

# Introduction to Measure Theory

JAN ERTL

July 7, 2019

"Mathematics is the art of giving the same name to different things."

— Henri Poincaré

## Acknowledgments

THANKS to Evan Chen, whose L<sup>A</sup>T<sub>E</sub>X-style I am using. My approach to the subject is inspired by Professors Wenjia Jing and Robert Fefferman, from whom I learned measure theory.

## §1 Measure Spaces

Try to recall how you had to compute areas for the first time in elementary school - you likely had an exercise of an irregular shape that you could decompose into subsets whose area you already knew. You might have also had such computation tasks for which it was easiest to calculate the area of a shape that contained the object of interest, and subtract those smaller chunks from this larger area that did not belong to the shape of interest.

In your calculation you implicitly assumed that if you can measure a (finite) set of shapes, you can measure their disjoint union, with the assigned area being the sum of the areas of well-known shapes. Likewise, you assumed that if you can measure an encompassing area, and you can measure certain subsets of it, you can also measure their complement relative to the space. These observations about elementary school area computations shall serve as a first motivation for the following definitions:

**Definition 1.1** (algebra,  $\sigma$ -algebra). Let  $X$  be a nonempty set. A family  $\mathcal{A}$  of subsets of  $X$  is said to be an **algebra** if

- $\emptyset \in \mathcal{A}$
- $\mathcal{A}$  is closed under complements and finite unions, i.e. if  $A \in \mathcal{A}$ , then  $A^c = X - A \in \mathcal{A}$ , and if  $A_1, A_2, \dots, A_N \in \mathcal{A}$ , then  $\bigcup_{n=1}^N A_n \in \mathcal{A}$

We say  $\mathcal{A}$  is a  **$\sigma$ -algebra** if it is also closed under countable unions, i.e.

- if  $A_1, A_2, \dots \in \mathcal{A}$ , then  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$

The additional condition on  $\sigma$ -algebras might be motivated by the observation that we can write many common sets, such as a circle, as the (almost) disjoint union of cubes inscribed in it.

**Remark 1.2.** In these notes, I will use  $\mathcal{A}$  for algebras, and  $\mathcal{S}$  for  $\sigma$ -algebras. Many other texts use  $\mathcal{F}$  to denote both because algebras used to be called fields.

**Exercise 1.3.** Let  $X = \mathbb{R}$ . Define

- $\mathcal{G}_1 = \{A : A \subset \mathbb{R}, A \text{ open}\}$
- $\mathcal{G}_2 = \{A : A \text{ is a finite union of intervals of the form } (a,b], (-\infty, b], \text{ and } (a,\infty) \}$
- $\mathcal{G}_3 = \{A : A \text{ is a countable union of } (a,b], (-\infty, b], \text{ and } (a,\infty) \}$

Which of these are algebras? Which are even  $\sigma$ -algebras?

**Remark 1.4.** If  $\mathcal{A}$  is a  $\sigma$ -algebra of  $X$ , then it is closed under countable intersection as well.

*Proof.* Suppose  $A_n \in \mathcal{A}, \forall n \in \mathbb{N}$ . Note that  $(\bigcap_{n=1}^{\infty} A_n)^c = \bigcup_{n=1}^{\infty} A_n^c \in \mathcal{A}$ , so  $\bigcap_{n=1}^{\infty} A_n$  itself is in  $\mathcal{A}$ . The equality follows from basic set theory (see Demorgan's Laws).  $\square$

### Example 1.5

Suppose  $X = \{\text{possible outcome of a coin flip}\} = \{H, T\}$ . Recall that in probability theory, an event is a set of outcomes, i.e. a subset of the sample space. Then note that the event space  $\{\emptyset, \{H\}, \{T\}, \{H, T\}\}$ , which happens to be the power set of the sample space, is a  $\sigma$ -algebra.

### Example 1.6

The **power set of  $X$**  is defined as  $\mathcal{P}(X) = \{A : A \subset X\}$ . It is easy to show that the power set is a  $\sigma$ -algebra. The set  $\{\emptyset, X\}$  is also a  $\sigma$ -algebra.

**Exercise 1.7.** Explain why these are the 'largest' and 'smallest' possible  $\sigma$ -algebras, respectively, in the following sense: if  $\mathcal{A}$  is any  $\sigma$ -algebra, then  $\{\emptyset, X\} \subset \mathcal{A} \subset \mathcal{P}(X)$ .

An important  $\sigma$ -algebra we will return to is the Borel- $\sigma$ -algebra on  $X$ , which is well-defined if  $X$  is a metric space. It is the smallest  $\sigma$ -algebra that contains all open sets of  $X$ . Let us make this precise.

**Definition 1.8** ( $\sigma$ -algebra generated by a family of subsets of  $X$ ). Given any family  $\mathcal{G}$  of subsets of  $X$ , we define

$$\sigma(\mathcal{G}) := \bigcap_{\alpha} \mathcal{S}_{\alpha}$$

where  $\{\mathcal{S}_{\alpha}\}$  is the collection of  $\sigma$ -algebras that include  $\mathcal{G}$ . Note that this is well-defined because  $\mathcal{P}(X) \in \{\mathcal{S}_{\alpha}\}$ . Furthermore,  $\sigma(\mathcal{G})$  is a  $\sigma$ -algebra. The proof is left as an easy exercise. We call  $\sigma(\mathcal{G})$  the  **$\sigma$ -algebra generated by  $\mathcal{G}$** . It is clear that it is the smallest  $\sigma$ -algebra containing  $\mathcal{G}$ .

### Example 1.9 (Binary infinite trees - Sequence space)

Let  $X = \{0, 1\}^{\mathbb{N}}$ , the space of infinite binary sequences. Given any finite sequence  $a = a_1 a_2 a_3 \dots a_m$ , define the set  $E_a$  to be the set of sequences whose first  $m$  members coincide with  $a$ . Let  $\mathcal{E}$  be the collection of all these sets. We refer to the pair  $(X, \sigma(\mathcal{E}))$  as **sequence space**. Note that you can also think of this as a subset of infinite coin flips.

**Exercise 1.10.** Prove the following Proposition:

Let  $\{\mathcal{S}_\alpha\}$  be a family of  $\sigma$ -algebras on  $X$ . Then  $\bigcap_\alpha \mathcal{S}_\alpha$  is also a  $\sigma$ -algebra.

We now have enough vocabulary to properly define the Borel  $\sigma$ -algebra:

**Definition 1.11** (Borel  $\sigma$ -algebra). Let  $X$  be a metric space, and let  $\mathcal{O}$  denote the collection of open sets of  $X$ .  $\sigma(\mathcal{O})$  is thus the smallest  $\sigma$ -algebra containing all open sets of  $X$ . We call it the **Borel  $\sigma$ -algebra of  $X$**  and denote it by  $\mathcal{B}(X)$ .

**Definition 1.12** (Measurable space). A pair  $(X, \mathcal{S})$  with  $X$  nonempty,  $\mathcal{S}$  a  $\sigma$ -algebra on  $X$  is called a **measurable space**.

**Definition 1.13** (Measure). Let  $(X, \mathcal{S})$  be a measurable space. A function  $\mu : \mathcal{S} \rightarrow [0, \infty]$  is called a (nonnegative) **measure** if it satisfies

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

$$(ii) \quad \mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i) \text{ for any } \{A_i\}_{i=1}^{\infty} \subset \mathcal{S} \text{ s.t. } A_i \cap A_j = \emptyset \forall i \neq j.$$

**Remark 1.14.** We can define signed measures  $\mu : \mathcal{S} \rightarrow [-\infty, \infty]$ . Our discussion here is restricted to non-negative measures, since they are natural in our applications of Lebesgue integration and probability theory.

**Definition 1.15** (Measure space). A triple  $(X, \mathcal{S}, \mu)$  is called a **measure space** if  $(X, \mathcal{S})$  is a measurable space, and  $\mu$  is a measure defined on it.

#### Example 1.16

If  $\mu(X) = 1$ , then we may say  $\mu$  is a probability measure, and the triple  $(X, \mathcal{S}, \mu)$  is a probability space.

#### Example 1.17

If for all  $B \subset \mathbb{N}$ ,  $\mu(B) = |B|$ , where  $|B|$  denotes the number of elements in  $B$ , then the triple  $(\mathbb{N}, \mathcal{P}(\mathbb{N}), \mu)$  is a measure space, and we call  $\mu$  the counting measure.

The following definition is incredible important and you will encounter it in many applications:

**Definition 1.18** (P holds almost everywhere). Let  $(X, \mathcal{S}, \mu)$  be a measure space. We say the property **P holds almost everywhere** in  $X$  (sometimes written as a.e.) if there is a set  $N$  with  $\mu(N) = 0$  such that P holds for all  $x \in X - N$ .

**Remark 1.19.** Note that at first glance, a more intuitive definition might be that P holds almost everywhere if  $\mu(\{x : \text{P does not hold}\}) = 0$ . However, this definition only works if the set in question is actually measurable, i.e.  $\{x : \text{P does not hold}\} \in \mathcal{S}$ .

**Definition 1.20.** Let  $(X, \mathcal{S}, \mu)$  be a measure space. A set  $N \in \mathcal{S}$  such that  $\mu(N) = 0$  is called a **null set**. A set  $M$  with  $M \subset N$  for some null set  $N$  is called a **negligible set**. If all negligible sets are measurable, i.e.  $M \in \mathcal{S}$  (and hence  $\mu(M) = 0$ ), we say the measure space is complete.

**Remark 1.21.** In this new language, we say P holds almost everywhere if the set on which it does not hold is negligible. In some applications, the term 'almost surely' (a.s.) is used instead of almost everywhere.

**Exercise 1.22.** Let  $(X, \mathcal{S}, \mu)$  be a measure space. Prove the following:

- $\mu$  is **monotone**: if  $A, B \in \mathcal{S}, A \subset B$ , then  $\mu(A) \leq \mu(B)$
- $\mu$  is **countably subadditive**: if  $\{A_i\}_{i=1}^{\infty} \subset \mathcal{A}$ , then  $\mu(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$

The following easy exercise will come in handy when defining conditional probabilities later on:

**Exercise 1.23.** Let  $(X, \mathcal{S}, \mu)$  be a measure space. Let  $B \in \mathcal{S}$ . Show that  $\lambda : \mathcal{S} \rightarrow [0, \infty]$  defined by  $\lambda(A) = \mu(A \cap B)$  is also a measure  $(X, \mathcal{S})$ .

**Example 1.24**

If  $\mu : \mathcal{S} \rightarrow [0, 1]$  is a probability measure, the exercise reveals that  $\lambda_B(A) = \frac{\mu(A \cap B)}{\mu(B)}$  is also a probability measure on the same space, if  $\mu(B) \neq 0$ . Conditional probability is well-defined in our framework for these cases! If  $\mu(B) = 0$ , we need to be more careful.

**Theorem 1.25**

Let  $\mu$  be a measure on  $(X, \mathcal{S})$ . Then it is continuous from below in the sense that:

- (i)  $(A_1 \subset A_2 \subset A_3 \subset \dots, A_i \in \mathcal{S}) \Rightarrow (\lim_{n \rightarrow \infty} \mu(A_n) = \mu(\cup_{i=1}^{\infty} A_i))$
- (ii)  $(A_1 \supset A_2 \supset A_3 \supset \dots, A_i \in \mathcal{S}, \mu(A_1) < \infty) \Rightarrow (\lim_{n \rightarrow \infty} \mu(A_n) = \mu(\cap_{i=1}^{\infty} A_i))$

*Proof.* We prove (i) and leave (ii) as an exercise. Let  $B_1 = A_1, B_i = A_i \cap A_{i-1}^c, i \geq 2$ .

$$\text{Then note } \begin{cases} A_n = A_{n-1} \cup B_n & \forall n \geq 1 \\ A_{n-1} \cap B_n = \emptyset, B_m \cap B_n = \emptyset & m \neq n \\ \cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} B_n \end{cases}$$

This means that  $\cup_{n=1}^{\infty} B_n$  is a countable union of pairwise disjoint sets in  $\mathcal{S}$ . Hence

$$\mu(\cup_{n=1}^{\infty} A_n) = \mu(\cup_{n=1}^{\infty} B_n) = \sum_{n=1}^{\infty} \mu(B_n) = \lim_{N \rightarrow \infty} \sum_{n=1}^N \mu(B_n) = \lim_{N \rightarrow \infty} \mu(A_N)$$

□

**Exercise 1.26.** Prove (ii).

A consequence of this continuity property leads to the Borel-Cantelli Lemma, an extremely handy tool in probability theory.

**Lemma 1.27 (Borel-Cantelli Lemma)**

Let  $\mu$  be a measure on a  $\sigma$ -algebra  $\mathcal{S}$ . Let the sequence  $\{E_i\}_{i=1}^{\infty} \subset \mathcal{S}$ , and suppose  $\sum_{n=1}^{\infty} \mu(E_n) < \infty$ . Then for  $D = \{x \in X : x \text{ belongs to infinitely many } E_n\}$  satisfies  $\mu(D) = 0$ .

*Proof.* The key observation here is that  $D = \bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} E_n$ . Fix  $\epsilon > 0$ . Since  $\sum_{n=1}^{\infty} \mu(E_n) < \infty$ , there is an  $m$  such that  $\sum_{n=m}^{\infty} \mu(E_n) < \epsilon$ , and then note

$$\mu(D) \leq \mu\left(\bigcup_{n \geq m} E_n\right) \leq \sum_{n=m}^{\infty} \mu(E_n) < \epsilon$$

□

**Example 1.28** (Borel-Cantelli application)

Why is this result so useful in probability applications? Think of the  $E_n$  as events in a probability space with probability measure  $\mu$ . Then  $\sum_{n=1}^{\infty} \mu(E_n) = \infty$  is a necessary condition for a positive probability of an event occurring infinitely often.

**Theorem 1.29** (Existence of non-measurable sets)

Assume the axiom of choice. There is no translation-invariant measure on  $(\mathbb{R}, \mathcal{P}(\mathbb{R}))$ !

*Proof.* We can return to this if people are interested and we have time at the end of class. □

**Theorem 1.30** (Banach-Tarski Paradox)

Informally: One can divide the unit sphere  $S \in \mathbb{R}^3$  into finitely many pieces and reassemble these using only rigid motions (translations and rotations) into two identical copies of the original sphere.

## §2 Construction of Lebesgue Measure

The preceding results show us that we cannot generally find a measure on a given  $\sigma$ -algebra. In particular, there are sets that defy the intuitive concepts of length or volume. Instead of giving up the concept of measure, we turn back to  $\sigma$ -algebras. We want to find  $\sigma$ -algebras for which there are well-defined measures that correspond to our intuitive notions of length, area and volume (and probability!). This means we purposefully declare some sets non-measurable, and exclude them from our measures' domain. At the same time, we do not want to discard too many sets. The following discussion shows that given an algebra for which we have an intuitive way to assign measure, we can indeed find a  $\sigma$ -algebra and measure large enough to include the original algebra and the  $\sigma$ -algebra it generates.

We follow an abstract approach to encompass many different measure spaces. However, the application that we are focused on is the real line  $\mathbb{R}$ . The measure we are constructing is called the Lebesgue measure, named after the French mathematician Henri Lebesgue.

We start the construction by defining a so called pre-measure on an algebra. You can think of assigning length, area or volume to simple intervals, rectangles and boxes.

**Definition 2.1** (Pre-measure).  $\nu : \mathcal{A} \rightarrow [0, \infty]$  is called a **pre-measure** if it satisfies:

- (i)  $\nu(\emptyset) = 0$

(ii)  $\nu(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \nu(A_n)$  for  $\{A_n\} \subset \mathcal{A}$ , pairwise disjoint

In one dimension, finding a

**Example 2.2** (Application to Lebesgue measure on  $\mathbb{R}$ )

For Lebesgue measure, we will take  $\mathcal{A} = \{A : A \text{ is a finite disjoint union of intervals of the form } (a, b], (-\infty, b], \text{ and } (a, \infty)\}$ . It was already an exercise to show that this is an algebra. It is easy to see that  $\nu$  defined by  $\nu(A) = \nu(\bigcup_{i=1}^N (a_i, b_i]) = \sum_{i=1}^N (b_i - a_i)$  is a pre-measure.

**Remark 2.3.** We can already see that this generalizes easily to two (and thus any finite) dimensions - instead of intervals, we may use cubes and boxes. We still need to be careful how to proceed in our construction.

**Definition 2.4** (Outer Measure). Given  $X$ , we call  $\mu^* : \mathcal{P}(X) \rightarrow [0, \infty]$  an **outer measure** if

- (i)  $\mu^*(\emptyset) = 0$ ,
- (ii) it is monotone: if  $A \subset B$ , then  $\mu^*(A) \leq \mu^*(B)$ ,
- (iii) and it is countably subadditive:  $\mu^*(\bigcup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$  for any  $\{A_i\}_{i=1}^{\infty}$ .

Based on our given pre-measure, we will now define a way to assign an outer measure. Note that by definition, the domain of outer measures is always the power set of the original space.

**Theorem 2.5** (Generation of outer measure)

Let  $\mathcal{A}$  be a collection of subset of  $X$ , containing  $\emptyset$  and  $X$ . Let  $\nu$  satisfy  $\nu(\emptyset) = 0$ . Then  $\mu^* : \mathcal{P}(X) \rightarrow [0, \infty]$  defined as

$$\mu^*(B) := \inf \left\{ \sum_{n \in \mathbb{N}} \nu(A_n) : \{A_n\} \subset \mathcal{A}, B \subset \bigcup_{n \in \mathbb{N}} A_n \right\}$$

constitutes an outer measure on  $X$ . We call it the **outer measure generated by  $\nu$** .

**Remark 2.6.** Note that in the theorem,  $\mathcal{A}$  need not be an algebra, and  $\nu$  need not be a measure. However, if they are, they satisfy the conditions and the theorem holds, which is exactly what we want.

**Example 2.7** (Application to Lebesgue measure on  $\mathbb{R}$ )

It is easy to show that the outer measure generated on  $\mathcal{P}(\mathbb{R})$  is

$$\mu^*(B) := \inf \left\{ \sum_{n=1}^{\infty} (b_n - a_n) : B \subset \bigcup_{i=1}^{\infty} (a_i, b_i] \right\}$$

We call this the **Lebesgue outer measure**. As before, the definition generalizes into higher (finite) dimensions by taking boxes instead of intervals.

*Proof.* Why is  $\mu^*$  well-defined? - We can always take  $A_n = X$  as a cover. Let us check each condition for an outer measure:

- (i) Note that the empty set is a finite cover of itself, so  $\mu^*(\emptyset) = 0$ .
- (ii) if  $A \subset B$ , then note that any cover of  $B$  is also a cover of  $A$ , so taking the infimum over all possible covers yields  $\mu^*(A) \leq \mu^*(B)$ .
- (iii) Take any  $\{B_n\}_{n=1}^\infty \subset \mathcal{P}(X)$ . Fix  $\epsilon > 0$ . For each  $B_n$ , there is a cover  $\{A_k^n\}_{k \in \mathbb{N}} \subset \mathcal{A}$  such that

$$\sum_{k \in \mathbb{N}} \nu(A_k^n) \leq \mu^*(B_n) + \frac{\epsilon}{2^n}$$

The union of all these covers  $\{A_k^n\}_{k \in \mathbb{N}, n \in \mathbb{N}}$ , then covers  $\{B_n\}_{n=1}^\infty$ , hence

$$\mu\left(\bigcup_{n \in \mathbb{N}} B_n\right) \leq \sum_{k \in \mathbb{N}, n \in \mathbb{N}} \nu(A_k^n) \leq \sum_{n \in \mathbb{N}} \left(\mu^*(B_n) + \frac{\epsilon}{2^n}\right) \leq \sum_{n \in \mathbb{N}} \mu^*(B_n) + \epsilon$$

Since  $\epsilon$  was arbitrary, this shows that  $\mu^*$  is countably subadditive.

□

### Theorem 2.8 (Carathéodory Construction)

Let  $\mu^*$  be an outer measure on  $X$ , and consider the collection  $\mathcal{M}$  of subsets  $E \subset X$  such that for every  $B \subset X$ ,

$$\mu^*(B) \geq \mu^*(B \cap E) + \mu^*(B \cap E^c) \quad (*)$$

Then  $\mathcal{M}$  is a  $\sigma$ -algebra, and  $\bar{\mu} : \mathcal{M} \rightarrow [0, \infty]$  defined by  $\bar{\mu} = \mu^*|_{\mathcal{M}}$  ( $\bar{\mu}(E) = \mu^*(E)$  for  $E \in \mathcal{M}$ ) is a measure on  $\mathcal{M}$ .

### Example 2.9 (Application to Lebesgue measure on $\mathbb{R}$ )

This Theorem yields that the triple  $(\mathbb{R}, \mathcal{M}, \bar{\mu} := \mu^*|_{\mathcal{M}})$  is a measure space. We call  $\bar{\mu}$  the **Lebesgue measure**. We call  $\mathcal{M}$  the collection of **Lebesgue measurable** sets.

**Exercise 2.10.** Explain why  $(*)$  in the preceding theorem could be replaced by

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

*Proof.* We need to check four properties.

1.  $\emptyset \in \mathcal{M}$ :  $(*)$  holds for  $E = \emptyset$
2.  $\mathcal{M}$  is closed under complements: This is obvious since  $(*)$  is symmetric.

3.  $\mathcal{M}$  is closed under countable union:

Let us show first that  $\mathcal{M}$  is closed under finite union: let  $E, F \in \mathcal{M}$ . We want to show that  $E \cup F \in \mathcal{M}$ . Observe that

$$B \cap (E \cup F) = (B \cap E \cap F^c) \cup (B \cap E^c \cap F) \cup (B \cap E \cap F)$$

and because  $F \in \mathcal{M}$ , we have both

$$\mu^*(B \cap E \cap F^c) + \mu^*(B \cap E \cap F) \leq \mu^*(B \cap E) \quad \text{and}$$

$$\mu^*(B \cap E^c \cap F^c) + \mu^*(B \cap E^c \cap F) \leq \mu^*(B \cap E^c)$$

Now using countable subadditivity, these inequalities, and finally the fact that  $E \in \mathcal{M}$ , we get

$$\begin{aligned} \mu^*(B \cap (E \cup F)) + \mu^*(B \cap (E \cup F)^c) &\leq \mu^*(B \cap E \cap F^c) + \mu^*(B \cap E \cap F) \\ &\quad + \mu^*(B \cap E^c \cap F^c) + \mu^*(B \cap E^c \cap F) \\ &\leq \mu^*(B \cap E) + \mu^*(B \cap E^c) \leq \mu^*(B) \end{aligned}$$

This shows closedness under finite union. Now take  $\{E_n\}_{n \in \mathbb{N}} \subset \mathcal{M}$ . Without loss of generality, we take the  $E_n$  to be pairwise disjoint. Define  $E = \bigcup_{n \in \mathbb{N}} E_n$ . By closedness under finite union, we have

$$\begin{aligned} \mu^*(B \cap (\bigcup_{n=1}^N E_n)) &= \mu^*(B \cap (\bigcup_{n=1}^N E_n) \cap E_N) + \mu^*(B \cap (\bigcup_{n=1}^N E_n) \cap E_N^c) \\ &= \mu^*(B \cap E_N) + \mu^*(B \cap (\bigcup_{n=1}^{N-1} E_n)) \text{ iterate...} \\ &= \sum_{n=1}^N \mu^*(B \cap E_n) \end{aligned}$$

Next, note that we once again have by (\*) that for any  $N$ ,

$$\begin{aligned} \mu^*(B) &= \mu^*(B \cap (\bigcup_{n=1}^N E_n)) + \mu^*(B \cap (\bigcup_{n=1}^N E_n)^c) \\ &\geq \sum_{n=1}^N \mu^*(B \cap E_n) + \mu^*(B \cap (\bigcup_{n=1}^{\infty} E_n)^c) \end{aligned}$$

where the inequality derives from the observation that  $(\bigcup_{n=1}^N E_n)^c$  is a decreasing chain. Note that the left-hand side of the equation is independent of  $N$ , so it also holds for the limit. Thus

$$\begin{aligned} \mu^*(B) &\geq \sum_{n=1}^{\infty} \mu^*(B \cap E_n) + \mu^*(B \cap (\bigcup_{n=1}^{\infty} E_n)^c) \\ &\geq \mu^*(B \cap (\bigcup_{n=1}^{\infty} E_n)) + \mu^*(B \cap (\bigcup_{n=1}^{\infty} E_n)^c) \end{aligned}$$

This completes the proof.  $\mathcal{M}$  is a closed under countable union, and is therefore a  $\sigma$ -algebra.



4. Finally, we need to show that  $\bar{\mu}$  is a measure: Take the first of the two lines above, and replace  $B$  by  $\bigcup_{n=1}^{\infty} E_n$ . This yields  $\mu^*(\bigcup_{n=1}^{\infty} E_n) \geq \sum_{n=1}^{\infty} \mu^*(E_n)$ . Note that countable subadditivity yields the reverse inequality. Hence equality holds, which means that  $\bar{\mu}$  is a measure.

□

**Remark 2.11.** We have not seen whether the algebra and pre-measure we started with were preserved, in the sense that  $\mathcal{A} \subset \mathcal{M}$  and  $\bar{\mu}|_{\mathcal{A}} = \nu$ . Thankfully, this is the case.

**Theorem 2.12** (Carathéodory Extension Theorem - Existence)

Let  $(\mathcal{A}, \nu)$  be an algebra - premeasure pair on  $X$ . Let  $\mu^*$  denote the outer measure generated by  $\nu$ , and  $\mathcal{M}$  denote the  $\sigma$ -algebra from the Carathéodory construction. Let  $\sigma(\mathcal{A})$  be the  $\sigma$ -algebra generated by  $\mathcal{A}$  and  $\mu := \mu^*|_{\sigma(\mathcal{A})}$ . Then  $\sigma(\mathcal{A}) \subset \mathcal{M}$ , and  $\mu|_{\mathcal{A}} = \mu^*|_{\mathcal{A}} = \nu$ .

**Example 2.13** (Application to Lebesgue measure on  $\mathbb{R}$ )

Carathéodory proves that not only is  $(\mathbb{R}, \mathcal{M}, \bar{\mu} := \mu^*|_{\mathcal{M}})$  a measure space, but that the Borel-algebra  $\mathcal{B}(\mathbb{R})$  is a subset of  $\mathcal{M}$ , and the Lebesgue measure corresponds to the intuition we have on intervals and their countably infinite combinations. We call  $\mu := \mu^*|_{\sigma(\mathcal{A})}$  the **Borel measure**.

**Exercise 2.14.** Why is it true that the Borel-algebra  $\mathcal{B}(\mathbb{R})$  is a subset of  $\mathcal{M}$ ? Hint: Carathéodory does most of the work - you only need to show that  $\sigma(\mathcal{A}) = \sigma(\mathcal{O})$ .  $\mathcal{A}$  here refers to the specific algebra we started with at the beginning of the section.

**Theorem 2.15** (Carathéodory Extension Theorem - Uniqueness)

In the notation of the previous theorem, if  $X$  can be written as the union of countably many sets of finite pre-measure in the original algebra, i.e.  $X = \bigcup_{n \in \mathbb{N}} A_n, A_n \in \mathcal{A}, \nu(A_n) < \infty$ , then the extension of the pre-measure is unique.

**Remark 2.16.** The condition on  $X$  above is called  **$\sigma$ -finiteness**.

**Example 2.17** (Application to Lebesgue measure on  $\mathbb{R}$ )

The Lebesgue measure is a natural extension, and is unique. This is all we have ever wanted in life.

**Example 2.18** (Other measures on  $\mathbb{R}$ )

There is a simple procedure to generate important measures on the real line:

Take any monotone increasing, right-continuous function  $F$ . To the half-open interval  $(a, b]$ , assign the (pre-)measure  $\nu((a, b]) = F(b) - F(a)$ . The procedure above allows  $\nu$  to be extended to a measure on  $\mathcal{B}(\mathbb{R})$ , i.e. the Borel sets. Some relevant examples:

- Lebesgue Measure: as above, take  $F(x) = x$ .
- Probability Measure: Let  $F$  be the distribution function of a random variable  $X$ , i.e.  $F(x) = P(X \leq x)$ . This will generate a probability measure, with  $\mu_F((a, b]) = F(b) - F(a) = P(a < X \leq b)$ . We will properly define random variables later.
- Dirac Measure: Let  $F$  be the so-called 'Heavyside' or Unit Step function, i.e.

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases}. \text{ Then the resulting measure assigns 1 to any set containing}$$

$$\text{the origin, and 0 to any other set: } \delta_0(A) = \begin{cases} 1 & 0 \in A \\ 0 & 0 \notin A \end{cases}.$$

### §3 Measurable Functions

In the last section, we have shown that we can indeed find a 'sufficiently large'  $\sigma$ -algebra to define a measure on. The next big project we will be working on is to generalize our notion of integrability: Recall that for traditional Riemann integrals of a function  $f$ , we partitioned the domain into small intervals, and defined step functions that would be greater (lower) than the function  $f$  on each interval. We called the infimum (supremum) of these the upper (lower) Riemann sum. We called the function Riemann-integrable if the lower and upper sums coincide. The Dirichlet function  $f : [0, 1] \rightarrow \mathbb{R}$  given by

$$f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases} \text{ is probably the most frequently cited example of a non-Riemann-}$$

integrable function. The upper sum is always equal to 1, since both the rationals and irrationals are dense: in every interval of the partition, no matter how fine it is, there are both rational and irrational numbers. Note, however, that based on our construction in Section 2, the rationals  $\mathbb{Q}$  have Lebesgue measure 0.

**Exercise 3.1.** Prove that every countable subset of the real line has Lebesgue measure 0.

Based on this observation, we can also conclude that the irrationals in  $[0, 1]$  have measure 1. Maybe we would want to define  $\int_0^1 f d\mu = \mu_{[0,1]}(\mathbb{Q}) \cdot 1 + \mu_{[0,1]}(\mathbb{Q}^c) \cdot 0 = 0$ ?

We will formalize this idea, and present the general theory of Lebesgue Integration in Section 4. In this section, we will build up the necessary vocabulary to get there.

**Definition 3.2** (Measurable Functions). Let  $(\mathbb{X}, \mathcal{M}_X, \mu_X)$ ,  $(\mathbb{Y}, \mathcal{M}_Y, \mu_Y)$  be measure spaces. A function  $f : \mathbb{X} \rightarrow \mathbb{Y}$  is  **$\mathcal{M}_X, \mathcal{M}_Y$ -measurable** if  $\forall B \in \mathcal{M}_Y$ , the set

$$f^{-1}(B) = \{x \in \mathbb{X} : f(x) \in B\}$$

is measurable, i.e.  $f^{-1}(B) \in \mathcal{M}_X$ .

**Theorem 3.3**

Let  $(\mathbb{X}, \mathcal{M}_X, \mu_X)$ ,  $(\mathbb{Y}, \mathcal{M}_Y, \mu_Y)$  be a measure spaces. Let  $\mathcal{G}_Y$  be a set that generates the  $\sigma$ -algebra  $\mathcal{M}_Y$ , i.e.  $\sigma(\mathcal{G}_Y) = \mathcal{M}_Y$ .

Then  $f : \mathbb{X} \rightarrow \mathbb{Y}$  is  $\mathcal{M}_X, \mathcal{M}_Y$ -measurable if and only if  $\forall B \in \mathcal{G}_Y, f^{-1}(B) \in \mathcal{M}_X$ .

*Proof.* Necessity follows by definition of  $\mathcal{M}_X, \mathcal{M}_Y$ -measurability.

For sufficiency, define the set of all sets whose pre-image under  $f$  is measurable:  $\mathcal{B} = \{B \in \mathcal{M}_Y : f^{-1}(B) \in \mathcal{M}_X\}$ . By this definition,  $\mathcal{B} \subset \mathcal{M}_Y$ , and we want to show equality holds. For this, check that  $\mathcal{B}$  contains  $\mathcal{G}_Y$ , and is a  $\sigma$ -algebra, hence  $\mathcal{M}_Y = \sigma(\mathcal{G}_Y) \subset \sigma(\mathcal{B}) = \mathcal{B}$ .  $\square$

**Remark 3.4.** Theorem 3.3 is useful because it allows us to only check a subset of  $\mathcal{M}_Y$  to confirm that a function is measurable. The following example presents a special class of measurable functions (real-valued measurable functions) which we will focus on for the remainder of these notes.

**Example 3.5**

Let  $(\mathbb{X}, \mathcal{M}, \mu)$  be a measure space. A real-valued function  $f : X \rightarrow \mathbb{R}$  is  **$\mathcal{M}$ -measurable** if  $\forall a \in \mathbb{R}$ , the set

$$\{x \in X : f(x) < a\} \quad (*)$$

is measurable ( $\in \mathcal{M}$ ).

If  $\mathbb{X}$  is a metric space, and  $\mathcal{M} = \mathcal{B}(\mathbb{X})$ , then we say  $f$  is **Borel-measurable**.

If it is clear which  $\sigma$ -algebra is referred to, we can say  $f$  is **measurable**.

**Remark 3.6.** Let  $\mathcal{M}$  be the Lebesgue-measurable sets. Take  $E$  to be a set in  $\mathcal{M}$ , but not in  $\mathcal{B}(\mathbb{R})$ . Then  $f(x) = \begin{cases} 1 & x \in E \\ 0 & x \notin E \end{cases}$  is  $\mathcal{M}$ -measurable, but not  $\mathcal{B}(\mathbb{R})$ -measurable.

**Exercise 3.7.** Explain why the set  $(*)$  could be replaced by any of the following:

$$\{x \in X : f(x) \leq a\}$$

$$\{x \in X : f(x) > a\}$$

$$\{x \in X : f(x) \geq a\}$$

What is the  $\sigma$ -algebra on  $\mathbb{R}$  that we are explicitly using here?

**Remark 3.8.** We could also write the set  $(*)$  as  $f^{-1}((-\infty, a))$ . We call  $f^{-1}$  the pre-image of  $f$ .  $f^{-1}$  need not be a function.

We will now show that the measurable functions are preserved under many common operations.

**Theorem 3.9**

Let  $f, g, \{f_n\}_{n \in \mathbb{N}}, \{g_n\}_{n \in \mathbb{N}}$  be real-valued measurable functions on  $(X, \mathcal{M})$ , and let  $F : (Im(f), Im(g)) \rightarrow \mathbb{R}$  be continuous. Then the following are measurable:

1.  $f + g, f \cdot g, \max(f, g), \min(f, g), |f|$
2.  $\sup_{n \in \mathbb{N}} f_n(x), \inf_{n \in \mathbb{N}} f_n(x)$
3.  $\limsup_{n \rightarrow \infty} f_n(x), \liminf_{n \rightarrow \infty} f_n(x)$
4.  $F(f(x), g(x))$

**Exercise 3.10.** Explain why 2. and 4. imply 1.

*Proof.* The exercise asks you to prove 1. given 2. and 4.

2.  $\sup_{n \in \mathbb{N}} f_n(x), \inf_{n \in \mathbb{N}} f_n(x)$ :

Note that  $\inf_{n \in \mathbb{N}} f_n(x) = -\sup_{n \in \mathbb{N}} -f_n(x)$ , so it suffices to show sup is measurable. Take any  $a \in \mathbb{R}$ . Note that

$$\{x : \sup_{n \in \mathbb{N}} f_n(x) \leq a\} = \bigcap_{n \in \mathbb{N}} \{x : f_n(x) \leq a\}$$

and the intersection of measurable sets is measurable (i.e. in the same  $\sigma$ -algebra). Why does this equality hold? If  $\sup_{n \in \mathbb{N}} f_n(x) \leq a$ , then clearly  $f_n(x) \leq a \forall n$  for a given  $x$ . If for a given  $x$ ,  $f_n(x) \leq a \forall n$ , just take the supremum of both sides, which yields  $\sup_{n \in \mathbb{N}} f_n(x) \leq a$ .

3.  $\limsup_{n \rightarrow \infty} f_n(x), \liminf_{n \rightarrow \infty} f_n(x)$ :

Again,  $\liminf_{n \in \mathbb{N}} f_n(x) = -\limsup_{n \in \mathbb{N}} -f_n(x)$ , so we focus on lim sup. We claim

$$\limsup_{n \rightarrow \infty} f_n(x) \stackrel{(1)}{=} \lim_{n \rightarrow \infty} \sup_{m \geq n} f_m(x) \stackrel{(2)}{=} \inf_{n \in \mathbb{N}} \sup_{m \geq n} f_m(x)$$

If we can show that the above equalities hold, 3. reduces to 2. We fix  $x$ , such that we can think of  $f_n(x)$  as  $a_n$ . (2) is straightforward once we realize that  $\sup_{m \geq n} a_m$  is decreasing in  $n$ . (1) requires a bit more work:

- (1) Denote  $\alpha = \limsup_{n \rightarrow \infty} a_n = \sup\{a : a_{n_k} \rightarrow a \text{ for some subsequence } \{n_k\}\}$ . Fix  $\epsilon > 0$ . Then there is a subsequence such that  $\alpha - \epsilon < \lim_{k \rightarrow \infty} a_{n_k}$ . Note also that  $a_{n_k} \leq \sup_{m \geq k} a_m$ , hence  $\lim_{k \rightarrow \infty} a_{n_k} \leq \lim_{k \rightarrow \infty} \sup_{m \geq k} a_m$ . We get

$$\alpha - \epsilon < \lim_{k \rightarrow \infty} a_{n_k} \leq \lim_{k \rightarrow \infty} \sup_{m \geq k} a_m, \text{ hence } \alpha \leq \lim_{k \rightarrow \infty} \sup_{m \geq k} a_m$$

We now want to show the reverse inequality. Note that by the definition of lim sup, only finitely many  $a_m$  can be larger than  $\alpha + \epsilon$ . Thus there is a (potentially large but) finite  $M$  such that  $a_m \leq \alpha + \epsilon, \forall m \geq M$ , hence  $\sup_{m \geq M} a_m \leq \alpha + \epsilon$ , and therefore

$$\lim_{k \rightarrow \infty} \sup_{m \geq k} a_m \leq \alpha + \epsilon, \text{ hence } \alpha \geq \lim_{k \rightarrow \infty} \sup_{m \geq k} a_m$$

This proves the first equality.

4.  $F(f(x), g(x))$ :

For any  $a \in \mathbb{R}$ , note that  $F^{-1}((-\infty, a))$  is an open set. Note in particular, that this set can be written as a countable union of cubes (formally, we are using  $\sigma$ -finiteness of  $\mathbb{R}^n$  here). Then

$$\begin{aligned} \{x : F(f, g)(x) < a\} &= \{x : (f(x), g(x)) \in F^{-1}((-\infty, a))\} \\ &= \{x : (f(x), g(x)) \in F^{-1}((-\infty, a))\} \\ &= \{x : (f(x), g(x)) \in \bigcup_{n \in \mathbb{N}} (a_n, b_n) \times (c_n, d_n)\} \\ &= \bigcup_{n \in \mathbb{N}} \{x : (f(x), g(x)) \in (a_n, b_n) \times (c_n, d_n)\} \\ &= \bigcup_{n \in \mathbb{N}} (\{x : f(x) \in (a_n, b_n)\} \cap \{x : g(x) \in (c_n, d_n)\}) \end{aligned}$$

Thus  $F(f, g)$  is measurable. □

**Remark 3.11.** Even if  $f, g$  are measurable  $f \circ g$  might not be measurable. Likewise  $\sup_{\alpha \in \mathcal{A}} f_\alpha$  for an uncountable index set might not be measurable.

Next, we will look at a particular family of functions, quite appropriately called simple functions, which are measurable under a straightforward condition.

**Definition 3.12** (Simple functions). A function  $f : X \rightarrow \mathbb{R}$  is called a **simple function** if the range of  $f$  is finite (i.e.  $f$  takes on only finitely many values).

**Example 3.13**

Any step function is a simple function. The Dirichlet function  $f(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & x \notin \mathbb{Q} \end{cases}$  is also simple function.

**Definition 3.14** (Indicator function). We call  $f(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$  the indicator function of  $A$ , and denote it by  $\chi_A$ .

**Remark 3.15** (Standard representation of simple functions). As discussed above, a simple function  $s : X \rightarrow \mathbb{R}$  takes on finitely many values  $\{c_1, c_2, \dots, c_N\}$ . Define  $E_i = \{x \in X : s(x) = c_i\}$ . Then it is clear that  $s(x) = \sum_{i=1}^N c_i \chi_{E_i}$ . It is obvious that  $s$  is measurable if and only if each of the  $E_i$  are measurable.

As described in the beginning of this section, we already have an idea what the integral of a simple function might be. Of course we are mainly interested in functions that are not simple. The following result shows why simple functions provide the first crucial step towards our new theory of integration.

**Theorem 3.16** (Approximation by simple functions)

For any  $f : X \rightarrow \mathbb{R}$ ,

- (1)  $\exists \{s_n\}$ , such that  $s_n \rightarrow f$  pointwise, i.e. for any  $x \in X$ ,  $\lim_{n \rightarrow \infty} s_n(x) = f(x)$ ,
- (2) if  $f$  is measurable,  $\{s_n\}$  may be taken to be measurable, and
- (3) if  $f \geq 0$ ,  $\{s_n\}$  may be taken as an increasing sequence (i.e.  $s_n \leq s_{n+1}$ )
- (4) if  $f$  is bounded, the convergence in (1) is uniform.

*Proof.* We proceed via a constructive proof for (3). Suppose  $f \geq 0$ , i.e.  $\text{Im}(f) \subset [0, \infty)$ . For each  $n$ , we partition the range:  $[0, \infty) = [0, \frac{1}{2^n}) \cup [\frac{1}{2^n}, \frac{2}{2^n}) \cup \dots \cup [\frac{n \cdot 2^n - 1}{2^n}, \frac{n \cdot 2^n}{2^n}) \cup [n, \infty)$ . We now find the corresponding sets in the domain. Define

$$E_i^n = f^{-1}([\frac{i-1}{2^n}, \frac{i}{2^n})) = \{x \in X : \frac{i-1}{2^n} \leq f(x) < \frac{i}{2^n}\}$$

$$E_\infty^n = f^{-1}([n, \infty)) = \{x \in X : n \leq f(x)\}$$

Note that the  $n$ th partition is a refinement of any earlier partition. Hence if we define

$$s_n(x) := \sum_{i=1}^{n \cdot 2^n} \frac{i-1}{2^n} \chi_{E_i^n} + n \chi_{E_\infty^n},$$

the  $s_n$  are non-negative and increasing. To see that they converge to  $f$  pointwise, fix any  $\epsilon > 0$ . Note that for any  $x$ , since  $f(x) \in [0, \infty)$  (i.e. finite), there is an  $M \in \mathbb{N}$  such that  $f(x) < M$ , so  $x \in E_i^M$  for some  $i$ . Note there is an  $N \geq M$  such that  $\frac{1}{2^N} < \epsilon$ . Then for any  $n \geq N$ ,  $|f(x) - s_n(x)| < \epsilon$ , so we have pointwise convergence. Note that the  $s_n$  are measurable if  $f$  is. Now for general  $f$ , note that we can decompose  $f = \max(0, f) + \min(0, f) = \max(0, f) - \max(0, -f) = f^+ - f^-$ , where  $f^+, f^-$  are non-negative. (3) provides sequences  $s_n^+, s_n^-$  that converge pointwise to  $f^+, f^-$ , respectively.  $s_n = s_n^+ - s_n^-$  then converges pointwise to  $f$ .  $\square$

**Exercise 3.17.** Prove (4).

## §4 Lebesgue Integration

We're ready to define our new integral! Wild stuff guys, wild stuff. Once again, we will proceed carefully: the next theorem defines the Lebesgue integral for non-negative simple functions, after which we will proceed to non-negative measurable functions, and finally any measurable function.

**Definition 4.1** (Lebesgue integral, nonnegative measurable simple function). If  $s(x) = \sum_{i=1}^N c_i \chi_{E_i}$ ,  $E_i$  measurable, then for any  $E \in \mathcal{M}$  we define

$$\int_E s d\mu = \sum_{i=1}^N c_i \mu(E \cap E_i)$$

as the Lebesgue integral of the simple function  $s$ .

**Definition 4.2** (Lebesgue integral, non-negative measurable function). If  $f$  is measurable,  $f \geq 0$ , then for any  $E \in \mathcal{M}$  we define

$$\int_E f d\mu = \sup \left\{ \int_E s d\mu : 0 \leq s \leq f, s \text{ simple, measurable} \right\}$$

as the Lebesgue integral of the function  $f$ .

**Definition 4.3** (Lebesgue integral, general measurable function). If  $f$  is measurable, then for any  $E \in \mathcal{M}$  we define

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

as the Lebesgue integral of the function  $f$ , if at least one of the two integrals is finite

#### Example 4.4

As a simple example of a function that is measurable, but for which we have not defined an integral, take the sign function  $\text{sgn}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}$ . Both integrals in the preceding definition would be infinite, and subtracting infinity from infinity constitutes terrible table manners.

#### Proposition 4.5

If  $a \leq f(x) \leq b, \forall x \in E$ , and  $\mu(E) < \infty$ , then

$$a\mu(E) \leq \int_E f d\mu \leq b\mu(E) \quad (*)$$

*Proof.* Note that  $a\chi_E$  is simple and measurable, and  $a\chi_E \leq f$ , hence

$$a\mu(E) = \int_E a\chi_E d\mu \leq \sup \left\{ \int_E s d\mu : 0 \leq s \leq f, s \text{ simple, measurable} \right\} = \int_E f d\mu.$$

On the other hand note that any simple  $s$  such that  $s \leq f$  must also satisfy  $s \leq b$ , so

$$\int_E s d\mu = \sum_{i=1}^N c_i \mu(E \cap E_i) \leq b\mu(E)$$

Taking the supremum on both sides yields the second inequality.  $\square$

#### Proposition 4.6

If  $f$  is measurable, and  $\mu(E) = 0$ , then

$$\int_E f d\mu = 0$$

*Proof.* 1. For  $s$  simple, we have  $\int_E s d\mu = \sum_{i=1}^N c_i \mu(E \cap E_i) \leq \sum_{i=1}^N c_i \mu(E) = 0$ .

2. For general  $f$ , we are taking a supremum over integrals of simple functions, which are all 0. Thus  $\int_E f d\mu = 0$ .  $\square$

**Proposition 4.7**

If  $f, g \geq 0$  are measurable on  $E$ ,  $f \leq g$  and  $E \in \mathcal{M}$ , then  $\int_E f d\mu \leq \int_E g d\mu$ .

*Proof.*  $\{s : 0 \leq s \leq f, s \text{ simple, measurable}\} \subset \{s : 0 \leq s \leq g, s \text{ simple, measurable}\}$ , hence  $\int_E f d\mu \leq \int_E g d\mu$ .  $\square$

**Proposition 4.8**

If  $c$  is a constant,  $E \in \mathcal{M}$ , and  $\int_E f d\mu$  is well defined, then  $c \int_E f d\mu = \int_E c f d\mu$ .

*Proof.* For simple functions, the property holds easily:

$$c \int_E s d\mu = c \sum_{i=1}^N c_i \mu(E \cap E_i) = \sum_{i=1}^N c \cdot c_i \mu(E \cap E_i) = \int_E c s d\mu.$$

Suppose  $c > 0, f \geq 0$ , such that  $c \cdot f \geq 0$  is well-defined. If  $s$  with  $0 \leq s \leq f$  is simple  $cs$  is also simple, hence

$$\begin{aligned} c \int_E f d\mu &= c \sup \left\{ \int_E s d\mu : 0 \leq s \leq f, s \text{ simple, measurable} \right\} \\ &= \sup \left\{ c \int_E s d\mu : 0 \leq s \leq f, s \text{ simple, measurable} \right\} \\ &= \sup \left\{ \int_E c s d\mu : 0 \leq s \leq f, s \text{ simple, measurable} \right\} \\ &= \sup \left\{ \int_E r d\mu : 0 \leq r \leq c f, r \text{ simple, measurable} \right\} \\ &= \int_E c f d\mu \end{aligned}$$

This finishes the proof for  $f \geq 0, c > 0$ . We can apply the above to  $f^+, f^-$ , and use the fact that  $\int_E f d\mu$  is well-defined to observe that the proposition holds for general  $f, c > 0$ . For  $c < 0$ , observe that  $c \cdot f = (-c) \cdot (-f)$ , so it suffices to prove the proposition for  $c = -1$ . This is easy to check.  $\square$

**Definition 4.9** (Integrability,  $\mathcal{L}^1(E, \mu)$ ). If both  $\int_E f^+ d\mu$  and  $\int_E f^- d\mu$  are finite, we say that  $f$  is **integrable** with respect to  $\mu$ . The space of integrable function is called  $\mathcal{L}^1$ .

**Remark 4.10.** Note that because  $|f| = f^+ + f^-$ , this also means that  $\int_E |f| d\mu < \infty$ . Hence the condition is sometimes called absolutely integrable.

**Definition 4.11** ( $\mathcal{L}^p$ -spaces). More generally,  $\mathcal{L}^p(X, \mu)$ -spaces are functional spaces (spaces whose members are functions) defined by the following condition:

$$f \in \mathcal{L}^p(E, \mu) \Leftrightarrow \int_E |f|^p d\mu < \infty$$

**Remark 4.12.**  $L^2 = \mathcal{L}^2 / \sim$  is an example of an inner product space, which will be the topic of next week.

**Exercise 4.13.** If  $f$  is measurable,  $|f| < M$  on  $E \in \mathcal{M}$  and  $\mu(E) < \infty$ , then  $f \in \mathcal{L}^1(\mu, E)$ .



**Exercise 4.14.** If  $f \in \mathcal{L}^1(\mu, E)$ , then  $f$  is finite almost everywhere on  $E$ .

**Exercise 4.15.**  $f, g \in \mathcal{L}^1(\mu, E)$ ,  $f \leq g$  on  $E \Rightarrow \int_E f d\mu \leq \int_E g d\mu$

**Exercise 4.16.** If  $f \in \mathcal{L}^1(\mu, E)$ ,  $A \in \mathcal{M}$ ,  $A \subset E \Rightarrow f \in \mathcal{L}^1(\mu, A)$

Riemann integrals satisfied nice properties, like additivity of integrals (over disjoint sets),  $\int_{A \cup B} f dx = \int_A f dx + \int_B f dx$ , and linearity,  $\int_A f + g dx = \int_A f dx + \int_A g dx$ . We can actually do even better than that - we want to check whether these properties are preserved in the Lebesgue integral.

#### Theorem 4.17

If  $s$  is any non-negative  $\mathcal{M}$ -measurable simple function, then  $\lambda(A) = \int_A s d\mu$  is a measure on  $\mathcal{M}$ .

**Remark 4.18.** This shows that the Lebesgue integral of a simple function is countably additive

*Proof.* Let  $A = \bigcup_{n=1}^{\infty} A_i$ ,  $A_i$  disjoint, then

$$\begin{aligned} \lambda(A) &= \sum_{i=1}^n c_i \cdot \mu(A \cap E_i) = \sum_{i=1}^n c_i \cdot \mu((\bigcup_{j=1}^{\infty} A_j) \cap E_i) \\ &= \sum_{i=1}^n c_i \cdot \mu(\bigcup_{j=1}^{\infty} (A_j \cap E_i)) \\ &= \sum_{i=1}^n c_i \cdot \sum_{j=1}^{\infty} \mu(A_j \cap E_i) \\ &= \sum_{j=1}^{\infty} \sum_{i=1}^n c_i \cdot \mu(A_j \cap E_i) \\ &= \sum_{j=1}^{\infty} \int_{A_j} s d\mu \\ &= \sum_{j=1}^{\infty} \lambda(A_j), \end{aligned}$$

so  $\lambda$  is indeed a measure on  $\mathcal{M}$ . □

#### Theorem 4.19

If  $f$  is any non-negative  $\mathcal{M}$ -measurable function, then  $\lambda(A) = \int_A f d\mu$  is a measure on  $\mathcal{M}$ .

*Proof.* Let  $A = \bigcup_{n=1}^{\infty} A_i$ ,  $A_i$  disjoint, then for all simple  $s$  such that  $0 \leq s \leq f$ , we have

$$\begin{aligned} \int_A s d\mu &= \sum_{n=1}^{\infty} \int_{A_n} s d\mu \\ &\leq \sum_{n=1}^{\infty} \int_{A_n} f d\mu = \sum_{n=1}^{\infty} \lambda(A_n) \end{aligned}$$

By taking the supremum, we get

$$\lambda(A) \leq \sum_{n=1}^{\infty} \lambda(A_n).$$

The reverse inequality is again a bit harder to show. Let  $\epsilon > 0$ . By definition of the integral, we can find  $s_n$  simple such that

$$\int_{A_n} s_n d\mu \geq \int_{A_n} f d\mu - \frac{\epsilon}{2^n}$$

Since the  $A_n$  are disjoint, and simple integrals are additive, we can define  $s = \sum_{n=1}^N s_n$  such that

$$\begin{aligned} \lambda(A) &\geq \lambda\left(\bigcup_{n=1}^N A_n\right) \\ &\geq \int_{\bigcup_{n=1}^N A_n} s d\mu = \sum_{n=1}^N \int_{A_n} s d\mu \\ &\geq \sum_{n=1}^N \int_{A_n} f d\mu - \epsilon \end{aligned}$$

$\epsilon$  was arbitrary, so we have  $\lambda(A) \geq \sum_{n=1}^{\infty} \lambda(A_n)$ , and taking the limit yields the desired inequality.

For general  $f$ , observe once more  $f = f^+ - f^-$ , and apply known results to these non-negative functions.  $\square$

#### Corollary 4.20

If  $A, B \in \mathcal{M}$ ,  $B \subset A$  and  $\mu(A - B) = 0$ , then if  $f \in \mathcal{L}^1$ ,

$$\int_A f d\mu = \int_B f d\mu.$$

**Exercise 4.21.** Prove the corollary.

#### Theorem 4.22

If  $f$  is measurable with  $|f| \leq g \in \mathcal{L}(E, \mu)$ , then  $f \in \mathcal{L}(E, \mu)$ .

*Proof.* Note  $0 \leq f^+, f^- \leq g$ , non-negative, hence  $\int_E f^{+/-} d\mu \leq \int_E g d\mu < \infty$ .  $\square$

#### Theorem 4.23 (Lebesgue Monotone Convergence Theorem)

Given  $E \in \mathcal{M}$  and  $\{f_n\}$  such that  $0 \leq f_1(x) \leq f_2(x) \leq \dots$ , let  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$ . Then

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

*Proof.* Note that  $\int_E f_n d\mu$  is an increasing sequence in  $[0, \infty]$ , so  $\alpha := \lim_{n \rightarrow \infty} \int_E f_n d\mu \in [0, \infty]$ . Since  $0 \leq f_n \leq f$ , we have  $\int_E f_n d\mu \leq \int_E f d\mu$ . Taking the supremum,

$$\alpha \leq \int_E f d\mu$$

To see the reverse inequality holds, take any simple function  $s$  such that  $0 \leq s \leq f$ , and any  $\epsilon \in (0, 1)$ . Take  $E_n = \{x \in E : f_n(x) > (1 - \epsilon)s(x)\}$ . Since  $f_n(x) \rightarrow f(x), \forall x \in E$ , and  $f_n$  are increasing, not that  $E_n \subset E_{n+1}$ , and  $E = \bigcup_{n=1}^{\infty} E_n$ . Hence

$$\int_E f_n d\mu \geq \int_{E_n} f_n d\mu \geq (1 - \epsilon) \int_{E_n} s d\mu$$

Taking limits on both sides, the LHS yields  $\alpha$ , while the RHS converges to  $(1 - \epsilon) \int_{E_n} s d\mu$  by Theorem 1.19 (i).  $\epsilon$  was arbitrary, so this indeed gives us

$$\alpha \geq \int_E s d\mu, \forall s : 0 \leq s \leq f$$

Taking the supremum over all these measurable simple functions, we get the desired inequality, which completes the proof.  $\square$

#### Corollary 4.24

Given  $E \in \mathcal{M}$  and  $\{f_n\}$  such that  $0 \leq f_1(x) \leq f_2(x) \leq \dots$ , let  $f(x) = \sum_{n=1}^{\infty} f_n(x)$ . Then

$$\sum_{n=1}^{\infty} \int_E f_n d\mu = \int_E f d\mu$$

*Proof.* Let  $g_m = \sum_{n=1}^m f_n(x)$ . Note that  $g_m \rightarrow f$ . Use linearity, apply MCT.  $\square$

#### Lemma 4.25 (Fatou's Lemma)

Given  $E \in \mathcal{M}$  and  $\{f_n\}$  non-negative, measurable, let  $f(x) = \liminf_{n \rightarrow \infty} f_n(x)$ . Then

$$\int_E f d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu.$$

*Proof.* Recall from our discussion of measurable functions that  $\liminf_{n \rightarrow \infty} f_n(x) = \lim_{n \rightarrow \infty} \inf_{m \geq n} f_m(x) = \lim_{n \rightarrow \infty} g_n(x)$ . Note further that  $0 \leq g_1(x) \leq g_2(x) \leq \dots$ ,  $g_n(x) \rightarrow f(x)$ , so by MCT,

$$\int_E g_n d\mu \rightarrow \int_E f d\mu.$$

Next, note that by definition of the  $g_n$ , we have  $0 \leq g_n(x) = \inf_{m \geq n} f_m(x) \leq f_n(x), \forall n$ . Taking the liminf on both sides yields

$$\liminf_{n \rightarrow \infty} \int_E g_n(x) \leq \liminf_{n \rightarrow \infty} \int_E f_n(x)$$

We combine the two central equations to generate the desired inequality:

$$\int_E f d\mu = \lim_{n \rightarrow \infty} \int_E g_n(x) \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu.$$

$\square$

**Theorem 4.26** (Lebesgue's dominated convergence theorem)

Given  $E \in \mathcal{M}$  and  $\{f_n\}$  measurable, and  $f_n(x) \rightarrow f(x)$  on  $E$ , if there exists a function  $g \in \mathcal{L}(E, \mu)$  such that  $|f_n(x)| \leq g(x)$ , then

$$\int_E f_n d\mu \rightarrow \int_E f d\mu.$$

*Proof.* Note  $g \geq 0$ ,  $|f_n| \leq g \in \mathcal{L}(E, \mu)$ , so  $f_n \in \mathcal{L}(E, \mu)$ ,  $\forall n \in \mathbb{N}$ . Note  $0 \leq f_n + g \rightarrow f + g$ , and apply Fatou's Lemma to get:

$$\int_E f d\mu + \int_E g d\mu = \int_E f + g d\mu \quad (\text{linearity})$$

$$\leq \liminf_{n \rightarrow \infty} \int_E f_n + g d\mu \quad (\text{Fatou})$$

$$= \liminf_{n \rightarrow \infty} \left( \int_E f_n d\mu + \int_E g d\mu \right) \quad (\text{linearity})$$

But  $g \in \mathcal{L}(E, \mu)$ , so  $\int_E f d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu$ . For the reverse inequality,

$$\int_E -f d\mu + \int_E g d\mu = \int_E -f + g d\mu \quad (\text{linearity})$$

$$\leq \liminf_{n \rightarrow \infty} \int_E -f_n + g d\mu \quad (\text{Fatou})$$

$$= \liminf_{n \rightarrow \infty} \left( - \int_E f_n d\mu + \int_E g d\mu \right) \quad (\text{linearity})$$

Again  $\int_E g d\mu$  is finite, and we have the following identity:

$$\begin{aligned} \int_E -f d\mu &\leq \liminf_{n \rightarrow \infty} \left( - \int_E f_n d\mu \right) = - \limsup_{n \rightarrow \infty} \left( \int_E f_n d\mu \right) \\ \Leftrightarrow \limsup_{n \rightarrow \infty} \left( \int_E f_n d\mu \right) &\leq \int_E f d\mu \end{aligned}$$

**Theorem 4.27**

Suppose  $f : E \subset \mathbb{R}$  is Riemann-integrable. Then  $f \in \mathcal{L}(E, \mu)$ , and  $\int_E f d\mu = \int_E f dx$ .

Hence  $\limsup_{n \rightarrow \infty} \int_E f_n d\mu \leq \int_E f d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu$ , so  $\int_E f_n d\mu \rightarrow \int_E f d\mu$ .  $\square$

**Exercise 4.28.** If  $f \in \mathcal{L}^1(\mu, E)$ , then  $f$  is finite almost everywhere on  $E$ .

**§5 Introduction to measure-theoretic probability**

This section, in many ways, is a reprise of Section 3. As it turns out, random variables are nothing but a special case of measurable functions. First, let us recall the definition of probability space:

**Definition 5.1** (Probability Space).  $(\Omega, \mathcal{F}, P)$  is a probability space if  $(\Omega, \mathcal{F}, P)$  is a measure space and  $\mu(\Omega) = 1$ . We call  $\Omega$  the sample space, elements  $\omega \in \Omega$  outcomes, and  $F \in \mathcal{F}$  events.

**Definition 5.2** (Random Variable). Let  $(\Omega, \mathcal{F}, P)$  be a probability space, and  $(E, \epsilon)$  a measurable space. A **random variable** is an  $\mathcal{F}$ -measurable function  $X : \Omega \rightarrow E$ , i.e.

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

If  $(E, \epsilon) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , we say  $X$  is a real random variable. In this section, we will only concern ourselves with real random variables, though generalizations (for example, complex random variables) do exist.

### Theorem 5.3

Let  $X, Y, \{X_n\}_{n \in \mathbb{N}}, \{Y_n\}_{n \in \mathbb{N}}$  be measurable functions on  $(\Omega, \mathcal{F})$ , and let  $F : (Im(X), Im(Y)) \rightarrow \mathbb{R}$  be continuous. Then the following are measurable:

1.  $X + Y, X \cdot Y, \max(X, Y), \min(X, Y), |X|$
2.  $\sup_{n \in \mathbb{N}} X_n(\omega), \inf_{n \in \mathbb{N}} X_n(\omega)$
3.  $\limsup_{n \rightarrow \infty} X_n(\omega), \liminf_{n \rightarrow \infty} X_n(\omega)$
4.  $F(X(\omega), Y(\omega))$

*Proof.* Note that this is just a special case of Theorem 3.9. □

**Remark 5.4.** In particular, note that this implies that expressions such as  $\frac{1}{n} \sum_{i=1}^n X_i$  or  $\frac{1}{n-1} \sum_{i=1}^n X_i^2$  are also random variables.

**Definition 5.5** (Indicator Random Variable). For any  $A \in \mathcal{F}$ , define  $I_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$ .

Note this is just a special case of an indicator function, with  $(\Omega, \mathcal{F}, P)$  a probability space.

**Definition 5.6** (Simple Random Variable). A simple random variable is a random variable that can be written as  $X(\omega) = \sum_{i=1}^n c_i I_{A_i}(\omega)$ , for  $\cup_{i=1}^n A_i = \Omega$ ,  $c_i \in \mathbb{R}$ . Again, this is just a special case of a simple function, with  $(\Omega, \mathcal{F}, P)$  a probability space.

### Example 5.7 (Random Walk, Repeated coin tosses)

**Remark 5.8.** In probability theory,  $X, Y$  and  $Z$  are most commonly used to denote random variables. To hammer home the point that random variables are nothing but measurable functions between measure spaces, I will use  $f$  and  $g$  in the upcoming definitions and theorems.

**Definition 5.9** ( $\sigma$ -algebra generated by a real random variable). Given a real random variable  $g : \Omega \rightarrow \mathbb{R}$ , we define the  $\sigma$ -algebra generated by  $g$ ,  $\mathcal{G}$  as:

$$\mathcal{G} = \sigma(\cup_{\alpha \in \mathbb{R}} \{\omega \in \Omega : g(\omega) \leq \alpha\}) = \sigma(\{g^{-1}(B) : B \in \mathcal{B}(\mathbb{R})\})$$

i.e. it is the smallest  $\sigma$ -algebra containing all sets of the form  $\{\omega \in \Omega : g(\omega) \leq \alpha\}$  for  $\alpha \in \mathbb{R}$ .

**Remark 5.10.** One interpretation of  $\mathcal{G}$  is that it is the coarsest (i.e. smallest)  $\sigma$ -algebra on  $\Omega$  that makes the random variable  $g$  measurable. This becomes more interesting if we look at sequences of random variables - stochastic processes. The  $\sigma$ -algebra generated by  $X_0, X_1, \dots, X_t$  can be interpreted as the information available up to (and including) time  $t$ . We will make this more formal later if we have time.

**Exercise 5.11.**

**Theorem 5.12**

**Lemma 5.13**

*Proof.*

□

**Definition 5.14.**