bounded support are dense.

Corollary 19.162 If $1 \leq p < \infty$, then

$$\{\phi \in L^p(\mathbb{R}^N) : \phi \text{ is simple and has bounded support}\}$$

is dense in $L^p(\mathbb{R}^N)$.

Definition 19.163 (*Span*). For a vector space V and W as a subset of V , we define the span of U by

$$\text{span}(U) := \left\{ \sum_{i=1}^n \lambda_i u_i : u_i \in U, \lambda_i \in \mathbb{K} \right\}.$$

This is the set of all linear combinations of elements in W . We also recall that the measure for this is finite or else it would not be in L^p . Recall that linear combinations are always finite sums.

We formulate above density results by saying that the span of indicator functions is a dense subset of $L^p(X)$. This is because simple functions are linear combinations of indicator functions.

$$\{\phi \in L^p(X) : \phi \text{ is simple}\} = \text{span}\{1_A : A \text{ is measurable and } \mu(A) < \infty\}$$

is dense in $L^p(X)$. If the measure is not finite, then it would not be in $L^p(X)$.

Looking at all the sets in \mathcal{A} may be too big and inconvenient, so we restrict it. In the specific case of $L^p(\mathbb{R}^N)$, we have

$$\text{span}\{1_A : A \subseteq \mathbb{R}^N \text{ is measurable and bounded}\}$$

is dense in $L^p(\mathbb{R}^N)$. The Lebesgue measure is always defined on \mathbb{R}^N unless specified otherwise. The proof is similar as before except we replace ϕ_n by $\phi_n 1_{B(0,n)}$. So if we fix a point, as ϕ_n goes to f , then this also goes to f .

The span of indicator functions on open sets is also dense in L^p . Using the definition of the Lebesgue measure involving coverings by rectangles, we can show that the span of indicator functions of rectangles are dense in $L^p(\mathbb{R}^N)$. Furthermore, we can replace rectangles by dyadic cubes. Once we have dyadic cubes, we can use the span of indicator functions of bounded open sets from dyadic decomposition or the outer regularity of the Lebesgue measure.

19.4.5 The space L^∞

The issue with this space is that we need a norm that works with such a space.

If $X = \{a_1, a_2, \dots, a_n\}$ is a finite set and μ is the counting measure, then for $1 \leq p < \infty$, we have for the p -norm

$$\|f\|_p = \left(\int_X |f|^p d\mu \right)^{1/p} = \left(\sum_{k=1}^n |f(a_k)|^p \right)^{1/p}$$

where we can just take the summation as it is the counting measure. We want to see the limit as $p \rightarrow \infty$.

To do so, we look at the maximum. Let $m := \max_{k=1,2,\dots,n} |f(a_k)|$.

$$m = (m^p)^{1/p} \leq \left(\sum_{k=1}^n |f(a_k)|^p \right)^{1/p} \leq (nm^p)^{1/p} = n^{1/p} m \rightarrow 1 \cdot m = m$$

where from analysis, $n^{1/p} m \rightarrow 1$ as $p \rightarrow \infty$. The first inequality comes from the fact that the norm of the maximum will be less than the norm of the sum. The second inequality comes from the fact that the norm of the sum will be less than the norm of the maximum scaled by the number of elements n . Hence by squeeze law,

$$\left(\sum_{k=1}^n |f(a_k)|^p \right)^{1/p} \rightarrow m = \max_{k=1,\dots,n} |f(a_k)|.$$

So by letting $p \rightarrow \infty$, we see that the p -norm is simply the maximum of the terms.

Hence,

$$\lim_{p \rightarrow \infty} \|f\|_p = \max_{a \in X} |f(x)| = \|f\|_\infty$$

which is the supremum norm. The reason we have ∞ in subscript is that it is limit of the p -norm as $p \rightarrow \infty$.

Definition 19.164 (*Supremum norm*). We define the **supremum norm** as

$$\|f\|_\infty := \sup_{x \in X} |f(x)|.$$

In general, we expect that

$$\lim_{p \rightarrow \infty} \|f\|_p = \sup_{x \in X} |f(x)| = \|f\|_\infty.$$

An issue is that $\sup_{x \in X} |f(x)|$ is sensitive to changes on sets of measure zero whilst $\lim_{p \rightarrow \infty} \|f\|_p$ is not. Modifying sets of measure 0 can make the supremum arbitrary large. Hence, we define a new notion of supremum which ignores sets of measure 0 and hence it is more robust.

Let $f : X \rightarrow \mathbb{R}$ be a real valued function defined on a set X .

Definition 19.165 (*Upper Bound for functions*). A real number a is called an **upper bound** for f if $f(x) \leq a$ for all $x \in X$. In other words,

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

is empty. That is, there is no preimage x of f , such that $f(x) > a$.

We now let U_f be the set of upper bounds of f . Then, the supremum of the function f is defined by

$$\sup f = \inf U_f$$

if the set of upper bounds U_f is nonempty, and $\sup = \infty$ otherwise.

Now working on a measure space with measurable function f , we say that a number a is called an **essential upper bound** of f if the measurable set $f^{-1}(a, \infty)$ is a set of measurable zero. In other words $f(x) \leq a$ for almost all $x \in X$ (so $f(x)$ can be greater than a if x is only in a set of measure 0). We now let

$$U_f^{ess} = \{a \in \mathbb{R} : \mu(f^{-1}(a, \infty)) = 0\}$$

Then, the **essential supremum** of the function f is defined in a similar manner where we say

$$ess - \sup f = \inf U_f^{ess}$$

if the set of upper bounds U_f^{ess} is nonempty, and $ess - \sup = \infty$ otherwise. From this, we can now state it more concisely.

Definition 19.166 (*Essential Supremum*). Let (X, \mathcal{A}, μ) be a measure space and $f : X \rightarrow \mathbb{R}$ be measurable. We call the quantity

$$ess - \sup_{x \in X} |f(x)| = \inf \{\alpha \in \mathbb{R} : \mu(\{x \in X : f(x) > \alpha\}) = 0\}$$

the **essential supremum** of f over X with respect to the measure μ .

Definition 19.167 (*Essential Supremum Norm*). For $f : X \rightarrow \mathbb{K}$, we define the **essential supremum norm** by

$$\|f\|_\infty := ess - \sup_{x \in X} |f(x)|.$$

Here, note that the p -norm where $p = \infty$ is simply taking the essential supremum over the modulus of the function.

Definition 19.168 (*Space of essentially bounded functions*). We let

$$\mathcal{L}^\infty(X, \mathcal{A}, \mu, \mathbb{K}) := \{f : X \rightarrow \mathbb{K} : f \text{ is measurable and } \|f\|_\infty < \infty\}$$

be the space of (essentially) bounded functions.

As before, we then turn $\|\cdot\|_\infty$ into a proper norm by

$$L^\infty(X) = \frac{\mathcal{L}^\infty}{\mathcal{X}}$$

as we consider functions that are equal a.e. to now be equal.

Note that the essential supremum depends on the measure. Hence, we must always specify the measure we are working with. If we are working on \mathbb{R}^N , we can assume the Lebesgue measure.

Proposition 19.169 (*Properties of the essential-supremum*). Let $f : X \rightarrow \mathbb{R}$ be measurable.

Then, $m = \text{ess} - \sup_{x \in X} f(x)$ if and only if

- (i) $\mu(\{x \in X : f(x) > m\}) = 0$ and
- (ii) $\mu(\{x \in X : f(x) > m - \epsilon\}) > 0$ for all $\epsilon > 0$.

Proof: (ii) First, we let $A = \{x \in X : f(x) > m\} = \bigcup_{n \in \mathbb{N}} A_n$ where

$$A_n := \{x \in X : f(x) > m + \frac{1}{n}\}.$$

These are the set of preimages that are greater than the infimum and hence has a measure of 0. The countable union of sets with measure zero will have a measure zero $\bigcup_{n \in \mathbb{N}} A_n$ as $\mu(A) = 0$ if and only if $\mu(A_n) = 0$ for all $n \in \mathbb{N}$, which is the case here. From the definition of the essential-supremum, $m \geq \text{ess} - \sup_{x \in X} |f(x)|$ if and only if $\mu(A) = 0$.

(ii) This is true by the definition of the infimum in the essential-supremum. (ii) is only satisfied if and only if $m \leq \text{ess} - \sup_{x \in X} |f(x)|$. ■

Proposition 19.170 (*Properties of the essential-supremum with regards to sets of measure 0*). Let $f : X \rightarrow \mathbb{R}$ be measurable.

Then, set $m = \text{ess} - \sup_{x \in X} f(x)$ and let $N \in \mathcal{A}$ be a set with measure 0 and

$$\{x \in X : f(x) > m\} \subseteq N.$$

Then

- (i) $\text{ess} - \sup_{x \in X} f(x) = \sup_{x \in X/N} f(x)$ (*Pointwise supremum*);
- (i) $\text{ess} - \sup_{x \in X} f(x) \leq \sup_{x \in X} f(x)$.

Proposition 19.171 (*Properties of the essential-supremum on functions*). Let $f_1, f_2 : X \rightarrow \mathbb{R}$ be measurable and bounded functions such that $f_1 = f_2$ almost everywhere. Then,

$$\text{ess} - \sup_{x \in X} f_1(x) = \text{ess} - \sup_{x \in X} f_2(x).$$

Proof:(Sketch). We define $N := N_1 \cup N_2 \cup \{f_1(x) \neq f_2(x)\}$.

Then, it follows as $f_1(x) = f_2(x)$ for all $x \in X/N$ as $\text{ess} - \sup_{x \in X} f(x) = \sup_{x \in X/N} f(x)$. ■

Proposition 19.172 (*Properties of the essential-supremum on Lebesgue Measure*). Suppose $U \subseteq \mathbb{R}^N$ is open with μ as the Lebesgue measure. If $f : U \rightarrow \mathbb{R}$ is continuous, then

$$\text{ess} - \sup_{x \in U} f(x) = \sup_{x \in U} f(x).$$

Proposition 19.173 (*Essential Supremum Norm*). The essential supremum norm on \mathcal{L}^∞ is a **pseudo-norm** due to the relaxed property of $\|f\|_\infty \leq 0$ with equality if and only if $f(x) = 0$ for almost all $x \in X$.

Remark 19.174 However, by removing sets of measure 0, then the essential-supremum norm is a proper norm.

Theorem 19.175 (*Completeness of \mathcal{L}^∞*). Let $\{f_n\}$ be a Cauchy sequence in $\mathcal{L}^\infty(X)$. Then there exists $f \in \mathcal{L}^\infty(X)$ such that $f_n \rightarrow f$ in $\mathcal{L}^\infty(X)$. In other words, $\|f_n - f\|_\infty \rightarrow 0$ as $n \rightarrow \infty$. Moreover, there exists a set N of zero measure such that $f_n \rightarrow f$ uniformly on X/N .

Uniform convergence also implies pointwise convergence, as we now have uniform convergence once we remove sets of measure 0.

We now let $[f]$ be the equivalence class of all functions equal to f almost everywhere. We define

$$L^\infty(X, \mathcal{A}, \mu, \mathbb{K}) := \{[f] : f \in \mathcal{L}^\infty(X, \mathcal{A}, \mu, \mathbb{K})\}$$

Theorem 19.176 The space $L^\infty(X)$ is a complete normed space with respect to the norm $\|\cdot\|_\infty$ (essential-supremum norm).

Just to introduce some further terminology.

Definition 19.177 (*Sequence Space*). A sequence space is a **vector space** whose elements are infinite sequences of real or complex numbers. We denote the vectors as $\{x_n\}_{n \in \mathbb{N}}$ for $x_n \in \mathbb{K}$. Note that each element in the vector is a number in \mathbb{K} , but the entire vector itself is in $\mathbb{K}^\mathbb{N}$.

Definition 19.178 (*Function Space*). A function space is a **vector space** whose elements are a set of functions between two fixed sets. Often, the domain and/or codomain will have additional structure which is inherited by the function space.

Hence, \mathcal{L}^∞ is a sequence space, where its elements are bounded sequences. \mathcal{L}^∞ is a **Banach space**.

Furthermore, L^∞ is a function space, where its elements are *essentially bounded measurable functions*.

MATH3969: Measure Theory and Fourier Analysis

20. Fubini's and Tonelli's Theorem

20.4.6 Tonelli's Theorem

We are now interested in looking at taking Lebesgue integrals over \mathbb{R}^2 . First, we recall some things.

Definition 20.179 (*Bounded Functions*). A function is bounded on a set X if there exists $m \in \mathbb{R}$ such that

$$|f(x)| \leq M$$

for all $x \in X$.

Proposition 20.180 (*Criterion for Lebesgue Integrability*). Let f be a bounded function on a set of finite measure E .

f is Lebesgue integrable on E if and only if f is measurable.

In particular, f is Lebesgue integrable if f is measurable and $\int |f(x)| d\mu(x) < \infty$.

Now, to evaluate the integral of a function $f \in L^1(\mathbb{R}^2)$, we look at

$$\int_{\mathbb{R}^2} f(x, y) dm_2.$$

In particular, we evaluate the iterated integrals of

$$\int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dx \right) dy = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} f(x, y) dy \right) dx.$$

We need 2 conditions for the above statement to make sense.

- (i) (Inner Integral) $x \rightarrow f(x, y)$ needs to be measurable for all $y \in \mathbb{R}$;
- (ii) (Outer Integral) $y \rightarrow \int_{\mathbb{R}} f(x, y) dx$ needs to be measurable for all $y \in \mathbb{R}$.

Recall that if we fix y , if the function is measurable, then it is integrable. Note that condition (ii) for outer integral is an integral with a parameter with respect to y .

The issue is that condition (i) is not true for all $x \in \mathbb{R}$ and hence (ii) will not make sense. Fixing y does not mean that the function is also measurable. Generally, $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ is generally not measurable. However, it is measurable for almost all y ! Furthermore, (ii) is not defined on all y but fortunately, integrals do not see sets of measure 0 and hence we can set the function f to 0 for these sets. In fact, this is equivalent to taking the integral of the original function.

Recall from earlier in the course, we look at the MCT and DCT. We have 2 cases where if we had non-negative functions, we could use the MCT without any other assumptions on the functions as $\infty + \infty$ is

defined. However, for general functions, we could have $\infty - \infty$ which is not defined. Hence, we need to restrict it to cases of integrable functions and then we could apply the DCT.

Likewise, we have 2 theorems for taking iterated integrals, one for non-negative functions, which does not require the functions to be integrable, and another theorem for general functions but we now require the functions to be integrable.

Theorem 20.181 (Tonelli). *Let $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow [0, \infty]$ be measurable as a function on \mathbb{R}^{n+m} . Then there exists sets $N \subseteq \mathbb{R}^n$ and $M \subseteq \mathbb{R}^m$ of measure zero such that*

- (i) $y \rightarrow f(x, y)$ is measurable as a function on \mathbb{R}^m for all $x \in \mathbb{R}^n/N$;
- (ii) $x \rightarrow f(x, y)$ is measurable as a function on \mathbb{R}^n for all $y \in \mathbb{R}^m/M$.

If we modify f on the set $N \times M$ of measure zero by setting $f(x, y) := 0$ for all $x \in N$ and all $y \in M$, then

- (i) $y \rightarrow \int_{\mathbb{R}^n} f(x, y) dx$ is measurable as a function on \mathbb{R}^m ;
- (ii) $x \rightarrow \int_{\mathbb{R}^m} f(x, y) dy$ is measurable as a function on \mathbb{R}^n .

Finally,

$$\int_{\mathbb{R}^{n+m}} f(x, y) d(x, y) = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} f(x, y) dx \right) dy = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f(x, y) dy \right) dx.$$

In Tonelli's theorem, we do not assume f to be integrable. The identity asserts that either all three expressions are finite and equal OR they are all infinite and therefore also equal.

We now look at only finite cases and remove the restriction that $f(x, y)$ is non-negative. We have 3 different forms of integrability conditions where we substitute the assumption of non-negativity earlier with this.

Theorem 20.182 (Fubini). *Let $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{K}$ be measurable as a function on \mathbb{R}^{n+m} . We then assume that **one** of the integrals are finite*

$$\int_{\mathbb{R}^{n+m}} |f(x, y)| d(x, y) = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} |f(x, y)| dx \right) dy = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} |f(x, y)| dy \right) dx.$$

Then there exists sets $N \subseteq \mathbb{R}^n$ and $M \subseteq \mathbb{R}^m$ of measure zero such that

- (i) $y \rightarrow f(x, y)$ is measurable as a function on \mathbb{R}^m for all $x \in \mathbb{R}^n/N$;
- (ii) $x \rightarrow f(x, y)$ is measurable as a function on \mathbb{R}^n for all $y \in \mathbb{R}^m/M$.

If we modify f on the set $N \times M$ of measure zero by setting $f(x, y) := 0$ for all $x \in N$ and all $y \in M$, then

- (i) $y \rightarrow \int_{\mathbb{R}^n} f(x, y) dx$ is measurable as a function on \mathbb{R}^m ;
- (ii) $x \rightarrow \int_{\mathbb{R}^m} f(x, y) dy$ is measurable as a function on \mathbb{R}^n ;

Finally,

$$\int_{\mathbb{R}^{n+m}} f(x, y) d(x, y) = \int_{\mathbb{R}^m} \left(\int_{\mathbb{R}^n} f(x, y) dx \right) dy = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^m} f(x, y) dy \right) dx.$$

Taking $\int |f|$ is well defined as taking the modulus of a function gives us a non-negative function, where we can use Tonelli's theorem. Tonelli's theorem tells us the 3 integrability conditions are the same.

Fubini and Tonelli's theorem applies to any product measure.

Remark 20.183 *Note that in probability theory, Tonelli's theorem always holds as PDFs are always non-negative.*

21. Translation of a function

21.5 Convolution and Approximations

21.5.1 Translation of a function

Let $f : \mathbb{R}^N \rightarrow \mathbb{K}$ and fix a vector $t \in \mathbb{R}^N$ fixed.

Definition 21.184 (*Translation Operator*). We define the translation operator τ_t

$$\tau_t f(x) := f(x - t).$$

Remark 21.185 The translation operator τ_t is both

- (1) Linear. $\tau_t(\lambda f + \mu g) = \lambda \tau_t f + \mu \tau_t g$.
- (2) Translation invariant of p -norm with respect to Lebesgue measure. $\|\tau_t f\|_p = \|f\|_p$.

Theorem 21.186 (*Continuity of translation on $L^p(\mathbb{R}^N)$*). Let $1 \leq p < \infty$ and $f \in L^p(\mathbb{R}^N)$. Then

$$\lim_{t \rightarrow 0} \|\tau_t f - f\|_p = 0.$$

Proposition 21.187 (*Measurability of translations*). Let $f : \mathbb{R}^N \rightarrow \mathbb{K}$ be measurable. Then the following functions are measurable as functions on $\mathbb{R}^N \times \mathbb{R}^N$.

- (i) $F_1(x, y) := f(x)$;
- (ii) $F_2(x, y) := f(y - x)$.

Proof:(Sketch).

- (i) $f^{-1}[U] \times \mathbb{R}^N$ is measurable.
- (ii) Let $F_2 := F_1 \circ T$, where $T(x, y) = (y - x, y)$. Then recall that Linear transformation under Lebesgue measure is measurable. ■

From this, if we look at $(x, t) \rightarrow \tau_t f(x) = f(x - t)$ as a function of $(x, t) \in \mathbb{R}^N \times \mathbb{R}^N$, then this translation operator is measurable.

MATH3969: Measure Theory and Fourier Analysis

22. Convex functions and Jensen's Inequality

22.5.2 Convex Functions

Definition 22.188 (*Convex Function*). We call a function $\phi : (a, b) \rightarrow \mathbb{R}$ convex if

$$\phi(\lambda s + (1 - \lambda)t) \leq \lambda\phi(s) + (1 - \lambda)\phi(t)$$

for all $s, t \in (a, b)$ and all $\lambda \in (0, 1)$.

We introduce some properties on convex functions.

Proposition 22.189 The function $\phi : (a, b) \rightarrow \mathbb{R}$ is convex if and only if

$$\frac{\phi(t) - \phi(s)}{t - s} \leq \frac{\phi(u) - \phi(t)}{u - t}$$

whenever $a < s < t < u < b$.

Proposition 22.190 (*Subtangent*). If $\phi : (a, b) \rightarrow \mathbb{R}$ is convex, then for every $t \in (a, b)$, there exists $m \in \mathbb{R}$ such that

$$\phi(s) \leq \phi(t) + m(s - t) \leq \phi(s)$$

for all $s \in (a, b)$.

Remark 22.191 This property asserts that through every point of the graph of a convex function, there is a line below the graph known as the **subtangent**. For a **smooth** convex function, the graph lies above the tangent at every point.

Proposition 22.192 Every convex function is continuous.

Theorem 22.193 (*Jensen's Inequality*). Let (X, \mathcal{A}, μ) be a probability space (that is, μ is a measure with $\mu(X) = 1$). If $\phi : (a, b) \rightarrow \mathbb{R}$ is convex and $f : X \rightarrow (a, b)$ measurable, then

$$\phi\left(\int_X f d\mu\right) \leq \int_X \phi \circ f d\mu.$$

Proof: Since ϕ depends on $[a, b]$, we have that $a < f < b$ and $\mu(X) = 1$. We have that

$$a < \int_X f d\mu < b.$$

Now, applying the subtangent property of convex functions by letting $t := \int_X f d\mu$ and $s := f(\tau)$. Recall that we have that

$$\phi(t) \leq \phi(t) + m(s - t) \leq \phi(s)$$

now become

$$\phi\left(\int_X f d\mu\right) + m(f(\tau) - \int_X f d\mu) \leq \phi(f(\tau))$$

for all $\tau \in (a, b)$.

ϕ is continuous as it is convex and if measurable, hence $\phi \circ f$ is measurable. We integrate this inequality and use the fact that $\mu(X) = 1$.

$$\begin{aligned} \mu(X)\phi\left(\int_X f d\mu\right) + m\left(\int_X f d\mu - \mu(X) \int_X f d\mu\right) &\leq \phi \circ f d\mu \\ \phi\left(\int_X f d\mu\right) + m(0 - 0) &\leq \phi \circ f d\mu \\ \phi\left(\int_X f d\mu\right) &\leq \phi \circ f d\mu. \end{aligned}$$

■

Remark 22.194 What Jensen's inequality says is that applying a convex function to the average gives a smaller result than averaging the convex function itself.

We can apply Jensen's inequality for arbitrary measure spaces. However, we need to normalise our measure before we can apply Jensen's inequality.

Corollary 22.195 Let (X, \mathcal{A}, μ) be an arbitrary measure space. Let $1 \leq p < \infty$, $g \in L^1(X)$ and $f : X \rightarrow \mathbb{K}$ be measurable. Then,

$$\left(\int_X |fg| d\mu\right)^p \leq \|g\|_1^{p-1} \int_X |f|^p |g| d\mu.$$

Proof: First, we normalise the measure so that for any $A \in \mathcal{A}$, we define

$$v(A) := \frac{1}{\|g\|_1} \int_A |g| d\mu.$$

Now, we have the convex function of raising the power to p , $t \rightarrow t^p$, which from Jensen's inequality says that

$$\left(\int_X |f| dv\right)^p \leq \int_X |f|^p dv.$$

where $\phi = (f)^p$. Furthermore, note that we have defined the integral in terms of our newly construct measure v .

From this, we have that (X, \mathcal{A}, v) is a probability space. Hence, going back to our proof for Jensen's inequality

$$\mu(X)\phi\left(\int_X f d\mu\right) + m\left(\int_X f d\mu - \mu(X) \int_X f d\mu\right) \leq \phi \circ f d\mu$$

Applying what we have constructed

$$\begin{aligned} \left(\frac{1}{\|g\|_1} \int_X |fg| d\mu\right)^p &\leq \frac{1}{\|g\|_1} \int_X |f|^p |g| d\mu. \\ &= \frac{1}{\|g\|_1^p} \left(\int_X |fg| d\mu\right)^p \leq \frac{1}{\|g\|_1} \int_X |f|^p |g| d\mu. \end{aligned}$$

We have that

$$\left(\int_X |fg| d\mu\right)^p \leq \|g\|_1^{p-1} \int_X |f|^p |g| d\mu.$$

■

22.5.3 Convolution

Convolution is a kind of multiplication of two functions on \mathbb{R}^N . It is akin to a smoothing operator, which averages a function out. It takes an average around the point x .

Definition 22.196 (*Convolution*). Let $f, g : \mathbb{R}^N \rightarrow \mathbb{K}$ be measurable functions. For $x \in \mathbb{R}^N$, we define

$$(f * g)(x) := \int_{\mathbb{R}^N} f(x - y)g(y)dy$$

whenever the integral exists. We call $f * g$ the convolution of f and g .

Proposition 22.197 The convolution operator is commutative, that is

$$(f * g)(x) = (g * f)(x)$$

.

Proof: We use the translation invariance of integrals on \mathbb{R}^N and substitution of $z = x - y$.

$$\begin{aligned} (f * g)(x) &= \int_{\mathbb{R}^N} f(x - y)g(y)dy = \int_{\mathbb{R}^N} -f(z)g(x - z)dz \\ &= \int_{\mathbb{R}^N} -g(x - z)f(z)dz = \int_{\mathbb{R}^N} g(x - z)f(z)dz = g * f(x). \end{aligned}$$

■

Theorem 22.198 (*Young's Inequality*). Let $1 \leq p \leq \infty$. If $f \in L^p(\mathbb{R}^N)$ and $g \in L^1(\mathbb{R}^N)$, then

$$(f * g) \in L^p(\mathbb{R}^N)$$

and

$$\|f * g\|_p \leq \|f\|_p \|g\|_1.$$

Theorem 22.199 Let $1 \leq p, q \leq \infty$ such that $1/p + 1/q = 1$. If $f \in L^p(\mathbb{R}^N)$ and $g \in L^q(\mathbb{R}^N)$, then

1. $(f * g)$ in the space of bounded continuous function $BC(\mathbb{R}^N)$;
2. $\|f * g\|_\infty \leq \|f\|_p \|g\|_q$.

Proof: Use Hölder's inequality and substitution of $z = x - y$.

■

Hence, the convolution of $f * g$ can be continuous even if f and g are not, hence, the convolution of two functions are usually smoother than the original functions.

We can now define multiplication on $L^1(\mathbb{R}^N)$ as a convolution. However, there is no identity, that is, a function $g \in L^1(\mathbb{R}^N)$ such that $f * g = f$ for all $f \in L^1(\mathbb{R}^N)$. Remember that L^1 is a vector space, but what we just described is multiplication between vectors, which does not necessarily require an identity element. So in the next section, we introduce there is **almost** such an identity.

MATH3969: Measure Theory and Fourier Analysis

23. Approximate Identities

23.5.4 Approximate Identities

Approximate identities helps us approximates dirac delta functions. Let $\phi \rightarrow [0, \infty)$ be measurable such that

$$\int_{\mathbb{R}^N} \phi dx = 1.$$

We then form the sequence

$$\phi_\epsilon(x) := \frac{1}{\epsilon^N} \phi\left(\frac{x}{\epsilon}\right).$$

Recall the translation formula that $\int_{im T} f dx = |\det T| \int_{\mathbb{R}^N} f \circ T dx$. We apply this to the linear map $x \rightarrow nx$, we get

$$\int_{\mathbb{R}^N} \phi_\epsilon dx = 1$$

for all $n \in \mathbb{N}$. Here, the matrix is $\frac{1}{\epsilon} I^N$, so the determinant is just the diagonal of the matrix and hence the $\frac{1}{\epsilon^N}$ when applying the linear transformation of Lebesgue integrals.

We look at a quick intuition on why this is useful. Fix $x \in \mathbb{R}^N$ and consider $f(x - y)\phi_\epsilon(y)$. We have $f(x - y)\phi_\epsilon(y) \approx f(x)\phi_\epsilon(y)$ as ϕ_ϵ concentrates near zero for a large n . Here, we have that

$$\begin{aligned} (f * \phi_\epsilon)(x) &= \int_{\mathbb{R}^N} f(x - y)\phi_\epsilon(y) dy \approx \int_{\mathbb{R}^N} f(x)\phi_\epsilon(y) dy \\ &= f(x) \int_{\mathbb{R}^N} \phi_\epsilon(y) dy = f(x)(1) = 1. \end{aligned}$$

Hence, this motivates the following definition.

Definition 23.200 (*Approximate Identity*). Let $\phi : \mathbb{R}^N \rightarrow [0, \infty)$ be measurable with

$$\int_{\mathbb{R}^N} \phi dx = 1$$

and set $\phi_\epsilon(x) := \frac{1}{\epsilon^N} \phi\left(\frac{x}{\epsilon}\right)$ for all $x \in \mathbb{R}^N$ and all $n \in \mathbb{N}$. Then the family $\{\phi_\epsilon\}_{\epsilon > 0}$ is called an approximate identity.

We now have something similar to the identity element, whereby taking the convolution between a function and the identity element gives us something similar to the function. This is useful in many cases such as the heat equation, where the convolution will get us back to the initial state of the system.

Theorem 23.201 (*First Approximation Theorem*). Let $\{\phi_n\}$ be an approximate identity and $1 \leq p < \infty$. Then for all $f \in L^p(\mathbb{R}^N)$, we have that

1. $f * \phi \in L^p(\mathbb{R}^N)$;

2. We have that the convolution

$$f * \phi_\epsilon \rightarrow f$$

in $L^p(\mathbb{R}^N)$ as $n \rightarrow \infty$.

What's the significance of this? In the space $L^1(\mathbb{R})$, for functions f and g in this space, we have closure under the operation of convolution of functions $f * g \in L^1(\mathbb{R}^N)$. However, there is no identity function h , such that $f * h = f$ for all $f \in L^1(\mathbb{R}^N)$. However, ϕ_ϵ approximates an identity such that

$$f * \phi_\epsilon \rightarrow f$$

as $\epsilon \rightarrow 0$.

Now we have a theorem on pointwise convergence.

Theorem 23.202 (*Pointwise Convergence*). Let $\{\phi_\epsilon\}$ be an approximate identity defined by $\phi_{-\epsilon} = \epsilon^N \phi(\epsilon x)$, and that one of the following conditions is satisfied:

(i) $f \in L^\infty(\mathbb{R}^N)$ and f is continuous at x_0 ;

(ii) $f \in L^1(\mathbb{R}^N)$ and f is continuous at x_0 , and there exists $\alpha > 0$ and $C > 0$ such that

$$|\phi(x)| \leq \frac{C}{1 + |x|^{N+\alpha}}$$

for all $x \in \mathbb{R}^N$.

Under **either** assumption,

$$(f * \phi_\epsilon)(x_0) \rightarrow f(x_0)$$

as $n \rightarrow \infty$.

24. Approximate Theorems

24.5.5 Approximate Theorems

We use approximate identities to help us show that L^p functions can be approximated by smooth functions with **compact support**, which will be used alot.

Definition 24.203 (*Test Function*). Let $U \subseteq \mathbb{R}^N$ be open. We define

$$C^\infty(U, \mathbb{K}) := \{f : U \rightarrow \mathbb{K} : f \text{ has partial derivatives of all orders (smooth)}\}$$

and we have the subspace of C^∞ as

$$C_c^\infty(U, \mathbb{K}) := \{f \in C^\infty(U, \mathbb{K}) : \text{supp}(f) \subseteq U \text{ and } \text{supp}(f) \text{ is compact.}\}$$

The elements of $C_c^\infty(U, \mathbb{K})$ are called **test functions**.

This subspace C_c^∞ is a dense subset of L^p for $p < \infty$. Hence, we have many such test functions. However, we do not have any test functions in \mathbb{C} .

Definition 24.204 (*Locally Integrable*). The integral of every bounded set is finite.

Convolutions with test functions are smooth functions.

Proposition 24.205 Let $f : \mathbb{R}^N \rightarrow \mathbb{K}$ be measurable so that $\int_B |f| dx < \infty$ for every bounded set $B \subseteq \mathbb{R}^N$, so the function is locally integrable. If $\phi \in C_c^\infty(\mathbb{R}^N)$, then $f * \phi \in C^\infty(\mathbb{R}^N)$ and

$$\frac{\partial}{\partial x_i}(\phi * f) = \frac{\partial \phi}{\partial x_i} * f$$

for $i = 1, 2, \dots, N$. If f has compact support, then $f * \phi \in C_c^\infty(\mathbb{R}^N)$ and

$$\text{supp}(f * \phi) \subseteq \text{supp}(f) + \text{supp}(\phi).$$

Hence, convolutions always preserves the properties of the **better** function. Furthermore, convolution has a compact support.

We now prove one of the most important density results.

Theorem 24.206 (*Test Functions are dense in L^p*). If $U \subseteq \mathbb{R}^N$ is open, then $C_c^\infty(U)$ is dense in $L^p(U)$ for all $1 \leq p < \infty$.

Many functions $f \in L^p$ are not smooth. Hence, we first analyse dense subsets and then extend its properties to a wider set of functions.

25. The Fourier Transform

25.6 The Fourier Transform

25.6.1 The Fourier Transform

A fourier transform is an **integral transform**. That is, it produces another integral from an integral.

Definition 25.207 (*Fourier Transform*). Let $f \in \mathcal{L}^1(\mathbb{R}^N, \mathbb{C})$. We call the function \hat{f} the Fourier transform which is given by

$$\hat{f}(t) = \int_{\mathbb{R}^N} f(x) e^{-2\pi i x \circ t} dx$$

for all $t \in \mathbb{R}^N$. Here, $x \circ t$ is the dot product of $x, t \in \mathbb{R}^N$.

The Fourier transform can be thought of as an integral with parameter, whereby t is the parameter. We can check whether the theorem of continuity of integral with parameter holds by looking at

- (i) $t \rightarrow f(x) e^{-2\pi i x \cdot t}$ is measurable for all $t \in \mathbb{R}^N$;
- (ii) $x \rightarrow f(x) e^{-2\pi i x \cdot t}$ is continuous for all $x \in \mathbb{R}^N$;
- (iii) $|f(x) e^{-2\pi i x \cdot t}| = |f(x)|$ is independent of t and integrable for all $t \in \mathbb{R}^N$.

From this, we have that \hat{f} is continuous. Furthermore, from the definition of \hat{f} and $f \in \mathcal{L}^1$, we have that

$$\|\hat{f}\|_{\infty} \leq \|f\|_1 < \infty$$

so \hat{f} is bounded.

Hence, the **Fourier transform is a bounded continuous function**.

We look at a function whose Fourier transform function is itself.

Proposition 25.208 (*Fourier Transform of the Gaussian Function*). Let $\phi(x) := e^{-\pi|x|^2}$ for all $x \in \mathbb{R}^N$. Then, $\hat{\phi} = \phi$ and

$$\int_{\mathbb{R}^N} \phi(x) dx = 1.$$

26. Properties of the Fourier transform

26.6.2 Properties of the Fourier transform

We know that the Fourier transform \hat{f} of any L^1 -function f is a bounded and continuous function. We want to show that this decays to 0, that is, $\hat{f}(t) \rightarrow 0$ as $|t| \rightarrow \infty$.

First, we can split \hat{f} into its real and imaginary parts, that is,

$$\hat{f}(t) = \int_{\mathbb{R}} f(x) \cos(2\pi xt) dx - i \int_{\mathbb{R}} f(x) \sin(2\pi xt) dx.$$

Proposition 26.209

$\hat{f}(t) \rightarrow 0$ if and only if the real and imaginary parts converge to zero.

Intuitively, this arises because when you take f , which is a bounded function and multiply it by an oscillating function such as \sin or \cos , then you have an oscillating function. In particular, it is the $e^{-2\pi ixt}$ which is oscillating as it just goes around the circle. The negative and positive area cancels and hence you have an area of 0 overall. These are known as **oscillatory integrals**.

Definition 26.210 (*Space of continuous functions vanishing at infinity*). We set

$$C_0(\mathbb{R}^N, \mathbb{K}) := \{f \in C(\mathbb{R}^N, \mathbb{K}) : \lim_{|x| \rightarrow \infty} f(x) = 0\}$$

which are continuous functions and go to 0 as $x \rightarrow \infty$.

Theorem 26.211 (*Riemann-Lebesgue Lemma*). If $f \in L^1(\mathbb{R}^N, C)$, then $\hat{f} \in C_0(\mathbb{R}^N, C)$ and $\|\hat{f}\|_{\infty} \leq \|f\|_1$.

Proof: We know that \hat{f} is continuous as $\|\hat{f}\|_{\infty} \leq \|f\|_1$. We only need to show $\hat{f}(t) \rightarrow 0$ as $|t| \rightarrow \infty$. ■

Proposition 26.212 Let $f, g \in L^1(\mathbb{R}^N)$. Then

$$\int_{\mathbb{R}^N} \hat{f} g dx = \int_{\mathbb{R}^N} f \hat{g} dx.$$

Proof: Use Fubini's theorem and the definition of the Fourier transform. ■

Theorem 26.213 (*Convolution Theorem*). If $f, g \in L^1(\mathbb{R}^N)$, recall that $f * g \in L^1(\mathbb{R}^N)$. Then, we have that

$$\widehat{f * g} = \hat{f} \hat{g}.$$

That is, the Fourier transform turns convolutions into pointwise products.

27. Further Properties of the Fourier transform

27.6.3 Further Properties of the Fourier transform

We are now interested in recovering f given its Fourier transform \hat{f} . There are some issues as \hat{f} is not necessarily integrable. Hence, what we need to do is to introduce a fast decaying factor to make the integral converge, and ensure that factor converges to 1.

Lemma 27.214 *Let $\phi \in L^1(\mathbb{R}^N)$ and set $\phi_\epsilon := \phi(\epsilon x)$ and $\psi_\epsilon(x) := \frac{1}{\epsilon^N} \hat{\phi}(\frac{x}{\epsilon})$. Then, we have*

$$\int_{\mathbb{R}^N} \hat{f}(t) \phi_\epsilon(t) e^{2\pi i x \cdot t} dt = (f * \psi_\epsilon)(x)$$

for all $n \in \mathbb{N}$ and $x \in \mathbb{R}^N$.

Here, $\phi_\epsilon \rightarrow 1$ pointwise is the fast decaying term whilst ψ_ϵ is an approximate identity. One good candidate for $\phi(x) = e^{-\pi|x|^2}$ because it is symmetric, coincides with its own Fourier transform, and has an integral of 1. Furthermore, we can set $\phi_\epsilon(x) := \frac{1}{\epsilon^N} \phi(\frac{x}{\epsilon})$.

Theorem 27.215 (*Fourier Inversion Formula*). *Let $f \in L^1(\mathbb{R}^N, \mathbb{C})$, $\phi_\epsilon := \frac{1}{\epsilon^N} \phi(\frac{x}{\epsilon})$, and $\phi(x) := e^{-\pi|x|^2}$. Then, we have that*

$$\int_{\mathbb{R}^N} \hat{f}(\psi) \phi(\epsilon t) e^{2\pi i x \cdot t} dt = (f * \phi_\epsilon)$$

for all $\epsilon > 0$ and for all $x \in \mathbb{R}^N$. Moreover,

$$\lim_{\epsilon \rightarrow 0^+} \int_{\mathbb{R}^N} \hat{f}(\psi) \phi(\epsilon t) e^{2\pi i x \cdot t} dt = f$$

pointwise limit.

If \hat{f} is integrable, then we have an inversion formula that is better.

Corollary 27.216 *Let $f \in L^1(\mathbb{R}^N)$. If $\hat{f} \in L^1(\mathbb{R}^N)$, then*

$$f(x) = \int_{\mathbb{R}^N} \hat{f}(t) e^{2\pi i x \cdot t} dt$$

for almost every $x \in \mathbb{R}^N$.

Fourier transforms uniquely determines f .

Corollary 27.217 *Suppose that $f, g \in L^1(\mathbb{R}^N)$ are such that $\hat{f} = \hat{g}$. Then $f = g$ almost everywhere.*

Remark 27.218 *This holds as continuous functions are either equal everywhere or not equal almost everywhere, which we can then alter the sets.*

MATH3969: Measure Theory and Fourier Analysis

28. Fourier Transform in L^2

28.6.4 Fourier Transform in L^2

We now want to define the Fourier transform on the $L^2(\mathbb{R}^N)$ space rather than just $L^1(\mathbb{R}^N)$. Recall that $L^2(\mathbb{R})$ is a Hilbert space and has an inner product. An issue we have is that we don't have $L^2(\mathbb{R}^N) \subseteq L^1(\mathbb{R}^N)$ or $L^1(\mathbb{R}^N) \subseteq L^2(\mathbb{R}^N)$ since $\mu(X) = \infty$.

So the trick on constructing a Fourier transform on $L^2(\mathbb{R}^N)$ is to define a map on a **dense subset** and then extend it to the closure of that subset.

We define the Fourier transform for $f \in L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$ by

$$\int_{\mathbb{R}^N} f(x) e^{-2\pi i x \cdot t} dx.$$

The test functions $C_c^\infty(\mathbb{R}^N)$ (smooth functions with compact supports) are dense in $L^1(\mathbb{R}^N)$ and in $L^2(\mathbb{R}^N)$. Hence, $L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$ is dense in $L^2(\mathbb{R}^N)$. We will first show the Fourier transform is continuous on $L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$ with respect to the L^2 -norm and then extend to all of $L^2(\mathbb{R}^N)$ so that it is a continuous function on $L^2(\mathbb{R}^N)$.

Theorem 28.219 (*Plancherel*). *If $f \in L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$, then*

$$\|\hat{f}\|_2 = \|f\|_2.$$

This shows the continuity of the Fourier transform on $L^2(\mathbb{R}^N)$.

Remark 28.220 *The Fourier transform $\mathcal{F} : L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N) \rightarrow L^2(\mathbb{R}^N)$ is a continuous linear operator. We can look at*

$$\|\hat{f} - \hat{g}\|_2 = \|\widehat{f - g}\|_2 = \|f - g\|_2$$

for all $f, g \in L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$. We can use the δ - ϵ argument by setting $\delta = \epsilon$ and this will show us uniform continuity.

Theorem 28.221 (*Fourier Transform on $L^2(\mathbb{R}^N)$*). *There exists a unique continuous linear map*

$$\mathcal{F} : L^2(\mathbb{R}^N) \rightarrow L^2(\mathbb{R}^N)$$

such that $\mathcal{F}f = \hat{f}$ for all $f \in L^1(\mathbb{R}^N) \cap L^2(\mathbb{R}^N)$. Moreover, $\|\mathcal{F}f\|_2 = \|f\|_2$ for all $f \in L^2(\mathbb{R}^N)$. If $\phi(x) := e^{-\pi|x|^2}$, then for every $f \in L^2(\mathbb{R}^N)$,

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^N} f(x) \phi(x/n) e^{-2\pi i x \cdot t} dx = \hat{f}$$

and

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}^N} \hat{f}(t) \phi(t/n) e^{2\pi i x \cdot t} dt = f$$

in $L^2(\mathbb{R}^N, \mathbb{C})$.

Theorem 28.222 *The Fourier transform $\mathcal{F} : L^2(\mathbb{R}^N) \rightarrow L^2(\mathbb{R}^N)$ is an isometric isomorphism. Moreover,*

$$(\mathcal{F}^{-1}f)(x) = (\mathcal{F}f)(-x) = \overline{(\mathcal{F}\bar{f})(x)}$$

almost everywhere, where the bars denote complex conjugate. Furthermore, from Plancherel, we have

$$\|\hat{f}\|_2 = \|f\|_2.$$

MATH3969: Measure Theory and Fourier Analysis

29. Hilbert Spaces

29.7 Absolute Continuity and the Radon-Nikodym Theorem

29.7.1 Hilbert Spaces

We need some Hilbert space theory as we require the projection theorem in order to help us derive further theorems.

Definition 29.223 (*Hilbert Space*). A Hilbert space is a **complete** inner product space.

Remark 29.224 Recall that an inner product induces a norm where $\|u\| := \sqrt{\langle u, u \rangle}$. Furthermore, a space is complete if every Cauchy sequence with respect to the norm converges.

A good example of an infinite dimensional Hilbert space is the $L^2(X)$ space. Recall that every $L^p(X)$ space is complete. This completeness property is what distinguishes Hilbert spaces from the usual inner product space.

The Hilbert spaces satisfies an useful identity.

Lemma 29.225 (*Parallelogram identity*). Let H be an inner product space. Then

$$\|u + v\|^2 + \|u - v\|^2 = 2\|u\|^2 + 2\|v\|^2$$

for all $u, v \in H$.

Theorem 29.226 (*Projection Theorem*). Let H be a Hilbert space and $M \subseteq H$ a closed subspace, so any sequence in M has its limit in M . Then, for every $u \in H$, there exists a unique $m_0 \in M$ such that

$$\|u - m_0\| = \min_{m \in M} \|u - m\|.$$

Moreover, m_0 satisfies this equation if and only if

$$\langle u - m_0, m \rangle = 0$$

for all $m \in M$. In other words, m_0 is orthogonal to every point in M .

The intuition behind this is that if we take a plane and a point in the space outside of the plane, there exists exactly 1 point on the plane that is closest to the outside point.

Definition 29.227 (*Bounded linear functional*). Let H be a Hilbert space over \mathbb{K} . A function $\phi : H \rightarrow \mathbb{K}$ is bounded linear functional on H if

1. ϕ is linear;

2. There exists a constant $C > 0$ such that

$$|\phi(f)| \leq C \|f\|$$

for all $f \in H$.

Remark 29.228 *Bounded linear functionals are linear maps and maps bounded sets of H onto bounded sets in \mathbb{K} . Furthermore, every bounded linear functional on H is continuous.*

The following theorem tells us that every linear functional on H can be represented by means of the inner product of H with a unique element.

Theorem 29.229 *(Riesz representation theorem). Let ϕ be arbitrary bounded linear functional on H . Then there exists a unique $g \in H$, so that $\phi(f) = \langle f, g \rangle$ for all $f \in H$.*

MATH3969: Measure Theory and Fourier Analysis

30. Radon-Nikodym Theorem

30.7.2 Absolute Continuity

We now want to define conditions for measures to exist with respect to a density. First, recall that if (X, \mathcal{A}, μ) is a measure space and $g : X \rightarrow [0, \infty)$ be a measurable function. Recall we can define a measure with respect to another measure by

$$\nu(A) := \int_A g d\mu = \int_X 1_A g d\mu$$

for all $A \in \mathcal{A}$. We call g the **density** of ν with respect to μ .

It is not possible to take 2 arbitrary measures and find a density. A good example is the Dirac measure concentrated at some $a \in X$ and the Lebesgue measure. Hence, we introduce a sufficient and necessary condition for the existence of such a density.

Definition 30.230 (*Absolute Continuity*). Let μ and ν be measures defined on the same σ -algebra of subsets of X . We call ν **absolutely continuous** with respect to μ if $\nu(A) = 0$ for every $A \in \mathcal{A}$ where $\mu(A) = 0$. We denote this as

$$\nu \ll \mu.$$

Remark 30.231 Absolute continuity can also be defined in terms of a $\delta - \epsilon$ argument. For every $\epsilon > 0$, there exists $\delta_\epsilon > 0$ such that

$$\sum_{k=0}^{\infty} |F(b_k) - F(a_k)| < \epsilon$$

for every collection of disjoint intervals (a_k, b_k) such that $\sum_{k=0}^{\infty} (b_k - a_k) < \delta$.

Remark 30.232 Absolute continuity is a stronger condition than uniform continuity.

Definition 30.233 (*Distribution Function*). The distribution function $F(t) : \mathbb{R} \rightarrow \mathbb{R}$ is defined as

$$F(t) := \mu((-\infty, t])$$

where $F(t)$ is right continuous and strictly increasing.

A measure is absolutely continuous if it has no jumps in the distribution function, which leads to the distribution function being continuous.

Remark 30.234 The Borel measure is absolutely continuous with respect to the Lebesgue measure if and only if the distribution function $F(t) := \mu((-\infty, t])$ is absolutely continuous.

30.7.3 Radon-Nikodym Theorem

We now look at a special case of a density existing for a pair of arbitrary measures.

Lemma 30.235 *Suppose μ and ν are finite measures. We define the new measure $\lambda := \mu + \nu$. Then there exists a non-negative function $h \in L^2(X, \lambda)$ such that*

$$\int_X f d\nu = \int_X f h d\lambda$$

for all $f \in L^2(X, \lambda)$.

Theorem 30.236 *(Radon-Nikodym Theorem). Suppose that $\mu(X) < \infty$ and $\nu(X) < \infty$ defined on the same σ -algebra \mathcal{A} . If $\nu \ll \mu$, then ν has a density $g : X \rightarrow [0, \infty)$ with respect to μ . Furthermore, the density is essentially unique, that is, any two density functions are equal μ -almost everywhere.*

Corollary 30.237 *The Radon-Nikodym is also known as a differentiation theorem and can be expressed as*

$$\int_A d\nu = \int_A \frac{d\nu}{d\mu} d\mu$$

where $\frac{d\nu}{d\mu}$ is the density function of the distribution function and $d\mu$ is the Lebesgue measure. This generalises the Fundamental Theorem of Calculus.

Definition 30.238 *(σ -finite measure). A measure is called σ -finite if there exists subsets $X_n \in \mathcal{A}$ such that $\mu(X_n) < \infty$ for all $n \in \mathbb{N}$ and $X = \bigcup_{n \in \mathbb{N}} X_n$.*

Remark 30.239 *In the Radon-Nikodym theorem, we assumed that μ and ν are finite measures. This condition can be relaxed to μ and ν being σ -finite instead.*

Proposition 30.240 *If h is a smooth function, then by the substitution formula*

$$\int_{\mathbb{R}} f(h) dh = \int_{\mathbb{R}} f(x) h'(x) dx$$

so that h' is the density of the measure dh with respect to the Lebesgue measure.

Proposition 30.241 *(Radon-Nikodym Derivative). Let $\mu \ll \nu$ and we can write*

$$\int_X f d\nu = \int_X f \frac{d\nu}{d\mu} d\mu.$$

If we let

$$\frac{d\nu}{d\mu} := g,$$

where g is the density function from the Radon-Nikodym theorem. Here, g is known as the **Radon-Nikodym derivative** of ν with respect to μ , which exists when $\mu \ll \nu$.

MATH3969: Measure Theory and Fourier Analysis

31. Introduction to Probability Theory

31.8 Probability Theory

31.8.1 Introduction to Probability Theory

We compare the different terminologies in measure theory and probability theory.

The sample space Ω is usually a finite set with the probability measure being the normalised counting measure where

$$P[A] = \frac{\text{Number of favourable outcomes}}{\text{Number of possible outcomes}} = \frac{\#A}{\#\Omega}$$

where $\#A$ denotes the cardinality of A .

Definition 31.242 (*Random Variable*). Let (σ, \mathcal{A}, P) be a probability space. A measurable function

$$X : \Omega \rightarrow \mathbb{R}$$

is called a random variable.

Remark 31.243 Recall that a measurable function is a function between two **measurable spaces** such that the preimage of any measurable set is measurable.

31.8.2 Definitions

Measure Theory	Probability Theory
(X, \mathcal{A}, μ) is a measure space	(Ω, \mathcal{A}, P) is a probability space
\mathcal{A} is a σ -algebra	\mathcal{A} is a σ -field
$A \in \mathcal{A}$ is called a measurable set	$A \in \mathcal{A}$ is called an event
f is a measurable function	X is a random variable
$\int_X f d\mu$ is an integral	$\mathbb{E}[X]$ is an expectation
$f \in L^p(X, \mathcal{A}, \mu)$	X has finite absolute p -th moment

nothing random about a random variable. However, the thing that is random about it is the choice of the sample point ω .

We are interested in $X(\omega)$. In particular, we are interested in the sets

1. $[X = \alpha] := \{\omega \in \Omega : X(\omega) = \alpha\};$
2. $[X < \alpha] := \{\omega \in \Omega : X(\omega) < \alpha\};$

3. $[X > \alpha] := \{\omega \in \Omega : X(\omega) > \alpha\};$
4. $[X \in A] := \{\omega \in \Omega : X(\omega) \in A\}.$

Recall that if we can show that $f^{-1}[[\alpha, \infty]]$ is measurable for all $\alpha \in \mathbb{Q}$, then this shows that $X : \Omega \rightarrow \mathbb{R}$ is measurable.

Definition 31.244 (*Expectation of random variable*). Let (Ω, \mathcal{A}, P) be a probability space and $X : \Omega \rightarrow \mathbb{R}$ be a random variable. We say that X has finite expectation if $X \in L^1(\Omega, \mathcal{A}, P; \mathbb{R})$ and call

$$E[X] := \int_{\Omega} X dp$$

the expectation of X .

Definition 31.245 (*Variance of random variable*). Let (Ω, \mathcal{A}, P) be a probability space and $X : \Omega \rightarrow \mathbb{R}$ be a random variable. We say that X has finite expectation if $X \in L^2(\Omega, \mathcal{A}, P; \mathbb{R})$ and call

$$\text{Var}(X) := E[(X - E[X])^2] := \|X - E[X]\|_2^2$$

the variance of X and $\sigma(X) := \sqrt{\text{Var}(X)}$ the standard deviation of X .

Remark 31.246 If $X \in L^p(\Omega, \mathcal{A}, P; \mathbb{R})$, we say that X has (absolute) finite p -th moment.

Remark 31.247 Recall that when we write dP , it is the Lebesgue integral with respect to the probability measure.

Remark 31.248 Anything with a finite variance has a finite moment since $P(\Omega) = 1 < \infty$. In particular, $L^p(\Omega, P) \subseteq L^1(\Omega, P)$ for all $p \geq 1$.

Remark 31.249 If $X \in L^2(\Omega, P)$, then

$$\text{Var}(X) = E[X^2] - E[X]^2$$

where we require a finite first moment.

MATH3969: Measure Theory and Fourier Analysis

32. Further Probability Theory

32.8.3 Distribution of a random variable

Now the issue we have is that when computing expectation or variance, we are integrating over Ω . However, what we want to do instead is to integrate over \mathbb{R} and we can do this through the **distribution of X** . In particular, we tend to hide the random variable and only use its distribution.

Recall that a function is Borel measurable if we had that any preimage of a Borel set is a Borel set. This then implies that the function is Lebesgue measurable. Furthermore, a measurable function has the preimage of every Borel set in the σ -algebra. In particular, we say that for every Borel set $B \subseteq \mathbb{R}$,

$$P_X[B] := P[X \in B] = P[\{\omega \in \Omega : X(\omega) \in B\}].$$

From this, we know that P_X is a Borel measure on \mathbb{R} as it is defined on a Borel σ -algebra. Then, if P_X is absolutely continuous with respect to the Lebesgue measure (dM), we have that from the Radon-Nikodym theorem, P_X has a density g which is a measurable non-negative function such that

$$P_X[B] = \int_B g(x) dM = \int_B g(x) dx$$

for every Borel set $B \subseteq \mathbb{R}$. The last equality comes from that fact that anything Lebesgue integrable is also Riemann integrable.

We go on to state these things more concisely.

Definition 32.250 (*Distribution of a random variable*). Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable on the probability space (Ω, \mathcal{A}, P) . The Borel measure P_X on \mathbb{R} given by

$$P_X[B] := P[X \in B] = \mu[\{\omega \in \Omega : X(\omega) \in B\}]$$

for all Borel sets $B \subseteq \mathbb{R}$ is called the *distribution of X* .

What have we done is started of with a probability space (X, \mathcal{A}, P) as the domain with \mathbb{R} as the codomain from the measurable random variable X . Then, we were able to generate the \mathcal{B} σ -algebra on \mathbb{R} from our random variable X . Finally, we constructed a Borel measure P_X , which is our original probability measure P restricted on the constructed Borel σ -algebra \mathcal{B} .

If P_X gives a measure one to a countable set of reals, then X is a discrete random variable. If P_X gives a measure 0 to every singleton and countable set, then X is a continuous random variable.

Definition 32.251 (*Distribution Function*). Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable on the probability space (Ω, \mathcal{A}, P) . The function

$$F_X(t) := P_X[(-\infty, t]] = P(X \leq t) = P(\{\omega \in \Omega : X(\omega) \leq t\}).$$

is called the *distribution function of X* .

Remark 32.252 *The distribution function is increasing and right continuous. Furthermore, $F_X(t) = 0$ as $t \rightarrow -\infty$ and $F_X(t) = 1$ as $t \rightarrow \infty$.*

Any F_X satisfying these properties are the distribution function of another random variable.

We tend to describe random variables in terms of their distribution function.

Definition 32.253 *(Distribution Density). Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable on the probability space (Ω, \mathcal{A}, P) . If P_X is absolutely continuous with respect to the Lebesgue measure and g is the Radon-Nikodym derivative, that is measurable and non-negative, then g is called the distribution density of X .*

$$P_X[B] = \int_B g(x)dx = \int_{\mathbb{R}} 1_B g(x)dx.$$

Remark 32.254 *Note that P is never absolutely continuous with P_X since they are defined on different σ -algebra. The Lebesgue measure is defined on the Lebesgue σ -algebra, which contains the Borel σ -algebra \mathcal{B} in which P_X is defined on.*

Usually for random variables, we only know its distribution but not the random variable in many cases. Hence, we want to be able to express expectation and variance in terms of the random variable's distribution.

Remark 32.255 *There is a injection between distributions and distribution functions. However, note that many different random variables can have the same distribution and distribution function.*

A key thing to note is that not all distributions have a density. In particular, if a distribution function is clearly not absolutely continuous with respect to the Lebesgue measure, then it has no density function.

Example 32.256 *We say that a random variable has a normal distribution if it has the Gaussian function as its density.*

We want to be able to substitute integrals with respect to Ω and P with integrals with respect to \mathbb{R} and P_X . We define some things to help us.

Lemma 32.257 *Let $X : \Omega \rightarrow \mathbb{R}$ be a random variable and $f : \mathbb{R} \rightarrow \mathbb{R}$ Borel measurable. Then $f \circ X : \Omega \rightarrow \mathbb{R}$ is measurable and therefore a random variable.*

Proof: Let $U \subseteq \mathbb{R}$ be open. As f is Borel measurable $f^{-1}[U]$ is a Borel set. As X is measurable, this implies that

$$(f \circ X)^{-1}[U] = X^{-1}[f^{-1}[U]]$$

is measurable for every open set $U \subseteq \mathbb{R}$. Hence, $f \circ X$ is measurable. ■

From this, we can prove the substitution formula for where we express integrals with respect to P over Ω in terms of integrals over \mathbb{R} in terms of P_X .

We have 2 formulations of the theorem requiring different conditions, one where we don't require integrability and one where we require integrability but not non-negativity.

Theorem 32.258 *Let (Ω, \mathcal{A}, P) be a probability space and $X : \Omega \rightarrow \mathbb{R}$ a random variable. Suppose that $f : \mathbb{R} \rightarrow \mathbb{R}$ is Borel measurable.*

(i) If f is non-negative, then

$$\int_{\Omega} f \circ X dp = \int_{\mathbb{R}} f dP_X$$

;

(ii) We have $f \circ X \in L^1(\Omega, P)$ if and only if $f \in L^1(\mathbb{R}, P_X)$. In that case,

$$\int_{\Omega} f \circ X dp = \int_{\mathbb{R}} f dP_X$$

.

Remark 32.259 Hence, we now have integrals of f with respect to the distribution.

From this, we can now express expectation and variance in terms of the distribution function.

Corollary 32.260 Let X be a random variable with distribution P_X . Then

$$E[X] = \int_{\mathbb{R}} x dP_x$$

and

$$Var(X) = \int_{\mathbb{R}} x^2 dP_X - E[X]^2.$$

If P_X has a distribution density g , then the above can be written as

$$E[X] = \int_{\mathbb{R}} xg(x)dx$$

and

$$Var(X) = \int_{\mathbb{R}} x^2 g(x)dx - E[X]^2.$$

Proof: We define the identity function $id : \mathbb{R} \rightarrow \mathbb{R}$ and let $f = id$.

$$\int_{\Omega} X dp = \int_{\Omega} id \circ X dp = \int_{\mathbb{R}} id(x) dP_X = \int_{\mathbb{R}} x dP_X.$$

■

Hence, we only need to know the distribution of X and not X itself in order to compute expectation and variance.

MATH3969: Measure Theory and Fourier Analysis

33. Conditional Expectation

33.8.4 Conditional Expectation

Let (Ω, \mathcal{A}, P) be a probability space and X a random variable with finite expectation. That is, $X \in L^1(\Omega, P)$. For $A \in \mathcal{A}$, we define

$$E[X|A] := \frac{1}{P[A]} \int_A X dp.$$

Let $A, B \in \mathcal{A}$ be disjoint sets such that $A \cup B = \Omega$. Then we have that

$$E[X|A] := \frac{1}{P[A]} \int_A X dp.$$

$$E[X|B] := \frac{1}{P[B]} \int_B X dp.$$

We then have that

$$E[X] = E[X|A]P[A] + E[X|B]P[B] = \int_{\Omega} E[X|A]1_A + E[X|B]1_B dP.$$

We define

$$X_0 := E[X|A]1_A + E[X|B]1_B dP.$$

Definition 33.261 (*Conditional Expectation*). Let (Ω, \mathcal{A}, P) be a probability space and $X : \Omega \rightarrow \mathbb{R}$ a random variable with finite expectation, that is, $X \in L^1(\Omega, \mathcal{A}, P)$. Moreover, let $\mathcal{A}_0 \subseteq \mathcal{A}$ be a σ -field. The random variable X_0 is called conditional expectation of X given \mathcal{A}_0 if

1. X_0 is measurable with respect to the σ -field \mathcal{A}_0 ;
2. $\int_A X_0 dP = \int_A X dP$ for all $A \in \mathcal{A}_0$.

If that is the case, we write

$$E[X|\mathcal{A}_0] := X_0.$$

Theorem 33.262 (*Conditional Expectation*). For every random variable $X \in L^1(\Omega, \mathcal{A}, P)$ and every σ -field $\mathcal{A}_0 \subseteq \mathcal{A}$, the conditional expectation $E[X|\mathcal{A}_0]$ exists and is unique.

Theorem 33.263 (*Properties of Conditional Expectation*). The conditional expectation has the following properties.

1. The map $X \rightarrow E[X|\mathcal{A}_0] : L^1(\Omega, \mathcal{A}, P) \rightarrow L^1(\Omega, \mathcal{A}_0, P)$ is linear;
2. $\|E[X|\mathcal{A}_0]\|_1 \leq \|X\|_1$;
3. If $X \in L^\infty(\Omega, \mathcal{A}, P)$, then $\|E[X|\mathcal{A}_0]\|_\infty \leq \|X\|_\infty$;

4. If $X \geq 0$, then $E[X|\mathcal{A}_0] \geq 0$;
5. If $X \leq Y$, then $E[X|\mathcal{A}_0] \leq E[Y|\mathcal{A}_0]$

Corollary 33.264 *If X is \mathcal{A}_0 -measurable, then*

$$E[X|\mathcal{A}_0] = X$$

almost everywhere.

Corollary 33.265 *If $\mathcal{A}_0 = \{\emptyset, \Omega\}$ is the smallest possible σ -field, then*

$$E[X|\mathcal{A}_0] = E[X].$$

Generally we work with random variables in $L^2(\Omega, \mathcal{A}, P)$. Recall that $L^2(\Omega, \mathcal{A}, P)$ is a Hilbert space and for $\mathcal{A}_0 \subseteq \mathcal{A}$, we have that every \mathcal{A}_0 -measurable function is \mathcal{A} -measurable. Then,

$$L^2(\Omega, \mathcal{A}_0, P) \subseteq L^2(\Omega, \mathcal{A}, P)$$

is a closed subspace.

Conditional expectation is then just the orthogonal projection into the closed subspace.

Proposition 33.266 *If $X \in L^2(\Omega, \mathcal{A}, P)$, then $E[X|\mathcal{A}_0]$ is the orthogonal projection of X onto the closed subspace $L^2(\Omega, \mathcal{A}_0, P)$ of $L^2(\Omega, \mathcal{A}, P)$.*

Proof: Let X_0 be the orthogonal projection of X onto $L^2(\mathcal{A}_0)$. Furthermore, $1_A \in L^2(\mathcal{A}_0)$. Then,

$$0 = \langle X - X_0, 1_A \rangle_{L^2} = \int_{\Omega} (X - X_0) 1_A dP = \int_A X dP - \int_A X_0 dP$$

. Since X_0 is \mathcal{A}_0 measurable and

$$\int_A X dP = \int_A X_0 dP$$

for all $A \subseteq \mathcal{A}_0$. By definition, X_0 is the conditional expectation $E[X|\mathcal{A}_0]$. ■

Proposition 33.267 *Let $X \in L^1(\Omega, \mathcal{A}, P)$ and $\mathcal{A}_0 \subseteq \mathcal{A}$ as σ -field. Then the following statements are equivalent.*

1. $X_0 = E[X|\mathcal{A}_0]$;
2. If X_0 is \mathcal{A}_0 and $E[X_0 Y] = E[XY]$ for all $Y \in L^\infty(\Omega, \mathcal{A}_0, P)$.

Moreover,

$$E[XY|\mathcal{A}_0] = Y E[X|\mathcal{A}_0]$$

for all $Y \in L^\infty(\Omega, \mathcal{A}_0, P)$.

Corollary 33.268 *Let $X, Y \in L^1(\Omega, \mathcal{A}, P)$ be two random variables. We define the σ – algebra*

$$\mathcal{A}_Y := \{Y^{-1}[B] : B \subseteq \mathbb{R} \text{ a Borel set}\}$$

to be a σ -field in Ω and $\mathcal{A}_Y \subseteq \mathcal{A}$. We then define the conditional expectation of a random variable on another random variable as

$$E[X|Y] := E[X|\mathcal{A}_Y].$$

Furthermore, from the definition of conditional expectation, we define the law of total expectation as

$$E[E[X|Y]] = E[X].$$