

Math 172: Lebesgue Integration and Fourier Analysis

Charlotte Chan

Spring Quarter 2011

These are lecture notes that I typed up for Professor Kannan Soundararajan's course (Math 172) on Lebesgue Integration and Fourier Analysis in Spring 2011. I should note that these notes are not polished and hence might be riddled with errors. If you notice any typos or errors, please do contact me at charchan@stanford.edu.

Contents

1	Motivation	3
2	The Notion of a Measure	7
3	The Outer Measure	10
4	Borel Measurable Sets	15
5	The Probability Space	18
6	The Lebesgue Integral	22
7	\mathcal{L}^1 -Theory	35
8	Lebesgue vs. Riemann	38
9	Random Variables and \mathcal{L}^p -spaces	42
10	Hilbert Spaces	46
11	Classical Fourier Series	49

Chapter 1

Motivation

We will discuss measure theory in this course; in particular the Lebesgue measure on \mathbb{R} and \mathbb{R}^n . In the latter half of the course we will move to discuss some Fourier Analysis.

For now, we give some motivation for measure theory via probability. We consider two questions that are quite similar in nature:

1. Toss a coin infinitely many times. Is it likely that there are infinitely many stopping points $2n$ after which you have n heads and n tails? (One would think yes.)
2. Roll a 6-sided die infinitely many times. Is it likely that there are infinitely many stopping points after which you have n 1's, 2's, \dots 6's? (The answer is no.) Is it likely that 1 never shows up? (One would think no.)

It is nontrivial to answer these questions, and in fact, it is not even clear what the question means. How can we make precise the notion of “likelihood”? We can first translate question 1 to a question about numbers in the interval $[0, 1)$.

Consider a Bernoulli sequence of heads and tails: $HTHHTHHHTTTH \dots$. We can assign H to 1 and T to 0 and then view this sequence as a binary expansion of a real number in the interval $[0, 1)$. Concretely, we can associate to each Bernoulli sequence the number

$$\sum_{n=1}^{\infty} \frac{1}{2^n} \begin{cases} 1 & \text{if } n\text{th toss is } H \\ 0 & \text{if } n\text{th toss is } T. \end{cases}$$

We have one small problem: This association is not bijective. That is, a sequence that ends in all H 's may be associated to the same real number as one that ends in all T 's. To get around this issue, we can just pick the binary expansion ending in all T 's if we have a choice. Hence we have a bijective map

$$\phi : [0, 1) \rightarrow \text{Bernoulli sequences except for those ending in all } H\text{'s}.$$

Note that there are only countably many Bernoulli sequences ending in all H 's.

Now we can look at what subsets of $[0, 1)$ are associated to particular coin toss specifications. As an easy first example, we can see that the set corresponding to the first toss being H is $[\frac{1}{2}, 1)$. The set corresponding to the first three tosses being HTT is $[\frac{1}{2}, \frac{1}{2} + \frac{1}{8})$. Note now that the probability of each of these events happening is exactly the length of the corresponding interval in $[0, 1)$. So for the coin tosses wherein the first 100 tosses have 50 heads and 50 tails, the corresponding set is a union of $\binom{100}{50}$ intervals, each of length $\frac{1}{2^{100}}$.

We have hence translated a question about coin tosses to a question about the “size” of some set. In a similar way, we can translate question 2, but using base 6 instead of base 2 in our expansion. Hence we associate each sequence of die rolls to the real number

$$\sum_{n=1}^{\infty} \frac{1}{6^n} \begin{cases} 0 & \text{if die} = 1 \\ \vdots & \vdots \\ 5 & \text{if die} = 6. \end{cases}$$

As before, we obtain a bijective map

$$\phi : [0, 1) \rightarrow \text{sequences except those that end in all 6's.}$$

To address the second part of question 2, we can ask what the set associated to the sequences wherein 1 never shows up looks like. What we get is a Cantor set obtained in the following way: We start with $[0, 1)$ and delete the first $1/6$ to illustrate all the sequences that do not start with 1. We have $[1/6, 1)$ remaining and we delete the first $1/6^2$ of each remaining $1/6$ -length interval. So we get 5 disjoint intervals, and then in the next iteration, 25 disjoint intervals, then 125, and so forth. This is the Cantor set which is, on the one hand very small (in terms of “length”) and on the other hand very large (in terms of number of elements—it is uncountable!).

It is natural to ask why we would bother with Lebesgue measures, and one place where this is very important is in integration. With Riemann integrals, we can integrate functions that are “sufficiently nice” (i.e. there are not too many discontinuities). But this way of integrating breaks down even for some really simple (but highly discontinuous) functions. Consider for instance the integral

$$\int_0^1 \chi_S(x) dx,$$

where χ_S is the characteristic function on S ; i.e. $\chi_S(x) = \begin{cases} 1 & \text{if } x \in S \\ 0 & \text{if } x \notin S. \end{cases}$ If

$S = \mathbb{Q} \cap [0, 1]$, then we are in big trouble and the Riemann integral does not exist. (The upper sum is 1 and the lower sum is 0.) Intuitively, we would expect that the probability of “hitting” a rational is 0, and so we would like to have a theory wherein the above integral is 0. This is another motivation for Lebesgue theory.

To give one last motivation, consider the space $L^2([0, 1])$, which is defined to be the space of functions $f : [0, 1] \rightarrow \mathbb{R}$ such that $\|f\|_2^2 := \int_0^1 |f(x)|^2 dx < \infty$. Consider a sequence of functions $f_1, f_2, \dots \in L^2([0, 1])$ such that $\|f_m - f_n\|_2 < \varepsilon$ if $m, n > N(\varepsilon)$. A natural question to ask is: Does there exist a function $f \in L^2([0, 1])$ such that $\lim_{n \rightarrow \infty} \|f - f_n\|_2 = 0$? If so, then we have a complete metric space, which is always nice. Lebesgue theory allows us to do this.

Now that we have (I think) sufficiently motivated the theory, we can talk more specifically about measures. First we can ask: What do we want from a measure? Given a set $S \subseteq \mathbb{R}$, we want to build up some notion of the size (i.e. measure) of S . Through class discussion, we have several points:

1. $S = [a, b]$, (a, b) , $[a, b)$, or $(a, b]$. Want $\mu(S) = b - a$?
2. Translation invariance: We would like $\mu(S+t) = \mu(S)$ for all $t \in \mathbb{R}$. (Here, $S+t = \{s+t : s \in S\}$.)
3. $\mu(S) \geq 0$.
4. If $A \subseteq B$, then $\mu(A) \leq \mu(B)$.
5. Finite additivity. If A and B are disjoint, then $\mu(A \cup B) = \mu(A) + \mu(B)$.
6. Countable additivity. If A_1, A_2, \dots is a disjoint countable sequence of sets, then

$$\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

7. All open sets should have a measure. Also, all closed sets should have a measure.

In this way, we can sort of get a feel for what things we would want to demand from a measure.

Now, ideally, we would like to have

- (i) All subsets of \mathbb{R} are measurable
- (ii) Countable additivity
- (iii) Translation invariance
- (iv) $\mu([a, b]) = b - a$

However, this does not exist. What we can do is to keep (ii), (iii), and (iv) and restrict ourselves to a nicer (but still large) class of subsets of \mathbb{R} . To see what we cannot possibly satisfy all four of the above, we consider the following example.

Consider the quotient of additive groups \mathbb{R}/\mathbb{Q} and let S be a set of representatives in $[0, 1]$ for each equivalence class in \mathbb{R}/\mathbb{Q} . (So we partition \mathbb{R} by the equivalence relation \sim defined by $x \sim y$ if $x - y \in \mathbb{Q}$.) Consider the intersection $\mathbb{Q} \cap [-1, 1]$. This is a countable set, and so we can enumerate the rationals

r_1, r_2, \dots . Notice that for $r_n \neq r_m$, we have $(S + r_n) \cap (S + r_m) = \emptyset$. Now consider the (countable) union

$$A := \bigcup_{n=1}^{\infty} (S + r_n).$$

This is a disjoint union. It is clear that $A \subseteq [-1, 2]$. Furthermore, for any $\alpha \in [0, 1]$, there is a representative $x \in S$ such that $q := x - \alpha \in \mathbb{Q}$. Also, $q \in [-1, 1]$, which means that $\alpha \in S + q \subseteq A$. Therefore $[0, 1] \subseteq A$.

Now suppose that S has a measure. By translation invariance, $\mu(S) = \mu(S + r_n)$ for all $n \in \mathbb{N}$, and by countable additivity, A must also have a measure. Since we showed above that $[0, 1] \subseteq A \subseteq [-1, 2]$, then we know $1 \leq \mu(A) \leq 3$. On the other hand, we have

$$\mu(A) = \sum_{n=1}^{\infty} \mu(S + r_n) = \sum_{n=1}^{\infty} \mu(S).$$

This is a contradiction and we have hence shown that S is unmeasurable.

In some sense, weird, unmeasurable sets like the above set exist because there are “too many” subsets of the real numbers. There are $2^{\mathbb{R}}$, and this is much bigger than \mathbb{R} !

Chapter 2

The Notion of a Measure

Consider a set X . We want to define a measure on some collection of subsets of X .

Definition 2.1. A *ring of subsets* of X is a nonempty collection \mathcal{R} with the properties:

- (i) If $A, B \in \mathcal{R}$, then $A \cup B \in \mathcal{R}$.
- (ii) If $A, B \in \mathcal{R}$, then $A - B := \{a \in A, a \notin B\} \in \mathcal{R}$.

From the definition, we can say more about the ring of subsets \mathcal{R} . We have:

- $\emptyset \in \mathcal{R}$ since $\emptyset = A - A \in \mathcal{R}$ for $A \in \mathcal{R}$.
- $A \cap B \in \mathcal{R}$ since $A \cap B = (A \cup B) - (A - B) - (B - A)$.
- Finite unions of elements of \mathcal{R} are again in \mathcal{R} .

Closely related to a ring of subsets is the notion of an *algebra of subsets*, which is a collection of sets closed under taking unions and complements. The only difference between a ring of subsets and an algebra of subsets is that X itself will always belong to an algebra (it is the complement of the empty set), but it may not necessarily belong to a ring of subsets.

Definition 2.2. A function $\mu : \mathcal{R} \rightarrow \mathbb{R}_{\geq 0}$ is *additive* if for $A, B \in \mathcal{R}$ such that $A \cap B = \emptyset$, we have

$$\mu(A \cup B) = \mu(A) + \mu(B).$$

Notice that from this definition we can say several things about additive functions that follow immediately from the definition:

- $\mu(\emptyset) = 0$.
- If $A \supseteq B$ and $A, B \in \mathcal{R}$, then $\mu(A) \geq \mu(B)$.
- For $A, B \in \mathcal{R}$, $\mu(A) + \mu(B) = \mu(A \cup B) - \mu(A \cap B)$.

- For $A_1, \dots, A_n \in \mathcal{R}$, we have

$$\begin{aligned}\mu(A_1 \cup \dots \cup A_n) &= (\mu(A_1) + \dots + \mu(A_n)) - (\mu(A_1 \cap A_2) + \dots + \mu(A_{n-1} \cap A_n)) \\ &\quad + (\text{sums of measures of intersections of three sets}) - \dots \\ &\leq \mu(A_1) + \dots + \mu(A_n).\end{aligned}$$

Example 2.3. Here is an easy example of a measure. Let X be a finite set and let \mathcal{R} be the collection of all subsets of X . We can let the measure be such that $\mu(S) = |S|$.

Example 2.4. Let $X = \mathbb{R}$ or \mathbb{R}^n . Let \mathcal{R} contain all finite intervals, i.e. $(a, b), (a, b], [a, b), [a, b]$. So then we get all finite unions of intervals in \mathcal{R} . We denote this ring by \mathcal{R}_{Leb} . (Note that \mathcal{R}_{Leb} is *not* an algebra!) We define a similar ring when $X = \mathbb{R}^n$ and denote it the same way. (That is, we take multi-intervals $[a_1, b_1] \times [a_2, b_2] \times \dots \times [a_n, b_n]$ and take finite unions of these.) For \mathbb{R} , define

$$\mu((a, b)) = b - a,$$

and for \mathbb{R}^n , define

$$\mu((a_1, b_1) \times \dots \times (a_n, b_n)) = \prod_{j=1}^n (b_j - a_j).$$

Note. For a moment, we return to a general X and a general μ . We say μ is *countably additive* if for A_1, A_2, \dots is a countable collection of mutually disjoint elements of \mathcal{R} with $A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$, we have

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

(Note that it may not always be that $\bigcup_{n=1}^{\infty} A_n \in \mathcal{R}$, and so it is important to choose the A_i to satisfy this.) If the above holds, then μ is a measure for \mathcal{R} .

Now we return to the example. We want to show that the μ we defined is indeed a measure. The note tells us that we need to show that it is countably additive.

Proposition 2.5. μ on \mathcal{R}_{Leb} is countably additive.

Proof. Consider $A \in \mathcal{R}_{\text{Leb}}$ and write $A = \bigcup_{n=1}^{\infty} A_n$ for $A_n \in \mathcal{R}_{\text{Leb}}$ with A_n disjoint. We want to show that $\mu(A) = \sum_{n=1}^{\infty} \mu(A_n)$.

Since $A \supseteq (A_1 \cup \dots \cup A_N)$, then $\mu(A) \geq \sum_{n=1}^N \mu(A_n)$, and so taking the limit, we get $\mu(A) \geq \sum_{n=1}^{\infty} \mu(A_n)$. The reverse inequality takes a bit more work. Notice that we can find a closed set $F \subseteq A$ such that $\mu(F) \geq \mu(A) - \varepsilon$ and an open subset $G \supseteq A$ such that $\mu(G) \leq \mu(A) + \varepsilon$. For each A_n , choose an open set $G_n \supseteq A_n$ such that $\mu(G_n) \leq \mu(A_n) + \frac{\varepsilon}{2^n}$. Now, $\bigcup G_n$ is an open cover

for G . But F is closed and bounded and hence compact, and so we have a finite subcover. So

$$\mu(F) \leq \mu(\text{finite subcover}) \leq \sum_{n=1}^{\infty} \mu(G_n) \leq \mu(A_n) + \frac{\varepsilon}{2^n} \leq \sum_{n=1}^{\infty} \mu(A_n) + \varepsilon.$$

But this holds for all ε , so we have $\mu(A) \leq \sum_{n=1}^{\infty} \mu(A_n)$. Since we proved first that we had the reverse inequality, equality must hold and we have shown that μ is countably additive. \square

Via the proposition, we have shown that μ is a measure. This completes our work in this example.

Chapter 3

The Outer Measure

In general, given a set X , a ring of subsets \mathcal{R} , and a measure μ on \mathcal{R} , we want to see how far we can extend the definition of μ to a larger class of sets. For example, in our earlier discussion, $\mathbb{Q} \notin \mathcal{R}_{\text{Leb}}$, but we would like to have some way of saying that $\mu(\mathbb{Q}) = 0$.

One way that we can do this is via the notion of an *outer measure* associated to μ . In this way, we can extend the measure to all subsets, but the problem is we lose countable additivity in the process. We define this outer measure, denoted μ^* in the following way: Take any subset $A \subseteq X$ and cover A by a countable union of sets in \mathcal{R} . Look at $\sum_{n=1}^{\infty} \mu(R_n)$ and take the infimum over all coverings. Explicitly, we have

$$\mu^*(A) = \inf_{\substack{\cup R_n \supseteq A \\ R_n \in \mathcal{R}}} \left(\sum_{n=1}^{\infty} \mu(R_n) \right).$$

If no such covering exists or if the sum does not converge, set $\mu^*(A) = \infty$.

Example 3.1. We have $\mu^*(\mathbb{Q} \cap [0, 1]) = 0$: We first enumerate the rationals r_1, r_2, \dots . Then we consider the union of the open intervals $(r_1 - \varepsilon, r_1 + \varepsilon), (r_2 - \frac{\varepsilon}{2}, r_2 + \frac{\varepsilon}{2}), \dots, (r_n - \frac{\varepsilon}{2^n}, r_n + \frac{\varepsilon}{2^n}), \dots$. Actually, even simpler, we could just consider the union of the measure-zero sets $\{r_n\}$ for each n .

Example 3.2. There are also uncountable sets that have measure zero. Let A be the Cantor set, for instance. The way we construct the Cantor set is by starting with the interval $[0, 1]$ and successively removing the middle third of each interval “piece.” Hence A is covered by $[0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$ and also by $[0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$, and so forth. Hence we can say that for each $n \in \mathbb{N}$,

$$\mu^*(A) \leq \left(\frac{2}{3} \right)^n,$$

and taking the limit to infinity, we get that $\mu^*(A) = 0$.

Similarly, this tells us that the outer measure of rolling a dice infinitely often and never getting 1 is 0. This solves the problems we started with in the first chapter.

We can say several things about the outer measure and we collect these thoughts into the following theorem.

Theorem 3.3. *Let X be a set, \mathcal{R} be a ring of subsets, and μ be a measure on \mathcal{R} .*

(i) *If $A \subseteq B \subseteq X$, then $\mu^*(A) \leq \mu^*(B)$.*

(ii) *For any countable collection of subsets of X , call them A_1, \dots, A_n, \dots , we have*

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

(iii) *If $A \in \mathcal{R}$, then $\mu^*(A) = \mu(A)$.*

Proof. (i) is true since cover of B is a cover of A .

To prove (ii), first notice that for each n , we can pick $A_{n,k}$ such that

$$A_n \subseteq \bigcup_{k=1}^{\infty} A_{n,k} \quad \text{and} \quad \sum_{k=1}^{\infty} \mu(A_{n,k}) \leq \mu^*(A_n) + \frac{\varepsilon}{2^n}.$$

Then since we have

$$A := \bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{n=1}^{\infty} \bigcup_{k=1}^{\infty} A_{n,k},$$

it follows that

$$\mu^*(A) \leq \sum_{n=1}^{\infty} \sum_{k=1}^{\infty} \mu(A_{n,k}) \leq \sum_{n=1}^{\infty} \mu^*(A_n) + \frac{\varepsilon}{2^n} = \sum_{n=1}^{\infty} \mu^*(A_n) + \varepsilon.$$

Finally for (iii), consider $A \in \mathcal{R}$. By covering A by itself, we get $\mu^*(A) \leq \mu(A)$. For the reverse inclusion, take a countable collection of elements of \mathcal{R} , call them A_n , such that

$$A \subseteq \bigcup_{n=1}^{\infty} A_n \quad \text{and} \quad \sum_{n=1}^{\infty} \mu(A_n) \leq \mu^*(A) + \varepsilon.$$

Out of every union of sets, we can extract a disjoint union in the following way:

- For A_1 , define $A'_1 = A_1$.
- For A_2 , define $A'_2 = A_2 - A_1$.
- For A_3 , define $A'_3 = A_3 - (A_1 \cup A_2)$.

For each n , define $A''_n = A'_n \cap A$. Then $A''_n \subseteq A'_n$ so $\mu(A''_n) \leq \mu(A'_n)$. Therefore we have

$$\mu(A) = \sum_{n=1}^{\infty} \mu(A''_n) \leq \sum_{n=1}^{\infty} \mu(A'_n) \leq \mu^*(A) + \varepsilon.$$

This completes the proof. \square

Let us now look at the case when $X = \mathbb{R}$, $\mathcal{R} = \mathcal{R}_{\text{Leb}}$, and μ is the Lebesgue measure. We know from the theorem that $\mu = \mu^*$ on \mathcal{R}_{Leb} and we know also that μ^* is countably subadditive. We also additionally know that $\mu^*((a, b)) = b - a$ and $\mu^*(S) = \mu^*(S + t)$. Here is a question: Can μ^* be countably additive? If it were, then μ^* would be a measure, and we would have found a measure on all subsets of X . However, we showed at the end of Chapter 1 that this was impossible, and so μ^* must not be countably additive. What is *really* surprising, though, is that in fact μ^* isn't even finitely additive! Sets that illustrate this are really weird, so we will not give an example, but the following proposition will give us that μ^* cannot be finitely additive.

Lemma 3.4. *If μ is finitely additive and countably subadditive, then μ is countably additive.*

Proof. Let A_1, \dots, A_n, \dots be a countable number of disjoint sets. By countable subadditivity, we know already that

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

On the other hand, we have

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

where the last equality holds by finite additivity. Taking the limit, we get

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

and so equality must hold. \square

We want to find a nice class of subsets where the outer measure will be countably additive. In the case of the Lebesgue measure, we are trying to enlarge \mathcal{R}_{Leb} to some ring where the outer measure will become a measure. We begin by first defining a notion of the “difference” between two sets.

Definition 3.5. For any subsets $A, B \subseteq X$, let $d(A, B) := \mu^*(S(A, B))$ where $S(A, B) = (A - B) \cup (B - A)$, the symmetric difference of A and B . We have several properties that are easy to check:

- $d(A, A) = 0$. (Note however that $d(A, B) = 0$ does not mean $A = B$.)

- $d(A, B) = d(B, A)$.
- $d(A, B) + d(B, C) \geq d(A, C)$ (triangle inequality).

Lemma 3.6. $|\mu^*(A) - \mu^*(B)| \leq d(A, B)$.

Proof. By the triangle inequality, we have

$$\mu^*(A) = d(A, \emptyset) \leq d(A, B) + d(B, \emptyset) = d(A, B) + \mu^*(B).$$

Reversing the roles of A and B , we get

$$\mu^*(B) \leq d(A, B) + \mu^*(A),$$

which gives the two inequalities

$$\mu^*(A) - \mu^*(B) \leq d(A, B), \quad \mu^*(B) - \mu^*(A) \leq d(A, B),$$

which means that

$$|\mu^*(A) - \mu^*(B)| \leq d(A, B).$$

□

Now look at sequences of sets in \mathcal{R} , call them $\{A_n\}$. We will say that $A_n \rightarrow A$ if $d(A, A_n) \rightarrow 0$ as $n \rightarrow \infty$. Let \mathcal{M}_F be the collection of sets A obtained as “limits” of sets in \mathcal{R} . (The subscript denotes that we are taking limits of sequences of sets of finite measure.)

Theorem 3.7. (i) \mathcal{M}_F is a ring.

(ii) If $A \in \mathcal{M}_F$ then $\mu^*(A) < \infty$.

(iii) μ^* is a measure on \mathcal{M}_F .

Example 3.8. Take the case when $X = \mathbb{R}$ and $\mathcal{R} = \mathcal{R}_{\text{Leb}}$. Then every set of outer measure 0 is in \mathcal{M}_F since $d(A, \emptyset) = \mu^*(A) = 0$. We see just from this that \mathcal{M}_F is HUGE! Indeed we proved already that the Cantor set has measure 0, and so the Cantor set (cardinality \mathbb{R}) and all its subsets (of which there are $2^{\mathbb{R}}$) are in \mathcal{M}_F ! Note that we also get, as the contrapositive, that all nonmeasurable sets have strictly positive outer measure.

Proof of Theorem 3.7. We first prove (ii). By Lemma 3.6, we know $|\mu^*(A) - \mu^*(B)| \leq d(A, B)$ so if $A_n \rightarrow A$, then

$$\mu^*(A) \leq \mu^*(A_n) + d(A, A_n) < \infty,$$

since $A_n \in \mathcal{R}$ and therefore must have finite measure.

To prove (i), we consider $A, B \in \mathcal{M}_F$. We want to show that $A \cup B, A - B \in \mathcal{M}_F$. By definition, there exists a sequence $\{A_n\}_{n=1,2,\dots}$ and a sequence $\{B_n\}_{n=1,2,\dots}$ such that $A_n \rightarrow A$ and $B_n \rightarrow B$. Since

$$((A \cup B) - (A_n \cup B_n)) \cup ((A_n \cup B_n) - (A \cup B)) \subseteq S(A, A_n) \cup S(B, B_n),$$

we can take their measures and get the inequality

$$\mu^*((A \cup B) - (A_n \cup B_n)) \cup ((A_n \cup B_n) - (A \cup B)) \leq \mu^*(S(A, A_n)) + \mu^*(S(B, B_n)),$$

This proves $A_n \cup B_n \rightarrow A \cup B$. Similarly, we can show that $A_n \cap B_n \rightarrow A \cap B$. Combining these two, we get $A_n - B_n \rightarrow A - B$. Therefore we have shown that $A \cup B$ and $A - B$ can both be approximated by sequences of elements of \mathcal{R} . Therefore $A \cup B, A - B \in \mathcal{M}_F$.

Finally, to complete the proof of this theorem, we need to show that μ^* is a measure. It is enough to show that μ^* is finitely additive (because of Lemma 3.4). We want to show: If $A, B \in \mathcal{M}_F$, then $\mu^*(A \cup B) + \mu^*(A \cap B) = \mu^*(A) + \mu^*(B)$. (Then if $A \cap B = \emptyset$, then $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$, as needed.) We can approximate A and B by sequences $\{A_n\}_{n=1,2,\dots}$ and $\{B_n\}_{n=1,2,\dots}$, respectively, and since these sequences are in \mathcal{R} , then we have

$$\mu^*(A_n \cup B_n) + \mu^*(A_n \cap B_n) = \mu^*(A_n) + \mu^*(B_n).$$

Taking the limit as $n \rightarrow \infty$, we get

$$\mu^*(A \cup B) + \mu^*(A \cap B) = \mu^*(A) + \mu^*(B). \quad \square$$

Definition 3.9. We say that A is *measurable* if $A = \bigcup_{n=1}^{\infty} A_n$ with $A_n \in \mathcal{M}_F$. We denote the class of measurable sets by \mathcal{M} .

Example 3.10. Let $X = \mathbb{R}$, $\mathcal{R} = \mathcal{R}_{\text{Leb}}$. Then $\mathbb{R} \in \mathcal{M}$ since $\mathbb{R} = \bigcup_{n=1}^{\infty} (-n, n)$.

Theorem 3.11. If $A \in \mathcal{M}$, then $\mu^*(A) < \infty$ if and only if $A \in \mathcal{M}_F$.

Proof. By definition, (\Leftarrow) is obvious. For the converse, let $A = \bigcup_{n=1}^{\infty} A_n$ where $A_n \in \mathcal{M}_F$. Without loss of generality we can assume that this is a disjoint union. By countable subadditivity, $\mu^*(A) \leq \sum_{n=1}^{\infty} \mu^*(A_n)$. On the other hand,

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

and so in fact we have $\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n)$. We can take a sequence of sets $\{\bigcup_{n=1}^N A_n\}_{N=1,2,\dots} \in \mathcal{M}_F$ with $d(A, \bigcup_{n=1}^N A_n) \rightarrow 0$. Then we can find sets $B_N \in \mathcal{R}$ that are very close to $\bigcup_{n=1}^N A_n$. In this way, we have approximated A by a sequence of elements of \mathcal{R} , which means that $A \in \mathcal{M}_F$. This completes the proof. \square

Chapter 4

Borel Measurable Sets

Definition 4.1. A σ -ring is a ring which is closed under countable unions.

Example 4.2. For \mathbb{R} , \mathcal{M}_F is *not* a σ -ring as it is only closed under finite unions.

Definition 4.3. A ring \mathcal{R} is called a *field* (or equivalently, an *algebra*) if $X \in \mathcal{R}$. A σ -ring \mathcal{R} is called a σ -field (or a σ -algebra) if $X \in \mathcal{R}$.

Theorem 4.4. \mathcal{M} is a σ -ring.

Proof. If $A_1, \dots, A_n, \dots \in \mathcal{M}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{M}$ since each A_n is a countable union of elements of \mathcal{M}_F and hence the union must also be a countable union of elements of \mathcal{M}_F . Hence if we can show that \mathcal{M} is a ring, then we have that it is a σ -ring.

Consider $A, B \in \mathcal{M}$. It is clear that $A \cup B \in \mathcal{M}$. We want to show that $A - B \in \mathcal{M}$. Let us first consider the case when $A \in \mathcal{M}_F$ (so $\mu^*(A) < \infty$). We can write $B = \cup_{n=1}^{\infty} B_n$ for $B_n \in \mathcal{M}_F$. Then

$$A - \cup_{n=1}^{\infty} B_n = A - (A \cap (\cup_{n=1}^{\infty} B_n)) = A - (\cup_{n=1}^{\infty} (A \cap B_n)).$$

We know that $\cup_{n=1}^{\infty} (A \cap B_n) \in \mathcal{M}$, and since this union is contained in A , which has finite outer measure, then $\cup_{n=1}^{\infty} (A \cap B_n) \in \mathcal{M}_F$. Therefore $A - B \in \mathcal{M}_F$. To finish, take $A = \cup_{n=1}^{\infty} A_n$ and $B \in \mathcal{M}$. Then

$$A - B = \cup_{n=1}^{\infty} (A_n - B),$$

and since $A_n - B \in \mathcal{M}_F$, then we have written $A - B$ as a countable union of elements of \mathcal{M}_F and so $A - B \in \mathcal{M}$. Therefore \mathcal{M} is a ring and hence a σ -ring. \square

Theorem 4.5. μ^* is countably additive on \mathcal{M} (allowing μ^* to take the value ∞).

Proof. Let A_1, \dots, A_n, \dots be pairwise disjoint sets in \mathcal{M} . We want to show

$$\mu^*(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu^*(A_n).$$

If $\mu^*(A) < \infty$, then this is true since in this case, $A, A_n \in \mathcal{M}_F$ and we already know that μ^* is countably additive on \mathcal{M}_F . If $\mu^*(A) = \infty$, then the RHS must also be ∞ by countable subadditivity. \square

In the case that we have \mathbb{R}^n with \mathcal{B}_{Leb} and the measure μ being the volume of a multi-interval, then on \mathcal{M} , $\mu^* = \mu_L$, and these are called the *Lebesgue measurable sets*. Notice that \mathcal{M} is a σ -field. Furthermore, it is σ -finite.

Definition 4.6. X is called σ -finite if $X = \cup_{n=1}^{\infty} X_n$ with $X_n \in \mathcal{M}_F$.

Theorem 4.7. All open and closed subsets of \mathbb{R}^n are Lebesgue measurable.

Proof. It is enough to show that open sets are measurable. Look at all the multi-intervals with rational endpoints. Call the collection of all these sets \mathcal{J} . There are countably many elements of \mathcal{J} . If U is an open subset of \mathbb{R}^n , then consider the set

$$U' := \bigcup_{\substack{I \in \mathcal{J} \\ I \subseteq U}} I.$$

Clearly the $U' \subseteq U$. Since U is open, then for any $x \in U$, there exists an $I \in \mathcal{J}$ such that $x \in I \subseteq U$, which means that $U \subseteq U'$. Hence $U = U'$ and we have proven that we can write any open subset of \mathbb{R}^n as a countable union of open multi-intervals. \square

Definition 4.8. Take a σ -field generated by all the open and closed subsets of \mathbb{R}^n . The elements of this σ -field are called the *Borel measurable sets* of \mathbb{R}^n .

Remark 4.9. There are many more Lebesgue measurable sets than Borel measurable sets; there are $2^{\mathbb{R}}$ Lebesgue measurable sets and \mathbb{R} Borel measurable sets.

Example 4.10. Is the Cantor set measurable? For every n , we have a collection C_n consisting of 2^n intervals each of length $\frac{1}{3^n}$. Then the Cantor set C is the intersection $C = \cap_{n=1}^{\infty} C_n$. This is Borel. Note, however, that there are lots of subsets of C that are not Borel!

Theorem 4.11. Every Lebesgue measurable set A can be written as $B \cup E$ where B is a Borel set and E is a set of measure 0.

Proof. We first prove that for any $\varepsilon > 0$, we can find a Borel set G with $G \supseteq A$ and $\mu^*(G - A) < \varepsilon$.

Because we can take complements, we can apply the above to A^c . That is, we can find a Borel set F with $F \subseteq A$ and $\mu^*(A - F) < \varepsilon$. So for every n , we can find a Borel set $F_n \subseteq A$ with $\mu^*(A - F_n) < \frac{1}{n}$. Then $\cup_{n=1}^{\infty} F_n$ is Borel and

$$\mu(A - \cup_{n=1}^{\infty} F_n) \leq \mu^*(A - F_n) < \frac{1}{n}.$$

Hence taking $B := \cup_{n=1}^{\infty} F_n$ and $E = A - B$, we have shown that A can be written as $B \cup E$ for a Borel set B and a measure 0 set E . \square

Now, it is one thing to understand proofs when they are presented, but what is also important is to be able to detect wrong proofs. Here is a “theorem.”

Fake Theorem 4.12. *If A is a set of finite outer measure, then for any $\varepsilon > 0$ there is a Borel set $G \supseteq A$ such that $\mu^*(G - A) < \varepsilon$.*

Proof. We can find $I_n \in \mathcal{R}_{\text{Leb}}$ such that $A \subseteq \bigcup_{n=1}^{\infty} I_n$ and $\sum_{n=1}^{\infty} \mu(I_n) < \mu^*(A) + \varepsilon$. (We may assume that I_n is a multi-interval.) But then $\bigcup_{n=1}^{\infty} I_n =: G$ is Borel, so $\mu^*(G - A) < \varepsilon$. \square

Where does this proof break down? Well, how can we conclude that $\mu^*(G - A) < \varepsilon$? We needed $\mu^*(G) - \mu^*(A) = \mu^*(G - A)$, but we don't necessarily have additivity, so we cannot actually conclude this. However, we can amend the hypotheses so that the above proof works.

Theorem 4.13. *If A is a measurable set, then for every $\varepsilon > 0$ there is a Borel set $G \supseteq A$ such that $\mu^*(G - A) < \varepsilon$. Applying this result for A^c , we can find a Borel set F with $A \supseteq F$ and $\mu^*(A - F) < \varepsilon$.*

Proof. We have already proven the above for when A has finite outer measure. For the case when A has infinite outer measure, we can write $A = \bigcup_{n=1}^{\infty} A_n$ where $A_n \in \mathcal{M}_F$. We can cover each A_n by some Borel set G_n such that $\mu^*(G_n - A_n) \leq \frac{\varepsilon}{2^n}$. Then $G := \bigcup_{n=1}^{\infty} G_n$ is Borel and $\mu^*(G - A) < \varepsilon$. This completes the proof. \square

Chapter 5

The Probability Space

Definition 5.1. A *probability space* is a triple (X, \mathcal{F}, μ) where X is a set, \mathcal{F} is a σ -field, and μ is a measure. We normalize this space so that $\mu(X) = 1$.

Example 5.2. (i) Let X be a finite set and let \mathcal{F} be the power set of X so $|\mathcal{F}| = 2^{|X|}$. Let μ be the counting measure, and since we want a normalized measure, we have $\mu(A) = |A|/|X|$ for any $A \in \mathcal{F}$.

(ii) Let $X = [0, 1]$ and let $\mathcal{F} = \mathcal{M} \cap [0, 1]$. We take μ to be the Lebesgue measure.

We now expand upon Example 5.2(ii). We can associate a Bernoulli sequence of coin tosses to a point in $[0, 1]$. We have two notions:

- The Strong Law of Large Numbers. If we toss a coin infinitely often, with probability 1, we will get H half the time and T half the time.
- The Weak Law of Large Numbers. Given $\varepsilon > 0$, if N is sufficiently large, then with probability very close to 1, after N coin tosses, $|(\text{number of heads}) - (\text{number of tails})| \leq \varepsilon N$.

We want to find some way to make sense of what these two laws mean in a rigorous way. The weak law is more accessible, so we work on that first.

Proof of Weak Law. The probability of tossing k heads and $N - k$ tails satisfying the desired inequality is $\frac{1}{2^N} \sum_{|2k - N| < 2\varepsilon N} \binom{N}{k}$. For every $\varepsilon > 0$, if N is sufficiently large, then this number tends to 1. (Note that we haven't yet proven that the claimed probability is the actual probability.) \square

Now we move onto the Strong Law, which takes significantly more work to make sense of.

Discussion of Strong Law. Take an $\omega \in [0, 1]$. Write $\omega = 0.b_1b_2\cdots$ where b_i is a binary digit. (So we can, for instance, associate H to 1 and T to 0.) Define

$$S_N(\omega) = \sum_{j=1}^N \left\{ \begin{array}{ll} 1 & \text{if } b_j = 1 \\ -1 & \text{if } b_j = 0 \end{array} \right\}.$$

We want to study $\frac{1}{N}S_N(\omega)$ and $\lim_{N \rightarrow \infty} \frac{1}{N}S_N(\omega)$. We want the limit to exist and to $= 0$.

But the problem is that there will definitely be cases when this limit does not exist. So to get around that problem by considering the set

$$\{\omega \in [0, 1] : \lim_{N \rightarrow \infty} \frac{1}{N}S_N(\omega) = 0\} =: \mathcal{G}.$$

Instead of demanding the limit to always exist and $= 0$, we can consider the set where this does happen and hope that this set (namely \mathcal{G}) has measure 1. It turns out that \mathcal{G} is Borel. We can obtain this information by looking at the complement. Define $\mathcal{B} =: \mathcal{G}^c$. We want to show that \mathcal{B} is Borel and $\mu(\mathcal{B}) = 0$.

We know that \limsup gives an upper bound for the limit and \liminf gives a lower bound. Hence if we are trying to consider an $\omega \in \mathcal{B}$, then $|\lim_{N \rightarrow \infty} \frac{S_N(\omega)}{N}| > \varepsilon$, so it is sufficient to ask that $|\limsup_{N \rightarrow \infty} \frac{S_N(\omega)}{N}| > \varepsilon$. So for every $\varepsilon > 0$, we have $\mathcal{B}_\varepsilon = \mathcal{B}_\varepsilon^+ \cup \mathcal{B}_\varepsilon^-$ where

$$\begin{aligned} \mathcal{B}_\varepsilon^+ &= \{\omega \in [0, 1] : \limsup_{N \rightarrow \infty} \frac{S_N(\omega)}{N} > \varepsilon\} \\ \mathcal{B}_\varepsilon^- &= \{\omega \in [0, 1] : \liminf_{N \rightarrow \infty} \frac{S_N(\omega)}{N} < -\varepsilon\}. \end{aligned}$$

Then

$$\mathcal{B} = \bigcup_{n=1}^{\infty} (\mathcal{B}_{1/n}^+ \cup \mathcal{B}_{1/n}^-).$$

We want to analyze $\mathcal{B}_{1/n}^+$. If $\limsup_{N \rightarrow \infty} \frac{S_N(\omega)}{N} > \frac{1}{n}$, then this means that $\frac{S_N(\omega)}{N} > \varepsilon$ for infinitely many N . Hence we have

$$\begin{aligned} \mathcal{B}_{1/n}^+ &= \{\omega \in [0, 1] : \text{infinitely many } k \text{ with } \frac{S_k(\omega)}{k} > \frac{1}{n}\} \\ &= \bigcap_{l=1}^{\infty} \left(\bigcup_{k=1}^{\infty} \{\omega \in [0, 1] : \frac{S_k(\omega)}{k} > \frac{1}{n}\} \right). \end{aligned}$$

But then this is an expression of $\mathcal{B}_{1/n}^+$ as a countable intersection of a countable union of a finite union of intervals, which is Borel. Therefore \mathcal{B} is a Borel set. It remains to be shown that $\mathcal{B}_{1/n}^+$ has measure zero. \square

The above argument is the content of the proof of the following more general fact.

Lemma 5.3 (Borel-Cantelli 1). *Let (X, \mathcal{F}, μ) be a probability space. For each $n \in \mathbb{N}$, we have “events” E_n , where E_n is a measurable set. Assume that $\sum \mu(E_n) < \infty$. Then $E = \{x : x \in E_n \text{ for infinitely many } n\}$ has measure zero.*

Proof. We have

$$E = \bigcap_{l=1}^{\infty} \left(\bigcup_{k=l}^{\infty} E_k \right),$$

and hence for any l ,

$$\mu(E) \leq \mu(\bigcup_{k=l}^{\infty} E_k) \leq \sum_{k=l}^{\infty} \mu(E_k),$$

which tends to 0 as $l \rightarrow \infty$. This completes the proof. \square

Discussion of Strong Law (continued). We now want to complete our discussion of the strong law of large number by showing that $\mathcal{B}_{1/n}^+$ has measure zero. Now consider the measure of the set $\{\omega \in [0, 1] : \frac{S_N(\omega)}{N} \geq \frac{1}{n}\}$, which we will call $\mu_n(N)$. If $\sum_{N=1}^{\infty} \mu_n(N) < \infty$, then we can apply Borel-Cantelli (Lemma 5.3). We have

$$\mu_n(N) = \frac{1}{2^N} \sum_{2k-N \geq \frac{N}{n}} \binom{N}{k} = \frac{1}{2^N} \sum_{k \geq \frac{N}{2}(1+\frac{1}{n})} \binom{N}{k}.$$

We want this to be very small for large N . (Note that to show that $\mathcal{B}_{1/n}^-$ has measure zero, we analyze $\frac{1}{2^N} \sum_{k \leq \frac{1}{2}N(1-\frac{1}{n})} \binom{N}{k}$, which, by the symmetry of Pascal’s triangle, is the same sum as the rightmost expression.) We have Stirling’s formula.

Stirling’s Formula. If N is large, then $N! \approx \sqrt{2\pi N} \left(\frac{N}{e}\right)^N$. So we can prove that if $k = \lambda N$, then

$$\binom{N}{k} \rightsquigarrow \frac{1}{\sqrt{N}} \left(\frac{1}{\lambda^\lambda (1-\lambda)^{1-\lambda}} \right)^N \quad (5.1)$$

Alternatively, for $x \geq 1$, we have

$$\frac{1}{2^N} \sum_{k \geq \frac{N}{2}\lambda} \binom{N}{k} \leq \frac{1}{2^N} \sum_{k \geq \frac{N}{2}\lambda} \binom{N}{k} \frac{x^k}{x^{N\lambda/2}} \leq \frac{1}{2^N} \sum_{k=0}^N \binom{N}{k} \frac{x^k}{x^{N\lambda/2}} = \left(\frac{1+x}{2x^{\lambda/2}} \right)^N.$$

If $\lambda = 1$, then $\frac{1+x}{2\sqrt{x}} \geq 1$. If $\lambda > 1$, we can use calculus to minimize $\frac{1+x}{2x^{\lambda/2}}$. We should get Equation 5.1. Intuitively, we have $x = 1 + \varepsilon$, and the numerator is $2 + \varepsilon$ and the denominator is $2(1 + \lambda\varepsilon/2) = 2 + \lambda\varepsilon$, so $\frac{1+x}{2x^{\lambda/2}} < 1$. This means

$$\left(\frac{1+x}{2x^{\lambda/2}} \right)^N \leq c(\lambda)^N$$

for some constant $c(\lambda) < 1$, choosing x appropriately.

We wanted $\sum_{N=1}^{\infty} \mu_n(N) < \infty$. From the above discussion, we have

$$\sum_{N=1}^{\infty} \mu_n(N) \leq \sum_{N=1}^{\infty} c(\lambda)^N < \infty.$$

Therefore by Borel-Cantelli, $\mathcal{B}_{1/n}^+$ (and $\mathcal{B}_{1/n}^-$) has measure zero, and this concludes our discussion of the strong law. \square

Example 5.4. Random walk on \mathbb{Z}^n : Start at the origin. At each step, move ± 1 in any of the n directions with equal probability. We have a theorem of Polya.

Theorem 5.5 (Polya). *With probability 1,*

- (a) *if $n \leq 2$, you will return to the origin infinitely often.*
- (b) *if $n \geq 3$, you will not return to the origin infinitely often.*

To make sense of what this means, we can, as usual, associate a subset of $[0, 1]$ to the probability of these events. For example, for $n = 3$, we can work in base 6, using each value as a different direction (up, down, north, south, east, west).

Definition 5.6. If E_1, E_2 are two measurable sets in a probability space (i.e. events), then E_1, E_2 are said to be independent if

$$\mu(E_1 \cap E_2) = \mu(E_1)\mu(E_2).$$

We can extend this definition for more sets: If $E_1, E_2, \dots, E_n, \dots$ are events, then they are independent if for any finite set $S \subseteq \mathbb{N}$,

$$\mu(\cap_{s \in S} E_s) = \prod_{s \in S} \mu(E_s).$$

Lemma 5.7 (Borel-Cantelli 2). *If E_1, E_2, \dots are independent and $\sum \mu(E_n)$ diverges, then $E := \{x : x \in E_n \text{ for infinitely many } n\}$ has measure 1.*

Proof. If E_1 and E_2 are independent, then E_1^c and E_2^c are also independent:

$$\begin{aligned} \mu(E_1^c)\mu(E_2^c) &= (1 - \mu(E_1))(1 - \mu(E_2)) = 1 - \mu(E_1) - \mu(E_2) + \mu(E_1)\mu(E_2) \\ &= 1 - (\mu(E_1) + \mu(E_2) - \mu(E_1 \cap E_2)) = 1 - \mu(E_1 \cup E_2) \\ &= \mu((E_1 \cup E_2)^c) = \mu(E_1^c \cap E_2^c). \end{aligned}$$

It is left as an exercise to prove that this is true for any finite collection. So we have that E_n^c are independent. We have

$$E = \cap_{k=1}^{\infty} (\cup_{n=k}^{\infty} E_n), \quad E^c = \cup_{k=1}^{\infty} (\cap_{n=k}^{\infty} E_n^c).$$

It is sufficient to show that $\cap_{n=k}^{\infty} E_n^c$ has measure zero. Indeed we have

$$\begin{aligned} \mu(\cap_{n=k}^{\infty} E_n^c) &\leq \mu(\cap_{n=k}^l E_n^c) = \prod_{n=k}^l \mu(E_n^c) = \prod_{n=k}^l (1 - \mu(E_n)) \\ &< \prod_{n=k}^l \exp(-\mu(E_n)) = \exp(-\sum_{n=k}^l \mu(E_n)) < \varepsilon, \end{aligned}$$

for large enough l . This completes the proof. \square

Chapter 6

The Lebesgue Integral

We would now like to work towards defining some notion of integration in a measure space (X, \mathcal{F}, μ) . Given a function $f : X \rightarrow \mathbb{R}$ and a set $E \subseteq X$, we want to have some notion of $\int_E f d\mu$ that satisfy some properties we would expect:

- For some constant c , $\int_E cf d\mu = c \int_E f d\mu$.
- $\int_E f_1 + f_2 d\mu = \int_E f_1 d\mu + \int_E f_2 d\mu$.
- For disjoint sets E_1, E_2 , $\int_{E_1 \cup E_2} f d\mu = \int_{E_1} f d\mu + \int_{E_2} f d\mu$.

We would like this to hold for all measurable sets E and all measurable functions f . In the case that $X = \mathbb{R}$, then this generalizes the Riemann integral. The first thing we must do is discuss the notion of a measurable function. We consider the measure space (X, \mathcal{F}, μ) , where \mathcal{F} is a σ -field, and $f : X \rightarrow \mathbb{R} \cup \{\infty, -\infty\}$. (We set $a \pm \infty = \pm\infty$, $b \cdot \infty = \infty$ for $b > 0$, $b \cdot \infty = -\infty$ for $b < 0$. We will not worry about $0 \cdot \infty$ or $\infty - \infty$.)

Theorem 6.1. *The following conditions are equivalent:*

- (i) *For all $a \in \mathbb{R}$, $\{x : f(x) \in (a, \infty)\}$ is measurable.*
- (ii) *For all $a \in \mathbb{R}$, $\{x : f(x) \in [a, \infty]\}$ is measurable.*
- (iii) *For all $a \in \mathbb{R}$, $\{x : f(x) < a\}$ is measurable.*
- (iv) *For all $a \in \mathbb{R}$, $\{x : f(x) \leq a\}$ is measurable.*
- (v) *For every Borel set B of \mathbb{R} , $\{x \in X : f(x) \in B\}$ is measurable.*

Definition 6.2. If any of the above conditions in Theorem 6.1 holds, then f is called a *measurable function*.

Proof. It is clear that $(i) \Leftrightarrow (iv)$ and $(ii) \Leftrightarrow (iii)$. To show that $(i) \Leftrightarrow (ii)$, notice that we can write

$$\begin{aligned}\{x : f(x) > a\} &= \bigcup_{n=1}^{\infty} \{x : f(x) \geq a + \frac{1}{n}\}, \\ \{x : f(x) \geq a\} &= \bigcap_{n=1}^{\infty} \{x : f(x) > a - \frac{1}{n}\}.\end{aligned}$$

Hence we have proven $(i) \Leftrightarrow (ii) \Leftrightarrow (iii) \Leftrightarrow (iv)$. It is clear that (v) implies $(i), \dots, (iv)$. We would like now to show $(i) \Rightarrow (v)$. Let \mathcal{C} be the collection of all sets $S \subseteq \mathbb{R}$ such that $\{x : f(x) \in S\}$ is measurable. (For example, if f is constant, then $\mathcal{C} = 2^{\mathbb{R}}$.) We want to show that \mathcal{C} contains all Borel sets. Now, we know that \mathcal{C} contains $(a, \infty], [a, \infty], [-\infty, a), [-\infty, a]$. But since functions preserve unions and intersections, if $A, B \in \mathcal{C}$, then $A \cup B \in \mathcal{C}$, and if $A_1, \dots, A_n, \dots \in \mathcal{C}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{C}$. Therefore \mathcal{C} contains all open sets, and since Borel sets are generated by open sets, then \mathcal{C} contains all Borel sets. This completes the proof. \square

Remark 6.3. It is worth pointing out that this definition seems a bit odd. It would seem more natural to define a measurable function as a function wherein the preimage of any measurable set is measurable. But the problem with setting this as the definition is that it is too restrictive and so we don't get a nice theory if we were to adopt this definition. We will show later in an example (Example 6.4(ix)) that weird things can happen if we take this to be the definition of a measurable function. More specifically, we will construct a continuous function wherein the preimage of a measurable set is not necessarily measurable.

Now we give a lot of examples of measurable functions.

Example 6.4. (i) Constant functions.

(ii) Consider $f : \mathbb{R}^n \rightarrow \mathbb{R}$. If f is continuous, then f is measurable.

(iii) Let $E \subseteq X$ and define the characteristic function of E :

$$\chi_E(x) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E. \end{cases}$$

Then χ_E is measurable if E is measurable.

(iv) Simple function s . We define this to be the measurable function which takes on finitely many real values. Let c_1, \dots, c_n be the distinct values of s . Then we can look at the set $E_j := \{x \in X : s(x) = c_j\}$. This set is measurable. Furthermore, the E_j are pairwise disjoint and they cover X .

(v) If f_1 and f_2 are measurable, then so are $\max(f_1, f_2)$ and $\min(f_1, f_2)$. This is true since we can write

$$\begin{aligned}\{x : \max(f_1, f_2)(x) > a\} &= \{x : f_1(x) > a\} \cup \{x : f_2(x) > a\}, \\ \{x : \min(f_1, f_2)(x) > a\} &= \{x : f_1(x) > a\} \cap \{x : f_2(x) > a\}.\end{aligned}$$

- (vi) If f is measurable, then $f^+ = \max(f, 0)$ and $f^- = \min(f, 0)$ are also measurable.
- (vii) If f_1 and f_2 are measurable, then $f_1 + f_2$ is measurable. This is true since we can write

$$\{x : f_1(x) + f_2(x) > a\} = \cup_{r \in \mathbb{Q}} (\{x : f_1(x) \geq r\} \cap \{x : f_2(x) > a - r\}).$$

In a similar way, we can also show that the (pointwise) product $f_1 f_2$ is measurable.

- (viii) Consider $f, g : \mathbb{R} \rightarrow \mathbb{R}$. When is the composition $f \circ g$ measurable?

Let us first look at the case when f is continuous and g is measurable. We want to know if $\{x : f(g(x)) > a\}$ is measurable. Since f is continuous and the inverse image of an open set is again open, we know that $\{x : f(x) > a\} =: B$ is Borel. Then $\{x : f(g(x)) > a\} = \{x : g(x) \in B\}$, which is measurable since g is measurable.

But from the above argument, we see that we need that intermediate step of Borel-ness. So in general, $f \circ g$ may not be measurable if f is only measurable (i.e. not continuous). In fact, even if g is continuous, the composition might not be measurable!

- (ix) Cantor's continuous function. We define a function $f : [0, 1] \rightarrow [0, 1]$. For $x \in [0, 1]$, write it in its ternary expansion: $x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$, $a_n = 0, 1, 2$. Take N to be the first time such that $a_N = 1$, and we will let $N = \infty$ if no such N exists (i.e. if x is in the Cantor set). Now define

$$f(x) = \sum_{n=1}^{N-1} \frac{a_n/2}{2^n} + \frac{1}{2^N}.$$

We observe some things about f . It is constant on any interval in the complement of the Cantor set. It is continuous (if x is very close to y , then many ternary terms will coincide). It is increasing. Now consider the function

$$g : [0, 1] \rightarrow [0, 2], \quad g(x) = x + f(x).$$

Then g is continuous and strictly increasing, and hence it is also bijective. So we have a strictly increasing, continuous inverse $g^{-1} : [0, 2] \rightarrow [0, 1]$. Now let C be the Cantor set. Then $[0, 1] - C$ is a countable union of intervals and $f([0, 1] - C)$ is a countable set (it takes on countably many values). Furthermore, $g([0, 1] - C)$ is a Borel set of measure 1 and therefore $g(C) = [0, 2] - g([0, 1] - C) =: A$ is also a Borel set of measure 1. Let us accept the following fact:

Fact. Every set of positive measure contains a nonmeasurable subset.

If $U \subseteq A$ and U is nonmeasurable, then there exists a $D \subseteq C$ with $g(D) = U$. We had a continuous function $g^{-1} : [0, 2] \rightarrow [0, 1]$. We also

have $D \subseteq C \subseteq [0, 1]$, a measurable set (of measure 0). The inverse image of D is $\{x : g^{-1}(x) \in D\} = \{g(x) : x \in D\} = U$. But this is a nonmeasurable set, so we've shown that even for a continuous function, we can construct something in which the inverse image of a measurable set is not measurable! Note that this also shows that we have sets that are Lebesgue measurable but not Borel measurable.

We can use this to find a measurable function $f : \mathbb{R} \rightarrow \mathbb{R}$ and a continuous function $g : \mathbb{R} \rightarrow \mathbb{R}$ such that $f \circ g$ is not measurable.

(x) If f_1, \dots, f_n, \dots are measurable functions, then

$$\begin{aligned}\bar{f}(x) &:= \sup_n f_n(x), \\ \underline{f}(x) &:= \inf_n f_n(x), \\ \limsup_n f_n(x) &= \inf_{k>1} \sup_{n \geq k} f_n(x), \\ \liminf_n f_n(x) &= \sup_{k>1} \inf_{n \geq k} f_n(x),\end{aligned}$$

are all measurable functions. If $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ exists, then this is measurable (since then $\limsup = \liminf = \lim$ if the limit exists).

Recall that our main goal for this section is to develop the notion of an integral $\int_E f d\mu$ for any measurable function f over any measurable set $E \subseteq X$. Recall also some desired properties that we would like:

- $\int_E f_1 + f_2 d\mu = \int_E f_1 d\mu + \int_E f_2 d\mu$.
- $\int_{E_1} f d\mu + \int_{E_2} f d\mu = \int_{E_1 \cup E_2} f d\mu$, for disjoint sets E_1 and E_2 .
- $\int_E 1 d\mu = \mu(E)$.
- $\int_E \chi_A d\mu = \mu(A \cap E)$.

Recall the notion of simple functions, as in Example 6.4(iv).

Definition 6.5. A function $s : X \rightarrow \mathbb{R}$ is simple if it takes a finite number of values. (Note that we do not allow s to evaluate to $\pm\infty$.) That is, there exist disjoint sets E_1, \dots, E_n such that $s(x) = c_j$ if $x \in E_j$.

If each $c_j \geq 0$, then $s \geq 0$, and from now on, we will consider only nonnegative simple functions.

Definition 6.6. For any nonnegative simple function s , we define

$$I_E(s) := \sum_{j=1}^n c_j \mu(E \cap E_j).$$

Following convention, we will take $0 \cdot \infty = 0$.

We have several properties of nonnegative simple functions s and $I_E(s)$ that come almost automatically:

- If s_1 and s_2 are nonnegative simple functions, then so are $s_1 s_2$ and $s_1 + s_2$.
- $I_E(s) \geq 0$.
- For $c \in \mathbb{R}$, $I_E(cs) = cI_E(s)$.
- $I_E(s_1 + s_2) = I_E(s_1) + I_E(s_2)$. (Write $X = E_1 \cup \dots \cup E_n = F_1 \cup \dots \cup F_m$ where $s_1(E_j) = c_j$ and $s_2(F_j) = d_j$. Then $s_1 + s_2$ takes value $c_i + d_j$ on $E_i \cap F_j$, and we have $X = \cup_{i,j} (E_i \cap F_j)$.)
- If $s_1 \leq s_2$, then $I_E(s_1) \leq I_E(s_2)$. (Since $s_2 = s_1 + (s_2 - s_1)$.)
- If $E \subseteq F$ are measurable sets, then $I_E(s) \leq I_F(s)$. (Since $\mu(E \cap E_j) \leq \mu(F \cap F_j)$ for all j .)
- If F_1 and F_2 are disjoint, then $I_{F_1}(s) + I_{F_2}(s) = I_{F_1 \cup F_2}(s)$. Since μ is countably additive, then this is also countably additive, and $\varphi_s(E) := I_E(s)$ is a measure on X .
- If E has measure zero, then $I_E(s) = 0$.

We will first deal with the theory of the integral of nonnegative functions, and then we will later extend this to all measurable functions via the identity $f = \max(f, 0) + \min(f, 0) = \max(f, 0) - \min(-f, 0)$.

Definition 6.7. Let $f : X \rightarrow \mathbb{R}$ be a nonnegative measurable function. Then we define

$$\int_E f d\mu = \sup_{0 \leq s \leq f} I_E(s).$$

Note that if f is a simple function, then $\int_E f d\mu = I_E(f)$.

We have some easily verifiable facts:

- If $f_1 \leq f_2$, then $\int_E f_1 d\mu \leq \int_E f_2 d\mu$. (This is true since if $s \leq f_1$, then $s \leq f_2$.)
- If E has measure zero, then $\int_E f d\mu = 0$.
- If $E \subseteq F$, then $\int_E f d\mu \leq \int_F f d\mu$. (For any $\varepsilon > 0$, we can choose a simple function s with $0 \leq s \leq f$ so that

$$\int_F f d\mu \geq \int_F s d\mu \geq \int_F s s \mu \geq \int_E f d\mu - \varepsilon.$$

Hence $\int_F f d\mu \geq \int_E f d\mu$.

Proposition 6.8 (Chebyshev's Inequality). *If $\int_E f d\mu < \infty$, then*

$$\mu(E_c) \leq \frac{1}{c} \int_E f d\mu, \text{ where } E_c = \{x \in E : f(x) \geq c\}.$$

Note that this also proves that if $\int_E f d\mu < \infty$, then $\mu\{x \in E : f(x) = \infty\} = 0$.

Proof.

$$c\mu(E_c) = \int_{E_c} c d\mu \leq \int_E f d\mu. \quad \square$$

Now we prove something more involved.

Proposition 6.9. *If $f \geq 0$ is measurable, then $\varphi(E) = \int_E f d\mu$ is a measure.*

Proof. We would like to prove that φ is countably additive. Let $E_1, E_2, \dots, E_n, \dots$ be disjoint sets and set $E = \cup_{n=1}^{\infty} E_n$. We want to show $\varphi(E) = \sum_{n=1}^{\infty} \varphi(E_n)$. We begin by showing that φ is finitely additive, and furthermore, note that it is enough to show that for disjoint sets E_1 and E_2 , we have $\int_{E_1 \cup E_2} f d\mu = \int_{E_1} f d\mu + \int_{E_2} f d\mu$.

For any simple function $0 \leq s \leq f$, we have

$$\int_{E_1 \cup E_2} f d\mu \geq \int_{E_1 \cup E_2} s d\mu = \int_{E_1} s d\mu + \int_{E_2} s d\mu.$$

Pick $\varepsilon > 0$ and let s_1 and s_2 be simple functions satisfying $0 \leq s_1, s_2 \leq f$ such that

$$\begin{aligned} \int_{E_1} s_1 d\mu &\geq \int_{E_1} f d\mu - \varepsilon, \\ \int_{E_2} s_2 d\mu &\geq \int_{E_2} f d\mu - \varepsilon, \end{aligned}$$

and take $s = \max(s_1, s_2)$. Then we have

$$\int_{E_1 \cup E_2} f d\mu \geq \int_{E_1} s d\mu + \int_{E_2} s d\mu \geq \int_{E_1} f d\mu + \int_{E_2} f d\mu - 2\varepsilon.$$

Since ε is chosen arbitrarily, then we in fact have the inequality

$$\int_{E_1 \cup E_2} f d\mu \geq \int_{E_1} f d\mu + \int_{E_2} f d\mu.$$

Now we want the reverse inequality. We have

$$\begin{aligned} \int_{E_1 \cup E_2} f d\mu &= \sup_{0 \leq s \leq f} \left(\int_{E_1 \cup E_2} s d\mu \right) = \sup_{0 \leq s \leq f} \left(\int_{E_1} s d\mu + \int_{E_2} s d\mu \right) \\ &\leq \sup_{0 \leq s \leq f} \int_{E_1} s d\mu + \sup_{0 \leq s \leq f} \int_{E_2} s d\mu = \int_{E_1} f d\mu + \int_{E_2} f d\mu. \end{aligned}$$

This shows that we have finite additivity. But notice also that this argument also holds for any countable union since 1 holds by countable additivity of simple functions. Therefore, we have countable subadditivity. Since finite additivity and countable subadditivity give countable additivity, then this completes the proof that φ is a measure. \square

We would now like to show that for measurable functions f and g and a measurable set E , we have

$$\int_E f + g \, d\mu = \int_E f \, d\mu + \int_E g \, d\mu.$$

This will take quite a bit of work, and we begin by proving some results about simple functions.

Theorem 6.10. *For any nonnegative measurable function f , we can find simple functions $0 \leq s_1 \leq s_2 \leq \dots$ with $\lim_{n \rightarrow \infty} s_n(x) = f(x)$. Furthermore, if f is bounded, then this sequence converges uniformly.*

Proof. Let $f : X \rightarrow \mathbb{R}_{\geq 0}$ be a measurable function. We can partition $\mathbb{R}_{\geq 0}$ in the following way:

$$\bigcup_{i=1}^{n \cdot 2^n - 1} \left[\frac{i}{2^n}, \frac{i+1}{2^n} \right] \cup [n, \infty].$$

For $0 \leq i \leq n \cdot 2^n - 1$, let

$$E_{i,n} = \left\{ x \in X : \frac{i}{2^n} \leq f(x) < \frac{i+1}{2^n} \right\}, \quad E_{n \cdot 2^n, n} = \{x : f(x) \geq n\}.$$

Now define the simple function s_n as

$$s_n(x) = \sum_{i=0}^{n \cdot 2^n - 1} \left(\frac{i}{2^n} \right) \chi_{E_{i,n}}(x) + n \cdot \chi_{E_{n \cdot 2^n, n}}(x).$$

It is clear that $s_n(x) \leq f(x)$. If $f(x) < n$, then $f(x) - \frac{1}{2^n} \leq s_n(x) \leq f(x)$. If $f(x) < \infty$, then $s_n \rightarrow f$. If $f(x) = \infty$, then $s_n \rightarrow \infty$ also (since then $s_n = n$). So $\lim_{n \rightarrow \infty} s_n(x) = f(x)$. Note that we have simultaneously proven that convergence is uniform if f is bounded.

It remains to show that the sequence $\{s_n\}$ is bounded. We first deal with case when $x \in E_{i,n}$, $0 \leq i \leq n \cdot 2^n - 1$. In this case, we have

$$\frac{2i}{2^{n+1}} = \frac{i}{2^n} \leq f(x) \leq \frac{i+1}{2^n} = \frac{2i+2}{2^{n+1}}.$$

Now, by definition,

$$s_{n+1}(x) = \begin{cases} \frac{2i}{2^{n+1}} & \text{if } \frac{2i}{2^{n+1}} \leq f(x) < \frac{2i+1}{2^{n+1}} \\ \frac{2i+1}{2^{n+1}} & \text{if } \frac{2i+1}{2^{n+1}} \leq f(x) \leq \frac{2i+2}{2^{n+1}}. \end{cases}$$

Hence indeed we have $s_n(x) \leq s_{n+1}(x)$ when $x \in E_{i,n}$ for $0 \leq i \leq n \cdot 2^n - 1$. Now consider the case when $x \in E_{n \cdot 2^n, n}$. We have $f(x) \geq n$ and

$$s_{n+1}(x) = \begin{cases} n+1 & \text{if } f(x) \geq n+1 \\ \geq n & \text{if } f(x) \in [n, n+1), \end{cases}$$

where the second inequality holds since we can then divide $[n, n+1)$ into segments of length $1/2^{n+1}$. Therefore $s_n \leq s_{n+1}$ everywhere and this proves monotonicity. This completes the proof. \square

In summary, what we have is that if $\{s_n\}$ is a monotone increasing sequence with limit f , then

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

Note that the condition that $\{s_n\}$ is monotone increasing is extremely important. The following example demonstrates this.

Example 6.11. Consider for instance a sequence of functions $\{f_n\}$ where

$$f_n(x) = \begin{cases} n & \text{if } x \in [\frac{1}{n}, \frac{2}{n}) \\ 0 & \text{otherwise.} \end{cases}$$

Then $\lim_{n \rightarrow \infty} f_n(x) = 0$. On the other hand, $\int_0^1 f_n dx = 1$.

Let us return to the question of how to prove that $\int_E f + g d\mu = \int_E f d\mu + \int_E g d\mu$. Why should we believe that such an identity is true? One direction of this inequality is not so hard to show. By Proposition 6.10, we know that there exist simple functions

$$\begin{aligned} s_1 &\leq s_2 \leq \cdots \leq s_n \leq \cdots \leq f, \\ t_1 &\leq t_2 \leq \cdots \leq t_n \leq \cdots \leq g. \end{aligned}$$

Summing these strings of inequalities gives a monotone increasing sequence of simple functions

$$s_1 + t_1 \leq s_2 + t_2 \leq \cdots \leq s_n + t_n \leq \cdots \leq f + g.$$

This means that we can choose a simple function s with $0 \leq s \leq f$ such that $\int_E s d\mu > \int_E f d\mu - \varepsilon$ and a simple function g with $0 \leq t \leq g$ such that $\int_E t d\mu > \int_E g d\mu - \varepsilon$. From this we have

$$\int_E f + g d\mu \geq \int_E s + t d\mu \geq \int_E f d\mu + \int_E g d\mu - 2\varepsilon,$$

which gives $\int_E f + g d\mu \geq \int_E f d\mu + \int_E g d\mu$. But how would we obtain the reverse inequality? We now present a theorem that will allow us to prove that the integral of a sum of measurable functions is the sum of the separate integrals.

Theorem 6.12 (Monotone Convergence). *If $f_1 \leq f_2 \leq \dots \leq f_n \leq \dots$ is a monotone sequence of nonnegative measurable functions and $f(x) = \lim_{n \rightarrow \infty} f_n(x)$, then $\int_E f \, d\mu = \lim_{n \rightarrow \infty} \int_E f_n \, d\mu$. (Note that f is automatically measurable since it is the limit of a sequence of measurable functions.)*

Proof. Let s be a simple function with $0 \leq s \leq f$. We would like to show that

$$\lim_{n \rightarrow \infty} \int_E f_n \, d\mu \geq \int_E s \, d\mu.$$

If we can do this, then we are done since we would then have

$$\lim_{n \rightarrow \infty} \int_E f_n \, d\mu \geq \sup_s \int_E s \, d\mu = \int_E f \, d\mu,$$

which is exactly the inequality we need to complete the previous argument. So now we will concentrate all of our efforts on proving this inequality.

Pick any $c \in \mathbb{R}$ with $0 < c < 1$. Let $E_n = \{x \in E : f_n(x) \geq cs\} \subseteq E$. If $E \not\supseteq \cup_n E_n$, then there exists some $x \in E$ such that $f_n(x) < cs(x)$ for all n . But then we have $s(x) \leq f(x) < cs(x)$, and since $c \in (0, 1)$, then $s(x) = 0$, which is a contradiction since f is nonnegative by assumption. Hence $E = \cup_n E_n$.

For any $c \in (0, 1)$ and for all n , we have

$$\int_E f_n \, d\mu \geq \int_{E_n} f_n \, d\mu \geq \int_{E_n} cs \, d\mu = c \int_{E_n} s \, d\mu,$$

and taking the limit as $n \rightarrow \infty$, we get

$$\lim_{n \rightarrow \infty} \int_E f_n \, d\mu \geq c \lim_{n \rightarrow \infty} \int_{E_n} s \, d\mu = \int_E s \, d\mu,$$

where the last equality via first defining disjoint sets $A_1 = E_1$, $A_2 = E_2 - E_1$, $A_3 = E_3 - E_2$, \dots , and then applying countable additivity. \square

What this proof gives us in particular is that if we have a monotone sequence of simple functions

$$0 \leq s_1 \leq s_2 \leq \dots \leq s_n \leq \dots \leq f,$$

where $s_n \rightarrow f$, which we have for any measurable function f by Proposition 6.10, then in fact

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

We can finally prove the following proposition.

Proposition 6.13. *If f and g are nonnegative measurable functions, then*

$$\int f + g \, d\mu = \int f \, d\mu + \int g \, d\mu.$$

Proof. By Proposition 6.10, we have simple functions

$$\begin{aligned} s_1 \leq s_2 \leq \cdots \leq s_n \leq \cdots \leq f, & \quad \lim_n s_n = f, \\ t_1 \leq t_2 \leq \cdots \leq t_n \leq \cdots \leq g, & \quad \lim_n t_n = g. \end{aligned}$$

Summing these, we have

$$s_1 + t_1 \leq s_2 + t_2 \leq \cdots \leq s_n + t_n \leq \cdots \leq f + g, \quad \lim_n s_n + t_n = f + g.$$

By the Monotone Convergence Theorem, we then have

$$\begin{aligned} \int f + g \, d\mu &= \lim_{n \rightarrow \infty} \int s_n + t_n \, d\mu \\ &= \lim_{n \rightarrow \infty} \int s_n \, d\mu + \lim_{n \rightarrow \infty} \int t_n \, d\mu \\ &= \int f \, d\mu + \int g \, d\mu. \end{aligned} \quad \square$$

Corollary 6.14. *For nonnegative measurable functions f_i , we have*

$$\int \sum_{i=1}^{\infty} f_i \, d\mu = \sum_{i=1}^{\infty} \int f_i \, d\mu.$$

Proof. We can take $g_1 = f_1$, $g_2 = f_1 + f_2$, $g_3 = f_1 + f_2 + f_3$, \dots , $g_n = \sum_{i=1}^n f_i$. This is a monotone increasing sequence and they converge pointwise to $\sum_{i=1}^{\infty} f_i$. Therefore by the Monotone Convergence Theorem, the result follows. \square

Recall Chebyshev's Inequality: If $\int_E f \, d\mu < \infty$, then

$$\mu(E_c) \leq \frac{1}{c} \int_E f \, d\mu, \text{ where } E_c = \{x \in E : f(x) \geq c\}.$$

In particular, this means that if $\int_X f < \infty$, then $\mu(\{x : f(x) = \infty\}) = 0$. We also have the following easy corollaries:

- If $\int_X f = 0$, then $f = 0$ except on a set of measure zero. (For any $\varepsilon > 0$, $\mu(\{x : f(x) > \varepsilon\}) \leq \frac{1}{\varepsilon} \int_X f = 0$. Since $\{x : f(x) > 0\} = \cup_{n=1}^{\infty} \{x : f(x) \geq \frac{1}{n}\}$, then $\mu(\{f(x) > 0\}) = 0$.)
- If $f \leq g$ and $\int_X f = \int_X g$, then $f = g$ except on a set of measure zero. Alternatively, we say that $f = g$ *almost everywhere*, or in shorthand, *a.e.*

Example 6.15. Let $E = \mathbb{Q} \cap [0, 1]$. Then $\int_{[0,1]} \chi_E = 0$, and so in some sense, we can think of χ_E as basically being “the same” as the function that is identically 0.

Now that we have developed the theory of the Lebesgue integral for non-negative measurable functions, we can extend this theory to all measurable functions.

Definition 6.16. For a measurable function $f : X \rightarrow \mathbb{R}$, we define $f^+ := \max(f, 0)$ and $f^- := \max(-f, 0)$. Then $f = f^+ - f^-$ and $|f| = f^+ + f^-$. We say that f is *Lebesgue integrable* if one of the two equivalent statements is true:

- (a) $\int_E f^+ d\mu$ and $\int_E f^- d\mu$ are both $< \infty$.
- (b) $\int_E |f| d\mu < \infty$.

(We can easily see that these two conditions are equivalent since $|f| = f^+ + f^-$.) In this situation, we define

$$\int f := \int f^+ - \int f^-.$$

Example 6.17. Note that on unbounded sets, we may have Riemann integrable functions that are not Lebesgue integrable. For instance, consider $\int_0^\infty \frac{\sin(x)}{x} dx$.

Note. We can also make sense of complex-valued measurable functions. Take $f : X \rightarrow \mathbb{C}$. We can write $f(x) = u(x) + iv(x)$ where u and v are real valued. Then we define

$$\int_E f d\mu := \int_E u d\mu + i \int_E v d\mu.$$

We say that f is *integrable* if $\Re f$ and $\Im f$ are both integrable. Since $|\Re f|, |\Im f| \leq |f| \leq |\Re f| + |\Im f|$, then we have that f is integrable if and only if $|f|$ is integrable.

From the definition of $\int f = \int f^+ - \int f^-$, for an integrable function $f : X \rightarrow \mathbb{R}$, we have several immediate properties:

- $\int_E cf d\mu = c \int_E f d\mu$, for $c \in \mathbb{R}$.
- $\int_E (f + g) d\mu = \int_E f d\mu + \int_E g d\mu$, where $g : X \rightarrow \mathbb{R}$ is also integrable. (We need to be careful in the proof here since $(f + g)^+ \neq f^+ + g^+$.)
- If $f \leq g$, then $\int f \leq \int g$.

Recall the example earlier that if $f_n(x) = \begin{cases} n & \text{on } [1/n, 2/n] \\ 0 & \text{otherwise} \end{cases}$, then we have

$\liminf \int f_n = 1$ but $\int \liminf f_n = 0$. So we see that \liminf and \int cannot always be interchanged. However, Fatou's lemma gives us an inequality comparing the results when we first take \liminf or \int .

Lemma 6.18 (Fatou's Lemma). *If f_1, f_2, \dots is a sequence of nonnegative measurable functions, then*

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

Proof. By definition, we have

$$\liminf_{n \rightarrow \infty} f_n(x) = \sup_k \inf_{n \geq k} f_n(x) = \sup_k g_k(x),$$

where we define $g_k(x) = \inf_{n \geq k} f_n(x)$. By construction, $\{g_k\}$ is monotone increasing. By the monotone convergence theorem,

$$\int \liminf f_n(x) d\mu = \int \lim_{k \rightarrow \infty} g_k(x) d\mu = \lim_{k \rightarrow \infty} \left(\int_E g_k(x) d\mu \right).$$

By construction, for all $n \geq k$,

$$\int_E g_k(x) d\mu \leq \int_E f_n(x),$$

and so $\int_E g_k(x) d\mu \leq \inf_{n \leq k} \left(\int_E f_n d\mu \right)$. Therefore we have

$$\int_E \liminf f_n(x) d\mu = \lim_{k \rightarrow \infty} \int_E g_k(x) d\mu \leq \liminf_{n \rightarrow \infty} \int_E f_n d\mu. \quad \square$$

Note that in the above proof, the place where we have a potential loss of equality is in comparing $\int_E g_k d\mu$ with $\int_E f_n d\mu$ for $n \geq k$. Note also that we used the monotone convergence theorem, which is something that we do not have for Riemann integrals. For instance, consider the following: Label the elements of the set $\mathbb{Q} \cap [0, 1]$ so that $r_1 < r_2 < \dots$ and consider the function $f_n := \chi(\{r_1, \dots, r_n\})$. Then certainly $f_1 \leq f_2 \leq \dots$ but $\lim_{n \rightarrow \infty} f_n$ is not Riemann integrable. (It turns out that if the limit is also Riemann integrable, then the conclusion of the monotone convergence theorem holds.)

We conclude this section with one final theorem and then we will move onto comparing the Lebesgue integral theory we have just developed with Riemann integration theory.

Theorem 6.19 (Lebesgue's Dominated Convergence Theorem). *If f_1, \dots, f_n, \dots are measurable functions with $|f_n| \leq g$ (f_n is dominated by g), $\int_E g d\mu < \infty$ (g is integrable), and $\lim_{n \rightarrow \infty} f_n = f$ (assuming this limit exists), then*

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

Proof. We prove this theorem via two applications of Fatou's lemma. Using the notation as in the theorem statement, we have $-g \leq f_n \leq g$. Hence $\{g + f_n\}$ and $\{g - f_n\}$ are both sequences of nonnegative functions. Applying Fatou's lemma to the first sequence, we have

$$\int g + \liminf_{n \rightarrow \infty} \int f_n = \liminf_{n \rightarrow \infty} \int (g + f_n) \geq \int \liminf_{n \rightarrow \infty} (g + f_n) = \int g + \liminf_{n \rightarrow \infty} \int f_n.$$

Since $\int g < \infty$, then we can cancel $\int g$ on both sides of this inequality and get

$$\liminf_{n \rightarrow \infty} \int_E f_n \geq \int_E \liminf_{n \rightarrow \infty} f_n. \quad (6.1)$$

Applying Fatou's lemma to the second sequence, $\{g - f_n\}$, we get

$$\int g - \limsup_{n \rightarrow \infty} \int f_n = \liminf_{n \rightarrow \infty} \int (g - f_n) \geq \int \liminf_{n \rightarrow \infty} (g - f_n) = \int g - \limsup_{n \rightarrow \infty} \int f_n.$$

Again cancelling $\int g$ on both sides, we have

$$\int \limsup f_n \geq \limsup \int f_n. \quad (6.2)$$

But since $\lim f_n$ exists, then $\liminf f_n = \lim f_n = \limsup f_n$, and so combining (6.1) and (6.2), we have

$$\liminf \int f_n \geq \int \lim f_n \geq \limsup \int f_n.$$

But then $\liminf \leq \limsup$, and so we must have equality. This completes the proof. \square

Chapter 7

\mathcal{L}^1 -Theory

Now we will study the space of measurable functions $f : X \rightarrow \mathbb{C}$ such that $\int_X |f| d\mu < \infty$. We call this space $\mathcal{L}^1(X, \mu)$. We define

$$\|f\|_1 = \int_X |f| d\mu.$$

(Recall in the previous section we defined this integral to be the sum of the integral of the real part and the integral of the imaginary part of f .) There are some immediate properties of $\|\cdot\|_1$.

1. $\|f\|_1 \geq 0$
2. For $c \in \mathbb{C}$, $\|cf\|_1 = |c| \cdot \|f\|_1$.
3. $\|f + g\|_1 \leq \|f\|_1 + \|g\|_1$.
4. $\|f\|_1 = \|\bar{f}\|_1$.

In general, a vector space with $\|\cdot\|$ satisfying the above properties is called a *normed vector space*. Hence $\mathcal{L}^1(X, \mu)$ is a normed vector space. But before we move on, notice that we have a small problem: If $\|f\|_1 = 0$, this only means that $f = 0$ almost everywhere. Hence to get around this issue, we will identify measurable functions f and g in $\mathcal{L}^1(X, \mu)$ that differ only on a set of measure 0. Now, to make $\mathcal{L}^1(X, \mu)$ into a metric space, we define

$$d(f, g) = \|f - g\|_1.$$

In general, a complete normed vector space is called a *Banach space*, so if we can show that $\mathcal{L}^1(X, \mu)$ is complete with respect to this notion of distance, then we will have proved the following theorem:

Theorem 7.1. $\mathcal{L}^1(X, \mu)$ is a Banach space.

What we need to prove here is that every Cauchy sequence in $\mathcal{L}^1(X, \mu)$ converges. That is, if f_1, f_2, \dots is a sequence of functions in $\mathcal{L}^1(X, \mu)$ such that for any ε there exists an $N(\varepsilon) \in \mathbb{N}$ such that if $m, n \geq N(\varepsilon)$, then $\|f_m - f_n\|_1 < \varepsilon$. Recall that $\{f_n\}$ converges to f in this space if $\|f_n - f\|_1 \rightarrow 0$ as $n \rightarrow \infty$. In fact, this limit is unique! Note however that this is not necessarily the same as $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ almost everywhere. For instance, recall the “bump” example in Example 6.11. We now proceed with the proof that $\mathcal{L}^1(X, \mu)$ is a Banach space.

Proof. Let f_1, f_2, \dots be a Cauchy sequence of \mathcal{L}^1 functions. For every $r \geq 1$, we can find an $n_r \in \mathbb{N}$ such that if $m, n \geq n_r$, then $\|f_m - f_n\|_1 \leq \frac{1}{2^r}$. Now consider the subsequence f_{n_1}, f_{n_2}, \dots , and notice that we can write this equivalently as

$$f_{n_1}, f_{n_1} + (f_{n_2} - f_{n_1}), f_{n_1} + (f_{n_2} - f_{n_1}) + (f_{n_3} - f_{n_2}), \dots$$

Let $g_1 = f_{n_1}, g_2 = f_{n_2} - f_{n_1}, \dots, g_r = f_{n_r} - f_{n_{r-1}}, \dots$. Then

$$\|g_1\|_1 < \infty, \|g_2\|_1 < \frac{1}{2}, \dots, \|g_r\|_1 < \frac{1}{2^r}, \dots$$

and hence $\sum_{k=1}^{\infty} \int |g_k| d\mu < \infty$. We would like to show that $\sum_{k=1}^{\infty} g_k$ converges almost everywhere. To obtain this, we prove the following lemma.

Lemma 7.2. *Let g_1, g_2, \dots be a sequence of integrable functions such that $\sum_{n=1}^{\infty} \int |g_n| d\mu < \infty$. Then*

1. $\sum_{n=1}^{\infty} g_n(x)$ converges absolutely almost everywhere and defines an integrable function.
2. $\int \sum_{n=1}^{\infty} g_n d\mu = \sum_{n=1}^{\infty} \int g_n d\mu$.

Proof of Lemma. Define $h_n = |g_n| \geq 0$. By the monotone convergence theorem, $\int \sum_{n=1}^{\infty} h_n = \sum_{n=1}^{\infty} \int h_n < \infty$. This implies that $\sum_{n=1}^{\infty} h_n(x) < \infty$ almost everywhere, which means that $\sum_{n=1}^{\infty} g_n(x)$ converges absolutely almost everywhere. Let $H(x) = \sum_{n=1}^{\infty} h_n(x)$. This is a nonnegative integrable function and since we have

$$\left| \sum_{n=1}^N g_n(x) \right| \leq \sum_{n=1}^N |g_n(x)| \leq H(x)$$

almost everywhere and $\int_X H d\mu < \infty$, then by the dominated convergence theorem,

$$\int \lim_{N \rightarrow \infty} \sum_{n=1}^N g_n(x) dx = \lim_{N \rightarrow \infty} \sum_{n=1}^N \int g_n d\mu = \sum_{n=1}^{\infty} \int g_n d\mu. \quad \square$$

So we have proved that if g_1, g_2, \dots is a sequence of functions in \mathcal{L}^1 and $\sum_{n=1}^{\infty} \|g_n\|_1 < \infty$, then $\sum_{n=1}^{\infty} g_n$ converges pointwise almost everywhere and defines an \mathcal{L}^1 function. We return now to the proof that \mathcal{L}^1 is complete.

We left off before the lemma with a sequence of functions g_1, g_2, \dots such that $\|g_1\|_1 < \infty, \|g_2\|_1 < \frac{1}{2}, \dots$ and in particular $\sum_{r=1}^{\infty} \|g_r\|_1 < \infty$. By the

lemma, $\sum_{r=1}^{\infty} g_r$ is convergence almost everywhere and defines an \mathcal{L}^1 function $f := \sum_{r=1}^{\infty} g_r$. Note that by construction f is the a.e. limit of the subsequence $\{f_{n_r}\}_{r=1,2,\dots}$. Finally, we must check that indeed $\|f - f_n\|_1 \rightarrow 0$ as $n \rightarrow \infty$. One way to do this is to use the triangle inequality and note that $\|f - f_n\|_1 \leq \|f - f_{n_r}\|_1 + \|f_{n_r} - f_n\|_1$, and each summand on the right hand side tends to 0 as n tends to infinity. Alternatively, we may use Fatou's lemma. This completes the proof that $\mathcal{L}^1(X, \mu)$ is a Banach space. \square

In the above proof, the main idea is really the lemma, which we restate in the following proposition.

Proposition 7.3. *A normed vector space is complete if and only if every absolutely summable series is summable.*

Chapter 8

Lebesgue vs. Riemann

We mentioned in Chapter 6 an example of where dominated convergence fails. That is, we discussed an instance of when we have a sequence of integrable functions whose limit is Lebesgue integrable but not Riemann integrable. We will now discuss the general situation of the relationship between Lebesgue integration theory and Riemann integration theory.

Consider a bounded function $f : [a, b] \rightarrow \mathbb{R}$. We discuss the Darboux integral of f . Take a partition of $[a, b]$: $P = \{x_0 = a < x_1 < \cdots < x_n = b\}$. We define the *mesh width* of P to be $m(P) = \max(x_{i+1} - x_i)$. Now let

$$\begin{aligned} L(f, P) &= \sum_{i=1}^N m_i(x_i - x_{i-1}), & m_i &= \inf_{x_{i-1} \leq x < x_i} f(x), \\ U(f, P) &= \sum_{i=1}^N M_i(x_i - x_{i-1}), & M_i &= \sup_{x_{i-1} \leq x < x_i} f(x), \end{aligned}$$

and define

$$\begin{aligned} L(f) &= \int_a^b f(x) dx = \sup_P L(f, P), \\ U(f) &= \int_a^b f(x) dx = \inf_P U(f, P). \end{aligned}$$

We say that f is *Riemann integrable* if $L(f) = U(f)$.

Theorem 8.1. *If $f : [a, b] \rightarrow \mathbb{R}$ is bounded and Riemann integrable, then f is Lebesgue integrable and*

$$\int_{[a,b]} f d\mu = \int_a^b f(x) dx.$$

We first establish some terminology. We say that a partition P' *refines* a partition P if P' contains all points in P and possibly more points. Then $L(f, P) \leq$

$L(f, P')$ and $U(f, P) \geq U(f, P')$. Also $L(f, P) \leq U(f, P)$. Furthermore, for any two partitions P_1 and P_2 , $L(f, P_1) \leq U(f, P_2)$. (We can choose a partition P that refines both P_1 and P_2 and get $L(f, P_1) \leq L(f, P) \leq U(f, P) \leq U(f, P_2)$. Now we may begin the proof.

Proof. Choose partitions P_k such that

1. P_k refines P_{k-1} .
2. $L(f, P_1) \leq L(f, P_2) \leq \dots$ converges to $L(f)$ and $U(f, P_1) \geq U(f, P_2) \geq \dots$ converges to $U(f)$.

Let L_k be the (simple) step function associated to $L(f, P_k)$ and U_k the (simple) step function associated to $U(f, P_k)$. When

$$L_1(x) \leq L_2(x) \leq \dots \leq f(x) \leq \dots \leq U_2(x) \leq U_1(x),$$

and hence setting $L(x) = \lim_{n \rightarrow \infty} L_n(x)$ and $U(x) = \lim_{n \rightarrow \infty} U_n(x)$, we have $L(x) \leq f(x) \leq U(x)$. Now notice that

$$L(f, P_k) = \int L_k d\mu \quad \text{and} \quad U(f, P_k) = \int U_k d\mu.$$

By the monotone convergence theorem,

$$\begin{aligned} \int L d\mu &= \lim_{n \rightarrow \infty} \int L_k d\mu = L(f) \\ \int U d\mu &= \lim_{n \rightarrow \infty} \int U_k d\mu = U(f). \end{aligned}$$

By assumption f is Riemann integrable, and so $\int (U - L) = \int U - \int L = 0$. Since $U - L \geq 0$, then this means that $U = L$ almost everywhere, and hence $U = f = L$ almost everywhere. Therefore f is measurable and Lebesgue integrable and $\int f d\mu = \int f(x) dx$. \square

Remark 8.2. A function is Riemann integrable if and only if the set of discontinuities is of measure zero.

Recall that we had the following example to demonstrate the failure of the dominated convergence theorem for Riemann integration theory.

Example 8.3. Let $f_n = \chi(\{r_1, \dots, r_n\})$, where r_1, r_2, \dots is an enumeration of the rationals in $[0, 1]$. Then $\|f_n\|_1 = 0$. Pointwise, $f_n \rightarrow \chi_{\mathbb{Q} \cap [0, 1]}$, but this is not Riemann integrable. On the other hand, in \mathcal{L}^1 , $f_n \rightarrow 0$ and 0 is Riemann integrable.

We have a better example.

Example 8.4. We construct what is known as the “fat Cantor set.”

1. Take $[0, 1]$, and throw away an open interval of length $\frac{1}{4}$ centered at $\frac{1}{2}$. Define

$$F_1 = [0, \frac{3}{8}] \cup [\frac{5}{8}, 1].$$

2. Now throw away an interval of length $\frac{1}{16}$ centered at the center of each interval. We define F_2 to be the resulting union of 4 intervals.
3. Iterate this process. Then F_n is the union of 2^n intervals. To get to F_{n+1} , throw away intervals of length $\frac{1}{4^{n+1}}$.

Now let

$$F = \bigcap_{n=1}^{\infty} F_n.$$

Since F is a countable intersection of closed intervals, then it too is closed and hence it is measurable and in fact even Borel. Now consider the complement of F in $[0, 1]$. We compute its measure:

$$\begin{aligned} \mu([0, 1] \setminus F) &= \frac{1}{4} + \frac{2}{4^2} + \frac{2^2}{4^3} + \cdots \\ &= \frac{1}{4} \left(1 + \frac{1}{2} + \frac{1}{4} + \cdots \right) = \frac{1}{2}. \end{aligned}$$

Notice that F does not contain any intervals. Now let $f_n = \chi_{F_n}$, the characteristic function of the set $F_n \subseteq [0, 1]$. This is Riemann integrable. In \mathcal{L}^1 , $\chi_{F_n} \rightarrow \chi_F$ as $n \rightarrow \infty$. In fact, we can take the Lebesgue integral of χ_F since it is just a simple function. We have

$$\int_{[0,1]} \chi_F d\mu = \mu(F) = \frac{1}{2}.$$

Since χ_F is discontinuous on a set of positive measure, then it is not Riemann integrable, and hence χ_{F_n} does not converge to a Riemann integrable function.

Now we would like to discuss the fundamental theorem of calculus in the context of Lebesgue integration theory. Recall from Riemann integration theory the following theorem.

Theorem 8.5 (Fundamental Theorem of Calculus). *1. If $\int_a^x f(t) dt = F(x)$, then $F'(x) = f(x)$.*

2. Also, $\int_a^x f(t) dt = f(x) - f(a)$.

Note that statement 2 can fail badly. Recall the Cantor function. Write $x = \sum_{n=1}^{\infty} \frac{a_n}{3^n}$. Let N denote the first j such that $a_j = 1$. Now consider the function

$$f(x) = \sum_{j=1}^{N-1} \frac{(a_j/2)}{2^j} + \frac{1}{2^N}.$$

This function is a continuous function $f : [0, 1] \rightarrow [0, 1]$ and it is differentiable on the complement of the cantor set with derivative identically 0. Then

$$0 = \int_{[0,1]} f' d\mu \neq f(1) - f(0) = 1,$$

and this illustrates that indeed the assertion 2 in the Fundamental Theorem of Calculus fails when we replace the Riemann integral with the corresponding Lebesgue integral.

Now consider a continuous, increasing function $f : [a, b] \rightarrow \mathbb{R}$ that is integrable and almost everywhere differentiable. Then

$$\begin{aligned} \int_a^x f'(t) dt &= \int_a^x \liminf_{n \rightarrow \infty} \left(\frac{f(t + 1/n) - f(t)}{1/n} \right) dt \\ &\leq \liminf_{n \rightarrow \infty} n \left(\int_a^x (f(t + 1/n) - f(t)) dt \right) \\ &\stackrel{1}{=} \liminf_{n \rightarrow \infty} n \left(\int_x^{x+1/n} f(t) dt - \int_a^{a+1/n} f(t) dt \right) = f(x) - f(a), \end{aligned}$$

where equality 1 holds by a change of variable, and the final equality holds by the continuity of f . This motivates the following theorem.

Theorem 8.6. *Let f be a continuous, nondecreasing function. Then the following are equivalent.*

1. f is absolutely continuous.
2. f is differentiable almost everywhere and $\int_a^x f'(t) dt = f(x) - f(a)$.
3. f maps sets of measure zero to sets of measure zero.

(Note that the final statement excludes functions like the Cantor function.)

Definition 8.7. A function f on $[a, b]$ is said to be *absolutely continuous* if for all $\varepsilon > 0$, there exists a $\delta > 0$ such that given any disjoint segments $(\alpha_1, \beta_1), \dots, (\alpha_n, \beta_n)$ with $\sum_{j=1}^n (\beta_j - \alpha_j) < \delta$, then $\sum_{j=1}^n |f(\beta_j) - f(\alpha_j)| < \varepsilon$.

Remark 8.8. The theorem tells us that μ and ν are absolutely continuous if and only if the sets of measure zero for μ are exactly the sets of measure zero for ν .

Chapter 9

Random Variables and \mathcal{L}^p -spaces

We can make precise the notion of a *random variable*. Let (X, \mathcal{F}, μ) be a probability space, with \mathcal{F} a σ -field. A measurable function f may be thought of as a random variable in the following way. We can denote the expectation of a random variable by $\mathbb{E}(f)$, and then

$$\mathbb{E}(f) = \int_X f \, d\mu.$$

By the linearity of the Lebesgue integral, we have $\mathbb{E}(f_1 + f_2) = \mathbb{E}(f_1) + \mathbb{E}(f_2)$.

Now let us discuss random walks in \mathbb{Z}^d . The main question here is the following: With what probability will a random walk return to the origin? We can make this precise by working in base (2d) to get a measure space. Consider the following function

$$X_n = \begin{cases} 1 & \text{at the } n\text{th step, we are back to 0} \\ 0 & \text{otherwise.} \end{cases}$$

Then $P_n := \mathbb{E}(X_n)$ is the probability of return after n steps. Then $X := \sum_{n=1}^{\infty} X_n$ counts the number of returns. By the monotone convergence theorem,

$$\mathbb{E}(X) = \sum_{n=1}^{\infty} \mathbb{E}(X_n) = \sum_{n=1}^{\infty} P_n.$$

Note that $P_n = 0$ if n is odd. Let ρ denote the probability that the random walk returns at least once to 0. Let ρ_k denote the probability that we return to 0 exactly k times. Then

$$\rho_0 = (1 - \rho), \quad \text{and} \quad \rho_k = \rho^k(1 - \rho).$$

From this, we have

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} k \cdot \rho^k (1 - \rho) = \rho(1 - \rho) \sum_{k=1}^{\infty} k \rho^{k-1} = \begin{cases} \frac{\rho(1-\rho)}{(1-\rho)^2} = \frac{\rho}{1-\rho} & \text{if } \rho < 1 \\ \infty & \text{if } \rho = 1. \end{cases}$$

Hence in the above discussion we have proven the following:

Theorem 9.1. *If $\rho < 1$, then $\mathbb{E}(\# \text{ of returns})$ is finite (which happens if and only if $\sum P_n$ converges). If $\rho = 1$, then $\mathbb{E}(\# \text{ of returns})$ is infinite (which happens if and only if $\sum P_n$ diverges). So $\mathbb{E}(\# \text{ of returns}) = \sum_{n=1}^{\infty} P_n$.*

Remark 9.2. This theorem is very similar to Borel-Cantelli 2, but it is slightly nicer as we do not need the independence of the events.

There is a beautiful theorem due to Polya that says that $\rho < 1$ if $d \geq 3$ and $\rho = 1$ if $d = 1, 2$. We first discuss the case when $d = 1$. We know $P_{2n+1} = 0$. Also, by the binomial theorem, using the largest coefficient in concluding the following inequality, we have

$$P_{2n} = \frac{1}{2^{2n}} \binom{2n}{n} \geq \frac{1}{2n+1}.$$

We can do a better job of estimating this, though. Recall Stirling's formula:

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.$$

Using this, we have

$$P_{2n} \approx \frac{\sqrt{2\pi 2n}}{2^{2n} (\sqrt{2\pi n})^2} \frac{\left(\frac{2n}{e}\right)^{2n}}{\left(\left(\frac{n}{e}\right)^n\right)^2} \approx \frac{c}{\sqrt{n}}.$$

(On the right hand side, $c = \frac{1}{\sqrt{\pi}}$.) Hence $\sum_{n=1}^{\infty} P_{2n}$ diverges, and this agrees with the $d = 1$ case of Polya's result.

Now consider the case $d = 2$. Then we have

$$\begin{aligned} P_{2n} &= \frac{1}{4^{2n}} \sum_{k=0}^n \binom{2n}{k} \binom{2n-k}{k} \binom{2n-2k}{n-k} \\ &= \frac{(2n)!(2n-2k)!}{k!k!(2n-2k)!(n-k)!(n-k)!} = \frac{(2n)!}{(k!)^2((n-k)!)^2}. \end{aligned}$$

Using the law of large numbers, we can see that the max is attained when $k \approx \frac{n}{2}$. Furthermore, there is an interval of length $\approx \sqrt{n}$ around $\frac{n}{2}$ where the expression is largest, and each such $k \approx \frac{\sqrt{n}}{(\sqrt{n})^4} \approx \frac{1}{n^{3/2}}$. Hence the number of k is $\approx \sqrt{n}$, which tends to $\frac{1}{n}$, so again $\sum_{n \rightarrow \infty} P_n = \infty$. In general, P_n is about size $\frac{c}{n^{d/2}}$ so $\sum P_n < \infty$ if $d \geq 3$.

Because of this discussion, we can now answer a question we started the course with... namely, if we toss a coin infinitely many times, how often will we

get exactly the same number of tails tossed as heads tossed? If we roll a die, how often will we get exactly the same number of each number rolled? We have shown the following sequence of equivalences:

$$\sum \mathbb{E}(X_n) < \infty \iff \mathbb{E}(X) < \infty \iff \rho < 1.$$

The last condition implies that the probability of the event happening infinitely often is 0, and hence we can conclude that the probability that there will be infinitely many times when each number on the dice is rolled the same number of times is 0.

Now let X be a probability space and let $f : X \rightarrow \mathbb{R}$ be a random variable. Then we define

$$\mu_f(B) = \mu(f^{-1}(B)),$$

where $B \subseteq \mathbb{R}$ is Borel. Then μ_f is a measure on $(\mathbb{R}, \mathcal{F})$, where \mathcal{F} is the σ -field of Borel sets. We call this the *probability distribution associated to f* .

We discussed earlier the Banach space (i.e. complete normed vector space) $\mathcal{L}^1(X, \mu)$. We may also discuss \mathcal{L}^p spaces for any $p \in \mathbb{N}$. We define

$$\mathcal{L}^p(X, \mu) = \{f : \int_X |f|^p d\mu < \infty\},$$

and we take the norm on \mathcal{L}^p to be

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

As in the case of \mathcal{L}^1 , we consider functions that are equivalent almost everywhere to be the same element in \mathcal{L}^p . This makes $\|f\|_p = 0$ if and only if $f = 0$ (almost everywhere), which is the first thing we need to check in checking that the claimed norm in fact makes \mathcal{L}^p into a normed vector space. The only nontrivial property that we need to check is the triangle inequality,

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

which is also known as Minkowski's inequality. This is actually a special case of Holder's inequality, which we state in the following proposition.

Proposition 9.3 (Holder's Inequality). *If $p, q \geq 1$ with $\frac{1}{p} + \frac{1}{q} = 1$ and $f \in \mathcal{L}^p(X, \mu)$ and $g \in \mathcal{L}^q(X, \mu)$, then $fg \in \mathcal{L}^1(X, \mu)$ and*

$$\|fg\|_1 = \int_X |fg| d\mu \leq \left(\int_X |f|^p d\mu \right)^{1/p} \left(\int_X |g|^q d\mu \right)^{1/q} = \|f\|_p \|g\|_q.$$

In the case when $p = q = 2$, Holder's inequality is just the usual Cauchy-Schwarz inequality:

Corollary 9.4 (Cauchy-Schwarz Inequality).

$$\left| \int fg d\mu \right| \leq \left(\int |fg| d\mu \right) \leq \left(\int |f|^2 d\mu \right)^{1/2} \left(\int |g|^2 d\mu \right)^{1/2}.$$

The proof of Holder's Inequality is left as a homework exercise.

Theorem 9.5 (Riesz-Fischer). \mathcal{L}^p is complete.

Proof (Sketch). Recall that a normed vector space is complete if and only if every absolutely summable series is summable. Hence we want to show that for a sequence $\{f_n\} \subseteq \mathcal{L}^p$, if $\sum \|f_n\|_p < \infty$, then the sequence of partial sums $F_N = \sum_{n=1}^N f_n$ converges in \mathcal{L}^p .

Let $g_k = \sum_{n=1}^k |f_n|$. (We sum $|f_n|$ pointwise to get g .) Then for all k ,

$$\|g_k\|_p \leq \sum_{n=1}^k \|f_n\|_p < C.$$

That is, $\|g_k\|_p$ is uniformly bounded by C . Therefore g_k^p is integrable, which means that g_k is finite almost everywhere. Notice that g_k is Cauchy in \mathcal{L}^p (we can just look at the tails) and $g_k \leq g_{k+1} \leq \dots$, and so $g_k(x)$ converges to a function g which is finite almost everywhere. This means that $F_N = \sum_{n=1}^N f_n$ converges absolutely pointwise almost everywhere. We can then use the dominated convergence theorem to show that F_N converges pointwise almost everywhere. This completes the proof. \square

For the sake of discussion, what happens to $\|f\|_p = (\int |f|^p d\mu)^{1/p}$ as $p \rightarrow \infty$? Well, if we had a finite list of numbers a_1, \dots, a_n and we looked at what happened to $(\sum a_n^p)^{1/p}$ as $p \rightarrow \infty$, the largest of the a_n would eventually dominate. Extending this argument, we have that $\|f\|_p$ tends to the *essential supremum* of $|f|$, which is the supremum of $|f|$ except on sets of measure zero. (This essential supremum is to account for the fact that we only care about the almost-everywhere behavior of a function.) The set \mathcal{L}^∞ is the set of all essentially bounded functions, and the norm $\|\cdot\|_\infty$ on this space is the essential supremum. One can check that this is a Banach space.

We will now focus our discussion on $p = 2$.

Chapter 10

Hilbert Spaces

The theory of \mathcal{L}^p -spaces in the case that $p = 2$ is worth discussing in particular because we have an extra structure, namely the inner product. For $f, g \in \mathcal{L}^2$, we define the inner product

$$\langle f, g \rangle := \int_X f \bar{g} d\mu,$$

and this makes \mathcal{L}^2 into an inner product space. Let us make this precise.

Definition 10.1. An *inner product space* is a complex vector space V with an inner product $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{C}$ satisfying

- (i) $\langle u_1 + u_2, v \rangle = \langle u_1, v \rangle + \langle u_2, v \rangle$ and $\langle u, v_1 + v_2 \rangle = \langle u, v_1 \rangle + \langle u, v_2 \rangle$.
- (ii) $\langle \lambda u, v \rangle = \lambda \langle u, v \rangle$ and $\langle u, \lambda v \rangle = \bar{\lambda} \langle u, v \rangle$, for $\lambda \in \mathbb{C}$.
- (iii) $\langle u, v \rangle = \overline{\langle v, u \rangle}$.
- (iv) $\langle u, u \rangle \geq 0$, with equality if and only if $u = 0$.

We can define a norm on V by $\|u\|^2 = \langle u, u \rangle$, and this induced norm makes V into a normed vector space.

Proposition 10.2 (Cauchy's Inequality). $|\langle u, v \rangle| \leq \|u\| \cdot \|v\|$.

Proof. Let $\lambda \in \mathbb{R}$. Then

$$\|\lambda u + v\|^2 = \langle \lambda u + v, \lambda u + v \rangle = \lambda^2 \|u\|^2 + 2\lambda \operatorname{Re} \langle u, v \rangle + \|v\|^2 \geq 0.$$

This means that the discriminant is ≤ 0 and hence

$$4|\langle u, v \rangle|^2 - 4\|u\|^2\|v\|^2 \leq 0,$$

and the desired result follows. □

So \mathcal{L}^2 is an inner product space and it is complete. Hence \mathcal{L}^2 is a Hilbert space. (In general, a *Hilbert space* is a Banach space with an inner product that induces the norm.) We will study Hilbert spaces in general and then later return to \mathcal{L}^2 .

Let H be a Hilbert space and let ϕ_1, ϕ_2, \dots be orthonormal vectors. That is, $\|\phi_n\|^2 = 1$ and $\langle \phi_m, \phi_n \rangle = 0$ if $m \neq n$. We first have some examples.

Example 10.3. 1. \mathbb{C}^N is a Hilbert space and example of an orthonormal basis is $(1, 0, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots$

2. $\ell^2 = \mathcal{L}^2(\mathbb{N}, \mu)$ where μ is the counting measure. This is the space of sequences $\{a_1, a_2, \dots\}$ such that $\sum_{n=1}^{\infty} |a_n|^2 < \infty$. Then the vectors $\phi_n = (0, 0, \dots, 0, 1, 0, \dots)$, where we have a 1 in the n th coordinate and 0's elsewhere, are orthonormal to each other.

The main question to motivate our study is the following: Can we write each $f \in H$ as a linear combination of the vectors ϕ_n ? We first have some definitions.

Definition 10.4. The n th Fourier coefficient of $f \in H$ is defined to be

$$c_n(f) := \langle f, \phi_n \rangle \in \mathbb{C}.$$

We also define a sort of “partial sum”:

$$S_N(f, \cdot) := \sum_{n=1}^N \langle f, \phi_n \rangle \cdot \phi_n = \sum_{n=1}^N c_n(f) \phi_n.$$

Definition 10.5. An orthonormal system is called *complete* if the only vector orthogonal to all ϕ_n is the zero vector.

Definition 10.6. A Hilbert space is called *separable* if a corresponding complete orthonormal system is countable.

Let $\{\phi_n\}$ be an orthonormal system (not necessarily complete). We have

$$\begin{aligned} 0 &\leq \|f - \sum_{n=1}^N c_n(f) \phi_n\|^2 = \langle f - \sum_{n=1}^N c_n(f) \phi_n, f - \sum_{n=1}^N c_n(f) \phi_n \rangle \\ &= \|f\|^2 - \sum_{n=1}^N \langle f, c_n(f) \phi_n \rangle - \sum_{n=1}^N \langle c_n(f) \phi_n, f \rangle + \|\sum_{n=1}^N c_n(f) \phi_n\|^2 \\ &= \|f\|^2 - 2 \sum_{n=1}^N |c_n(f)|^2 + \sum_{m,n=1}^N \langle c_m(f) \phi_m, c_n(f) \phi_n \rangle \\ &= \|f\|^2 - 2 \sum_{n=1}^N |c_n(f)|^2 + \sum_{n=1}^N |c_n(f)|^2 = \|f\|^2 - \sum_{n=1}^N |c_n(f)|^2. \end{aligned}$$

This proves Bessel's Inequality.

Proposition 10.7 (Bessel's Inequality). *If $f \in H$, then*

$$\|f\|^2 \geq \sum_{n=1}^{\infty} |c_n(f)|^2.$$

In particular, $c_n(f) \rightarrow 0$ as $n \rightarrow \infty$.

Given all this, what can we say about the convergence of the partial sums $S_N(f) = \sum_{n=1}^N c_n(f)\phi_n$? Does the sequence $\{S_N(f)\}$ converge to some element in H ? Does it converge to f ? In what sense? Since H is complete, it is enough to show that $\{S_N(f)\}$ is a Cauchy sequence. That is, we want to show that

$$\|S_n - S_m\|^2 \rightarrow 0, \text{ as } m, n \rightarrow \infty.$$

But $\sum |c_n(f)|^2 < \infty$, and hence $\|S_n - S_m\|^2 = \sum_{k=m+1}^n |c_k(f)|^2 \rightarrow 0$ as $m, n \rightarrow \infty$. So we have shown that $\{S_N(f)\}$ converges in the \mathcal{L}^2 sense. Note that this does not mean that $S_N = \sum_{n=1}^N c_n(f)\phi_n(x)$ converges for a given value of x .

Suppose now that $\{\phi_n\}$ is a complete orthonormal system and suppose that we know $S_N \rightarrow g$ for $g \in H$. Then $c_n(f) = \langle S_N, \phi_n \rangle \rightarrow \langle g, \phi_n \rangle$ as $N \rightarrow \infty$. More generally, if $u_n \rightarrow u$ and $v_n \rightarrow v$ in H , then $\langle u_n, v_n \rangle \rightarrow \langle u, v \rangle$ as $n \rightarrow \infty$, and this can be easily checked by multiple applications of the triangle inequality. Now, if g and f have the same Fourier coefficients, then all the Fourier coefficients of the difference $f - g$ are 0, which means that $f - g = 0$, where we make this last conclusion by using the assumption that the orthonormal system is complete. Hence in the case that $\{\phi_n\}$ is a complete orthonormal system, we have

$$\|f\|^2 = \sum_{n=1}^{\infty} |c_n(f)|^2.$$

In other words, Bessel's inequality becomes an equality.

From the above, we see that if H is a Hilbert space and $\{\phi_n\}$ is a complete orthonormal system, then we can identify $f \in H$ with the sequence of coefficients $c_n(f) = \langle f, \phi_n \rangle$. That is, we make the identification

$$f \rightarrow (c_1(f), c_2(f), \dots, c_n(f), \dots),$$

and this gives us an identification between the Hilbert space H and ℓ^2 . From this we may conclude that every separable Hilbert space either looks like ℓ^2 or like \mathbb{C}^N !

Chapter 11

Classical Fourier Series

We now move on to discuss a theory known as the classical Fourier series. We consider functions

$$f : [-\pi, \pi] \rightarrow \mathbb{C}.$$

(Alternatively, we may think of f as being complex-valued and 2π -periodic on \mathbb{R} .) Identifying $[-\pi, \pi]$ with the circle (or one-dimensional torus), we can equivalently think of these functions as

$$f : T \rightarrow \mathbb{C}, \text{ where } T = \mathbb{R}/2\pi\mathbb{Z}.$$

The space we would like to consider is

$$\mathcal{L}^2(T, \mu) = \{f : \int_{-\pi}^{\pi} |f|^2 d\mu < \infty\},$$

where μ is the Lebesgue measure. It will turn out later that this is in fact a Hilbert space. We have an inner product on this space:

$$\begin{aligned} \langle \cdot, \cdot \rangle : \mathcal{L}^2(T, \mu) \times \mathcal{L}^2(T, \mu) &\rightarrow \mathbb{C}, \\ (f, g) &\mapsto \int_{-\pi}^{\pi} f(x) \overline{g(x)} dx. \end{aligned}$$

We define an orthonormal basis as follows:

$$\phi_n(x) = \frac{e^{inx}}{\sqrt{2\pi}}, \text{ for } n \in \mathbb{Z},$$

and we may check that indeed $\langle \phi_n, \phi_n \rangle = \|\phi_n(x)\|^2 = 1$ and $\langle \phi_m, \phi_n \rangle = 0$ if $m \neq n$. To compute the n th Fourier coefficient of f , we need to compute the inner product:

$$c_n(f) = \langle f(x), e^{inx} \rangle = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) e^{-inx} dx =: \hat{f}(n).$$

We will use the book's notation in this exposition (even though it might be better notation to develop theory using functions $[0, 1] \rightarrow \mathbb{C}$ and to use the orthonormal basis $e^{2\pi i n x}$). The partial sums are

$$S_n(f, x) = \sum_{n=-N}^N \hat{f}(n) \frac{e^{inx}}{\sqrt{2\pi}}.$$

We know from our previous discussion that $S_N(f, \cdot) \rightarrow g$ in \mathcal{L}^2 . We would next like to know that $\{\phi_n\}$ is complete, which will allow us to say that in fact $S_N \rightarrow f$ in \mathcal{L}^2 , not just to any function g . This will also tell us that $\mathcal{L}^2(T, \mu)$ is a Hilbert space.

Theorem 11.1. *ϕ_n is a complete orthonormal system for $\mathcal{L}^2(T, \mu)$.*

We will in fact prove a stronger result, which is due to Fejer.

Theorem 11.2 (Fejer). *If f is continuous on T , then there is a sequence of functions*

$$T_N(x) = \sum_{n=-N}^N t_n e^{inx}$$

with $T_N(x) \rightarrow f(x)$ uniformly for all $x \in T$ as $N \rightarrow \infty$. In particular, $\|T_N - f\| \rightarrow 0$.

Remark 11.3. Since T is compact, then the hypothesis that f is continuous in fact means that f is uniformly continuous, i.e. for every $\varepsilon > 0$ there exists a δ such that $|f(x) - f(y)| < \varepsilon$ whenever $|x - y| < \delta$.

Proof. We expand out, by hand, the partial sums:

$$\begin{aligned} S_n(f, x) &= \sum_{k=-n}^n \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx \right) \frac{e^{ikx}}{\sqrt{2\pi}} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) \left(\sum_{k=-n}^n e^{ik(x_0-x)} \right) dx \\ &= \frac{1}{2\pi} \int_{x_0+\pi}^{x_0-\pi} f(x_0-y) D_n(y) d(x_0-y) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x_0-t) D_n(t) dt, \end{aligned}$$

where $D_n(x_0 - x) = \sum_{k=-n}^n e^{ik(x_0-x)}$. We have a general form for this, and it is called the Dirichlet kernel. Expanding out the Dirichlet kernel expression gives us:

$$\begin{aligned} D_n(t) &= \sum_{k=-n}^n e^{ikt} = e^{-int} (1 + e^{it} + \cdots + e^{2int}) \\ &= \begin{cases} \frac{e^{-int}(e^{(2n+1)it} - 1)}{e^{it} - 1} & \text{if } t \neq 0 \\ 2n + 1 & \text{if } t = 0. \end{cases} \end{aligned}$$

So when $t \neq 0$, then

$$D_n(t) = \frac{e^{(n+\frac{1}{2})it} - e^{-(n+\frac{1}{2})it}}{e^{it/2} - e^{-it/2}} = \frac{\sin((n+\frac{1}{2})t)}{\sin(\frac{1}{2}t)}.$$

In particular,

$$D_n(t) \leq \frac{1}{|\sin(t/2)|} \text{ for all } t \neq 0.$$

Notice that as $n \rightarrow \infty$, $D_n(t)$ tends to infinity at $t = 0$ and is finite at $t \neq 0$. Also, $\int_{-\pi}^{\pi} D_n(t) dt = 2\pi$ and

$$S_n(f, x_0) - f(x_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(x_0 - t) - f(x_0)) D_n(t) dt.$$

But we have a problem:

$$\int_{-\pi}^{\pi} |D_n(t)| dt \sim c \log(n),$$

which means that as $n \rightarrow \infty$, this integral tends to infinity, and so we can't say anything about the difference $S_n(f, x_0) - f(x_0)$. This is where Fejer's argument comes in.

Instead of looking at the convergence of S_n , we want to sort of average over these S_n 's and look at the convergence of the result. Let

$$\sigma_N(f, x_0) = \frac{1}{N+1} \sum_{n=0}^N S_n(f, x_0).$$

Then we have

$$\begin{aligned} \sigma_N(f, x_0) &= \frac{1}{N+1} \sum_{n=0}^N \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x_0 - t) D_n(t) dt \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x_0 - t) \kappa_N(t) dt, \end{aligned}$$

where

$$\kappa_N(t) = \frac{1}{N+1} \sum_{n=0}^N D_n(t).$$

This is called the Fejer kernel, and we can expand this by hand:

$$\begin{aligned} \kappa_N(t) &= \frac{1}{N+1} \sum_{n=0}^N D_n(t) = \frac{1}{N+1} \sum_{n=0}^N \frac{e^{-int}(e^{2(n+1)it} - 1)}{e^{it} - 1} \\ &= \frac{1}{(N+1)(e^{it} - 1)} \sum_{n=0}^N (e^{i(n+1)t} - e^{-int}) \quad (*) \\ &= \frac{1}{N+1} \sum_{n=0}^N \sum_{|k| \leq n} e^{ikt} = \frac{1}{N+1} \sum_{|k| \leq N} e^{ikt} \left(\sum_{N \geq n \geq |k|} 1 \right) \\ &= \frac{1}{N+1} \sum_{|k| \leq N} (N+1 - |k|) e^{ikt}. \end{aligned}$$

Summing (*) directly, we have

$$\begin{aligned}
& \frac{1}{(N+1)(e^{it} - 1)} \left(\frac{e^{it}(e^{i(N+1)t} - 1)}{e^{it} - 1} - \frac{1 - e^{-i(N+1)t}}{1 - e^{-it}} \right) \\
&= \frac{1}{(N+1)(e^{it} - 1)(1 - e^{-it})} (e^{i(N+1)t} - 2 + e^{-i(N+1)t}) \\
&= \frac{1}{N+1} \left(\frac{e^{i(N+1)t/2} - e^{-i(N+1)t/2}}{e^{it/2} - e^{-it/2}} \right)^2 \\
&= \frac{1}{N+1} \left(\frac{\sin\left(\frac{(N+1)t}{2}\right)}{e^{it/2} - e^{-it/2}} \right)^2.
\end{aligned}$$

Notice that

$$\int_{-\pi}^{\pi} \kappa_n(t) dt = 2\pi,$$

and that for any $\delta > 0$,

$$\int_{\pi \geq |t| \geq \delta} \kappa_n(t) dt \leq \frac{1}{N+1} \int_{\pi \geq |t| \geq \delta} \frac{1}{(\sin(t/2))^2} dt =: \frac{1}{N+1} c(\delta).$$

From this, we have

$$\sigma_N(f, x_0) - f(x_0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (f(x_0 - t) - f(x_0)) \kappa_N(t) dt.$$

If $|t| \leq \delta$, then

$$\sigma_N(f, x_0) - f(x_0) \leq \frac{\varepsilon}{2\pi} \int_{|t| \leq \delta} \kappa_N(t) dt \leq \varepsilon,$$

and if $|t| > \delta$, then

$$\sigma_N(f, x_0) - f(x_0) \leq 2 \max |f| \frac{c(\delta)}{2\pi(N+1)}.$$

Therefore we have shown that for sufficiently large N , $\sigma_N(f, x_0) - f(x_0) \leq 2\varepsilon$, and this proves Fejer's theorem. \square

Remark 11.4. The main idea in the proof of this theorem is that we can improve convergence behavior by taking the sequence of averages. This is a useful tool in other proofs as well.

From this, we know that if $[a, b] \subseteq [-\pi, \pi]$, then the characteristic function $\chi_{[a, b]}$ can be approximated by a suitable trigonometric polynomial $T_N = \sum_{n=-N}^N t_n e^{-inx}$. So we may construct a sequence $\{T_N\}$ such that $\|\chi - T_N\| \rightarrow 0$ as $N \rightarrow \infty$. So the characteristic function of a finite union of intervals can be approximated by trigonometric polynomials.

Now let $f \in \mathcal{L}^2$ be a function that is orthogonal to all the $\phi_k = \frac{e^{ikx}}{\sqrt{2\pi}}$. We would like to show that $f = 0$ almost everywhere. Let $A \subseteq (-\pi, \pi)$ be the set on which $f \geq 0$. Then A is measurable and has finite measure. We can find a trigonometric polynomial T_N such that $\|T_N - \chi_A\| < \varepsilon$. Then necessarily

$$\langle f, T_N \rangle = 0$$

by assumption. On the other hand, we have

$$\langle f, T_N \rangle = \langle f, \chi_A \rangle + \langle f, T_N - \chi_A \rangle \leq \langle f, \chi_A \rangle + \|f\| \cdot \|T_N - \chi_A\|,$$

which means that $\langle f, \chi_A \rangle = 0$, and hence $\int_A f d\mu = 0$, and so $f = 0$ almost everywhere on A . This shows that ϕ_n is indeed a complete orthonormal system.

Proposition 11.5 (Parseval/Plancharel).

$$\|f\|^2 = \int_{-\pi}^{\pi} |f(x)|^2 dx = \sum_{k=-\infty}^{\infty} \left| \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx \right|^2 = \sum_{k=-\infty}^{\infty} |c_n(f)|^2.$$

Using Plancharel's theorem, we can prove identities like $\sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{\pi^2}{6}$. This completes our discussion of classical Fourier analysis. The final topic of the course will be Fourier transforms.

1,300