

Chapter 0

Preliminaries from Measure Theory

0.1. Algebras and σ -Algebras

Throughout this section, we use X to denote an arbitrary set. If A is a subset of X , we write $A^c = X \setminus A$ for the set-complement of A relative to X . We also define $\mathcal{P}(X)$ to be the power set of X , that is, the set formed by all subsets of X .

Definition 1 (Algebras and σ -algebras). Let $\mathcal{A} \subseteq \mathcal{P}(X)$ be a collection of subsets of X . Then $\mathcal{A}$ is called an *algebra over X* if it satisfies:

- (i) (Contains the empty set as an element) We have $\emptyset \in \mathcal{A}$;
- (ii) (Closed under complements) If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$;
- (iii) (Closed under finite unions) If $A_1, \dots, A_k \in \mathcal{A}$, then $A_1 \cup \dots \cup A_k \in \mathcal{A}$.

Moreover, $\mathcal{A}$ is called a *σ -algebra over X* if in addition to being an algebra, it also satisfies:

- (iv) (Closed under countable unions) If $\{A_n\}_{n \in I} \subseteq \mathcal{A}$, $I \subseteq \mathbb{N}$, is a countable family of sets in $\mathcal{A}$, then $\bigcup_{n \in I} A_n \in \mathcal{A}$.

If $\mathcal{A}$ is an algebra (or a σ -algebra) of subsets of X , then a subset of X is said to be *$\mathcal{A}$ -measurable* if it belongs to $\mathcal{A}$.

Here are some first examples of algebras over a set X .

Example 2 (Algebras).

- The collection of subsets of X which are either finite or co-finite (meaning that their complement is finite) is an algebra.
- The collection of all finite unions of intervals of the form $(-\infty, b], (a, b], (a, \infty)$, for $a, b \in \mathbb{R}$, is an algebra on the real numbers $\mathbb{R}$.

Note that any σ -algebra is an algebra but the converse is not true. Indeed, the second algebra provided in Example 2 above is not a σ -algebra.

A σ -algebra is also closed under countable intersections, that is, given a σ -algebra $\mathcal{A}$ and a countable family of sets $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{A}$, $I \subseteq \mathbb{N}$, we have that $\bigcap_{n \in I} A_n \in \mathcal{A}$. This follows from De Morgan's law $(\bigcap_{n \in I} A_n)^c = \bigcup_{n \in I} A_n^c \in \mathcal{A}$, and property (iv). Here are some basic examples of σ -algebras.

Example 3 (σ -Algebras).

- $\mathcal{P}(X)$ and $\{\emptyset, X\}$ are σ -algebras.
- For any subset $A \subseteq X$, $\mathcal{A} = \{\emptyset, A, X \setminus A, X\}$ is a σ -algebra.
- The collection of subsets of X which are either countable or co-countable (meaning that their complement is countable) is a σ -algebra.
- Given two σ -algebras $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{P}(X)$, we have that $\mathcal{A}_1 \cap \mathcal{A}_2$ is also a σ -algebra. More generally, for any (possibly uncountable) family of σ -algebras $\mathcal{A}_i \subseteq \mathcal{P}(X)$, $i \in I$, the intersection $\bigcap_{i \in I} \mathcal{A}_i$ is a σ -algebra.

There is a very natural way of generating σ -algebras from a collection of subsets:

Definition 4 (σ -algebra generated by a collection of subsets). Let $\mathcal{F} \subseteq \mathcal{P}(X)$ be a collection of subsets of X . The smallest σ -algebra containing $\mathcal{F}$, that is, the intersection of all σ -algebras containing $\mathcal{F}$ is called the σ -algebra generated by $\mathcal{F}$, and is usually denoted by $\sigma(\mathcal{F})$.

There are many families of subsets that generate useful σ -algebra, we will cover in this section some of them. Here are two simple examples of generated σ -algebras.

Example 5 (Generated σ -algebras).

- The collection of subsets of X which are countable or whose complements are countable is the σ -algebra generated by the singletons of X .
- Let X_1, X_2 be two sets, and $\mathcal{A}_1, \mathcal{A}_2$ be σ -algebras on X_1 and X_2 respectively. We define $\mathcal{A}_1 \otimes \mathcal{A}_2$ to be the σ -algebra on the Cartesian product $X = X_1 \times X_2$ generated by all the subsets of the form $A_1 \times A_2 \subseteq X$, where $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$. Note that $\mathcal{A}_1 \otimes \mathcal{A}_2$ is called the *product σ -algebra* generated by $\mathcal{A}_1$ and $\mathcal{A}_2$.

A special case of σ -algebras generated by a collection of subsets are the σ -algebras generated by the open subsets with respect to some topology. This type of σ -algebra will be one of our main focus in Ergodic Theory as we will study the dynamical properties of dynamical systems on topological spaces such as the torus.

Definition 6 (Borel σ -algebra over any topological space). Let (X, τ) be a topological space. The σ -algebra generated by the open subsets of X is called the *Borel σ -algebra on X* and we usually denote it by $\mathcal{B}_X$, or simply $\mathcal{B}$. Its elements are called the *Borel measurable* subset of X .

We give here two examples of such σ -algebras that will be used latter in this section.

Example 7 (Borel σ -algebra).

- Consider $\mathbb{R}^d$ endowed with its usual topology. Then, the Borel- σ -algebra on

$\mathbb{R}^d$ is the σ -algebra generated by the open balls $B_r(\mathbf{x}) = \{\mathbf{y} \in \mathbb{R}^d : |\mathbf{x} - \mathbf{y}| < r\}$. It contains all closed subsets of $\mathbb{R}^d$, but not all subsets of $\mathbb{R}^d$.

- Consider a finite set Σ , usually called the *alphabet*, containing n elements, usually referred to as the *letters* of Σ . The collection of all infinite strings in these letters is defined as the product space $\Sigma^{\mathbb{N}}$. Observe that the natural topology on Σ is the discrete topology, whose basis consists of singletons, i.e. individual letters. The Borel σ -algebra on $\Sigma^{\mathbb{N}}$ is the σ -algebra generated by the algebra of cylinder sets, where cylinder sets consist of the open sets of $x \in \Sigma^{\mathbb{N}}$ (with respect to the product topology of $\Sigma^{\mathbb{N}}$) that have finitely many coordinates fixed.

We now define the notion of monotone class, which gives rise to another characterization of σ -algebras.

Definition 8 (Monotone class). A monotone class $\mathcal{M} \subseteq \mathcal{P}(X)$ is a collection of subsets of X having the following properties:

- (i) if $A_1, A_2, \dots \in \mathcal{M}$ and $A_1 \subseteq A_2 \subseteq \dots$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{M}$
- (ii) if $B_1, B_2, \dots \in \mathcal{M}$ and $B_1 \supseteq B_2 \supseteq \dots$, then $\bigcap_{n=1}^{\infty} B_n \in \mathcal{M}$

Note that both $\mathcal{P}(X)$ and $\{\emptyset, X\}$ are monotone classes. Thus any collection of subsets is contained in a monotone class. The following theorem gives an alternative characterization of the σ -algebra generated by an algebra.

Theorem 9 (Monotone Class Theorem). Let $\mathcal{A} \subseteq \mathcal{P}(X)$ be an algebra and let $\mathcal{S}$ be the smallest monotone class containing $\mathcal{A}$. Then we have $\sigma(\mathcal{A}) = \mathcal{S}$.

0.2. Measures and Measure Spaces

A *measure* is a function that assigns a non-negative number to certain subsets of a set X in a manner consistent with the algebra of Boolean set operations, including unions, intersections, and complements. Measures provide the mathematical foundation for modeling quantities such as mass, length, area, volume, and, most importantly, probability. The subsets to which a measure can be assigned are called the *measurable sets*.

Definition 10 (Measurable space and measurable set). An ordered pair $(X, \mathcal{A})$, where X is a set and $\mathcal{A} \subseteq \mathcal{P}(X)$ is a σ -algebra, is called a *measurable space*, and any set $A \in \mathcal{A}$ is called a *measurable set*.

Definition 11 (Measure and measure space). A *measure* μ on a measurable space $(X, \mathcal{A})$ is a set function $\mu: \mathcal{A} \rightarrow [0, \infty]$ such that:

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

- (ii) For any countable (or finite) sequence of pairwise disjoint sets $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}$, we have

$$\mu\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu(A_n). \quad (\sigma\text{-additivity})$$

If $(X, \mathcal{A})$ measurable space and μ is a measure on it then the triple $(X, \mathcal{A}, \mu)$ is called a *measure space*.

The main structure of interest in classical ergodic theory is that of a probability space.

Definition 12 (finite and σ -finite measure space, probability space). A measure space $(X, \mathcal{A}, \mu)$ is said to be a finite measure space if μ satisfies $\mu(X) < \infty$, and if in addition $\mu(X) = 1$, $(X, \mathcal{A}, \mu)$ is called a probability space.

$(X, \mathcal{A}, \mu)$ is called a σ -finite measure space if X is a countable union of elements of $\mathcal{A}$ of finite measure.

We now state different useful properties about measures.

Proposition 13. Given a measure space $(X, \mathcal{A}, \mu)$, we have the following properties:

- (i) (Finite unions) For any positive integer n and disjoint sets $A_1, A_2, \dots, A_n \in \mathcal{A}$, using the fact that $\mu(\emptyset) = 0$, we have

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

- (ii) (Monotonicity) If $A, B \in \mathcal{A}$ and $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
 (iii) (Countable subadditivity) For any countable family of sets $\{A_n\}_{n \in I} \subseteq \mathcal{A}$, $I \subseteq \mathbb{N}$, not necessarily disjoint, we have

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

- (iv) (Continuity) If $A_1 \subseteq A_2 \subseteq \dots \in \mathcal{A}$, then

$$\lim_{k \rightarrow \infty} \mu(A_k) = \mu\left(\bigcup_{n=1}^{\infty} A_n\right),$$

and if $A_1 \supseteq A_2 \supseteq \dots \in \mathcal{A}$, and $\mu(A_1) < \infty$, then

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

Throughout the course we will extensively use the notion of "almost everywhere" (or "for almost every"). In short, a property holds almost everywhere on a set X if the subsets of elements for which it doesn't hold has zero measure. During the course, as we deal with probability measures, one way of seeing this notion is as follows: If we pick at random an element $x \in X$, then the probability that x satisfies

the given property is 1. Here is the formal definition.

Definition 14 (Almost everywhere). Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that a property holds μ -almost everywhere on X (sometimes abbreviated as μ -a.e.) if it the set of elements for which the property does not hold has zero measure with respect to μ .

Examples

Below, we provide several examples of important measure spaces, many of which will appear again as we delve deeper into ergodic theory throughout the course.

Null measure. Let X be a non-empty set, let $\mathcal{A}$ be a σ -algebra on X and define $\mu(A) = 0, \forall A \in \mathcal{A}$. Then $(X, \mathcal{A}, \mu)$ is a measure space and μ is referred to as the *null measure* on $(X, \mathcal{A})$.

Counting measure. Let X be a set, and for any $A \in \mathcal{P}(X)$ define $\mu(A) = |A|$, where $|A|$ denotes the cardinality of A . Then $(X, \mathcal{P}(X), \mu)$ is a measure space and μ is called the *counting measure* on X . This measure is finite when X is finite, it is σ -finite when X is countable, and it is not σ -finite when X is uncountable.

Dirac δ -measure. Let $(X, \mathcal{A})$ be a measurable space, and $x \in X$. Then we define the *Dirac measure* δ_x by

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

The Dirac measure is a probability measure, and it represents the almost sure outcome x in the measurable space.

Restriction of a measure. Let $(X, \mathcal{A}, \mu)$ be a measure space and $A \in \mathcal{A}$. We define the measure ν by $\nu(B) = \mu(B \cap A), \forall B \in \mathcal{A}$, to be the restriction of μ to A . Then $(X, \mathcal{A}, \nu)$ is a measure space and $\nu(B) = 0, \forall B \in \mathcal{A}$ with $B \subseteq X \setminus A$.

Conditional measure. Let $(X, \mathcal{A}, \mu)$ be a measure space and $A \in \mathcal{A}$ with $\mu(A) > 0$. We define for every $B \in \mathcal{A}$,

$$\mu|_A(B) = \mu(B|A) = \frac{\mu(A \cap B)}{\mu(A)}.$$

The set function $B \mapsto \mu|_A(B)$ is a measure on $\mathcal{A}$ called the *conditional measure with respect to A* . If μ is a finite measure (resp. probability measure) then the conditional measure with respect to A is also a finite measure (resp. probability measure).

Product measure. Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be two measure spaces. Let $\mathcal{A} = \mathcal{A}_1 \otimes \mathcal{A}_2$ be the product σ -algebra on the product space $X = X_1 \times X_2$. We define the product measure $\mu = \mu_1 \times \mu_2$ (also sometimes denoted $\mu_1 \otimes \mu_2$) to be the unique measure on the measurable space $(X, \mathcal{A})$ which satisfies $\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2)$ for every $A_1 \in \mathcal{A}_1$ and $A_2 \in \mathcal{A}_2$.

Probability measure on $\Sigma^{\mathbb{N}}$. Let $X = \Sigma^{\mathbb{N}}$ be the set of all infinite strings whose letters are in the finite alphabet Σ as previously defined. Denote by $\mathcal{A}$ the Borel σ -algebra on $\Sigma^{\mathbb{N}}$ generated by the cylinder sets. Let μ_0 be any probability measure on Σ . We define $\mu = \mu_0^{\mathbb{N}}$ to be the product measure on $\Sigma^{\mathbb{N}}$, which is the unique measure satisfying for every cylinder set I ,

$$\mu(I) = \prod_{i \in F} \mu_0(\{x_i\})$$

where F is the finite set of the indices of the fixed coordinates of I .

Borel measure. In order to define Borel measures, we recall two definitions from topology.

Definition 15. A topological space X is Hausdorff if for any distinct points $x, y \in X$, there exists open neighborhoods U, V of x and y respectively such that U and V are disjoint.

Definition 16. A topological space X is locally compact if every $x \in X$ has a compact neighborhood.

Now, let X be a locally compact Hausdorff topological space and $\mathcal{B}$ the Borel σ -algebra defined on X . Then, any measure μ defined on $\mathcal{B}$ is called a Borel measure. If $\mu(X) = 1$, we say that μ is a Borel probability measure.

Radon measure. Let X be a locally compact Hausdorff topological space, $\mathcal{B}$ the Borel σ -algebra defined on X , and μ a finite Borel measure on $\mathcal{B}$. If in addition μ is tight, in the sense that for all $\varepsilon > 0$ there exists a compact set $K \subseteq X$ such that $\mu(X \setminus K) < \varepsilon$ (or equivalently $\mu(K) \geq \mu(X) - \varepsilon$), μ is called a Radon measure. These conditions guarantee that the measure is compatible in some sense with the topology of the space. We state the following theorem, without providing a proof, which gives a sufficient condition for a Borel measure to be Radon under a commonly encountered topological assumption on X :

Theorem 17. Let X be a locally compact Hausdorff topological space in which every open set is σ -compact (that is, a countable union of compact sets). Then every Borel measure on X that is finite on compact sets is a Radon measure.

It follows that every finite Borel measure on a compact metric Hausdorff topological space X is automatically Radon, as X verifies the conditions of the theorem.

Notice that if we only assume X to be a locally compact, second-countable Hausdorff topological space, then because it follows that X is σ -compact, X still verifies the conditions of the theorem.

An useful property of the Radon measure is that it makes the map $f \mapsto \int f d\mu$, where $f \in L^1(X)$, continuous (recalls about integration theory are given in the next section). The following measures are examples of Radon measures: the Lebesgue measure on an Euclidean space, the Haar measure on any locally compact topological group, the Dirac measure on any topological space.

Lebesgue measure. The Lebesgue measure is the unique measure μ on the Borel- σ -algebra $\mathcal{B}_{\mathbb{R}}$ such that for every interval $I \subseteq \mathbb{R}$, the measure $\mu(I)$ is the length of I .

We observe that the restriction of μ to the Borel- σ -algebra $\mathcal{B}_{[0,1]}$ of subsets of $[0, 1]$ is the so called uniform distribution from probability theory.

We can generalize this idea to higher dimensions. Indeed, for the lower dimensions $n = 1, 2$, the Lebesgue measure coincides with the notions of area and volume. For higher dimensions, it is also called n -dimensional volume.

More generally, if we consider the measurable space $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$, the Lebesgue measure μ on $\mathcal{B}_{\mathbb{R}^n}$ is the unique measure such that if A is a cartesian product of intervals $I_1 \times \cdots \times I_n$, then A is Lebesgue measurable (in the sense that we can attribute a Lebesgue measure to A) and $\mu(A) = \prod_{i=1}^n l(I_i)$, where l denotes the length of the interval I_i , i.e, l is the Lebesgue measure in one dimension.

We list, without proof, some of the properties of the Lebesgue measure on $\mathcal{B}_{\mathbb{R}^n}$:

- (i) (translation invariance) If $A \subseteq \mathbb{R}^n$ is Lebesgue measurable, and $x \in \mathbb{R}^n$, then $A + x = \{y \in \mathbb{R}^n : y + x \in A\}$ is Lebesgue measurable and $\mu(A + x) = \mu(A)$. In particular, $A \subseteq \mathbb{R}^n$ is Lebesgue measurable if, and only if, all translates of A is Lebesgue measurable.
- (ii) (dilation and scaling) Let $c > 0$, $A \subseteq \mathbb{R}^n$ be Lebesgue measurable, and let $cA = \{cy \in \mathbb{R}^n : y \in A\}$, then cA is Lebesgue measurable and $\mu(cA) = c^n \mu(A)$.
- (iii) More generally, if T is a linear transformation and A is a Lebesgue measurable subset of $\mathbb{R}^n$, then $T(A)$ is a Lebesgue measurable set of measure $|\det(T)|\mu(A)$.
- (iv) Finite or countable sets are Lebesgue measurable and have Lebesgue measure 0, and there exist uncountable Lebesgue measurable sets of measure 0. As an example, we can consider the Cantor set (when $n = 1$). Moreover, there exists sets which are not Lebesgue measurable.

Finally, note that the Haar measure (to be seen in the section about topological groups) on a locally compact Hausdorff topological group can be thought of as the natural generalization of the Lebesgue measure to a general locally compact Hausdorff topological group.

Atomic, non-atomic, and continuous measures. In order to define discrete and continuous measures we will need the following definition.

Definition 18 (Atom). Given a measure space $(X, \mathcal{A}, \mu)$, a set $A \in \mathcal{A}$ is called an atom if:

- (i) $\mu(A) > 0$, and
- (ii) For any measurable set $B \subseteq A$ with $\mu(B) < \mu(A)$ we have $\mu(B) = 0$.

A σ -finite measure μ on a measurable space $(X, \mathcal{A})$ is called *purely atomic* if every measurable set of positive measure contains an atom. On the contrary, a σ -finite measure which has no atoms is called *non-atomic*. Equivalently, μ is *non-atomic* if for every measurable set A such that $\mu(A) > 0$ there exists a measurable subset B of A such that $0 < \mu(B) < \mu(A)$.

Finally, a σ -finite measure μ is called *continuous* if for any $A \in \mathcal{A}$ and any $c \in \mathbb{R}$ such that $0 < c < \mu(A)$, there exists a measurable subset B of A such that $\mu(B) = c$. Note that any continuous measure is non-atomic.

There are two important existence theorems for measures, the Carathéodory Extension Theorem and the Riesz Representation Theorem.

Definition 19 (Pre-measure). Let $\mathcal{A}$ be an algebra on a set X . A set function $\mu_0: \mathcal{A} \rightarrow [0, \infty]$ is called a *pre-measure* on $(X, \mathcal{A})$ if $\mu_0(\emptyset) = 0$ and, for every countable (or finite) sequence $A_1, A_2, \dots \in \mathcal{A}$ of pairwise disjoint sets whose union lies in $\mathcal{A}$, we have

$$\mu_0\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mu_0(A_n). \quad (\sigma\text{-additivity})$$

Theorem 20 (Carathéodory). Let $\mathcal{A}$ be an algebra on a set X . Any pre-measure μ_0 on $\mathcal{A}$ extends to a measure μ on the σ -algebra $\sigma(\mathcal{A})$ generated by $\mathcal{A}$. Moreover, if μ_0 is σ -finite then this extension is unique and σ -finite.

Let X be a locally compact Hausdorff space. We write $C(X)$ for the space of all continuous complex-valued functions on X . Within this space, we distinguish the subspace $C_0(X)$ of functions *vanishing at infinity*, meaning

$$f \in C_0(X) \iff \forall \varepsilon > 0, \{x \in X : |f(x)| \geq \varepsilon\} \text{ is compact.}$$

Equipped with the uniform norm $\|f\|_{\infty} = \sup_{x \in X} |f(x)|$, $C_0(X)$ is a Banach space.

When X is compact, every continuous function automatically vanishes at infinity, so $C_0(X) = C(X)$. In the non-compact case, $C_0(X)$ forms a proper subspace of $C(X)$, consisting precisely of those continuous functions that tend to 0 at infinity.

Theorem 21 (Riesz-Markov-Kakutani representation theorem). Let X be a locally compact Hausdorff topological space in which every open set is σ -compact (cf. Theorem 17). For any continuous linear functional $l: C_c(X) \rightarrow \mathbb{C}$ there exists a unique

complex-valued Radon measure μ on X such that

$$l(f) = \int_X f(x) \, d\mu(x),$$

for all $f \in C_c(X)$.

0.3. Measurable Functions and Integrals

Throughout this section we let $(X, \mathcal{A}, \mu)$ be a measure space. Natural classes of measurable functions on X are built up from simpler functions, just as the σ -algebra $\mathcal{A}$ may be built up from simpler collections of sets. Given a set $A \subseteq X$, we denote by $1_A: X \rightarrow \{0, 1\}$ the indicator function of A , that is,

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A, \end{cases} \quad \forall x \in X.$$

Definition 22 (Simple function). A function $f: X \rightarrow \mathbb{R}$ is called *simple* if

$$f(x) = \sum_{j=1}^m c_j 1_{A_j}(x), \quad \forall x \in X,$$

where $c_j \in \mathbb{R}$ and the $A_j \in \mathcal{A}$ are disjoint sets $\forall j = 1, \dots, m$. The *integral* of f is then defined to be

$$\int f \, d\mu = \sum_{j=1}^m c_j \mu(A_j). \quad (0.3.1)$$

Definition 23 (Measurable function). A function $g: X \rightarrow \mathbb{R}$ is called measurable if $g^{-1}(A) \in \mathcal{A}$ for any (Borel) measurable set $A \in \mathcal{B}_{\mathbb{R}}$.

Note that simple functions are always measurable functions. Below, we outline several methods for generating new measurable functions from existing measurable ones.

Proposition 24. Let $f, g: X \rightarrow \mathbb{R}$ be measurable, and $c \in \mathbb{R}$. Then, the following functions are measurable:

- (i) cf
- (ii) $f + g$
- (iii) fg
- (iv) $|f|$
- (v) $\min\{f, g\}$ and $\max\{f, g\}$
- (vi) $\operatorname{Re}(f)$ and $\operatorname{Im}(f)$, where we understand $\operatorname{Re}(f)$ (resp. $\operatorname{Im}(f)$) as the unique function such that $\operatorname{Re}(f(x)) = \operatorname{Re}(f)(x)$ (resp. $\operatorname{Im}(f(x)) = \operatorname{Im}(f)(x)$) $\forall x \in X$.

The integral of simple functions has already been defined in (0.3.1). Our next

goal is to extend this definition to all measurable functions. To achieve this, we rely on the following key approximation result.

Proposition 25. *Let $g: X \rightarrow [0, \infty)$ be a measurable function taking non-negative values. There exists a pointwise increasing sequence of simple functions $(f_n)_{n \in \mathbb{N}}$ (in the sense that $f_n(x) \leq f_{n+1}(x)$ for all $x \in X$ and $n \in \mathbb{N}$) such that $\lim_{n \rightarrow \infty} f_n(x) = g(x)$ for each $x \in X$.*

Definition 26 (Integral of non-negative measurable function). Let $g: X \rightarrow \mathbb{R}_{\geq 0}$ be a measurable function taking non-negative values, and let $(f_n)_{n \in \mathbb{N}}$ be a pointwise increasing sequence of simple functions converging to g as guaranteed by Proposition 25. Then the *integral* of g is defined to be

$$\int g \, d\mu = \lim_{n \rightarrow \infty} \int f_n \, d\mu.$$

Moreover, g is called *integrable* if $\int g \, d\mu < \infty$.

Observe that the expression $\int g \, d\mu$ defined above is guaranteed to exist since $f_n(x) \leq f_{n+1}(x)$ for all $n \in \mathbb{N}$ and $x \in X$. One can show that this is well-defined, i.e., that it is independent of the choice of the sequence of simple functions.

We now extend the notion of integral for any measurable functions.

Definition 27 (Integral of measurable function). Given a measurable function $g: X \rightarrow \mathbb{R}$, g has in general a unique decomposition $g = g_+ - g_-$, where $g_+(x) = \max\{g(x), 0\}$ and $g_-(x) = \max\{-g(x), 0\}$ for every $x \in X$. Note that both g_+ and g_- are measurable. The function g is said to be *integrable* if both g_+ and g_- are integrable, and the *integral* of g is defined as

$$\int g \, d\mu = \int g_+ \, d\mu - \int g_- \, d\mu.$$

Consider now a measurable complex-valued function $g: X \rightarrow \mathbb{C}$, which we can decompose as $g = \operatorname{Re}(g) + i\operatorname{Im}(g)$, where both $\operatorname{Re}(g)$ and $\operatorname{Im}(g)$ are measurable. Then g is said to be integrable if $|g| = \sqrt{\operatorname{Re}(g)^2 + \operatorname{Im}(g)^2} \geq 0$ satisfies :

$$\int |g| \, d\mu < +\infty$$

and the integral of g is then defined as :

$$\int g \, d\mu = \int \operatorname{Re}(g) \, d\mu + i \int \operatorname{Im}(g) \, d\mu \in \mathbb{C}$$

so that we have :

$$\begin{aligned} \operatorname{Re}\left(\int g \, d\mu\right) &= \int \operatorname{Re}(g) \, d\mu \\ \operatorname{Im}\left(\int g \, d\mu\right) &= \int \operatorname{Im}(g) \, d\mu \end{aligned}$$

Note that since we have $|Re(g)| \leq |g|$ and $|Im(g)| \leq |g|$, the condition $\int |g| d\mu < +\infty$ implies that both $Re(g)$ and $Im(g)$ are integrable.

Here is a way of determining if a given function is integrable or not.

Proposition 28. *Let $f, g: X \rightarrow \mathbb{R}$. If f is integrable and g is measurable with $|g| \leq f$, then g is integrable.*

Being integrable is preserved under restriction to a measurable set, and we give the definition of the integrable restricted to a measurable set:

Definition 29. Let $f: X \rightarrow \mathbb{R}$ be an integrable function, and A be a measurable set. The integral of f over A is defined as

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

0.4. L^p Spaces

We now recall some definitions and facts about L^p spaces, which are function spaces defined using a natural generalization of the p -norm for finite-dimensional vector spaces. L^p spaces form an important class of Banach spaces in functional analysis, and of topological vector spaces. In the course of Ergodic Theory we will use various results about functional analysis and in particular about L^p spaces. Further recalls about functional analysis are given in the next section.

Definition 30 ($\mathcal{L}^p$ spaces). Let $(X, \mathcal{A}, \mu)$ be a measure space. For $1 \leq p < \infty$, we define the set $\mathcal{L}^p(X, \mathcal{A}, \mu)$ (sometimes also denoted $\mathcal{L}^p(\mu)$) to be the set of all measurable functions $f: X \rightarrow \mathbb{C}$ such that $\int |f|^p d\mu < \infty$.

Definition 31 (L^p spaces). We define an equivalence relation on $\mathcal{L}^p(\mu)$ by $f \sim g$ if $\int |f - g|^p d\mu = 0$ and we write $L^p(\mu) = \mathcal{L}^p(\mu) / \sim$ for the space of equivalence classes. Elements of $L^p(\mu)$ will be described as functions rather than equivalence classes, but it is important to remember that this is an abuse of notation.

Furthermore, we define the norm $\|\cdot\|_{L^p}$ by:

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

We now give the definition of the $L^p(\mu)$ in the case $p = \infty$.

Definition 32 (Essential supremum). The *essential supremum* is the generalization to measurable functions of the supremum of a continuous function, and is defined by

$$\text{ess sup } f = \inf\{\alpha \in \mathbb{R} : \mu(\{x \in X : f(x) > \alpha\}) = 0\}.$$

Let the norm $\|\cdot\|_{L^\infty}$ be given by

$$\|f\|_{L^\infty} = \text{ess sup } |f|.$$

The space $\mathcal{L}^\infty(\mu)$ is then defined to be the set of measurable functions f such that $\|f\|_{L^\infty} < \infty$. Once again, $L^\infty(\mu)$ is defined to be $\mathcal{L}^\infty(\mu)/\sim$.

Proposition 33. *For every $1 \leq p \leq \infty$, the space $L^p(\mu)$ is complete with respect to the norm $\|\cdot\|_{L^p}$, and hence is a Banach space.*

Proposition 34. *For $1 \leq p < q \leq \infty$ we have $L^p(\mu) \supseteq L^q(\mu)$ for any finite measure space.*

Finally we turn to integration of functions of several variables.

Theorem 35 (Fubini–Tonelli). *Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two measure spaces and let f be a non-negative integrable function on the product space $(X \times Y, \mathcal{A} \otimes \mathcal{B}, \mu \otimes \nu)$. Then, for μ -almost every $x \in X$ the function $y \mapsto f(x, y)$ is integrable, and for ν -almost every $y \in Y$ the function $x \mapsto f(x, y)$ is integrable, and we have*

$$\int_{X \times Y} f(x, y) d(\mu \times \nu)(x, y) = \int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x) = \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y).$$

0.5. Convergence Theorems

The most important distinction between integration on L^p spaces as defined above and Riemann integration on bounded Riemann-integrable functions is that the L^p functions are closed under several natural limiting operations, allowing for the following important convergence theorems. We start with the Monotone Convergence Theorem.

Theorem 36 (Monotone Convergence Theorem). *Suppose $f_1 \leq f_2 \leq \dots$ is a pointwise increasing sequence of non-negative real-valued measurable functions on the measure space $(X, \mathcal{A}, \mu)$ which converges almost everywhere to a function f on X . Then f is measurable and*

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

In particular, if $\lim_{n \rightarrow \infty} \int f_n d\mu < \infty$, then f is integrable.

When the $f_n, n \in \mathbb{N}$, are integrable, the assumption that f_n is non-negative for every $n \in \mathbb{N}$ can be dropped by considering instead the non-negative sequence of measurable function $g_n = f_n - f_1$, which is also a pointwise increasing sequence. Next, we state Fatou's lemma, which is not only needed to prove the dominated

convergence theorem below but it includes also a statement of the behavior of the integral under pointwise (or almost everywhere) convergence: The integral is lower semi-continuous under almost everywhere convergence.

Theorem 37 (Fatou's lemma – liminf version). *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of non-negative real-valued measurable functions on the measure space $(X, \mathcal{A}, \mu)$. Then, $f = \liminf_{n \rightarrow \infty} f_n$ is measurable and*

$$\liminf_{n \rightarrow \infty} \int f_n \, d\mu \geq \int f \, d\mu = \int \liminf_{n \rightarrow \infty} f_n \, d\mu$$

In particular, if f_n is integrable for every $n \in \mathbb{N}$, then f is also integrable.

For completeness, we also add the reverse version of Fatou's lemma.

Corollary 38 (Fatou's lemma – limsup version). *Let $(f_n)_{n \in \mathbb{N}}$ be a sequence of real-valued measurable functions on the measure space $(X, \mathcal{A}, \mu)$. If there exists an integrable function g on X such that $f_n \leq g$, $\forall n \in \mathbb{N}$, then :*

$$\limsup_{n \rightarrow \infty} \int f_n \, d\mu \leq \int \limsup_{n \rightarrow \infty} f_n \, d\mu$$

Contrary to the Monotone Convergence Theorem, the hypothesis that f_n is non-negative for every $n \in \mathbb{N}$ cannot be dropped.

Finally, we state the Dominated Convergence Theorem, which formulates sufficient conditions under which almost everywhere convergence yields an integrable function and such that limit and integral are interchangeable. Note that this is an important difference with Riemann integral.

Theorem 39 (Dominated Convergence Theorem). *Let $(X, \mathcal{A}, \mu)$ be a measure space. If $h: X \rightarrow \mathbb{R}$ is a non-negative integrable function, and $(f_n)_{n \in \mathbb{N}}$ is a sequence of measurable complex-valued functions on $(X, \mathcal{A}, \mu)$ which are dominated by h in the sense that $|f_n| \leq h$, $\forall n \in \mathbb{N}$, and $\lim_{n \rightarrow \infty} f_n = f$ exists almost everywhere, then f is integrable and*

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