Measure Theory

Notes for Pure Math 451

Kenneth R. Davidson

University of Waterloo

CONTENTS

CHAPTER 1

Measures

1.1. Introduction

The idea of Riemann integration is the following: Convergence of the upper and lower sums to the same limit is required for Riemann integrability. The integral

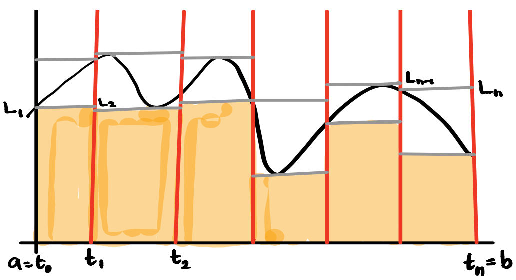

FIGURE 1.1. Riemann Integral

is approximated by Riemann sums $\int_a^b f(x)\,dx \approx \sum_{i=1}^{n} L_i(t_i - t_{i-1})$. The function $f(x)$ has to be 'almost continuous' in order for this to succeed. It is too restrictive. Moreover there are no good limit theorems. (Again one needs uniform convergence for most results,)

Lebesgue's approach was to chop up the range instead. Then one gets the

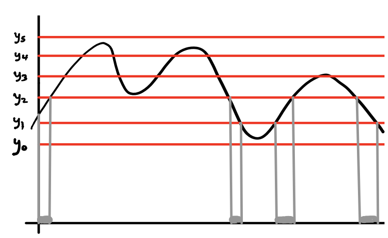

FIGURE 1.2. Lebesgue integral

approximations $\int_a^b f(x)\,dx \approx \sum_{i=1}^{n} y_i \big|\{x : y_{i-1} < f(x) \leq y_i\}\big|$. The notation

$|A|$ is meant to convey the *length* of a set A. The complication is that, even for continuous functions, the sets $\{x : y_{i-1} < f(x) \leq y_i\}$ are neither open nor closed, and can be complicated. It is a Borel set, and that will be seen to be good enough. Lebesgue's idea was to extend the notion of length or measure from open intervals to a function m on a much larger class of sets so that

(1) $m((a,b)) = b - a$

(2) $m(A + x) = m(A)$ for sets A and $x \in \mathbb{R}$ (translation invariance)

(3) if A is the disjoint union $A = \dot{\bigcup}_{n=1}^{\infty} A_n$, then $m(A) = \sum_{n=1}^{\infty} m(A_n)$. (countable additivity)

It turns out that it is not possible to define such a function on all subsets of $\mathbb{R}$. To see this, we define an equivalence relation on $[0,1)$ by $x \sim y$ if $x - y \in \mathbb{Q}$. Then use the Axiom of Choice to select a set A which contains exactly one element from each equivalence class. Enumerate $\mathbb{Q} \cap [0,1) = \{0 = r_0, r_1, r_2, \dots\}$ and define

$$A_n = A + r_n (\text{mod } 1) = \big((A + r_n) \cup (A + r_n - 1)\big) \cap [0,1) \quad \text{for} \quad n \geq 0.$$

Now by translation invariance and finite additivity,

$$\begin{aligned} m(A) &= m(A + r_n) \\ &= m\big((A + r_n) \cap [0,1)\big) + m\big((A + r_n) \cap [1,2)\big) \\ &= m\big((A + r_n) \cap [0,1)\big) + m\big((A + r_n - 1) \cap [0,1)\big) \\ &= m(A_n). \end{aligned}$$

Observe that $A_m \cap A_n = \varnothing$ if $m \neq n$. By construction, $\bigcup_{n \geq 1} A_n = [0,1)$. Hence by countable additivity,

$$1 = m([0,1)) = \sum_{n=0}^{\infty} m(A_n) = \sum_{n=0}^{\infty} m(A).$$

There is no value that can be assigned to $m(A)$ to make sense of this.

The way we deal with this is to declare that the set A is not measureable. We will only assign a value to $m(A)$ for a class of 'nice' sets.

1.1.1. Remark. Banach showed that it is possible to define a finitely additive, translation invariant function on all subsets of $\mathbb{R}$. However in $\mathbb{R}^n$ for $n \geq 3$, even this is not possible. The Banach-Tarski paradox shows that if A and B are two bounded subsets of $\mathbb{R}^3$ with interior, then there is a finite decomposition of each: $A = \dot{\bigcup}_{i=1}^{n} A_i$ and $B = \dot{\bigcup}_{i=1}^{n} B_i$ and rigid motions (a combination of a rotation and a translation) which carry A_i onto B_i. For example let A be the sphere of radius 1 and let B be the sphere of radius 10^{10}. Clearly this does not preserve volume, in spite of what common sense suggests. This decomposition requires the Axiom of Choice, and can't be accomplished by hand.

1.2. σ-Algebras

1.2.1. DEFINITION. If X is a set, an *algebra of subsets* is a non-empty collection of subsets of X, i.e., $\mathcal{A} \subset \mathcal{P}(X)$, which contains the empty set, $\varnothing$, and is closed under complements and finite unions. A *σ-algebra* is an algebra $\mathcal{B}$ of subsets of X which is closed under countable unions. We say that $(X, \mathcal{B})$ is a *measure space*.

1.2.2 SIMPLE PROPERTIES OF σ-ALGEBRAS. Let $\mathcal{A}$ be an algebra of subsets $\mathcal{A} \subset \mathcal{P}(X)$; and let $\mathcal{B}$ be a σ-algebra of subsets of X.

(1) $\mathcal{A}$ is non-empty, so suppose that $E \in \mathcal{A}$. Then E^c, $\varnothing$ and $X = \varnothing^c$ all belong to $\mathcal{A}$.

(2) If $E, F \in \mathcal{A}$, then $E \cup F$, $E \cap F = (E^c \cup F^c)^c$, $E \setminus F = E \cap F^c$ and $E \triangle F = (E \setminus F) \cup (F \setminus E)$ belong to $\mathcal{A}$.

(3) If $E_n \in \mathcal{B}$, then $\bigcap_{n\geq 1} E_n = \left(\bigcup_{n\geq 1} E_n^c\right)^c \in \mathcal{B}$.

(4) If $E_n \in \mathcal{B}$, then $\bigcup_{n\geq 1} E_n = \dot{\bigcup}_{n\geq 1} F_n$ where $F_n = E_n \setminus \bigcup_{i=1}^{n-1} E_i$ are disjoint elements of $\mathcal{B}$.

(5) If $\mathcal{B}_\lambda$ are σ-algebras for $\lambda \in \Lambda$, then $\mathcal{B} = \bigcap_{\lambda\in\Lambda} \mathcal{B}_\lambda$ is a σ-algebra. So if $\mathcal{E} \subset \mathcal{P}(X)$ is any collection of subsets, the intersection of all σ-algebras containing $\mathcal{E}$ is the unique smallest σ-algebra containing $\mathcal{E}$. This is called the σ-algebra generated by $\mathcal{E}$.

1.2.3. DEFINITION. If X is a topological space, the *Borel sets* are the elements of the σ-algebra Bor_X or $\mathrm{Bor}(X)$ generated by the collection of open subsets of X.

A *G_δ set* is a countable intersection of open sets. An *F_σ set* is a countable union of closed sets.

1.2.4. DEFINITION. A *measure* on $(X, \mathcal{B})$ is a map $\mu : \mathcal{B} \to [0, \infty) \cup \{\infty\} =: [0, \infty]$ such that $\mu(\varnothing) = 0$ and is *countably additive*: if $A = \dot{\bigcup}_{n\geq 1} A_n$ where $A_n \cap A_m = \varnothing$ if $m \neq n$, then $\mu(A) = \sum_{i\geq 1} \mu(A_i)$.

A measure μ if *finite* if $\mu(X) < \infty$ and *σ-finite* if $X = \bigcup_{n\geq 1} E_n$ such that $\mu(E_n) < \infty$ for $n \geq 1$. It is called a *probability measure* if $\mu(X) = 1$. A measure μ is *semi-finite* if for every $F \in \mathcal{B}$ with $\mu(F) \neq 0$, there is an $E \in \mathcal{B}$ with $E \subset F$ and $0 < \mu(E) < \infty$.

1.2.5 SIMPLE PROPERTIES OF MEASURES. Let μ be a measure on $(X, \mathcal{B})$.

(1) (*monotonicity*) If $E, F \in \mathcal{B}$ with $E \subset F$, then

$$\mu(F) = \mu(E) + \mu(F \setminus E) \geq \mu(E).$$

(2) (*subadditivity*) If $E_n \in \mathcal{B}$, write $\bigcup_{n\geq 1} E_n = \dot{\bigcup}_{n\geq 1} F_n$ where $F_n \subset E_n$. Then
$$\mu(E) = \sum_{n\geq 1} \mu(F_n) \leq \sum_{n\geq 1} \mu(E_n).$$

(3) (*continuity from below*) If $E_n \in \mathcal{B}$ with $E_n \subset E_{n+1}$ for $n \geq 1$, then setting $E_0 = \varnothing$,
$$\begin{aligned}\mu\big(\bigcup_{n\geq 1} E_n\big) &= \mu\big(\dot{\bigcup}_{n\geq 1} E_n \setminus E_{n-1}\big) = \sum_{n\geq 1} \mu(E_n \setminus E_{n-1}) \\ &= \lim_{n\to\infty} \sum_{i=1}^{n} \mu(E_i \setminus E_{i-1}) = \lim_{n\to\infty} \mu(E_n).\end{aligned}$$

(4) (*continuity from above*) If $E_n \supset E_{n+1}$ for $n \geq 0$ and $\mu(E_0) < \infty$, then
$$\mu\big(\bigcap_{n\geq 1} E_n\big) = \lim_{n\to\infty} \mu(E_n).$$

1.2.6 Simple Examples of Measures. Let μ be a measure on $(X, \mathcal{B})$.

(1) *Counting measure* on $(X, \mathcal{P}(X))$ is $\mu_c(E) = \begin{cases} |E| & \text{if } E \text{ is finite} \\ \infty & \text{if } E \text{ is infinite.}\end{cases}$

This measure is semi-finite. It is σ-finite if and only if X is countable.

(2) *point mass* on $(X, \mathcal{P}(X))$ for $x \in X$ is $\delta_x(E) = \begin{cases} 1 & \text{if } x \in E \\ 0 & \text{if } x \notin E.\end{cases}$

This is a probability measure.

(3) Define μ on $(X, \mathcal{P}(X))$ by $\mu(E) = \begin{cases} 0 & \text{if } E = \varnothing \\ \infty & \text{if } E \neq \varnothing\end{cases}$. This is a valid measure, but it is essentially useless. Note that it is not even semifinite.

1.2.7. Definition. A measure μ on $(X, \mathcal{B})$ is *complete* if $E \in \mathcal{B}$, $\mu(E) = 0$ and $F \subset E$ implies that $F \in \mathcal{B}$.

1.2.8. Theorem. *If $(X, \mathcal{B}, \mu)$ is a measure, then*
$$\overline{\mathcal{B}} = \{E \cup F : E, N \in \mathcal{B},\ F \subset N,\ \mu(N) = 0\}$$
is a σ-algebra and $\bar{\mu}(E \cup F) := \mu(E)$ is a complete measure on $(X, \overline{\mathcal{B}})$ such that $\bar{\mu}|_{\mathcal{B}} = \mu$.. Moreover this is the smallest σ-algebra containing $\mathcal{B}$ on which μ extends to a complete measure.

Proof. Note that $(E \cup F)^c = (E \cup N)^c \cup (N \setminus (E \cup F))$. So $\overline{\mathcal{B}}$ is closed under complements. Also if $E_i \cup F_i \in \overline{\mathcal{B}}$ and $F_i \subset N_i$ where $N_i \in \mathcal{B}$ and $\mu(N_i) = 0$, then

$$\bigcup_{i\geq 1} E_i \cup F_i = \bigcup_{i\geq 1} E_i \cup \bigcup_{i\geq 1} F_i = E \cup F$$

where $F = \bigcup_{i\geq 1} F_i \subset \bigcup_{i\geq 1} N_i =: N$, $N \in \mathcal{B}$ and $\mu(N) = 0$ by subadditivity. So $\overline{\mathcal{B}}$ is a σ-algebra. Moreover this is clearly the smallest σ-algebra containing $\mathcal{B}$ and all subsets of null sets.

Next we show that $\bar{\mu}$ is well-defined. Suppose that $A = E_1 \cup F_1 = E_2 \cup F_2$ where $E_i \in \mathcal{B}$, $F_i \subset N_i$ and $\mu(N_i) = 0$. Let $E = E_1 \cap E_2 \in \mathcal{B}$. Then

$$E \subset E_i \subset A \subset (E_1 \cup N_1) \cap (E_2 \cup N_2) \subset (E_1 \cap E_2) \cup N_1 \cup N_2.$$

Therefore $\mu(E) \leq \mu(E_i) \leq \mu(E) + \mu(N_1) + \mu(N_2) = \mu(E)$. So $\mu(E_1) = \mu(E_2)$. Thus $\bar{\mu}(A) = \mu(E_i)$ is well-defined. In particular, if $A \in \mathcal{B}$, then $\bar{\mu}(A) = \mu(A)$. So $\bar{\mu}|_{\mathcal{B}} = \mu$.

To see that $\bar{\mu}$ is countably additive, suppose that $A_i = E_i \cup F_i$ are disjoint, and $F_i \subset N_i$ where N_i are null sets. Then $A := \dot{\bigcup}_{i\geq 1} A_i = \big(\dot{\bigcup}_{i\geq 1} E_i\big) \cup \big(\dot{\bigcup}_{i\geq 1} F_i\big)$. Moreover $F = \bigcup_{i\geq 1} F_i \subset \bigcup_{i\geq 1} N_i =: N$, and $\mu(N) = 0$ by subadditivity. So

$$\bar{\mu}(A) = \mu\big(\dot{\bigcup}_{i\geq 1} E_i\big) = \sum_{i\geq 1} \mu(E_i) = \sum_{i\geq 1} \bar{\mu}(A_i).$$

Thus $\bar{\mu}$ is a measure.

If $E \in \mathcal{B}$, then $\bar{\mu}(E) = \mu(E)$, so $\bar{\mu}|_{\mathcal{B}} = \mu$; i.e. $\bar{\mu}$ extends the definition of μ. To see that $\bar{\mu}$ is complete, suppose that $\bar{\mu}(M) = 0$ and $G \subset M$. Then $M = E \cup F$ where $E, N \in \mathcal{B}$, $F \subset N$ and $\mu(N) = 0$. Also $\mu(E) = \bar{\mu}(M) = 0$. Thus $G \subset M \subset E \cup N$, and $\mu(E \cup N) = 0$. Hence $G \in \overline{\mathcal{B}}$. Thus $\bar{\mu}$ is a complete measure.

Clearly $\overline{\mathcal{B}}$ is the smallest σ-algebra containing $\mathcal{B}$ and all subsets of null sets. Moreover it is clear that the only way to extend the definition of μ to $\overline{\mathcal{B}}$ and be a measure is to set $\bar{\mu}(F) = 0$ when $F \subset N$ and $\mu(N) = 0$. Thus, this is the unique smallest complete measure extending μ. ■

1.2.9. DEFINITION. If μ is a measure on $(X, \mathcal{B})$, then $(X, \overline{\mathcal{B}}, \bar{\mu})$ is called the *completion* of μ.

1.2.10. DEFINITION. Let μ be a measure on $(X, \mathcal{B})$. A property about points in X is true *μ-almost everywhere* if it is true except on a set of measure 0 (a *null set*). Write a.e.(μ) or just a.e. if the measure is understood.

For example, the statement about $f : X \to \mathbb{R}$ saying that $f = 0$ a.e.(μ) means that there is a μ-null set N so that $f(x) = 0$ for $x \in X \setminus N$. It does not say that $\{x : f(x) \neq 0\}$ is measurable, only that it is contained in the set N of measure 0. However if μ is a complete measure, then all subsets of null sets are measurable. So in this case, it is true that $\{x : f(x) \neq 0\}$ is measurable.

1.3. Construction of Measures

We need to develop some machinery so that we can define more interesting measures such as Lebesgue measure on the real line. Indeed, the construction of a family of measures related to Lebesgue measure will be our main application of this machinery at this time.

1.3.1. DEFINITION. Let X be a non-empty set. An *outer measure* on X is a map $\mu^* : \mathcal{P}(X) \to [0, \infty]$ such that

(1) $\mu^*(\varnothing) = 0$.

(2) (monotonicity) if $A \subset B$, then $\mu^*(A) \leq \mu^*(B)$.

(3) (sub-additivity) if $\{A_i : i \geq 1\}$ is a countable collection of sets, then $\mu^*\big(\bigcup_{i\geq 1} A_i\big) \leq \sum_{i\geq 1} \mu^*(A_i)$.

In other words, μ^* is a monotone, subadditive function on subsets of X. The following easy proposition shows that outer measures can easily be produced.

1.3.2. PROPOSITION. *Suppose that $\{\varnothing, X\} \subset \mathcal{E} \subset \mathcal{P}(X)$ and $\mu : \mathcal{E} \to [0, \infty]$ is a function with $\mu(\varnothing) = 0$. For $A \in \mathcal{P}(X)$, define*

$$\mu^*(A) = \inf\Big\{ \sum_{i\geq 1} \mu(E_i) : E_i \in \mathcal{E},\ A \subset \bigcup_{i\geq 1} E_i \Big\}.$$

Then μ^* is an outer measure.

PROOF. Note that $\mu^*(\varnothing) = 0$ by definition, and that monotonicity is immediate. To establish sub-additivity, suppose that $\{A_i : i \geq 1\} \subset \mathcal{P}(X)$. There is nothing to prove unless $\sum_{i\geq 1} \mu^*(A_i) < \infty$. In this case, let $\varepsilon > 0$. For each i, find sets $E_{ij} \in \mathcal{E}$ for $j \geq 1$ so that $A_i \subset \bigcup_{j\geq 1} E_{ij}$ and $\sum_{j\geq 1} \mu(E_{ij}) < \mu^*(A_i) + 2^{-i}\varepsilon$. Then $A = \bigcup_{i\geq 1} A_i$ is covered by $\bigcup_{i\geq 1} \bigcup_{j\geq 1} E_{ij}$ and so

$$\mu^*(A) \leq \sum_{i\geq 1} \sum_{j\geq 1} \mu(E_{ij}) < \sum_{i\geq 1} \big(\mu^*(A_i) + 2^{-i}\varepsilon\big) = \sum_{i\geq 1} \mu^*(A_i) \ + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, this establishes the claim. So μ^* is an outer measure. ■

1.3.3. EXAMPLE. Let $X = \mathbb{R}$ and let $\mathcal{E}$ denote the collection of all bounded open intervals. Define $\rho((a, b)) = b - a$. This determines an outer measure which will be used to construct Lebesgue measure.

1.3.4. DEFINITION. Let μ^* be an outer measure on X. A subset $A \subset X$ is *μ^*-measurable* if

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \quad \text{for all} \quad E \subset X.$$

Note that subadditivity shows that $\mu^*(E) \le \mu^*(E \cap A) + \mu^*(E \cap A^c)$. So we only need to show that $\mu^*(E) \ge \mu^*(E \cap A) + \mu^*(E \cap A^c)$ when $\mu^*(E) < \infty$. We are selecting those sets A for which this is always additive.

The main result about outer measures is the following important result.

1.3.5. CARATHÉODORY'S THEOREM. *Let μ^* be an outer measure on X. Then the collection $\mathcal{B}$ of all μ^*-measurable sets is a σ-algebra, and $\mu = \mu^*|_{\mathcal{B}}$ is a complete measure.*

PROOF. It is clear from the definition that if A is measurable, then so is A^c. Suppose that $A, B \in \mathcal{B}$ and let $E \subset X$. Then

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap A) + \mu^*(E \cap A^c)\\ &= \mu^*(E \cap A) + \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c)\\ &\ge \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c).\end{aligned}$$

The first line follows since A is μ^*-measurable. The second line follows since B is μ^*-measurable. The last line follows from subadditivity of μ^*. This is the non-trivial direction, so this is an equality. Thus $A \cup B$ is μ^*-measurable.

Combining these two observations shows that $\mathcal{B}$ is an algebra of sets, and thus is closed under finite unions and intersections.

Next suppose that A and B are disjoint. Then

$$\begin{aligned}\mu^*(E \cap (A \dot\cup B)) &= \mu^*(E \cap (A \dot\cup B) \cap A) + \mu^*(E \cap (A \dot\cup B) \cap A^c)\\ &= \mu^*(E \cap A) + \mu^*(E \cap B).\end{aligned}$$

By induction, we see that if $A_1, \dots, A_n$ are pairwise disjoint μ^*-measurable sets, then for any $E \subset X$,

$$\mu^*(E \cap \dot{\bigcup}_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu^*(E \cap A_i).$$

Now we consider countable unions. Let $A_i \in \mathcal{B}$ for $i \ge 1$. Set $B_n = \bigcup_{i=1}^n A_i$ and $B = \bigcup_{i\ge 1} A_i$. Set $A_i' = A_i \backslash B_{i-1}$ so that $B_n = \dot{\bigcup}_{i=1}^n A_i'$ and $B = \dot{\bigcup}_{i\ge 1} A_i'$. We know that $B_n, A_n' \in \mathcal{B}$. Take any $E \subset X$. Using the additivity from the previous paragraph,

$$\begin{aligned}\mu^*(E) &= \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c)\\ &\ge \mu^*(E \cap B_n) + \mu^*(E \cap B^c)\\ &= \sum_{i=1}^{n} \mu^*(E \cap A_i') + \mu^*(E \cap B^c).\end{aligned}$$

This is true for all $n \geq 1$ so we can take limits and get

$$\begin{aligned}\mu^*(E) &\geq \sum_{i\geq 1} \mu^*(E \cap A_i') + \mu^*(E \cap B^c) \\ &\geq \mu^*(E \cap B) + \mu^*(E \cap B^c) \\ &\geq \mu^*(E)\end{aligned}$$

The last two lines use subadditivity. When $\mu^*(E) < \infty$, we obtain equality, and thus B is μ^*-measurable. Therefore $\mathcal{B}$ is a σ-algebra.

Next suppose that the A_i are pairwise disjoint. So $A_i' = A_i$ in the previous paragraph. Taking $E = B$, we get that

$$\mu(B) \geq \sum_{i\geq 1} \mu(A_i) \geq \mu(B).$$

Therefore μ is countably additive on $\mathcal{B}$.

Finally suppose that $\mu^*(A) = 0$. Then for $E \subset X$,

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

Hence this is an equality, showing that A is μ^*-measurable and $\mu(A) = 0$. Any subset $F \subset A$ also has $\mu^*(F) = 0$, and thus F is μ^*-measurable. Therefore $(X, \mathcal{B}, \mu)$ is a complete measure. ∎

1.4. Premeasures

1.4.1. DEFINITION. A *premeasure* is a function $\mu : \mathcal{A} \to [0, \infty]$ defined on an algebra $\mathcal{A} \subset \mathcal{P}(X)$ of sets such that $\mu(\varnothing) = 0$ and whenever $A_i \in \mathcal{A}$ are pairwise disjoint and $A = \dot{\bigcup}_{i\geq 1} A_i \in \mathcal{A}$, then $\mu(A) = \sum_{i\geq 1} \mu(A_i)$. In particular, premeasures are additive: $\mu(A_1 \dot\cup A_2) = \mu(A_1) + \mu(A_2)$ if $A_1, A_2 \in \mathcal{A}$ are disjoint.

A premeasure is an improvement on the arbitrary function used in Proposition 1.3.2 to define an outer measure. In that earlier construction, the outer measure need not reflect much about the original function μ. However in the case of a premeasure, the Carathéodory construction yields a measure that extends the premeasure.

1.4.2. THEOREM. *If μ is a premeasure on an algebra $\mathcal{A} \subset \mathcal{P}(X)$, then applying Carathéodory's Theorem to the outer measure μ^* yields a complete measure $(X, \mathcal{B}, \bar{\mu})$ such that $\mathcal{B} \supset \mathcal{A}$ and $\bar{\mu}|_{\mathcal{A}} = \mu$.*

PROOF. By Proposition 1.3.2, μ^* is an outer measure. So an application of Carathéodory's Theorem 1.3.5, there is a complete measure $(X, \mathcal{B}, \bar{\mu})$ defined on the σ-algebra $\mathcal{B}$ of μ^*-measurable sets.

Suppose that $A \in \mathcal{A}$ and $A_i \in \mathcal{A}$ such that $A \subset \bigcup_{i\geq 1} A_i$. Define

$$B_i = A \cap A_i \setminus \bigcup_{j=1}^{i-1} A_j \quad \text{for} \quad i \geq 1,$$

which belong to $\mathcal{A}$ since it is an algebra. By construction, the B_i are pairwise disjoint and $A = \dot{\bigcup}_{i\geq 1} B_i$. Therefore, since μ is a premeasure,

$$\mu(A) = \sum_{i\geq 1} \mu(B_i) \leq \sum_{i\geq 1} \mu(A_i).$$

Therefore

$$\mu^*(A) = \inf\Big\{ \sum_{i\geq 1} \mu(A_i) : A \subset \bigcup_{i\geq 1} A_i \Big\} = \mu(A).$$

Now we show that A is μ^*-measurable. Let $E \subset X$ with $\mu^*(E) < \infty$, and let $\varepsilon > 0$. We can find $A_i \in \mathcal{A}$ so that $E \subset \bigcup_{i\geq 1} A_i$ and $\sum_{i\geq 1} \mu(A_i) < \mu^*(E) + \varepsilon$. Notice that $E \cap A \subset \bigcup_{i\geq 1} A_i \cap A$ and $E \cap A^c \subset \bigcup_{i\geq 1} A_i \cap A^c$. Therefore by the additivity of μ,

$$\begin{aligned} \mu^*(E \cap A) + \mu^*(E \cap A^c) &\leq \sum_{i\geq 1} \mu(A_i \cap A) + \sum_{i\geq 1} \mu(A_i \cap A^c) \\ &= \sum_{i\geq 1} \mu(A_i) < \mu^*(E) + \varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, $\mu^*(E \cap A) + \mu^*(E \cap A^c) \leq \mu^*(E)$, which is the non-trivial inequality; so this is an equality. Hence A is μ^*-measurable. We conclude that $\mathcal{A} \subset \mathcal{B}$ and for $A \in \mathcal{A}$, $\bar{\mu}(A) = \mu^*(A) = \mu(A)$. So $\bar{\mu}|_{\mathcal{A}} = \mu$. ■

Here is further detail about the outer measure construction which explains when the extension is unique.

1.4.3. Proposition. *Let μ be a premeasure on an algebra $\mathcal{A} \subset \mathcal{P}(X)$ and let $(X, \mathcal{B}, \bar{\mu})$ be the measure of Theorem 1.4.2. Let ν be any measure on a σ-algebra $\mathcal{C}$ satisfying $\mathcal{A} \subset \mathcal{C} \subset \overline{\mathcal{B}}$ such that $\nu|_{\mathcal{A}} = \mu$. Then $\nu(E) \leq \bar{\mu}(E)$ for all $E \in \mathcal{C}$, with equality if $\bar{\mu}(E) < \infty$. Moreover, if E is σ-finite, then $\nu(E) = \bar{\mu}(E)$. So if μ is σ-finite, then $\bar{\mu}|_{\mathcal{C}}$ is the unique extension of μ to a measure on $\mathcal{C}$.*

Proof. Let $E \in \mathcal{C}$. If $E \subset \bigcup_{i\geq 1} A_i$ for $A_i \in \mathcal{A}$, then by subadditivity,

$$\nu(E) \leq \sum_{i\geq 1} \nu(A_i) = \sum_{i\geq 1} \mu(A_i).$$

Now take the inf over all such covers of E to obtain $\nu(E) \leq \bar{\mu}(E)$.

If $\bar{\mu}(E) < \infty$ and $\varepsilon > 0$, we can choose the A_i so that $\sum_{i\geq 1} \mu(A_i) < \bar{\mu}(E)+\varepsilon$. Let $B_i = A_i \setminus \bigcup_{j=1}^{i-1} A_j$, which are in $\mathcal{A}$; and $A = \bigcup_{i\geq 1} A_i = \dot{\bigcup}_{i\geq 1} B_i$. Then

$$\nu(A) = \sum_{i\geq 1} \nu(B_i) = \sum_{i\geq 1} \mu(B_i) = \bar{\mu}(A).$$

Therefore

$$\nu(E) + \nu(A \setminus E) = \nu(A) = \bar{\mu}(A) = \bar{\mu}(E) + \bar{\mu}(A \setminus E) < \bar{\mu}(E) + \varepsilon.$$

So $\nu(A \setminus E) \leq \bar{\mu}(A \setminus E) < \varepsilon$. Whence $\nu(E) \geq \bar{\mu}(A) - \varepsilon \geq \bar{\mu}(E) - 2\varepsilon$. Let $\varepsilon \to 0$ to get $\nu(E) = \bar{\mu}(E)$.

Now if $E = \bigcup_{n\geq 1} E_n$ where $\bar{\mu}(E_n) < \infty$, we may replace E_n with $\bigcup_{i=1}^{n} E_i$ so that $E_n \subset E_{n+1}$. Then by continuity from below,

$$\nu(E) = \lim_{n\to\infty} \nu(E_i) = \lim_{n\to\infty} \bar{\mu}(E_i) = \bar{\mu}(E).$$

Finally if μ is σ-finite, then every set $E \in \mathcal{C}$ is sigma-finite. Thus $\bar{\mu}|_{\mathcal{C}}$ is the unique extension of μ. ■

1.5. Lebesgue-Stieltjes measures on $\mathbb{R}$

Suppose that μ is a Borel measure on $\mathbb{R}$ such that $\mu(K) < \infty$ if K is compact. Then we define a function $F : \mathbb{R} \to \mathbb{R}$ by

$$F(x) = \begin{cases} \mu([0, x]) & \text{if} \quad x \geq 0 \\ -\mu((x, 0)) & \text{if} \quad x < 0. \end{cases}$$

Note that F is monotone increasing (i.e. non-decreasing if you prefer double negatives). For $x \geq 0$ and $x_{n+1} \leq x_n$ with $x_n \downarrow x$, the continuity from above Property 1.2.5(4) shows that

$$F(x) = \mu(\bigcap_{n\geq 1} [0, x_n]) = \lim_{n\to\infty} \mu([0, x_n]) = \lim_{n\to\infty} F(x_n).$$

This uses the fact that $\mu([0, x_1]) < \infty$. Similarly if $x < 0$ and $0 > x_n \downarrow x$, then the continuity from below property shows that

$$F(x) = -\mu(\bigcup_{n\geq 1} (x_n, 0)) = \lim_{n\to\infty} -\mu((x_n, 0)) = \lim_{n\to\infty} F(x_n).$$

Thus F is right continuous.

The function F may fail to be left continuous. For example, for the point mass δ_0, $F(x) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases}$. If we knew that there was Lebesgue measure on the line, it would determine the function $F(x) = x$.

The Lebesgue-Stieltjes construction works in the other direction. Let $F:\mathbb{R}\to\mathbb{R}$ be a monotone increasing, right continuous function. We will extend the definition to include $F(\infty)=\lim_{x\to\infty}F(x)$ and $F(-\infty)=\lim_{x\to-\infty}F(x)$, and these values may include ∞ or $-\infty$ respectively.

Set $\mathcal{A}$ to be the algebra of sets consisting of all finite unions of half open intervals $(a,b]$, where we allow $b=\infty$ and $a=-\infty$; so that (a,∞), $(-\infty,b]$ and $(-\infty,\infty)$ belong to $\mathcal{A}$. Since $(a,b]^c=(-\infty,a]\cup(b,\infty)$ is in $\mathcal{A}$, it is easy to check that $\mathcal{A}$ is an algebra. Define

$$\mu_F\big(\dot{\bigcup}_{i=1}^{n}(a_i,b_i]\big)=\sum_{i=1}^{n}F(b_i)-F(a_i)$$

1.5.1. LEMMA. *μ_F is a premeasure.*

PROOF. First let's show that μ_F is well defined. To this end, suppose that $I=(a,b]=\dot{\bigcup}_{i=1}^{n}(a_i,b_i]$. Then after rearrranging if necessary, we may suppose that $a=a_1<b_1=a_2<\cdots<b_{n-1}=a_n<b_n=b$. Then if we set $a_{n+1}:=b_n=b$,

$$\sum_{i=1}^{n}F(b_i)-F(a_i)=\sum_{i=1}^{n}F(a_{i+1})-F(a_i)=F(b)-F(a).$$

Thus $\mu_F(I)$ does not depend on the decomposition into finitely many pieces. This readily extends to a finite union of intervals.

Now we consider the restricted version of countable additivity. Suppose that $I=(a,b]=\dot{\bigcup}_{i\ge1}(a_i,b_i]$. This is more difficult to deal with because these intervals cannot be ordered, end to end, as in the case of a finite union. One direction is easy: since $I=\dot{\bigcup}_{i=1}^{n}(a_i,b_i]\dot{\cup}\big(I\setminus\bigcup_{i=1}^{n}(a_i,b_i]\big)$ and the last set belongs to $\mathcal{A}$,

$$\mu_F(I)=\sum_{i=1}^{n}\mu_F((a_i,b_i])+\mu_F\big(I\setminus\bigcup_{i=1}^{n}(a_i,b_i]\big)\ge\sum_{i=1}^{n}\mu_F((a_i,b_i]).$$

Taking a limit as $n\to\infty$ yields that $\mu_F(I)\ge\sum_{i=1}^{\infty}\mu_F((a_i,b_i])$.

Conversely, first suppose that $a,b\in\mathbb{R}$. Fix $\varepsilon>0$. By right continuity, there is a $\delta>0$ so that $F(a+\delta)<F(a)+\varepsilon$. Likewise, there are $\delta_i>0$ so that $F(b_i+\delta_i)<F(b_i)+2^{-i}\varepsilon$. The compact interval $[a+\delta,b]$ has the open cover $\{(a_i,b_i+\delta_i):i\ge1\}$; so there is a finite subcover $(a_{i_1},b_{i_1}+\delta_{i_1}),\dots,(a_{i_p},b_{i_p}+\delta_{i_p})$.

Therefore $\sum_{j=1}^{p} F(b_{i_j}+\delta_{i_j}) - F(a_{i_j}) \ge F(b) - F(a+\delta)$. Consequently,

$$\begin{aligned}\sum_{i\ge1} \mu_F((a_i,b_i]) &= \sum_{i\ge1} F(b_i) - F(a_i)\\ &\ge \sum_{j=1}^{p} F(b_{i_j}) - F(a_{i_j})\\ &\ge \sum_{j=1}^{p} F(b_{i_j}+\delta_{i_j}) - 2^{-i_j}\varepsilon - F(a_{i_j})\\ &\ge F(b) - F(a+\delta) - \varepsilon\\ &\ge F(b) - F(a) - 2\varepsilon.\end{aligned}$$

Now let $\varepsilon \downarrow 0$ to get $\sum_{i\ge1} \mu_F((a_i,b_i]) \ge F(b) - F(a)$; whence we have equality. Now if $b = \infty$, we still have $\sum_{i\ge1} \mu_F((a_i,b_i]) \ge \mu_F((a,N]) = F(N) - F(a)$ for $N \in \mathbb{N}$. Letting $N \to \infty$, we get $\sum_{i\ge1} \mu_F((a_i,b_i]) \ge F(\infty) - F(a)$. Similarly we can handle $a = -\infty$.

Finally if $A = \dot{\bigcup}_{k=1}^{m} (a_k,b_k]$ is written as a disjoint union of half open intervals, we can split the union into m pieces and use the argument for a single interval on each one. Thus we obtain countable additivity (provided the union remains in $\mathcal{A}$). So μ_F is a premeasure. ■

1.5.2. THEOREM. *If $F : \mathbb{R} \to \mathbb{R}$ is an increasing, right continuous function, then there is a complete measure $(\mathbb{R}, \mathcal{B}, \bar{\mu}_F)$ that extends μ_F. The σ-algebra $\mathcal{B}$ contains the σ-algebra $\mathrm{Bor}_{\mathbb{R}}$ of Borel sets. The restriction of $\bar{\mu}_F$ to $\mathrm{Bor}_{\mathbb{R}}$ is the unique Borel measure μ_F on $\mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $a < b$.*

Conversely, if μ is a Borel measure on $\mathbb{R}$ such that $\mu(K) < \infty$ for compact sets $K \subset \mathbb{R}$, then there is an increasing, right continuous function $F : \mathbb{R} \to \mathbb{R}$ such that $\mu = \mu_F$. Also given two increasing, right continuous functions F, G, $\mu_F = \mu_G$ if and only if $F - G$ is constant.

PROOF. By Lemma 1.5.1, μ_F is a premeasure. Thus by Theorem 1.4.2, there is a complete measure $\bar{\mu}_F$ on a σ-algebra $\mathcal{B}$ containing $\mathcal{A}$ which extends μ_F. The σ-algebra $\mathcal{B}$ contains the σ-algebra generated by $\mathcal{A}$. So it contains all intervals $(a,b) = \bigcup_{n\ge1}(a, b-\frac{1}{n}]$. Hence it contains all open sets (because every open subset of $\mathbb{R}$ is the countable union of intervals), and thus all Borel sets. The restriction μ_F of $\bar{\mu}_F$ to $\mathrm{Bor}_{\mathbb{R}}$ is a Borel measure with $\mu_F((a,b]) = F(b) - F(a)$ for all $a < b$. Since μ_F is σ-finite, Proposition 1.4.3 shows that this Borel measure is unique.

Conversely, we showed that every Borel measure which is finite on bounded intervals determines an increasing, right continuous function $F : \mathbb{R} \to \mathbb{R}$ such that $\mu_F((a,b]) = F(b) - F(a)$ for all $a < b$. By the uniqueness of the construction, we see that $\mu = \mu_F$. Finally, if $\mu_G = \mu_F$, then

$$\mu_G((a,b]) = G(b) - G(a) = F(b) - F(a) \quad \text{for} \quad a < b.$$

It follows that $G(x) = F(x) + (G(0) - F(0))$; i.e., $F - G$ is constant. ■

1.5.3. COROLLARY. *Lebesgue measure is the complete measure $m = \bar{\mu}_F$ for the function $F(x) = x$. The σ-algebra $\mathcal{L}$ of Lebesgue measurable sets contains all Borel sets, and $m((a, b)) = b - a$ for all $a < b$ in $\mathbb{R}$.*

Lebesgue measure has some special properties. If $E \subset \mathbb{R}$ and $s \in \mathbb{R}$, let $E + s = \{x + s : x \in E\}$ and $sE = \{sx : x \in E\}$.

1.5.4. THEOREM. *Lebesgue measure is translation invariant: $m(E + s) = m(E)$ for all Lebesgue measurable sets $E \subset \mathbb{R}$. Also $m(sE) = |s|m(E)$.*

PROOF. The set of open intervals is invariant under translation, and hence so is $\mathrm{Bor}_{\mathbb{R}}$. The measure $m_s(E) = m(E + s)$ agrees with m on open intervals. By Theorem 1.5.2, the measures are determined by the functions

$$F(x) = \begin{cases} m((0, x]) & x \ge 0 \\ -m((x, 0]) & x < 0 \end{cases} \quad \text{and} \quad G(x) = \begin{cases} m_s((0, x]) & x \ge 0 \\ -m_s((x, 0]) & x < 0 \end{cases}.$$

However as we have observed, these functions are equal. So $m_s = m$. Hence $m(E + s) = m(E)$ for all Borel sets $E \subset \mathbb{R}$. Now Proposition 1.4.3 shows that there is a unique extension of $m|_{\mathrm{Bor}_{\mathbb{R}}}$ to the σ-algebra $\mathcal{L}$ of Lebesgue measurable sets.

If $s = 0$, then $m(0E) = 0 = 0m(E)$, so suppose that $s \ne 0$. Let $n_s(E) = |s|^{-1}m(sE)$. Observe that

$$n_s((a, b)) = |s|^{-1}|sb - sa| = b - a = m((a, b)).$$

Arguing as in the previous paragraph, we see that $n_s = m$. Hence $m(sE) = |s|n_s(E) = |s|m(E)$ for all measurable sets E. ■

1.5.5. REMARK. A point $\{a\}$ has Lebesgue-Stieltjes measure

$$\mu_F(\{a\}) = \lim_{n\to\infty} \mu_F((a - \tfrac{1}{n}, a]) = F(a) - \lim_{x\to a^-} F(x) = F(a) - F(a^-).$$

Thus $\mu_F(a) > 0$ if and only if F has a jump discontinuity at a. In this case, we say that a is an *atom* of μ_F. The only discontinuities of a monotone function are jump discontinuities. There are at most countably many jump discontinuities. To see this, consider how many jumps of size at least δ there can be in $(a, b]$; at most $(F(b) - F(a))/\delta$ which is finite. So there are at most $n(F(n) - F(-n))$ points in $(-n, n]$ with a jump of more than $\frac{1}{n}$. The union of this countable collection of finite sets is countable, and contains all of the jumps. Note, however that the discontinuities can be dense, say at every rational point.

Suppose that the discontinuities occur at a_i with jump $\alpha_i > 0$ for $i \ge 1$. Recall that δ_x is the point mass that sets $\delta(A) = 1$ if $x \in A$ and $\delta_x(A) = 0$ otherwise.

Then $\mu_a = \sum_{i\geq 1} \alpha_i \delta_{a_i}$ is called an *atomic measure*. When we subtract it from μ_F, we get a measure $\mu_c = \mu_F - \mu_a$ which has no atomic part. There are increasing, right continuous functions F_a and F_c so that $\mu_a = \mu_{F_a}$ and $\mu_c = \mu_{F_c}$. Moreover F_c is continuous and F_a is entirely determined by its jumps. We can show that (up to a constant)

$$F_a(x) = \begin{cases} \sum_{\{i:0\leq a_i \leq x\}} \alpha_i & \text{if} \quad x \geq 0 \\ -\sum_{\{i:x<a_i<0\}} \alpha_i & \text{if} \quad x < 0 \end{cases}.$$

This converges for every x because $\sum_{\{i:0\leq a_i\leq x\}} \alpha_i \leq F(x) - F(0) < \infty$ when $x \geq 0$, and similarly $\sum_{\{i:x<a_i<0\}} \alpha_i \leq F(0) - F(x)$ when $x < 0$.

We conclude this section with some regularity properties of $\bar{\mu}_F$-measurable sets.

1.5.6. THEOREM. *Let $\bar{\mu}_F$ be a Lebesgue-Stieltjes measure. For $E \subset \mathbb{R}$, the following are equivalent.*

(1) *E is $\bar{\mu}_F$-measurable.*

(2) *For all $\varepsilon > 0$, there is an open set $U \supset E$ such that $\mu_F^*(U \setminus E) < \varepsilon$.*

(3) *For all $\varepsilon > 0$, there is an closed set $C \subset E$ such that $\mu_F^*(E \setminus C) < \varepsilon$.*

(4) *There is a G_δ set $G \supset E$ so that $\mu_F^*(G \setminus E) = 0$.*

(5) *There is an F_σ set $F \subset E$ so that $\mu_F^*(E \setminus F) = 0$.*

In this case,

$$\bar{\mu}_F(E) = \inf\{\mu_F(U) : E \subset U \textit{ open}\} = \sup\{\mu_F(K) : E \supset K \textit{ compact}\}.$$

PROOF. First assume that E is bounded. Fix $\varepsilon > 0$. Since E is measurable, $\bar{\mu}_F(E) = \mu_F^*(E) = \inf \mu_F(A)$ where $E \subset A = \bigcup_{i\geq 1}(a_i, b_i]$. Now $\bar{\mu}_F(E) < \infty$ so we can choose A so that $\mu_F(A) < \bar{\mu}_F(E) + \varepsilon/2$. For each $i \geq 1$, choose $c_i > b_i$ so that $F(c_i) < F(b_i) + 2^{-i-1}\varepsilon$. Let $U = \bigcup_{i\geq 1}(a_i, c_i)$. Since E is measurable, $\mu_F^*(A) = \mu_F^*(E) + \mu_F^*(A \setminus E)$. Thus $\mu_F^*(A \setminus E) < \varepsilon/2$. Also

$$\mu_F(U \setminus A) = \mu_F\Big(\bigcup_{i\geq 1}(b_i, c_i)\Big) \leq \sum_{i\geq 1} F(c_i) - F(b_i) < \varepsilon/2.$$

So $\bar{\mu}_F^*(U \setminus E) = \mu_F(U \setminus A) + \mu_F^*(A \setminus E) < \varepsilon$.

Now if E is unbounded, let $E_n = E \cap (n-1, n]$ for $n \in \mathbb{Z}$. Find an open set $U_n \supset E_n$ with $\mu_F^*(U_n \setminus E_n) < 2^{-2|n|-1}\varepsilon$. Then $E \subset U := \bigcup_{n\in\mathbb{Z}} U_n$ is an open set, and $\mu_F^*(U \setminus E) \leq \sum_{n\in\mathbb{Z}} \mu_F^*(U_n \setminus E_n) < \frac{\varepsilon}{2} + 2\sum_{n\geq 1} 2^{-2n-1}\varepsilon < \varepsilon$. So $(1) \Longrightarrow (2)$.

$(2) \Longrightarrow (4)$. Find open sets $U_n \supset E$ with $\mu_F^*(U_n \setminus E) < \frac{1}{n}$. Set $G = \bigcap_{n\geq 1} U_n$. This is a G_δ set containing E such that $\mu_F^*(G \setminus E) \leq \mu_F^*(U_n \setminus E) < \frac{1}{n}$ for all $n \geq 1$. Hence $\mu_F^*(G \setminus E) = 0$.

(4) $\implies$ (1). Since $\bar{\mu}_F$ is complete and $\mu_F^*(G \setminus E) = 0$, the set $G \setminus E$ is $\bar{\mu}_F$-measurable. Since μ_F is a Borel measure, G is also $\bar{\mu}_F$-measurable. Therefore $E = G \setminus (G \setminus E)$ is $\bar{\mu}_F$-measurable.

(1) $\implies$ (3). If E is measurable, so is E^c. By the equivalence of (1) and (2), there is an open set $U \supset E^c$ such that $\mu_F^*(U \setminus E^c) < \varepsilon$. Let $C = U^c$, which is closed and $C \subset E$. Then $\mu_F^*(E \setminus C) = \mu_F^*(U \setminus E^c) < \varepsilon$.

(3) $\implies$ (5). Select closed sets $C_n \subset E$ with $\mu_F^*(E \setminus C_n) < \frac{1}{n}$. Let $F = \bigcup_{n \geq 1} C_n$. This is an F_σ set contained in E such that $\mu_F^*(E \setminus F) < \mu_F^*(E \setminus C_n) < \frac{1}{n}$ for all $n \geq 1$. Hence $\mu_F^*(E \setminus F) = 0$.

(5) $\implies$ (1). Since $\bar{\mu}_F$ is complete and $\mu_F^*(E \setminus F) = 0$, the set $E \setminus F$ is $\bar{\mu}_F$-measurable. Since μ_F is a Borel measure, F is also $\bar{\mu}_F$-measurable. Therefore $E = F \cup (E \setminus F)$ is $\bar{\mu}_F$-measurable.

For the final statement, the first statement holds by (2), and the second holds by (4) provided that we use closed sets. If E is bounded, the closed sets are compact. So suppose that E is unbounded. Set $E_n = E \cap [-n, n]$. Then

$$\bar{\mu}_F(E) = \sup_{n \geq 1} \bar{\mu}_F(E_n) = \sup_{n \geq 1} \sup_{E_n \supset K \text{ compact}} \mu_F(K). \qquad \blacksquare$$

CHAPTER 2

Functions

2.1. Measurable Functions

2.1.1. DEFINITION. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measure spaces. A function $f : X \to Y$ is *measurable* if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$.

In particular, $f : X \to \mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$ is measurable if $f^{-1}(B) \in \mathcal{A}$ for all $B \in \mathrm{Bor}_{\mathbb{F}}$; i.e. for all Borel sets $B \subset \mathbb{F}$.

We are most often interested in scalar valued functions, i.e. with range in $\mathbb{R}$ or $\mathbb{C}$. We always use the σ-algebra of Borel sets, not Lebesgue measurable sets, when defining measurable functions.

It is not necessary to verify the measurability condition on all Borel sets, only on a generating family. That is, if $\mathcal{E} \subset \mathcal{B}$ generates $\mathcal{B}$ as a σ-algebra and $f^{-1}(E) \in \mathcal{A}$ for all $E \in \mathcal{E}$, then f is measurable. This follows from the easy facts:

$$f^{-1}(E^c) = (f^{-1}(E))^c \quad \text{and} \quad f^{-1}\Big(\bigcup_{n \geq 1} E_n\Big) = \bigcup_{n \geq 1} f^{-1}(E_n).$$

So the following proposition is immediate.

2.1.2. PROPOSITION. *Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measure spaces.*

(a) *If $f : X \to Y$, then $\{B \in \mathcal{B} : f^{-1}(B) \in \mathcal{A}\}$ is a σ-algebra.*

(b) *If $f : X \to \mathbb{R}$, the following are equivalent:*
  (1) *f is measurable.*
  (2) *$\{x \in X : f(x) < \alpha\}$ is measurable for all $\alpha \in \mathbb{R}$.*
  (3) *$\{x \in X : f(x) \leq \alpha\}$ is measurable for all $\alpha \in \mathbb{R}$.*
  (4) *$\{x \in X : f(x) > \alpha\}$ is measurable for all $\alpha \in \mathbb{R}$.*
  (5) *$\{x \in X : f(x) \geq \alpha\}$ is measurable for all $\alpha \in \mathbb{R}$.*

(c) *If $f : X \to \mathbb{C}$, then f is measurable if and only if* $\mathrm{Re}\, f$ *and* $\mathrm{Im}\, f$ *are measurable $\mathbb{R}$-valued functions.*

2.1.3. COROLLARY. *Let X be a topological space, and consider (X, Bor_X). Continuous functions $f : X \to \mathbb{F}$ are (Borel) measurable.*

PROOF. Since f is continuous, if $U \subset \mathbb{F}$ is open, then $f^{-1}(U)$ is open and thus Borel. ■

The following records some basic operations that preserve measurability.

2.1.4. PROPOSITION. *Let $(X, \mathcal{B})$ be a measure space.*

(a) *Suppose that $f, g : X \to \mathbb{F}$ are measurable. Then $f + \lambda g$ for $\lambda \in \mathbb{F}$ and fg are measurable; and if $g \neq 0$, then f/g is measurable.*

(b) *If $f_n : X \to \mathbb{R}$ are measurable, then $\sup f_n$, $\inf f_n$, $\limsup f_n$ and $\liminf f_n$ are measurable. Thus, if $f = \lim f_n$ exists pointwise, then f is measurable.*

PROOF. Clearly if $\lambda \neq 0$, then g is measurable if and only if λg is measurable. So consider $f + g$. Note that

$$\{x : f(x) + g(x) < \alpha\} = \bigcup_{r \in \mathbb{Q}} \{x : f(x) < r\} \cap \{x : g(x) < \alpha - r\} \in \mathcal{B}.$$

So $f + g$ is measurable. Now consider fg. If $\alpha > 0$, then

$$\begin{aligned}\{x : f(x)g(x) < \alpha\} = &\big(\{x : f(x) \geq 0\} \cap \{x : g(x) \leq 0\}\big) \\ &\cup \big(\{x : f(x) \leq 0\} \cap \{x : g(x) \geq 0\}\big) \\ &\cup \bigcup_{r \in \mathbb{Q}^+} \{x : |f(x)| < r\} \cap \{x : |g(x)| < \alpha/r\}\end{aligned}$$

This lies in $\mathcal{B}$. Also

$$\begin{aligned}\{x : f(x)g(x) < 0\} = &\big(\{x : f(x) > 0\} \cap \{x : g(x) < 0\}\big) \\ &\cup \big(\{x : f(x) < 0\} \cap \{x : g(x) > 0\}\big)\end{aligned}$$

And finally, if $\alpha < 0$, then

$$\begin{aligned}\{x : f(x)g(x) < \alpha\} = &\bigcup_{r \in \mathbb{Q}^+} \big(\{x : f(x) > r\} \cap \{x : g(x) < \alpha/r\}\big) \\ &\cup \bigcup_{r \in \mathbb{Q}^+} \big(\{x : f(x) < \alpha/r\} \cap \{x : g(x) > r\}\big)\end{aligned}$$

Thus fg is measurable. Finally, if $g \neq 0$, then for $\alpha > 0$, $\{x : 1/g(x) < \alpha\} = \{x : g(x) > \alpha^{-1}\} \cup \{x : g(x) < 0\}$; also $\{x : 1/g(x) < 0\} = \{x : g(x) < 0\}$; and if $\alpha < 0$, $\{x : 1/g(x) < \alpha\} = \{x : \alpha^{-1} < g(x) < 0\}$. So $1/g$ is measurable. Thus f/g is measurable by the product result.

Now suppose that f_n are measurable for $n \geq 1$. Then

$$\{x : \sup f_n(x) > \alpha\} = \bigcup_{n \geq 1} \{x : f_n(x) > \alpha\} \in \mathcal{B}.$$

So $\sup f_n$ is measurable. Similarly $\inf f_n$ is measurable. Therefore $\limsup f_n = \inf_{k\geq 1} \sup_{n\geq k} f_n$ is measurable, and similarly for $\liminf f_n$. So $f = \lim f_n$ exists pointwise means that $f = \limsup f_n = \liminf f_n$. So f is measurable. ■

When the measure is complete, changing things on a set of measure 0 has no important consequence.

2.1.5. PROPOSITION. *Let μ be a complete measure on $(X, \mathcal{B})$. Then if f is measurable and $g = f$ a.e.(μ), then g is measurable. Also if f_n are measurable and f is a function such that f_n converge to f a.e.(μ), then f is measurable.*

PROOF. Suppose that N is a null set such that $f = g$ on $X \setminus N$. For any Borel set $B \subset \mathbb{R}$, $f^{-1}(B) \triangle g^{-1}(B) \subset N$. As these two sets differ by a subset of a null set, they differ by a null set by completeness. So one is measurable because the other is.

Suppose that $f(x) = \lim f_n(x)$ except on a null set N. Then $f = \limsup f_n$ except on N, and hence is measurable by the previous paragraph. ■

2.2. Simple Functions

2.2.1. DEFINITION. Let $(X, \mathcal{B})$ be a measure space. A *simple function* is a function $\varphi : X \to \mathbb{C}$ of the form

$$\varphi(x) = \sum_{i=1}^{n} a_i \chi_{E_i} \quad \text{where } a_i \in \mathbb{C}^*,\ E_i \in \mathcal{B} \text{ and } E_i \cap E_j = \varnothing \text{ for } i \neq j.$$

A simple function is just a measurable function with finite range. Indeed, the range of φ is contained in $\{a_i : 1 \leq i \leq n\} \cup \{0\}$. Conversely, if the range of a measurable function φ is contained in $\{a_i : 1 \leq i \leq n\} \cup \{0\}$ with $a_i \neq a_j \neq 0$ for $i \neq j$, then we can set $E_i = \varphi^{-1}(\{a_i\})$. These are disjoint, and $\varphi = \sum_{i=1}^{n} a_i \chi_{E_i}$. This is called the *standard form* of a simple function.

Note that the set of all simple functions is an algebra, meaning that it is a vector space and is closed under products. The subset of simple functions such that $\mu(E_i) < \infty$ for all i is a subalgebra.

The main result of this section is that every measurable function is a limit of simple functions in a nice way.

2.2.2. PROPOSITION.
(a) *If $f : X \to [0, \infty]$ is measurable, then there is a sequence of simple functions φ_n so that $\varphi_n \leq \varphi_{n+1}$, $\lim_{n\to\infty} \varphi_n(x) = f(x)$, and convergence is uniform on $\{x : f(x) \leq R\}$ for any $R \in \mathbb{R}$.*

(b) *If $f : X \to \mathbb{C}$ is measurable, then there is a sequence of simple functions φ_n so that $|\varphi_n| \le |\varphi_{n+1}|$, $\lim_{n\to\infty} \varphi_n(x) = f(x)$, and convergence is uniform on $\{x : |f(x)| \le R\}$ for any $R \in \mathbb{R}$.*

PROOF. (a) For $n \ge 1$, let $A_{k,n} = \left\{x : \frac{k}{2^n} \le f(x) < \frac{k+1}{2^n}\right\}$ for $0 \le k < 4^n$ and $A_{4^n,n} = \{x : f(x) \ge 2^n\}$. Define simple functions

$$\varphi_n = \sum_{k=1}^{4^n} \frac{k}{2^n} \chi_{A_{k,n}}.$$

Observe that

$$A_{k,n} = A_{2k,n+1} \dot\cup A_{2k+1,n+1} \quad \text{for } 0 \le k < 4^n \quad \text{and} \quad A_{4^n,n} = \dot\bigcup_{k=2\cdot 4^n}^{4^{n+1}} A_{k,n+1}.$$

Moreover for $0 \le k < 4^n$,

$$\tfrac{k}{2^n}\chi_{A_{k,n}} \le \tfrac{2k}{2^{n+1}}\chi_{A_{2k,n+1}} + \tfrac{2k+1}{2^{n+1}}\chi_{A_{2k+1,n+1}} \quad \text{and} \quad 2^n\chi_{A_{4^n,n}} \le \sum_{k=2\cdot 4^n}^{4^{n+1}} \tfrac{k}{2^{n+1}}\chi_{A_{k,n+1}}.$$

Thus $\varphi_n \le \varphi_{n+1} \le f$. Moreover, if $f(x) \le 2^N$, then $f(x) - \varphi_n(x) \le 2^{-n}$ for all $n \ge N$; and if $f(x) = \infty$, then $\varphi_n(x) = 2^n$. Thus $\lim_{n\to\infty} \varphi_n = f$ pointwise. Moreover convergence is uniform on $\{x : |f(x)| \le R\}$ since our estimate is uniformly good once $2^n \ge R$.

(b) Let $f = g_1 - g_2 + ig_3 - ig_4$ where $g_1 = \max\{\operatorname{Re} f, 0\}$, $g_2 = \max\{-\operatorname{Re} f, 0\}$, $g_3 = \max\{\operatorname{Im} f, 0\}$ and $g_4 = \max\{-\operatorname{Im} f, 0\}$. For each $1 \le i \le 4$, apply part (a) to get sequences of simple functions $\psi_{i,n}$ increasing to g_i. Then the sequence

$$\varphi_n = \psi_{1,n} - \psi_{2,n} + i\psi_{3,n} - i\psi_{4,n}$$

works. Details are left to the reader. ∎

2.3. Two Theorems about Measurable Functions

Littlewood had three principles of Lebesgue measure:

- every measurable set of finite mesure is almost a finite union of intervals (use Theorem 1.5.6(2) and throw out the very small intervals).
- Every measurable function is almost continuous (Lusin's Theorem).
- A pointwise convergence sequence of measurable functions is almost uniformly convergent (Egorov's Theorem).

2.3.1. EGOROV'S THEOREM. *Let $(X, \mathcal{B}, \mu)$ be a finite measure space, i.e., $\mu(X) < \infty$. Suppose that $f_n : X \to \mathbb{C}$ are measurable functions and $f_n \to f$ a.e.(μ). Then for $\varepsilon > 0$, there is a set $E \in \mathcal{B}$ with $\mu(X \setminus E) < \varepsilon$ so that $f_n \to f$ uniformly on E.*

PROOF. Observe that $f_n \to f$ uniformly on E provided that for each $m \geq 1$, there is an integer N_m so that $|f_n(x) - f(x)| < \frac{1}{m}$ for all $x \in E$ and all $n \geq N_m$. Define

$$A_{m,N} = \{x : |f_n(x) - f(x)| < \tfrac{1}{m} \ \forall n \geq N\} = \bigcap_{n \geq N} \{x : |f_n(x) - f(x)| < \tfrac{1}{m}\}.$$

Note that $A_{m,1} \subset A_{m,2} \subset \dots$ and that

$$\bigcup_{n \geq 1} A_{m,n} \supset \{x : f_n(x) \to f(x)\} = X \setminus N$$

where N is a null set. Since X has finite measure, we can choose an integer N_m so that $\mu(A_{m,N_m}) > \mu(X) - 2^{-m}\varepsilon$. Hence $\mu(A^c_{m,N_m}) < 2^{-m}\varepsilon$. Let $E = \bigcap_{m \geq 1} A_{m,N_m}$. Then,

$$\mu(E^c) = \mu\big(\bigcup_{m \geq 1} A^c_{m,N_m}\big) \leq \sum_{m \geq 1} \mu(A^c_{m,N_m}) < \sum_{m \geq 1} 2^{-m}\varepsilon = \varepsilon.$$

By the first line of the proof, f_n converges uniformly to f on E. ■

2.3.2. LUSIN'S THEOREM. *Let $f : [a,b] \to \mathbb{C}$ be a Lebesgue measurable function, and let $\varepsilon > 0$. Then there is a continuous function $g \in C[a,b]$ so that $m(\{x : f(x) \neq g(x)\}) < \varepsilon$.*

PROOF. If $\varphi = \sum_{i=1}^m a_i \chi_{E_i}$ is a simple function and $\delta > 0$, we can find compact sets $A_i \subset E_i$ so that $m\big(\bigcup_{i=1}^m (E_i \setminus A_i)\big) < \delta$. Observe that $K = \bigcup A_i$ is compact and $\varphi|_K$ is continuous as a function on K (because it is locally constant).

Choose simple functions φ_n converging pointwise to f. Find compact sets K_n so that $\varphi_n|_{K_n}$ is continuous and $m\big([a,b] \setminus K_n\big) < 2^{-n-1}\varepsilon$. Then $K_0 = \bigcap K_n$ is compact, each $\varphi_n|_{K_0}$ is continuous, and

$$m\big([a,b] \setminus K_0\big) \leq \sum m\big([a,b] \setminus K_n\big) < \varepsilon/2.$$

By Egorov's Theorem, there is a measurable set $E \subset K_0$ with $m(K_0 \setminus E) < \varepsilon/4$ so that so that φ_n converge uniformly to f on E. Then by Theorem 1.5.6, there is a closed set $K \subset E$ with $m(E \setminus K) < \varepsilon/4$. Since $\varphi_n|_K$ are continuous, and converge uniformly to $f|_K$, we see that $f|_K$ is continuous. Also

$$m\big([a,b] \setminus K\big) < m\big([a,b] \setminus K_0\big) + m(K_0 \setminus E) + m(E \setminus K) < \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon.$$

It remains to extend $f|_K$ to a continuous function g on $[a,b]$. This is a (non-trivial) exercise in a basic real analysis course. Just make g piecewise linear on each open component of the complement. Once achieved, $\{x : f(x) \neq g(x)\}$ is contained in $[a,b] \setminus K$, thus has measure less than ε. ■

CHAPTER 3

Integration

The goal of this chapter is to develop a theory of integration with respect to an arbitrary measure.

3.1. Integrating positive functions

We begin by considering the integral only of positive functions. Moreover, we insist that the function be measurable. We fix a measure space $(X, \mathcal{B}, \mu)$. Let

$$L^+ = \{f : X \to [0, \infty] \text{ measurable}\}.$$

Notice that L^+ only depends on the measure space $(X, \mathcal{B})$ and not on μ. The simple functions have a very natural definition of integral. We bootstrap that into a definition for positive measurable functions. Remember that simple functions only take finite values.

3.1.1. DEFINITION. If $\varphi = \sum_{i=1}^n a_i \chi_{E_i}$ is a simple function in L^+, set

$$\int \varphi \, d\mu := \sum_{i=1}^n a_i \mu(E_i).$$

We use the convention that $0 \cdot \infty = 0 = \infty \cdot 0$. For general $f \in L^+$, define

$$\int f \, d\mu := \sup \Big\{ \int \varphi \, d\mu : 0 \le \varphi \le f, \ \varphi \text{ simple} \Big\}.$$

Also define $\displaystyle\int_A f \, d\mu := \int f \chi_A \, d\mu$ for any measurable set $A \in \mathcal{B}$.

The convention that $0 \cdot \infty = 0$ means that $f = 0\chi_X$ has $\displaystyle\int f \, d\mu = 0$ even if $\mu(X) = \infty$; since the integral of a simple function should not be changed by adding a zero function to it. The second identity $\infty \cdot 0 = 0$ is also needed to coincide with the notion that the value of f on a set of measure 0 should not affect the integral. So $\displaystyle\int \infty \chi_{\{x\}} \, d\mu = 0$ is $\mu(x) = 0$. At this stage, we allow ∞ both as a value of f and of the integral. Later, in some cases, we impose further restrictions.

In various books, you will see the following notations which all mean the same thing:

$$\int f\,d\mu, \quad \int f(x)\,d\mu(x), \quad \int f(x)\,\mu(dx), \quad \int f.$$

Of course, the last one will only make sense if the measure μ is known. We will use this when convenient to simplify the notation.

3.1.2. LEMMA. *Let φ, ψ be simple functions in L^+; and let $f, g \in L^+$.*

(a) *$\int \varphi\,d\mu$ is well defined.*

(b) *If $c \geq 0$, then $\int c\varphi\,d\mu = c\int \varphi\,d\mu$.* (b′) *$\int cf\,d\mu = c\int f\,d\mu$.*

(c) *$\int \varphi + \psi\,d\mu = \int \varphi\,d\mu + \int \psi\,d\mu$.*

(d) *If $\varphi \leq \psi$, then $\int \varphi\,d\mu \leq \int \psi\,d\mu$.* (d′) *If $f \leq g$, $\int f\,d\mu \leq \int g\,d\mu$.*

(e) *$\nu(A) := \int_A \varphi\,d\mu$ for $A \in \mathcal{B}$ defines a measure on $(X, \mathcal{B})$.*

PROOF. (a) Suppose that $\varphi = \sum_{i=1}^m a_i\chi_{E_i} = \sum_{j=1}^n b_j\chi_{F_j}$. We can disregard zeros, so we may suppose that $a_i \neq 0 \neq b_j$. Moreover since the E_i and F_j are pairwise disjoint collections, we have that $\{a_i : 1 \leq i \leq m\} = \{b_j : 1 \leq j \leq n\}$. Let $c_1, \dots, c_p$ be the distinct non-zero values of φ, and set $A_k = \{x : \varphi(x) = c_k\}$ for $1 \leq k \leq p$. Then

$$A_k = \bigcup\{E_i : a_i = c_k\} = \bigcup\{F_j : b_j = c_k\}.$$

Therefore

$$\begin{aligned}
\sum_{i=1}^n a_i\mu(E_i) &= \sum_{k=1}^p c_k \sum_{a_i=c_k} \mu(E_i) \\
&= \sum_{k=1}^p c_k\mu\Big(\bigcup_{a_i=c_k} E_i\Big) = \sum_{k=1}^p c_k\mu(A_k) \\
&= \sum_{k=1}^p c_k\mu\Big(\bigcup_{b_j=c_k} F_j\Big) = \sum_{j=1}^n b_j\mu(F_j).
\end{aligned}$$

(b) is trivial, and (b′) follows because

$$\{\varphi : \varphi \in L^+,\ \text{simple},\ \varphi \leq cf\} = \{c\varphi : \varphi \in L^+,\ \text{simple},\ \varphi \leq f\}.$$

For (c), suppose that $\varphi = \sum_{i=0}^m a_i\chi_{E_i}$ and $\psi = \sum_{j=0}^m b_j\chi_{F_j}$, where we add in the term $a_0 = b_0 = 0$ and $E_0 = \big(\bigcup_{i=1}^m E_i\big)^c$ and $F_0 = \big(\bigcup_{j=1}^n F_j\big)^c$ in order that $X = \bigcup_{i=0}^m E_i = \bigcup_{j=0}^n F_j$. Define $A_{ij} = E_i \cap F_j$ for $0 \leq i \leq m$ and $0 \leq j \leq n$. Then

$$\varphi = \sum_{i=0}^m \sum_{j=0}^n a_i\chi_{A_{ij}}, \quad \psi = \sum_{i=0}^m \sum_{j=0}^n b_i\chi_{A_{ij}} \text{ and } \varphi + \psi = \sum_{i=0}^m \sum_{ij=0}^n (a_i + b_j)\chi_{A_{ij}}.$$

Therefore using (a), we get

$$\begin{aligned}\int \varphi+\psi\,d\mu &= \sum_{i=0}^{m}\sum_{j=0}^{n}(a_i+b_j)\mu(A_{ij})\\ &= \sum_{i=0}^{m} a_i \sum_{j=0}^{n}\mu(A_{ij}) + \sum_{j=0}^{n} b_j \sum_{i=0}^{m}\mu(A_{ij})\\ &= \sum_{i=0}^{m} a_i\mu\big(\bigcup_{j=0}^{n} A_{ij}\big) + \sum_{j=0}^{n} b_j\mu\big(\bigcup_{i=0}^{m} A_{ij}\big)\\ &= \sum_{i=0}^{m} a_i\mu(E_i) + \sum_{j=0}^{n} b_j\mu(F_j) = \int \varphi\,d\mu + \int \psi\,d\mu.\end{aligned}$$

(d) Write φ and ψ as in (c), and note that

$$\varphi = \sum_{i=0}^{m}\sum_{j=0}^{n} a_i\chi_{A_{ij}} \le \psi = \sum_{i=01}^{m}\sum_{j=0}^{n} b_i\chi_{A_{ij}}$$

means that $a_i \le b_j$ if $A_{ij} \neq \varnothing$. Therefore by (a),

$$\int \varphi\,d\mu = \sum_{i=0}^{m}\sum_{j=0}^{n} a_i\mu(A_{ij}) \le \sum_{i=0}^{m}\sum_{j=0}^{n} b_j\mu(A_{ij}) = \int \psi\,d\mu.$$

(d′) follows because

$$\{\varphi : \varphi \in L^+,\ \text{simple } \varphi \le f\} \subset \{\varphi : \varphi \in L^+,\ \text{simple } \varphi \le g\}.$$

(e) Clearly $\nu(\varnothing) = 0$ and ν takes values in $[0,\infty]$. Suppose that A_j are pairwise disjoint sets in $\mathcal{B}$ with $A = \dot{\bigcup}_{j\ge 1} A_j$. Then

$$\begin{aligned}\nu(A) &= \int_A \varphi\,d\mu = \sum_{i=1}^{n} a_i\mu(E_i\cap A)\\ &= \sum_{i=1}^{n} a_i\mu\big(\dot{\bigcup}_{j\ge 1} E_i\cap A_j\big) = \sum_{i=1}^{n} a_i \sum_{j\ge 1}\mu(E_i\cap A_j)\\ &= \sum_{j\ge 1}\sum_{i=1}^{n} a_i\mu(E_i\cap A_j) = \sum_{j\ge 1}\int_{A_j}\varphi\,d\mu = \sum_{j\ge 1}\nu(A_j).\end{aligned}$$

Thus ν is countably additive, and hence is a measure. ■

3.2. Two limit theorems

This section establishes two fundamental limit theorems, and extends the additivity of the integral to arbitrary functions in L^+.

3.2.1. LEMMA. *Let $\varphi \geq 0$ be a simple function,. Let $A_n \in \mathcal{B}$ with $A_n \subset A_{n+1}$ and $\bigcup_{n\geq 1} A_n = X$. Then*

$$\lim_{n\to\infty} \int_{A_n} \varphi \, d\mu = \int \varphi \, d\mu.$$

PROOF. Let $\nu(A) = \displaystyle\int_A \varphi \, d\mu$ be the measure defined in Lemma 3.1.2(e). Then

$$\lim_{n\to\infty} \int_{A_n} \varphi \, d\mu = \lim_{n\to\infty} \nu(A_n) = \nu(X) = \int \varphi \, d\mu$$

by the property of continuity from below. ∎

3.2.2. MONOTONE CONVERGENCE THEOREM. *Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $f_n \in L^+$ such that $f_n \leq f_{n+1}$ for $n \geq 1$. Define $f(x) = \lim_{n\to\infty} f_n(x)$. Then*

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

PROOF. Since $f_n(x)$ is increasing, the limit always exists in $[0, \infty]$. Also since $f_n \leq f_{n+1} \leq f$, we have that

$$\int f_n \, d\mu \leq \int f_{n+1} \, d\mu \leq \int f \, d\mu.$$

Thus $\lim_{n\to\infty} \int f_n \, d\mu$ exists in $[0, \infty]$ and is no larger than $\int f \, d\mu$.

Conversely, suppose that $\varphi \leq f$ is a non-negative simple function, and let $\varepsilon > 0$. Define $A_n = \{x : f_n(x) \geq (1-\varepsilon)\varphi(x)\}$. Since $\varphi(x) < \infty$ and $f_n(x) \to f(x) \geq \varphi(x)$, we have $A_n \subset A_{n+1}$ and $\bigcup_{n\geq 1} A_n = X$. Therefore by Lemma 3.2.1,

$$(1-\varepsilon) \int \varphi \, d\mu = \lim_{n\to\infty} \int_{A_n} (1-\varepsilon)\varphi \, d\mu \leq \lim_{n\to\infty} \int_{A_n} f_n \, d\mu \leq \lim_{n\to\infty} \int f_n \, d\mu.$$

Let $\varepsilon \downarrow 0$ to get that $\int \varphi \, d\mu \leq \lim_{n\to\infty} \int f_n \, d\mu$. Taking the supremum yields $\int f \, d\mu \leq \lim_{n\to\infty} \int f_n \, d\mu$. So equality holds. ∎

3.2.3. COROLLARY. *The integral is countably additive on L^+. That is, if $f_n \in L^+$ for $n \geq 1$, then*

$$\int \sum_{n\geq 1} f_n \, d\mu = \sum_{n\geq 1} \int f_n \, d\mu.$$

PROOF. Apply MCT to the sequence $g_k = \sum_{n=1}^{k} f_n$. ∎

3.2.4. COROLLARY. *If $f \in L^+$, $\nu(A) = \int_A f \, d\mu$ is a measure on $(X, \mathcal{B})$.*

PROOF. Countable additivity is a special case of the previous corollary; since if $A = \dot{\bigcup}_{n\geq 1} A_n$, then $f\chi_A = \sum_{n\geq 1} f\chi_{A_n}$. ∎

3.2.5. LEMMA. *If $f \in L^+$, then $\int f\,d\mu = 0$ if and only if $f = 0$ a.e.(μ).*

PROOF. If $f = 0$ a.e.(μ) and $0 \leq \varphi \leq f$ where $\varphi = \sum_{i=1}^n a_i\chi_{E_i}$ is a simple function, then $a_i > 0$ implies $\mu(E_i) = 0$. Therefore $\int \varphi\,d\mu = \sum_{i=1}^n a_i\mu(E_i) = 0$. Hence $\int f\,d\mu = \sup_{0\leq\varphi\leq f} \int \varphi\,d\mu = 0$.

Conversely, suppose that $\int f\,d\mu = 0$. Let $E = \{x : f(x) > 0\}$ and $E_n = \{x : f(x) \geq \frac{1}{n}\}$. Then $\varphi_n = \frac{1}{n}\chi_{E_n} \leq f$. So $0 = \int f\,d\mu \geq \int \varphi_n\,d\mu = \frac{1}{n}\mu(E_n)$. Therefore $\mu(E) = \lim_{n\to\infty} \mu(E_n) = 0$; i.e. $f = 0$ a.e.(μ). ∎

3.2.6. COROLLARY. *If $f, f_n \in L^+$ such that $f_n \leq f_{n+1}$ a.e.(μ) for $n \geq 1$ and $f(x) = \lim_{n\to\infty} f_n(x)$ a.e.(μ), then*

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

PROOF. Define

$$E_n = \{x : f_n(x) > f_{n+1}(x)\} \text{ for } n \geq 1 \quad \text{and} \quad E_0 = \{x : \lim_{n\to\infty} f_n(x) \neq f(x)\}.$$

Then $\mu(E_n) = 0$ for $n \geq 0$ by hypothesis. Let $E = \bigcup_{n\geq 0} E_n$. Then $\mu(E) = 0$, $f_n\chi_{E^c} \leq f_{n+1}\chi_{E^c}$ and $f\chi_{E^c} = \lim_{n\to\infty} f_n\chi_{E^c}$. So by the MCT and Lemma 3.2.5,

$$\int f\,d\mu = \int f\chi_{E^c}\,d\mu = \lim_{n\to\infty} \int f_n\chi_{E^c}\,d\mu = \lim_{n\to\infty} \int f_n\,d\mu. \qquad \blacksquare$$

3.2.7. EXAMPLE. Things don't work out quite as well for limits which are not monotone. For example, let $f_n = \frac{1}{n}\chi_{[0,n]} \in L^+(m)$. Then $f_n \to 0$ uniformly on $\mathbb{R}$. However,

$$\lim_{n\to\infty} \int f_n\,dm = \lim_{n\to\infty} 1 = 1 \neq 0 = \int 0\,dm.$$

The following theorem provides an important inequality.

3.2.8. FATOU'S LEMMA. *Let $(X, \mathcal{B}, \mu)$ be a measure space. Suppose $f_n \in L^+$ for $n \geq 1$. Then*

$$\int \liminf f_n\,d\mu \leq \liminf \int f_n\,d\mu.$$

PROOF. Let $g_n(x) = \inf_{k\geq n} f_k(x) \leq f_n(x)$. Then $g_n \leq g_{n+1}$ for $n \geq 1$, and $\liminf f_n = \lim_{n\to\infty} g_n$. Therefore by MCT,

$$\int \liminf f_n \, d\mu = \lim_{n\to\infty} \int g_n \, d\mu = \liminf \int g_n \, d\mu \le \liminf \int f_n \, d\mu. \qquad \blacksquare$$

3.3. Integrating complex valued functions

3.3.1. DEFINITION. Let $(X, \mathcal{B}, \mu)$ be a measure space. A measurable function $f : X \to \mathbb{C}$ is *integrable* if

$$\|f\|_1 = \int |f| \, d\mu < \infty.$$

The set of all integrable functions will be (provisionally) called $L^1(\mu)$. Write $f = f_1 - f_2 + if_3 - if_4$ where $f_1 = \operatorname{Re} f \vee 0 = \max\{\operatorname{Re} f, 0\}$, $f_2 = (-\operatorname{Re} f) \vee 0$, $f_3 = \operatorname{Im} f \vee 0$ and $f_3 = (-\operatorname{Im} f) \vee 0$. If f is integrable, define

$$\int f \, d\mu = \int f_1 \, d\mu - \int f_2 \, d\mu + i \int f_3 \, d\mu - i \int f_4 \, d\mu.$$

Note that since $f_i \in L^+$ and $f_i \le |f|$, we have $\int f_i \, d\mu \le \int |f| \, d\mu < \infty$. So $\int f \, d\mu$ is defined as a complex number.

The reason for the temporary nature of our definition of $L^1(\mu)$ is that this is not a normed vector space, due to the fact that if $f = 0$ a.e.(μ), then $\|f\|_1 = 0$. So $\|\cdot\|_1$ is just a pseudo-norm. That is, it satisfies $\|\lambda f\|_1 = |\lambda| \, \|f\|_1$ and the triangle inequality, but is not positive definite. We will rectify this soon by identifying functions which agree a.e. into an equivalence class representing an element of $L^1(\mu)$.

3.3.2. PROPOSITION. *$L^1(\mu)$ is a vector space and $\|\cdot\|_1$ is a pseudo-norm. That is, it is positive homogeneous: $\|\lambda f\|_1 = |\lambda| \, \|f\|_1$ for $\lambda \in \mathbb{C}$, and the triangle inequality holds, but $\|f\|_1 = 0$ if and only if $f = 0$ a.e.(μ). The map $\int$ taking f to $\int f \, d\mu$ is linear, and*

$$\Big| \int f \, d\mu \Big| \le \int |f| \, d\mu = \|f\|_1.$$

PROOF. Showing that $L^1(\mu)$ is a vector space is straightforward. Multiplication by a real number or an imaginary number merely scales and shuffles the functions f_i, so homogeneity for these scalars is easy. In general, you need to write out $\lambda = a + ib$ and chase some details which are left to the reader.

Lemma 3.2.5 shows that $\|f\|_1 = 0$ if and only if $f = 0$ a.e.(μ). The triangle inequality is also straightforward: if $f, g \in L^1(\mu)$, then

$$\|f + g\|_1 = \int |f + g| \, d\mu \le \int |f| + |g| \, d\mu = \|f\|_1 + \|g\|_1.$$

Linearity follows from linearity of the integral in L^+, and is left to the reader. Let $f \in L^1(\mu)$ and choose θ so that $\int f = e^{i\theta}\left|\int f\right|$. Then

$$\left|\int f\right| = \int e^{-i\theta} f = \operatorname{Re}\int e^{-i\theta} f = \int \operatorname{Re} e^{-i\theta} f.$$

Write $\operatorname{Re} e^{-i\theta} f = g_1 - g_2$, where $g_1 = \operatorname{Re} e^{-i\theta} f \vee 0$ and $g_2 = -(\operatorname{Re} e^{-i\theta} f) \vee 0$. Then

$$0 \le \int \operatorname{Re} e^{-i\theta} f = \int g_1 - g_2 \le \int g_1 + g_2 \le \int |f|.$$

Combining these two formulae, we get the desired inequality. ■

We now arrive at the most important limit theorem in measure theory. With this result and MCT, you can deal with most situations that arise.

3.3.3. Lebesgue Dominated Convergence Theorem. *Suppose that $f_n, g \in L^1(\mu)$, $g \ge 0$, such that $\lim_{n\to\infty} f_n = f$ a.e.(μ) and $|f_n| \le g$ a.e.(μ) for $n \ge 1$. Then $f \in L^1(\mu)$ and*

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

Proof. First assume that f_n are real valued. Then we apply Fatou's Lemma 3.2.8 to the sequences $g \pm f_n$ in L^+ to get

$$\int g + f\,d\mu \le \liminf \int g + f_n\,d\mu = \int g\,d\mu + \liminf \int f_n\,d\mu$$

and

$$\int g - f\,d\mu \le \liminf \int g - f_n\,d\mu = \int g\,d\mu - \limsup \int f_n\,d\mu.$$

Since $\int g\,d\mu < \infty$, it can be cancelled off and we obtain that

$$\limsup \int f_n\,d\mu \le \int f\,d\mu \le \liminf \int f_n\,d\mu.$$

It follows that $\int f\,d\mu = \lim_{n\to\infty}\int f_n\,d\mu$. In particular, applying this to $|f|$ and $|f_n|$ we obtain that

$$\|f\|_1 = \int |f|\,d\mu = \lim_{n\to\infty}\int |f_n|\,d\mu \le \int g\,d\mu < \infty.$$

So $f \in L^1(\mu)$ and thus is integrable.

In general, split f as the sum of its real and imaginary parts, use the real result, and then recombine. ■

3.3.4. EXAMPLE. Consider $\displaystyle\lim_{n\to\infty}\int_a^b \frac{n}{1+n^2x^2}\,dx$ for $0 \le a < b \le \infty$. Observe that $f_n(x) = \frac{n}{1+n^2x^2} = \frac{1}{\frac{1}{n}+nx^2}$. So $\displaystyle\lim_{n\to\infty} f_n(x) = 0$ for all $x > 0$. The function $h(t) = t + \frac{1}{t}x^2$ attains its minimum at $t = x$. Thus if $0 \le x \le 1$,

$$\sup_{n\ge 1} f_n(x) \le \sup_{0<t\le 1} \frac{1}{h(t)} = \frac{1}{h(x)} = \frac{1}{2x}.$$

On the other hand, if $x \ge 1$, then

$$\sup_{n\ge 1} f_n(x) \le \sup_{0<t\le 1} \frac{1}{h(t)} = \frac{1}{h(1)} = \frac{1}{1+x^2}.$$

So if we set $g(x) = \begin{cases} \frac{1}{2x} & \text{if } \ 0 < x \le 1 \\ \frac{1}{1+x^2} & \text{if } \ x \ge 1 \end{cases}$, then $0 \le f_n \le g$. The problem is that $g \notin L^1(0,\infty)$ since $\int_0^1 \frac{1}{2x}\,dx = +\infty$. However if $0 < a \le 1$, then

$$\int_a^\infty g(x)\,dx = \int_a^1 \tfrac{1}{2x}\,dx + \int_1^\infty \tfrac{1}{1+x^2}\,dx = \tfrac{1}{2}\ln x\Big|_a^1 + \tan^{-1}(x)\Big|_1^\infty = \tfrac{1}{2}\ln a^{-1} + \tfrac{\pi}{2} < \infty.$$

Hence as long as $a > 0$, the Lebesgue dominated convergence theorem (LDCT) applies. So $\lim_{n\to\infty}\int_a^b f_n\,d\mu = 0$. However if $a = 0$,

$$\int_0^b f_n(x)\,dx = \int_0^b \frac{n}{1+n^2x^2}\,dx = \tan^{-1}(nx)\Big|_0^b = \tan^{-1}(nb).$$

So

$$\lim_{n\to\infty}\int_0^b f_n(x)\,dx = \lim_{n\to\infty}\tan^{-1}(nb) = \frac{\pi}{2} \ne \int_0^b 0\,dx = 0.$$

Thus the domination condition by an *integrable* function is crucial.

3.4. The space $L^1(\mu)$

For the purposes of this section, let $\mathcal{L}^1(\mu)$ denote the vector space of all μ-integrable functions.

3.4.1. DEFINITION. Given $(X, \mathcal{B}, \mu)$, let $\mathcal{N} = \{f \in \mathcal{L}^1(\mu) : \|f\|_1 = 0\} = \{f \text{ measurable} : f = 0 \text{ a.e.}(\mu)\}$. Put an equivalence relation on $\mathcal{L}^1(\mu)$ by $f \sim g$ if $f = g$ a.e.(μ); i.e. if $f - g \in \mathcal{N}$. The normed vector space space $L^1(\mu) = \mathcal{L}^1/\mathcal{N}$ consists of the equivalence classes with the induced norm $\|[f]\|_1 = \|f\|_1$. By abuse of notation, we will frequently write $f \in L^1(\mu)$ when we mean $[f]$.

3.4.2. THEOREM. *$L^1(\mu)$ is a complete normed vector space (a Banach space).*

PROOF. It is easy to check that $\mathcal{N}$ is a subspace of the vector space $\mathcal{L}^1(\mu)$, so that the quotient $L^1(\mu) = \mathcal{L}^1/\mathcal{N}$ is a vector space. Notice that by Proposition 3.3.2, $f = g$ a.e.(μ) if and only if $f - g \in \mathcal{N}$ if and only if $\|f - g\|_1 = 0$. If $[f]$ is an element of $L^1(\mu)$ and $f \sim f'$, then $h = f' - f = 0$ a.e. Thus $\int |f'|\,d\mu = \int |f|\,d\mu$. So $\|[f]\|_1$ is well defined. The construction has quotiented out all non-zero elements of zero norm, and thus the norm on $L^1(\mu)$ is positive definite. It is easily seen to be positive homogeneous. The triangle inequality follows immediately from the triangle inequality for $\mathcal{L}^1(\mu)$. Thus $L^1(\mu)$ is a normed vector space.

Now suppose that $([f_n])_{n\ge 1}$ is a Cauchy sequence in $L^1(\mu)$. Choose representatives $f_n \in [f_n]$ so that we can evaluate them at points. For each $k \ge 1$, there is an n_k so that if $m, n \ge n_k$, then $\|f_m - f_n\|_1 < 2^{-k}$. In particular, $\|f_{n_{k+1}} - f_{n_k}\|_1 < 2^{-k}$. Let $g = |f_{n_1}| + \sum_{k\ge 1} |f_{n_{k+1}} - f_{n_k}|$. By the MCT,

$$\begin{aligned}\|g\|_1 = \int g\,d\mu &= \int |f_{n_1}|\,d\mu + \sum_{k\ge 1} \int |f_{n_{k+1}} - f_{n_k}|\,d\mu \\ &= \|f_{n_1}\|_1 + \sum_{k\ge 1} \|f_{n_{k+1}} - f_{n_k}\|_1 < \|f_{n_1}\|_1 + \sum_{k\ge 1} 2^{-k} < \infty.\end{aligned}$$

Hence g is integrable, and therefore is finite a.e.(μ), say on $X \setminus N$ for $\mu(N) = 0$. It follows that the sum

$$f(x) = f_{n_1}(x) + \sum_{k\ge 1} f_{n_{k+1}}(x) - f_{n_k}(x) = \lim_{k\to\infty} f_{n_k}(x)$$

converges absolutely on $X \setminus N$. We define $f|_N = 0$. Moreover $|f| \le g$ so that f is integrable, and $f = \lim_{k\to\infty} f_{n_k}$ a.e.(μ). Also

$$|f_{n_k}| \le \|f_{n_1}\|_1 + \sum_{i=1}^{k-1} \|f_{n_{i+1}} - f_{n_i}\| \le g.$$

Therefore $|f - f_{n_k}| \le 2g$; and so by LDCT,

$$0 = \|f - f\|_1 = \lim_{k\to\infty} \|f - f_{n_k}\|_1.$$

That is, f_{n_k} converges to f in $L^1(\mu)$. Finally a standard argument shows that the whole Cauchy sequence (f_n) converges to f in norm. ∎

3.5. Comparison with the Riemann Integral

We will do a speedy review of part of the Riemann integral theory. Given a bounded real valued function f on a finite interval $[a, b]$, we consider a partition $\mathcal{P} = \{a = x_0 < x_1 < \cdots < x_n = b\}$ of $[a, b]$. Define real numbers

$$M_j = \sup_{x_{j-1}\le x\le x_j} f(x) \quad\text{and}\quad m_j = \inf_{x_{j-1}\le x\le x_j} f(x) \quad\text{for}\quad 1 \le j \le n.$$

Then

$$l_{\mathcal{P}}(x) = m_1\chi_{\{a\}} + \sum_{j=1}^{n} m_j\chi_{(x_{j-1},x_j]} \le f \le M_1\chi_{\{a\}} + \sum_{j=1}^{n} M_j\chi_{(x_{j-1},x_j]} = u_{\mathcal{P}}(x).$$

The upper and lower sums are

$$L(f,\mathcal{P}) = \sum_{j=1}^{n} m_j(x_j - x_{j-1}) \quad \text{and} \quad U(f,\mathcal{P}) = \sum_{j=1}^{n} M_j(x_j - x_{j-1}).$$

If $\mathcal{P}$ and $\mathcal{Q}$ are two partitions, let $\mathcal{P} \vee \mathcal{Q}$ denote the partition using the points in $\mathcal{P} \cup \mathcal{Q}$. It is easy to show that

$$L(f,\mathcal{P}) \le L(f,\mathcal{P}\vee\mathcal{Q}) \le U(f,\mathcal{P}\vee\mathcal{Q}) \le U(f,\mathcal{Q}).$$

It follows that

$$\sup_{\mathcal{P}} L(f,\mathcal{P}) \le \inf_{\mathcal{Q}} U(f,\mathcal{Q}).$$

A function f is *Riemann integrable* if for every $\varepsilon > 0$, there is a partition $\mathcal{P}$ so that

$$U(f,\mathcal{P}) - L(f,\mathcal{P}) = \sum_{j=1}^{n} (M_j - m_j)(x_j - x_{j-1}) < \varepsilon.$$

In this case, one defines

$$\int_a^b f(x)\,dx := \sup_{\mathcal{P}} L(f,\mathcal{P}) = \inf_{\mathcal{Q}} U(f,\mathcal{Q}).$$

If one defines the *mesh* of a partition as $\operatorname{mesh}(\mathcal{P}) = \max_{1\le j\le n} x_j - x_{j-1}$, then for any $\varepsilon > 0$, there is a $\delta > 0$ so that if $\operatorname{mesh}(\mathcal{P}) < \delta$, then $U(f,\mathcal{P}) - L(f,\mathcal{P}) < \varepsilon$. Hence we can take a nested sequence of partitions $\mathcal{P}_1 \subset \mathcal{P}_n \subset \mathcal{P}_{n+1} \subset \dots$ with $\operatorname{mesh}\mathcal{P}_n \to 0$ and conclude that

$$\int_a^b f(x)\,dx = \lim_{n\to\infty} L(f,\mathcal{P}_n) = \lim_{n\to\infty} U(f,\mathcal{P}_n).$$

3.5.1. THEOREM. *Every Riemann integrable function f on $[a,b]$ is Lebesgue integrable, and $\int f\,dm = \int_a^b f(x)\,dx$.*

PROOF. Notice that a Riemann integrable function is approximated from above and below by a special class of simple functions, the piecewise constant functions, $l_{\mathcal{P}} \le f \le u_{\mathcal{P}}$. Moreover the Lebesgue and Riemann integrals agree on these piecewise constant functions. Take a nested sequence of partitions $\mathcal{P}_n$ as above with $\operatorname{mesh}\mathcal{P}_n \to 0$ and observe that

$$l_{\mathcal{P}_n} \le l_{\mathcal{P}_{n+1}} \le f \le u_{\mathcal{P}_{n+1}} \le u_{\mathcal{P}_n}.$$

Define $l(x) = \lim_{n\to\infty} l_{\mathcal{P}_n}(x)$ and $u(x) = \lim_{n\to\infty} u_{\mathcal{P}_n}(x)$. Note that $l \le f \le u$. These functions are Lebesgue measurable. Moreover, $l_{\mathcal{P}_n} - L_{\mathcal{P}_1} \ge 0$ and increase to $l - l_{\mathcal{P}_1}$. Since $l_{\mathcal{P}_1}$ is integrable, the MCT shows that

$$\begin{aligned}\int l\,dm &= \int l_{\mathcal{P}_1}\,dm + \lim_{n\to\infty}\int l_{\mathcal{P}_n} - l_{\mathcal{P}_1}\,dm \\ &= \lim_{n\to\infty}\int l_{\mathcal{P}_n}\,dm = \lim_{n\to\infty} L(f,\mathcal{P}_n) = \int_a^b f(x)\,dx.\end{aligned}$$

Similarly, $\displaystyle\int u\,dm = \int_a^b f(x)\,dx$. In particular, $\displaystyle\int u - l\,dm = 0$. By Lemma 3.2.5, $u - l = 0$ a.e.(m). Therefore $l = f = u$ a.e.(m), so that f is measurable and

$$\int f\,dm = \int u\,dm = \int_a^b f(x)\,dx. \qquad \blacksquare$$

3.5.2. REMARKS. There is one situation where the Riemann integral can do something that the Lebesgue integral can't. That is the improper Riemann integrals in which a function which is not Riemann integrable and be integrated as a limit. The typical example is $\displaystyle\int_0^\infty \frac{\sin x}{x}\,dx$. The integrand extends to be continuous at $x = 0$, so that is not an issue. In the Riemann theory, this integral is evaluated as

$$\int_0^\infty \frac{\sin x}{x}\,dx = \lim_{r\to\infty}\int_0^r \frac{\sin x}{x}\,dx.$$

This can be shown by a number of techniques to equal $\frac{\pi}{2}$. The reason it is not Lebesgue integrable is that this integral exists only as a conditional limit, and

$$\int_0^\infty \left|\frac{\sin x}{x}\right| dx = \lim_{r\to\infty}\int_0^r \left|\frac{\sin x}{x}\right| dx = +\infty.$$

So this function is neither Lebesgue nor Riemann integrable. However if a function f is Lebesgue integrable, then so is $|f|$.

Let's examine this example in more detail.

$$\int_0^\infty \frac{\sin x}{x}\,dx = \sum_{n=0}^\infty \int_{n\pi}^{(n+1)\pi} \frac{\sin x}{x}\,dx = \sum_{n=0}^\infty (-1)^n a_n$$

where for $n \ge 1$,

$$a_n = \int_{n\pi}^{(n+1)\pi} \frac{|\sin x|}{x}\,dx \le \frac{1}{n\pi}\int_0^\pi \sin x\,dx = \frac{2}{n\pi}$$

and

$$a_n = \int_{n\pi}^{(n+1)\pi} \frac{|\sin x|}{x}\,dx \ge \frac{1}{(n+1)\pi}\int_0^\pi \sin x\,dx = \frac{2}{(n+1)\pi}.$$

It follows that $a_n \geq \frac{2}{(n+1)\pi} \geq a_{n+1}$ for $n \geq 1$; so $a_n \to 0$ monotonely. Therefore $\sum_{n=0}^{\infty}(-1)^n a_n$ converges by the alternating series test. It also shows that

$$\int_0^\infty \Big|\frac{\sin x}{x}\Big|\,dx = \sum_{n=0}^{\infty}\int_{n\pi}^{(n+1)\pi}\frac{|\sin x|}{x}\,dx = \sum_{n=0}^{\infty} a_n.$$

Since $a_n \geq \frac{2}{(n+1)\pi}$ for $n \geq 1$, this series diverges by comparison with the harmonic series.

The other place where an improper Riemann integral is requires is the integration of unbounded functions. The Riemann theory works only for bounded functions. A function like $f(x) = x^{-a}$ for $0 < a < 1$ has an improper Riemann integral on $[0,1]$. Since this function is positive and continuous with a finite integral, it is Lebesgue integrable. It is possible again to construct a function with alternating sign on $(0,1]$ that blows up near 0 so that the improper Riemann integral exists as

$$\int_0^1 f(x)\,dx = \lim_{\varepsilon\to 0}\int_\varepsilon^1 f(x)\,dx$$

but so that the absolute value is not integrable. An example is $\displaystyle\int_0^1 \frac{1}{x}\sin\frac{1}{x}\,dx$, which after the change of variables $u = \frac{1}{x}$ converts this to the integral $\displaystyle\int_1^\infty \frac{\sin u}{u}\,du$ which we have already analyzed.

When you have to integrate specific functions, you will usually find yourself falling back on the many techniques that have been developed for the Riemann integral. Almost all explicit functions that you will need to integrate will be Riemann integrable or at least locally Riemann integrable.

The power of the Lebesgue integral lies in the limit theorems MCT and LDCT. These are much stronger than anything available in the Riemann theory. Also we obtain the completeness of $L^1(\mu)$ and many other related spaces that allow the power of functional analysis to apply.

There is a nice characterization of Riemann integrable functions due to Lebesgue.

3.5.3. THEOREM. *If $f : [a,b] \to \mathbb{R}$ is a bounded function, then f is Riemann integrable if and only if f is continuous except on a set of measure 0.*

PROOF. Define functions

$$U(x) = \inf_{\delta>0}\sup_{|y-x|<\delta} f(y) \quad\text{and}\quad L(x) = \sup_{\delta>0}\inf_{|z-x|<\delta} f(z).$$

It is an easy exercise to show that f is continuous at x if and only if $L(x) = U(x)$. The quantity $\omega(f,x) = U(x) - L(x)$ is called the *oscillation* of f at x.

Suppose that f is Riemann integrable. We use the notation from the proof of Theorem 3.5.1. Choose an increasing family of partitions $\mathcal{P}_n$ with $\operatorname{mesh}(\mathcal{P}_n) \to 0$, and define $l_{\mathcal{P}_n}, l, u_{\mathcal{P}_n}$ and u as before. Let

$$A = \{x : l(x) \neq u(x)\} \cup \bigcup_{n \geq 1} \mathcal{P}_n.$$

The set $\{x : l(x) \neq u(x)\}$ has measure 0, and $\bigcup_{n \geq 1} \mathcal{P}_n$ is countable. Therefore $m(A) = 0$. Moreover

$$\lim_{n \to \infty} u_{\mathcal{P}_n}(x) - l_{\mathcal{P}_n}(x) = u(x) - l(x) = 0 \quad \text{for} \quad x \in [a,b] \setminus A.$$

If $x \in [a,b] \setminus A$, choose n so that $u_{\mathcal{P}_n}(x) - l_{\mathcal{P}_n}(x) < \varepsilon$. Then there are adjacent points $x_{i-1} < x_i$ in $\mathcal{P}_n$ so that $x_{i-1} < x < x_i$. Let $\delta = \min\{x - x_{i-1}, x_i - x\}$. Then since $u_{\mathcal{P}_n}$ and $l_{\mathcal{P}_n}$ are constant on $(x_{i-1}, x_i]$ and are upper and lower bounds for f, respectively, it follows that

$$U(x) - L(x) \leq \sup_{\substack{|y-x|<\delta \\ |z-x|<\delta}} f(y) - f(z) \leq u_{\mathcal{P}_n}(x) - l_{\mathcal{P}_n}(x) < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we have $U(x) = L(x)$, and thus f is continuous on $[a,b] \setminus A$, which is a.e.(m).

Conversely suppose that f is continuous except on a set A of Lebesgue measure 0. Then $U(x) = L(x)$ for $x \in [a,b] \setminus A$. Note that $A = \bigcup_{n \geq 1} A_n$ where $A_n = \{x \in [a,b] : U(x) - L(x) = \omega(f,x) \geq 2^{-n}\}$. The set A_n is closed because its complement is open: if $\omega(f,x) < r$, then there is a $\delta > 0$ so that

$$\sup_{\substack{|y-x|<\delta \\ |z-x|<\delta}} f(y) - f(z) < r. \quad \text{So if } |x' - x| = d < \delta, \qquad \sup_{\substack{|y-x'|<\delta-d \\ |z-x'|<\delta-d}} f(y) - f(z) < r.$$

Thus $\omega(f,x') < r$ as well. Cover A_n with a countable family of open intervals of total length at most 2^{-n}. Then since A_n is compact, there is a finite subcover $I_1, \dots, I_m$. Let $B = [a,b] \setminus \bigcup_{i=1}^m I_i$. For each $x \in B$, since $U(x) - L(x) < 2^{-n}$, there is an open inteval $J_x \ni x$ so that the oscillation over $\overline{J_x}$ is less than 2^{-n}. The collection $\{J_x : x \in B\}$ is an open cover of the compact set B. Therefore there is a finite subcover $J_{x_1}, \dots, J_{x_p}$.

Let $\mathcal{P}$ be the partition consisting of the endpoints of $I_1, \dots, I_m, J_{x_1}, \dots, J_{x_p}$ (together with a, b if necessary). Any of the intervals $[x_{i-1}, x_i]$ contained in the union $\bigcup_{j=1}^p \overline{J_{x_j}}$ will have oscillation less than 2^{-n}, meaning $u_{\mathcal{P}_n}(x) - l_{\mathcal{P}_n}(x) < 2^{-n}$ on $(x_{i-1}, x_i]$. The remaining intervals are contained in $\bigcup_{i=1}^m \overline{I_i}$ and thus have total length at most 2^{-n}. The only thing we can say is that $u_{\mathcal{P}_n}(x) - l_{\mathcal{P}_n}(x) \leq 2\|f\|_\infty$ on these intervals. Therefore we obtain the estimate

$$\begin{aligned} U(f, \mathcal{P}_n) - L(f, \mathcal{P}_n) &< 2^{-k} \sum_{j=1}^p m(J_{x_j}) + 2\|f\|_\infty \sum_{i=1}^m m(I_i) \\ &< 2^{-k} \big(b - a + 2\|f\|_\infty\big). \end{aligned}$$

As $n \to \infty$, this converges to 0. Hence f is Riemann integrable. ■

3.6. Product Measures

3.6.1. DEFINITION. If X_λ for $\lambda \in \Lambda$ are non-empty sets, then the *product space* is

$$X = \prod_{\lambda\in\Lambda} X_\lambda = \{(x_\lambda) : x_\lambda \in X_\lambda,\ \lambda \in \Lambda\} = \{f : \Lambda \to \dot\cup_{\lambda\in\Lambda} X_\lambda : f(\lambda) = x_\lambda \in X_\lambda\}.$$

The maps $\pi_\lambda : X \to X_\lambda$ given by $\pi_\lambda(x) = x_\lambda$ are the *coordinate projections*.

If $(X_\lambda, \mathcal{B}_\lambda)$ are σ-algebras, then the *product σ-algebra* $(\prod_{\lambda\in\Lambda} X_\lambda, \bigotimes_{\lambda\in\Lambda} \mathcal{B}_\lambda)$ is the σ-algebra of subsets of X generated by the sets $\{\pi_\lambda^{-1}(A) : A \in \mathcal{B}_\lambda.\ \lambda \in \Lambda\}$.

3.6.2. REMARK. When $\Lambda = \{1, \dots, n\}$ is finite, there is no problem defining the product space. In this case, the product σ-algebra is generated by all of the "cubes" $A_1 \times \cdots \times A_n$ for $A_i \in \mathcal{B}_i$, $1 \le i \le n$.

When Λ is infinite, the Axiom of Choice is often needed to guarantee that the product space is non-empty. In this case, the product σ-algebra is generated by sets $\pi_\lambda^{-1}(A_\lambda) = A_\lambda \times \prod_{\mu\in\Lambda\setminus\{\lambda\}} X_\mu$. The intersection of finitely many yields a cube of the form $A_{\lambda_1} \times \cdots \times A_{\lambda_n} \times \prod_{\mu\in\Lambda\setminus\{\lambda_i : 1\le i\le n\}} X_\mu$. One can also take the intersection of countably many such slices, but if Λ is uncountable, the cubes formed with proper subsets from each X_λ will generally not belong to the product σ-algebra. Even when Λ is countable, these infinite cubes many turn out to be measure 0 and hence negligible.

The following proposition shows that in the familiar case of Borel sets on metric spaces, and finite products, we get the desired result. The separability hypothesis is crucial.

3.6.3. PROPOSITION. *If (X_i, d_i) are separable metric spaces for $1 \le i \le n$, then*

$$\bigotimes_{i=1}^n \mathrm{Bor}(X_i) = \mathrm{Bor}(\prod_{i=1}^n X_i).$$

PROOF. The product space $X = \prod_{i=1}^n X_i$ is also a metric space with the metric $d((x_i), (y_i)) = \max\{d_i(x_i, y_i)\}$. With this (or any equivalent) metric, the coordinate projections are continuous. Thus if U_i is open in X_i, the set $\pi_i^{-1}(U_i)$ is open in X. These sets generate $\bigotimes_{i=1}^n \mathrm{Bor}(X_i)$, and thus it is contained in $\mathrm{Bor}(X)$.

Conversely, since each X_i is separable, so is X. Therefore X is second countable. Indeed if $\{x_j : j \ge 1\}$ is a dense subset, then $\{b_r(x_j) : j \ge 1,\ r \in \mathbb{Q}_+\}$ is a countable neighbourhood base for X. Now $b_r(x_j) = \prod_{i=1}^n b_r(x_{j,i})$, where $\pi_i(x_j) = x_{j,i}$, belongs to $\bigotimes_{i=1}^n \mathrm{Bor}(X_i)$. As a σ-algebra is closed under countable unions and every open set in X is the union of those (countably many) sets in the neighbourhood base that it contains, it follows that every open subset of X belongs to $\bigotimes_{i=1}^n \mathrm{Bor}(X_i)$. Thus $\mathrm{Bor}(X)$ is contained in $\bigotimes_{i=1}^n \mathrm{Bor}(X_i)$. ■

3.6.4. COROLLARY.

$$\mathrm{Bor}(\mathbb{R}^n) = \bigotimes_{i=1}^{n} \mathrm{Bor}(\mathbb{R}) \quad \textit{and} \quad \mathrm{Bor}(\mathbb{C}) = \mathrm{Bor}(\mathbb{R}) \otimes \mathrm{Bor}(\mathbb{R}).$$

Now let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{B}', \nu)$ be two measure spaces. Let $\mathcal{A}$ be the collection of all finite unions of disjoint rectangles of the form $A \times B$ for $A \in \mathcal{B}$ and $B \in \mathcal{B}'$. This is an algebra since it is closed under finite unions and complements. Indeed, $(A \times B)^c = A^c \times Y \,\dot\cup\, A \times B^c$ and

$$A_1 \times B_1 \,\cup\, A_2 \times B_2 = A_1 \times B_1 \,\dot\cup\, (A_2 \setminus A_1) \times B_2 \,\dot\cup\, (A_1 \setminus A_2) \times (B_2 \setminus B_1).$$

Now define a set function

$$\pi\Big(\dot{\bigcup}_{i=1}^{n} A_i \times B_i\Big) = \sum_{i=1}^{n} \mu(A_i)\nu(B_i).$$

To apply Carathéodory's Theorem, we need the following lemma.

3.6.5. LEMMA. *π is a premeasure on $\mathcal{A}$.*

PROOF. We need to show that if $A \times B = \dot{\bigcup}_{i \geq 1} A_i \times B_i$, then

$$\pi(A \times B) = \mu(A)\nu(B) = \sum_{i \geq 1} \mu(A_i)\nu(B_i) = \sum_{i \geq 1} \pi(A_i \times B_i).$$

Note that

$$\chi_A(x)\chi_B(y) = \sum_{i \geq 1} \chi_{A_i}(x)\chi_{B_i}(y).$$

If we fix $y \in Y$, we obtain a sum of non-negative measurable functions on X. By the Monotone Convergence Theorem,

$$\begin{aligned} \mu(A)\chi_B(y) &= \int \chi_A(x)\, d\mu(x)\, \chi_B(y) \\ &= \sum_{i \geq 1} \int \chi_{A_i}(x)\, d\mu(x)\, \chi_{B_i}(y) = \sum_{i \geq 1} \mu(A_i)\chi_{B_i}(y). \end{aligned}$$

These functions are non-negative measurable functions on Y, so a second application of the MCT yields

$$\begin{aligned} \mu(A)\nu(B) &= \mu(A) \int \chi_B(y)\, d\nu(y) \\ &= \sum_{i \geq 1} \mu(A_i) \int \chi_{B_i}(y)\, d\nu(y) = \sum_{i \geq 1} \mu(A_i)\nu(B_i). \end{aligned}$$

Hence π is a premeasure. ■

Now we can apply Carathéodory's Theorem via Theorem 1.4.2 to obtain the following, called the *product measure*.

3.6.6. THEOREM. *Let $(X,\mathcal{B},\mu)$ and $(Y,\mathcal{B}',\nu)$ be two measure spaces. There is a complete measure $(X\times Y,\overline{\mathcal{B}\otimes\mathcal{B}'},\mu\times\nu)$ on $X\times Y$ such that $\overline{\mathcal{B}\otimes\mathcal{B}'}\supset\mathcal{B}\otimes\mathcal{B}'$ and $\mu\times\nu(A\times B)=\mu(A)\nu(B)$ for all $A\in\mathcal{B}$ and $B\in B'$.*

We need some more refined information about the sets in $\overline{\mathcal{B}\otimes\mathcal{B}'}$.

3.6.7. DEFINITION. Let $\mathcal{A}$ be the algebra of finite unions of rectangle in $\mathcal{B}\otimes\mathcal{B}'$ as above. Define

$$\mathcal{A}_\sigma=\big\{E=\bigcup_{i\ge1}A_i : A_i\in\mathcal{A}\big\}$$

and

$$\mathcal{A}_{\sigma\delta}=\big\{G=\bigcap_{j\ge1}E_j : E_j\in\mathcal{A}_\sigma\big\}.$$

3.6.8. LEMMA. *If $E\in\overline{\mathcal{B}\otimes\mathcal{B}'}$ and $\mu\times\nu(E)<\infty$, then there is a set $G\in\mathcal{A}_{\sigma\delta}$ such that $E\subset G$ and $\mu\times\nu(G\setminus E)=0$.*

PROOF. By definition of the outer measure π^* which defines $\mu\times\nu$, we have that

$$\mu\times\nu(E)=\inf\Big\{\sum_{i\ge1}\mu\times\nu(A_i) : A_i\in\mathcal{A}\text{ and }E\subset\bigcup_{i\ge1}A_i\Big\}.$$

Thus we can choose $E_j=\bigcup_{i\ge1}A_{j,i}\supset E$ in $\mathcal{A}_\sigma$ so that $\mu\times\nu(E_j)<\mu\times\nu(E)+\frac{1}{j}$. Hence $\mu\times\nu(E_j\setminus E)<\frac{1}{j}$. Define $G=\bigcap_{j\ge1}E_j\in\mathcal{A}_{\sigma\delta}$. Then $E\subset G$ and $\mu\times\nu(G\setminus E)=0$. ■

3.7. Product Integration

An important result from the Riemann theory for integration in multiple variables is that integration over a nice region in $\mathbb{R}^n$ is equal to an iterated integral in which the integration is done one variable at a time. There is an important analogue of this for general measures. It actually comes in two flavours.

3.7.1. FUBINI'S THEOREM. *Let $(X,\mathcal{B},\mu)$ and $(Y,\mathcal{B}',\nu)$ be complete measures. If $f\in L^1(\mu\times\nu)$, then*

(1) (i) $f_x(y) = f(x, y) \in L^1(\nu)$ *for* a.e. $x(\mu)$.

(ii) $f^y(x) = f(x, y) \in L^1(\mu)$ *for* a.e. $y(\nu)$.

(2) (i) $\displaystyle\int_Y f_x(y)\, d\nu =: F(x) \in L^1(\mu)$.

(ii) $\displaystyle\int_X f^y(x)\, d\mu =: G(y) \in L^1(\nu)$.

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

3.7.2. Tonelli's Theorem. *Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{B}', \nu)$ be complete measures, and suppose that $\mu \times \nu$ is σ-finite. If $f \in L^+(\mu \times \nu)$, then*

(1) (i) $f_x(y) = f(x, y) \in L^+(\nu)$ *for* a.e. $x(\mu)$.

(ii) $f^y(x) = f(x, y) \in L^+(\mu)$ *for* a.e. $y(\nu)$.

(2) (i) $\displaystyle\int_Y f_x(y)\, d\nu =: F(x) \in L^+(\mu)$.

(ii) $\displaystyle\int_X f^y(x)\, d\mu =: G(y) \in L^+(\nu)$.

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

3.7.3. Remarks. (1) The hypothesis that $\mu \times \nu$ is σ-finite in Tonelli's Theorem is critical, as an example will show. However it is not needed for Fubini's Theorem because when $f \in L^1(\mu \times \nu)$, we have that $\displaystyle\int |f|\, d\mu \times \nu = L < \infty$. It follows that $C_n = \{(x, y) \in X \times Y : |f(x, y)| \geq \frac{1}{n}\}$ has $\mu \times \nu(C_n) \leq nL$. Thus the restriction of $\mu \times \nu$ to $C = \{(x, y) \in X \times Y : f(x, y) \neq 0\} = \bigcup_{n \geq 1} C_n$ is σ-finite.

(2) Also the hypothesis that $f \in L^1(\mu \times \nu)$ or $f \in L^+(\mu \times \nu)$ is also critical as examples will show. The failure to check this condition leads to common misuses of these theorems.

(3) Normal practice is to omit the parentheses and if there is no confusion, also the variables, in multiple integrals. So we write

$$\iint f\, d\nu\, d\mu \quad \text{or} \quad \iint f(x, y) d\nu(y)\, d\mu(x) \quad \text{for} \quad \int_X \Big(\int_Y f(x, y) d\nu(y) \Big)\, d\mu(x).$$

If $E \subset X \times Y$, define $E_x = \{y \in Y : (x, y) \in E\}$ for $x \in X$; and let $E^y = \{x \in X : (x, y) \in E\}$ for $y \in Y$.

3.7.4. Lemma. *If $E \in \mathcal{A}_{\sigma\delta}$ and $\mu \times \nu(E) < \infty$, then $g(x) = \nu(E_x)$ is μ-measurable, $g \in L^+ \cap L^1(\mu)$, and $\int g\,d\mu = \mu \times \nu(E)$. Similarly, $h(y) = \mu(E^y)$ is ν-measurable, $h \in L^+ \cap L^1(\nu)$, and $\int h\,d\nu = \mu \times \nu(E)$.*

Proof. If $E = A \times B$ for $A \in \mathcal{B}$ and $B \in \mathcal{B}'$, then $E_x = \begin{cases} B & \text{if} \quad x \in A \\ \varnothing & \text{if} \quad x \notin A. \end{cases}$
Therefore $g(x) = \nu(B)\chi_A$ is μ-measurable and $g \geq 0$. Moreover,

$$\int g\,d\mu = \int \nu(B)\chi_A\,d\mu = \mu(A)\nu(B) = \mu \times \nu(E) < \infty.$$

Thus $g \in L^+ \cap L^1(\mu)$.

Next, if $E \in \mathcal{A}_\sigma$, we can write $E = \bigcup_{i\geq 1} A_i \times B_i$ for $A_i \in \mathcal{B}$ and $B_i \in \mathcal{B}'$. We can rewrite this is a disjoint union because $A_n \times B_n \setminus \bigcup_{i=1}^{n-1} A_i \times B_i \in \mathcal{A}$ and thus can be written as a finite disjoint union of rectangles, which are clearly also disjoint from $\bigcup_{i=1}^{n-1} A_i \times B_i$. Proceeding recursively, E can be rewritten as a disjoint union $E = \bigcup_{i\geq 1} C_i \times D_i$ for $C_i \in \mathcal{B}$ and $D_i \in \mathcal{B}'$. Let $g(x) = \nu(E_x)$ and $g_i(x) = \nu(C_i)\chi_{D_i}$. Then $g(x) = \sum_{i\geq 1} g_i(x)$. Hence g is measurable, and thus belongs to $L^+(\mu)$. By the MCT,

$$\int g\,d\mu = \sum_{i\geq 1} \int g_i\,d\mu = \sum_{i\geq 1} \mu \times \nu(C_i \times D_i) = \mu \times \nu(E) < \infty.$$

Moreover, we now have $g \in L^+ \cap L^1(\mu)$.

Finally, suppose that $E = \bigcap_{n\geq 1} E_n$ where $E_n \supset E_{n+1}$ all lie in $\mathcal{A}_\sigma$ such that $\mu \times \nu(E_1) < \infty$ and $\mu \times \nu(E_n) \downarrow \mu \times \nu(E)$. Let $g_n(x) = \nu((E_n)_x)$ and $g(x) = \nu(E_x)$. Then

$$0 \leq g \leq g_{n+1} \leq g_n \leq g_1;$$

and $g(x) = \inf g_n(x) = \lim g_n(x)$. Hence g is μ-measurable. It is dominated by g_1 which is integrable, so $g \in L^+ \cap L^1(\mu)$. By the LDCT,

$$\int g\,d\mu = \lim_{n\to\infty} \int g_n\,d\mu = \lim_{n\to\infty} \mu \times \nu(E_n) = \mu \times \nu(E).$$

The last equality follows by continuity from above since $\mu \times \nu(E_1) < \infty$.

Similarly we can interchange the role of x and y. ■

3.7.5. Lemma. *If $E \in \overline{\mathcal{B} \otimes \mathcal{B}'}$ has $\mu \times \nu(E) = 0$, then $\nu(E_x) = 0$ a.e.(μ) and $\mu(E^y) = 0$ a.e.(ν).*

Proof. By Lemma 3.6.8, there is a set $G \in \mathcal{A}_{\sigma\delta}$ such that $E \subset G$ and $\mu \times \nu(G) = 0$. Therefore by the previous lemma, $f(x) = \nu(G_x) \in L^+ \cap L^1(\mu)$ and $\int f\,d\mu = 0$. Therefore $f = 0$ a.e.(μ). Now $E_x \subset G_x$, and since ν is a complete measure, $g(x) = \nu(E_x) = 0$ a.e.(μ). It could happen that E_x is not measurable on a set of ν-measure 0. A function which differs from a measurable function on a set of measure 0 is also measurable.

Similarly we can interchange the role of x and y. ■

3.7.6. COROLLARY. *If $E \in \overline{\mathcal{B} \otimes \mathcal{B}'}$ and $\mu \times \nu(E) < \infty$, then E_x is ν-measurable for* a.e.(μ) $x \in X$, *$g(x) = \nu(E_x)$ is μ-measurable, $g \in L^+ \cap L^1(\mu)$, and $\displaystyle\int g\,d\mu = \mu \times \nu(E)$. Similarly, E^y is μ-measurable for* a.e.(ν) $y \in Y$, *$h(y) = \mu(E^y)$ is ν-measurable, $h \in L^+ \cap L^1(\nu)$, and $\displaystyle\int h\,d\nu = \mu \times \nu(E)$.*

PROOF. Find $G \in \mathcal{A}_{\sigma\delta}$ so that $E \subset G$ and $\mu \times \nu(G \setminus E) = 0$. Then by Lemma 3.7.4, $f(x) = \nu(G_x) \in L^+ \cap L^1(\mu)$, and $\displaystyle\int f\,d\mu = \mu \times \nu(G)$. By Lemma 3.7.5, $0 \le f - g = 0$ a.e.(μ). The result follows.

Similarly we can interchange the role of x and y. ■

Now we are ready to prove the theorems.

PROOF OF FUBINI'S THEOREM. The various integrals and iterated integrals are linear operations, so we can write $f \in L^1(\mu \times \nu)$ as $f = f_1 - f_2 + i f_3 - i f_4$ where $f_i \in L^+ \cap L^1(\mu \times \nu)$ by letting $f_1 = \operatorname{Re} f \vee 0$, etc. So we may suppose that $f \in L^+ \cap L^1(\mu \times \nu)$. There are simple functions φ_n with $0 \le \varphi_n \le \varphi_{n+1} \le f$ so that $f = \lim_{n\to\infty} \varphi_n$ and each φ_n is supported on a set of finite measure (such as the set C_n in Remark 3.7.3(1)).

By Corollary 3.7.6 and linearity, Fubini's Theorem is valid for simple functions with finite measure support. By the MCT,

$$\int f\,d\mu \times \nu = \lim_{n\to\infty} \int \varphi_n\,d\mu \times \nu = \lim_{n\to\infty} \int_X \int_Y \varphi_n(x,y) d\nu(y)\,d\mu(x).$$

The functions $F_n(x) = \displaystyle\int_Y \varphi_n(x,y) d\nu(y)$ are positive, measurable and monotone increasing to $F(x) = \displaystyle\int_Y f(x,y) d\nu(y)$. Hence this is a measurable function, and by MCT,

$$\lim_{n\to\infty} \int_X F_n(x)\,d\mu(x) = \int_X F(x)\,d\mu(x).$$

Thus

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

Interchanging the role of X and Y yields the other iterated integral. ■

PROOF OF TONELLI'S THEOREM. Let $f \in L^+(\mu \times \nu)$. Since $\mu \times \nu$ is σ-finite, there are measurable sets $C_n \subset C_{n+1}$ with $\mu \times \nu(C_n) < \infty$ and $X \times Y = \bigcup_{n \ge 1} C_n$.

For $n \geq 1$, let $f_n = (f \wedge n)\chi_{C_n}$, where $(f \wedge n)(x) = \min\{f(x), n\}$. Then $0 \leq f_n \leq n\chi_{C_n}$, and so $f_n \in L^+ \cap L^1(\mu \times \nu)$. Now by Fubini's Theorem and MCT, we have

$$\int f \, d\mu \times \nu = \lim_{n\to\infty} \int f_n \, d\mu \times \nu = \lim_{n\to\infty} \int_X \int_Y f_n(x,y) d\nu(y) \, d\mu(x).$$

The proof is completed as before. The functions $F_n(x) = \displaystyle\int_Y f_n(x,y) d\nu(y)$ are positive, measurable and monotone increasing to $F(x) = \displaystyle\int_Y f(x,y) d\nu(y)$. Hence this is a measurable function, and by MCT,

$$\lim_{n\to\infty} \int_X F_n(x) \, d\mu(x) = \int_X F(x) \, d\mu(x).$$

Thus

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

Interchanging the role of X and Y yields the other iterated integral. ∎

There is a straightforward variant of these results when the measures are not complete. The difference is that given measures $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{B}', \nu)$, we form $\mu \times \nu$ and restrict it to the σ-algebra $\mathcal{B} \otimes \mathcal{B}'$. I will also call this $\mu \times \nu$.

3.7.7. Fubini-Tonelli Theorem without Completeness. *Let $(X, \mathcal{B}, \mu)$ and $(Y, \mathcal{B}', \nu)$ be measures with product $(X \times Y, \mathcal{B} \otimes \mathcal{B}', \mu \times \nu)$. If $f \in L^1(\mu \times \nu)$ (or if $\mu \times \nu$ is σ-finite and $f \in L^+(\mu \times \nu)$), then*

(1) (i) $f_x(y) = f(x,y) \in L^1(\nu)$ *(or $L^+(\nu)$) for all* $x \in X$

(ii) $f^y(x) = f(x,y) \in L^1(\mu)$ *(or $L^+(\mu)$) for all* $y \in Y$.

(2) (i) $\displaystyle\int_Y f_x(y) \, d\nu =: F(x) \in L^1(\mu)$ *(or $L^+(\mu)$).*

(ii) $\displaystyle\int_X f^y(x) \, d\mu =: G(y) \in L^1(\nu)$ *(or $L^+(\nu)$).*

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

In Assignment 4, Q5, you showed that if $E \in \mathcal{B} \otimes \mathcal{B}'$, then $E_x \in \mathcal{B}'$ and $E^y \in \mathcal{B}$ for all $x \in X$ and $y \in Y$. This extends to simple functions, so that (1i) and (1ii) hold for simple functions. This extends to limits, and since every measurable function is a limit of simple functions, we obtain (1i) and (1ii). The remainder of the proof is identical to the proof in the complete case.

We now present a few counterexamples that show the limits of these theorems.

3.7.8. EXAMPLE. Consider counting measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}), m_c)$. Form the product, which is $(\mathbb{N}^2, \mathcal{P}(\mathbb{N}^2), m_c \times m_c)$, where $m_c \times m_c$ is just counting measure on $\mathbb{N}^2$. Consider

$$f(m,n) = \begin{cases} 1 & \text{if} \quad n = m \\ -1 & \text{if} \quad n = m+1 \\ 0 & \text{otherwise.} \end{cases}$$

Think of this as an $\mathbb{N}$ by $\mathbb{N}$ array

$$\begin{matrix} 1 & -1 & 0 & 0 & 0 & \dots \\ 0 & 1 & -1 & 0 & 0 & \dots \\ 0 & 0 & 1 & -1 & 0 & \dots \\ 0 & 0 & 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix}$$

The first iterated integral first sums the rows, each absolutely summable with total 0:

$$\int_X \int_Y f(m,n)\, dm_c(n)\, dm_c(m) = \sum_{m\geq 1} \Big(\sum_{n \geq 1} f(m,n) \Big) = \sum_{m \geq 1} 0 = 0.$$

The second iterated integral first sums the columns, each absolutely summable. However the first column sums to 1, the rest to 0:

$$\int_Y \int_X f(m,n)\, dm_c(m)\, dm_c(n) = \sum_{n\geq 1} \Big(\sum_{m \geq 1} f(m,n) \Big) = 1 + \sum_{n \geq 2} 0 = 1.$$

These measures are σ-finite since $\mathbb{N}^2$ is countable and points have measure 1. Tonelli's theorem is not applicable because f is not positive. More importantly, Fubini's theorem does not apply because f is not integrable. If it were, then $|f|$ would have a finite integral, but clearly $\displaystyle\iint |f|\, dm_c \times m_c = \sum_{m\geq 1}\sum_{n \geq 1} |f(m,n)| = \infty$.

3.7.9. EXAMPLE. Consider Lebesgue measure $(X = [0,1], \mathcal{B}, m)$ and counting measure on $(Y = [0,1], \mathcal{P}([0,1]), m_c)$. We have $([0,1]^2, \overline{\mathcal{B} \otimes \mathcal{P}([0,1])}, m \times m_c)$ as the product. Consider $D = \{(x,x) : x \in [0,1]\}$, and let $f = \chi_D$. Observe that

$$D = \bigcap_{k \geq 1} \bigcup_{j=1}^{2^k} [\frac{j-1}{2^k}, \frac{j}{2^k}]^2$$

is a measurable set. Consider the iterated integrals. When integrating $f_x(y) = \delta_x$ with respect to counting measure, we get 1. Therefore

$$\int_X \int_Y f(x,y)\, dm_c(y)\, dm(x) = \int_X 1\, dm(x) = 1.$$

When integrating $f^y(x) = \delta_y$ with respect to Lebesgue measure, we get 0. Therefore

$$\int_Y \int_X f(x,y)\,dm(x)\,dm_c(y) = \int_Y 0\,dm_c(y) = 0.$$

What went wrong? Here $f \geq 0$ is measurable, so belongs to $L^+(m \times m_c)$. The problem is that m_c on $[0,1]$ is not σ-finite because $[0,1]$ is uncountable.

Finally let's compute $\displaystyle\iint f\,dm \times m_c = m \times m_c(D)$. This is computed using the outer measure obtained by covering D with a countable number of rectangles, say $D \subset \bigcup_{i \geq 1} A_i \times B_i$ where A_i are Lebesgue measurable. Note that this union cannot cover D if for each $i \geq 1$, either $m(A_i) = 0$ or B_i is finite. Indeed if $m(A_i) = 0$ for $i \in J_1$ and B_i is finite for $i \in J_2$, then

$$\bigcup_{i \in J_1} A_i \times B_i \subset A \times [0,1] \quad \text{and} \quad \bigcup_{i \in J_2} A_i \times B_i \subset [0,1] \times B$$

where $A = \bigcup_{i \in J_1} A_i$ has $m(A) = 0$, and $B = \bigcup_{i \in J_2} B_i$ is countable. But their union omits $A^c \times B^c$, which contains $(A \cup B)^c \times (A \cup B)^c$ and so intersects D. Therefore there is some i_0 so that $m(A_{i_0}) > 0$ and B_{i_0} is infinite, and hence $m_c(B_{i_0}) = \infty$. But then the premeasure of the rectangle is $\pi(A_{i_0} \times B_{i_0}) = m(A_{i_0})m_c(B_{i_0}) = \infty$. Since D is measurable, we have that $m \times m_c(D) = \infty$. So neither iterated integral yields the value of the product integral even though both exist.

3.7.10. Example. Let $X = Y = [0,1]$ with Lebesgue measure. If we assume the Continuum Hypothesis, there is a well-ordering $\prec$ on $[0,1]$ with the property that $\{y \in [0,1] : y \prec x\}$ is countable for all x. Let $E = \{(x,y) : y \prec x\} \subset [0,1]^2$. Let $f = \chi_E$. Clearly $f \geq 0$. For each $x \in [0,1]$, $E_x = \{y \in [0,1] : y \prec x\}$ is countable. And for each $y \in [0,1]$, $E^y = \{x \in [0,1] : y \prec x\}$ is the complement of a countable set. In particular, these sets are Lebesgue measurable. So we can evaluate the iterated integrals

$$\int_X \int_Y f(x,y)\,dm(y)\,dm(x) = \int_X m(E_x)\,dm(x) = 0$$

and

$$\int_Y \int_X f(x,y)\,dm(x)\,dm(y) = \int_Y m(E^y)\,dm_c(y) = 1.$$

What went wrong? The measures are finite, and hence σ-finite. If E were measurable, then its measure would be finite and so f would belong to $L^+ \cap L^1(m^2)$ Then both Fubini's Theorem and Tonelli's Theorem would apply! It must be the case that E is not measurable. This is in spite of the fact that E_x and E^y are Borel for all x and y.

There is a variant of this example which avoids the Continuum Hypothesis, but requires more detailed knowledge of ordinals. Let Ω be the first uncountable ordinal. This is a well-ordered uncountable set with respect to an order $\prec$ with the property that $\{y \in \Omega : y \prec x\}$ is countable for all $x \in \Omega$. The open sets in the order topology are generated by $\{y \in \Omega : y \prec x\}$ and $\{y \in \Omega : x \prec y\}$ for $x \in \Omega$. It is not hard to see that these sets are either countable of have countable complement (co-countable). Moreover one can check that $\mathrm{Bor}(\Omega)$ consists or all countable and co-countable sets. Put a measure μ on $(\Omega, \mathrm{Bor}(\Omega))$ by

$$\mu(A) = \begin{cases} 0 & \text{if} \quad A \text{ is countable} \\ 1 & \text{if} \quad A \text{ is co-countable.} \end{cases}.$$

This is a finite measure. Form the product space $(\Omega^2, \mathrm{Bor}(\Omega^2), \mu \times \mu)$. Again we set $E = \{(x, y) \in \Omega^2 : y \prec x\}$ and $f = \chi_E$. We have the same contradiction. It must be the case that E is not a measurable set.

3.8. Lebesgue Measure on $\mathbb{R}^n$

Let $(\mathbb{R}, \mathcal{L}, m)$ be Lebesgue measure on the real line. Then we can form $m^n = m \times \cdots \times m$ on $(\mathbb{R}^n, \overline{\mathcal{L} \otimes \cdots \otimes \mathcal{L}}) = (\mathbb{R}^n, \overline{\mathrm{Bor}(\mathbb{R}^n)})$. This is the completion of a Borel measure on $\mathbb{R}^n$ such that

$$m^n(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} m(A_i) \quad \text{for} \quad A_i \in \mathcal{L}.$$

This is the unique complete measure with this property. Let $\mathcal{L}^n$ denote the σ-algebra of Lebesgue measurable sets in $\mathbb{R}^n$.

Even though m is a complete measure, the σ-algebra $\mathcal{L} \otimes \mathcal{L}$ is not complete. For example, if $E \subset \mathbb{R}$ is not measurable and $B \subset \mathcal{L}$ has $m(B) = 0$, then

$$E \times B \subset \mathbb{R} \times B = \bigcup_{n \in \mathbb{Z}} [n, n+1) \times B.$$

Since $m^2([n, n+1) \times B) = m(B) = 0$, we have that $m(\mathbb{R} \times B) = 0$. Then since m^2 is complete, $m^2(E \times B) = 0$ as well. This set does not belong to $\mathcal{L} \otimes \mathcal{L}$.

Note that when there is no confusion, Lebesgue measure on $\mathbb{R}^n$ is commonly written as m instead of m^n. We will adopt this practice.

The first result is to show that certain regularity properties of Lebesgue measure on the line transfer to m^n.

3.8.1. Proposition. *Let $E \in \mathcal{L}^n$, and let $\varepsilon > 0$.*

(1) *There is an open set $U \supset E$ such that $m(U \setminus E) < \varepsilon$; and*

$$m(E) = \inf\{m(U) : E \subset U,\ U \text{ open}\}.$$

(2) *There is a closed set $C \subset E$ so that $m(E \setminus C) < \varepsilon$; and*

$$m(E) = \sup\{m(K) : K \subset E,\ K \text{ compact}\}.$$

(3) *There is an F_σ set F and a G_δ set G so that $F \subset E \subset G$ and*

$$m(E \setminus F) = 0 = m(G \setminus E).$$

(4) *If $m(E) < \infty$, then there is a finite set of pairwise disjoint rectangles $R_1, \dots, R_N$ whose sides are intervals so that $m\big(E \triangle \bigcup_{i=1}^N R_i\big) < \varepsilon$.*

PROOF. (1) First assume that E is bounded. The construction of m^n comes from the outer measure on the algebra $\mathcal{A}$ of finite unions of disjoint rectangles. Thus there is a countable union of (bounded) rectangles R_i covering E such that $\sum_{i \geq 1} m(R_i) < m(E) + \varepsilon/2$. For each $i \geq 1$, we can use Theorem 1.5.6 to enclose the sides of R_i in open sets with slight increase in measure to obtain an open rectangle $U_i \supset R_i$ with $m(U_i) < m(R_i) + 2^{-i-1}\varepsilon$. Then $E \subset U = \bigcup_{i \geq 1} U_i$ and

$$m(U) \leq \sum_{i \geq 1} m(E_i) < \sum_{i \geq 1} m(R_i) + 2^{-i-1}\varepsilon < m(E) + \varepsilon.$$

Hence $m(U \setminus E) < \varepsilon$.

Now if E is unbounded, divide $\mathbb{R}^n$ into countably many disjoint unit cubes labeled C_i for $i \geq 1$. Then $E = \bigcup_{i \geq 1} E \cap C_i$. For each i, find an open set $U_i \supset E \cap C_i$ so that $m(U_i \setminus (E \cap C_i) < 2^{-i}\varepsilon$. Then $U = \bigcup_{i \geq 1} U_i$ does the job.

The other parts of this proposition are left as an exercise. ■

3.8.2. LEMMA. *If f is Lebesgue measurable, there is a G_δ set G with $m(G) = 0$ so that $f\chi_{G^c}$ is Borel measurable.*

PROOF. Write $f = f_1 - f_2 + if_3 - if_4$ where f_i are positive and measurable. This reduces the problem to the case of a positive function f.

Let $\{r_n : n \geq 1\}$ be a dense subset of $[0, \infty)$. Then $E_n = f^{-1}([0, r_n))$ are measurable. Select F_σ sets B_n and null sets N_n so that $E_n = B_n \cup N_n$. Then $N = \bigcup_{n \geq 1} N_n$ is a null set. Let G be a G_δ set so that $N \subset G$ and $m(G) = 0$. Then $(f\chi_{G^c})^{-1}([0, r_n)) = E_n \cup G = B_n \cup G$ is Borel for all r_n. The countable union of Borel sets is Borel, and thus $f\chi_{G^c}$ is Borel. ■

As for Lebesgue measure on the line, Lebesgue measure on $\mathbb{R}^n$ is translation invariant. It also behaves well under a linear change of variables just as the Riemann integral does. Let GL_n denote the group of invertible $n \times n$ real matrices. It is a fact from linear algebra that every invertible matrix is the product of a number of elementary matrices [1,§3.2, Corollary 3]. The elementary matrices have three types: multiplying the ith row by a non-zero scalar c

$$R_i(x_1, \dots, x_i, \dots, x_n) = (x_1, \dots, cx_i, \dots, x_n),$$

adding a multiple of the jth row to the ith row for $i \neq j$, $1 \leq i, j \leq n$

$$A_{ij}(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = (x_1, \dots, x_i + cx_j, \dots, x_j, \dots, x_n),$$

and interchanging two rows i, j for $i \neq j$, $1 \leq i, j \leq n$

$$S_{ij}(x_1, \dots, x_i, \dots, x_j, \dots, x_n) = (x_1, \dots, x_j, \dots, x_i, \dots, x_n).$$

3.8.3. THEOREM. *Let m be Lebesgue measure on $\mathbb{R}^n$.*

(1) *m is translation invariant: $m(E + x) = m(E)$ for $E \in \mathcal{L}^n$ and $x \in \mathbb{R}^n$.*

(2) *If $T \in \mathrm{GL}_n$ and f is measurable, then $f \circ T$ is measurable. If $f \in L^1(m)$ or $L^+(m)$, then*
$$\int f\, dm = |\det T| \int f \circ T\, dm.$$
In particular, if $E \in \mathcal{L}^n$, then $m(T(E)) = |\det T| m(E)$.

(3) *m is invariant under rotation: $m(U(E)) = m(E)$ for $E \in \mathcal{L}^n$ and U unitary.*

PROOF. (1) Note that translation of a cube preserves the measure. Thus the premeasure is translation invariant. It follows that the outer measure is translation invariant, and hence so is m.

(2) If f is Borel measurable, then so is $f \circ T$ because T is continuous and hence Borel by Corollary 2.1.3. If the result is true for invertible matrices S and T, then

$$\begin{aligned} \int f\, dm &= |\det S| \int f \circ S\, dm \\ &= |\det S|\, |\det T| \int (f \circ S) \circ T\, dm \\ &= |\det ST| \int f \circ ST\, dm. \end{aligned}$$

Hence it is true for their product. Therefore it suffices to establish the result for elementary matrices.

Using either the Fubini or Tonelli Theorem, the integral is equal to the iterated integral. For multiplying a row by a non-zero constant, the iterated integral reduces to the one variable fact. Adding a multiple of one row to another, integrate first with respect to x_i and use translation invariance. Interchanging two coordinates is equivalent to interchanging the order of integration, which results in no change, again by the Fubini or Tonelli theorem.

For E Borel, apply the result to χ_E to get

$$\begin{aligned} m(E) = \int \chi_E\, dm &= |\det T^{-1}| \int \chi_E \circ T^{-1}\, dm \\ &= |\det T|^{-1} \int \chi_{T(E)}\, dm = |\det T|^{-1} m(T(E)). \end{aligned}$$

The result for a Lebesgue measurable function f now follows because there is a G_δ set G so that $f\chi_{G^c}$ is Borel measurable. Now $(f\chi_{G^c}) \circ T = f \circ T\chi_{T^{-1}(G)^c}$.

By the Borel result, $m(T^{-1}(G)) = 0$. So $f \circ T = (f\chi_{G^c}) \circ T$ a.e.(m). The result follows.

(3) If U is unitary, then $|\det U| = 1$. In particular every rotation is unitary; and thus m is rotation invariant. ■

3.9. Infinite Product Measures

We briefly mention the subject of infinite products. Suppose that we are given probability measures $(X_i, \mathcal{B}_i, \mu_i)$; i.e. $\mu_i(X_i) = 1$ for $i \geq 1$. We wish to define a measure μ on $(\prod_{i\geq 1} X_i, \bigotimes_{i\geq 1} \mathcal{B}_i)$ so that

$$\mu\Big(A_i \times \cdots \times A_n \times \prod_{i>n} X_i\Big) = \prod_{i=1}^{n} \mu_i(A_i)$$

for all $n \geq 1$ and $A_i \in \mathcal{B}_i$.

Let $\mathcal{A}$ denote the algebra of sets generated by the *cylinders* $\pi_i^{-1}(A_i)$ for $i \geq 1$ and $A_i \in \mathcal{B}_i$, where as usual π_i is the coordinate projection of $X = \prod_{i\geq 1} X_i$ onto X_i. One can show as in section 3.6 that elements of $\mathcal{A}$ can be written as a finite *disjoint* union of rectangles of the form

$$R(A_1, \dots, A_n) := A_i \times \cdots \times A_n \times \prod_{i>n} X_i.$$

Define $\mu\big(R(A_1, \dots, A_n)\big) = \prod_{i=1}^{n} \mu_i(A_i)$ and extend this to finite disjoint unions by additivity. Note that we have already shown that there are measures $\mu^{(n)} = \mu_1 \times \cdots \times \mu_n$ on $(\prod_{i=1}^{n} X_i, \bigotimes_{i=1}^{n} \mathcal{B}_i)$ so that $\mu^{(n)}(A_i \times \cdots \times A_n) = \mu(R)$ for all rectangles in $\bigotimes_{i=1}^{n} \mathcal{B}_i \times \prod_{i>n} X_i$. If we let $\mu^{(m,n]} = \mu_{m+1} \times \cdots \times \mu_n$ for $1 \leq m < n$, we have that $\mu^{(n)} = \mu^{(m)} \times \mu^{(m,n]}$. for $1 \leq m < n$.

3.9.1. THEOREM. *μ is a premeasure on $\mathcal{A}$.*

PROOF. It suffices to show that if R, R_i for $i \geq 1$ are rectangles and $R = \dot{\bigcup}_{i\geq 1} R_i$, then $\mu(R) = \sum_{i\geq 1} \mu(R_i)$. There is an integer m_0 so that

$$R \in \bigotimes_{j=1}^{m_0} \mathcal{B}_j \times \prod_{j>m_0} X_j =: \mathcal{B}^{(m_0)}.$$

Then choose integers $m_0 \leq m_n \leq m_{n+1}$ so that $R_i \in \mathcal{B}^{(m_n)}$ for $1 \leq i \leq n$. Define $F_n = R \setminus \dot{\bigcup}_{i=1}^{n} R_i$. Observe that $F_n \in \mathcal{A} \cap \mathcal{B}^{(m_n)}$, $F_1 \supset F_n \supset F_{n+1} \supset \dots$ and $\bigcap_{n\geq 1} F_n = \varnothing$. Therefore

$$\mu(R) = \mu^{(m_n)}(R) = \sum_{i=1}^{n} \mu^{(m_n)}(R_i) + \mu^{(m_n)}(F_n) = \sum_{i=1}^{n} \mu(R_i) + \mu(F_n)$$

Our result will follow if we show that $\lim_{n\to\infty} \mu(F_n) = 0$.

Assuming that this is false, we have that $\mu(F_n) \geq \varepsilon > 0$ for all $n \geq 1$. We will produce a contradiction. Given $F \in \mathcal{A}$ and $x_i \in X_i$ for $1 \leq i \leq p$, define

$$F(x_1, \dots, x_p) = \{x \in \prod_{i>p} X_i : (x_1, \dots, x_p, x) \in F\}.$$

Then define

$$G_{1,n} = \big\{x_1 \in X_1 : \mu^{(1,m_n]}(F_n(x_1)) > \varepsilon/2\big\}.$$

Note that

$$\begin{aligned}\varepsilon \leq \mu(F_n) &= \int_{X_1} \mu^{(1,m_n]}(F_n(x_1))\, d\mu_1(x_1)\\ &\leq \int_{G_{1,n}} 1\, d\mu_1(x_1) + \int_{G_{1,n}^c} \frac{\varepsilon}{2}\, d\mu_1(x_1) \leq \mu^{(1)}(G_{1,n}) + \frac{\varepsilon}{2}.\end{aligned}$$

Therefore $\mu^{(1)}(G_{1,n}) \geq \frac{\varepsilon}{2}$. Now $G_{1,n} \supset G_{1,n+1}$ and hence $\mu\big(\bigcap_{n\geq 1} G_{1,n}\big) \geq \frac{\varepsilon}{2}$. Choose a point $a_1 \in \bigcap_{n\geq 1} G_{1,n}$. Then

$$\mu^{(1,m_n]}(F_n(a_1)) \geq \frac{\varepsilon}{2} \quad \text{for all} \quad n \geq 1.$$

Recursively we construct points $a_i \in X_i$ so that

$$\mu^{(k,m_k]}(F(a_1, \dots, a_k)) \geq \frac{\varepsilon}{2^k} \quad \text{for all} \quad n \geq 1.$$

Suppose that $a_1, \dots, a_{k-1}$ have been defined with this property. Define

$$G_{k,n} = \big\{x_k \in X_k : \mu^{(k,m_n]}(F_n(a_1, \dots, a_{k-1}, x_k)) > \frac{\varepsilon}{2^k}\big\}.$$

Arguing as above,

$$\begin{aligned}\frac{\varepsilon}{2^{k-1}} &\leq \mu(^{(k-1,m_n]}(F_n(a_1, \dots, a_{k-1}))\\ &= \int_{X_k} \mu^{(k,m_n]}(F_n(a_1, \dots, a_{k-1}, x_k))\, d\mu_k(x_k)\\ &\leq \int_{G_{k,n}} 1\, d\mu_k(x_k) + \int_{G_{k,n}^c} \frac{\varepsilon}{2^k}\, d\mu_k(x_k) \leq \mu^{(k)}(G_{k,n}) + \frac{\varepsilon}{2^k}.\end{aligned}$$

Therefore $\mu_k\big(\bigcap_{n\geq 1} G_{k,n}\big) \geq \frac{\varepsilon}{2^k}$. Pick a point $a_k \in \bigcap_{n\geq 1} G_{k,n}$. This does the job.

It follows that for all $k \geq 1$ and $n \geq 1$, there is a point x so that $(a_1, \dots, a_k, x)$ is in F_n. In particular, there is a point y so that $(a_1, \dots, a_{m_n}, y) \in F_n$. But F_n is a union of cylinder sets of level no greater than m_n, and hence $(a_1, \dots, a_{m_n}, y) \in F_n$ for all $y \in \prod_{i>m_n} X_i$. Therefore $a = (a_1, a_2, a_3, \dots) \in \bigcap_{n\geq 1} F_n$. This is a contradiction since this intersection in empty.

It follows that $\lim_{n\to\infty} \mu(F_n) = 0$ and hence $\mu(R) = \sum_{i\geq 1} \mu(R_i)$. ∎

As an immediate consequence of Theorem 1.4.2, we obtain the desired measure.

3.9.2. Corollary. *There is a unique complete measure* $\mu = \prod_{i\geq 1} \mu_i$ *on* $(\prod_{i\geq 1} X_i, \overline{\bigotimes_{i\geq 1} \mathcal{B}_i})$ *so that*

$$\mu\Big(A_i \times \cdots \times A_n \times \prod_{i>n} X_i\Big) = \prod_{i=1}^{n} \mu_i(A_i)$$

for all $n \geq 1$ *and* $A_i \in \mathcal{B}_i$.

3.9.3. Example. Fix $0 < p < 1$. Take $X_i = \mathbf{2} = \{0, 1\}$ and let μ_i be the probability measure on $(X_i, \mathcal{P}(X_i))$ such that $\mu_i(0) = p$ and $\mu_i(1) = 1 - p$. Then $X = \prod_{i\geq 1} X_i$ is homeomorphic to the Cantor set C. Let μ_p denote the infinite product measure $\prod_{i\geq 1} \mu_i$.

Identify the Cantor ternary set with $\{x = \big(0.(2\varepsilon_1)(2\varepsilon_2)(2\varepsilon_3)\dots\big)_{\text{base } 3}\}$ for $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots) \in X$. The homeomorphism of X onto C is given by

$$h(\varepsilon) = \big(0.(2\varepsilon_1)(2\varepsilon_2)(2\varepsilon_3)\dots\big)_{\text{base } 3} \quad \text{for} \quad \varepsilon = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots) \in X.$$

We obtain a Borel measure on C by setting $\nu_p(A) = \mu_p(h^{-1}(A))$.

Let C_n denote the subset of $[0, 1]$ after n operations of removing the middle thirds has been accomplished; so that there remain 2^n intervals of length 3^{-n}. With this explicit homeomorphism, the rectangle $R(\varepsilon_1, \dots, \varepsilon_n)$ corresponds to one of the intervals of C_n intersected with C. For example, $R(0, 0) = [0, \frac{1}{9}] \cap C$, $R(0, 1) = [\frac{2}{9}, \frac{1}{3}] \cap C$, $R(1, 0) = [\frac{2}{3}, \frac{7}{9}] \cap C$ and $R(1, 1) = [\frac{8}{9}, 1] \cap C$.

Consider the analogue of the Cantor function which is defined on $[0, 1]$ by $f_p(x) = \nu_p(C \cap [0, x])$. Then $f(0) = 0$ and $f_p(1) = 1$. Observe that this is a monotone increasing function. Moreover on the middle third $[\frac{1}{3}, \frac{2}{3}]$, it takes the value p. Then on the interval $[\frac{1}{9}, \frac{2}{9}]$, it takes the value p^2 and on $[\frac{7}{9}, \frac{8}{9}]$, it takes the value $p + p(1 - p)$. Indeed, at level n, we have defined $f_p(x)$ on $[0, 1] \setminus \text{int}(C_n)$. If $[a, b]$ is a component of C_n, then the middle third is removed to form C_{n+1} and

$$f_p(x) = f_p(a) + p(f_p(b) - f_p(a)) \quad \text{for} \quad a + \tfrac{b-a}{3} \leq x \leq a + \tfrac{2(b-a)}{3}.$$

Notice that f_p takes all of the values $p^k(1 - p)^{n-k}$ for $0 \leq k \leq n$ and $n \geq 1$. These values are dense in $[0, 1]$. As for the usual Cantor function, it follows that f_p has no jump discontinuities. So f_p is continuous.

The measure ν_p is the Lebesgue-Stieltjes measure for the function f_p by Theorem 1.5.2. So ν_p is a probability measure supported on C. Since f_p take the same value at each endpoint of a removed interval, we see that $\nu_p(\mathbb{R} \setminus C) = 0$. The continuity of f_p means that ν_p has no atoms.

CHAPTER 4

Differentiation and Signed Measures

4.1. Differentiation

Question: Is there a measure theoretic analogue of the Fundamental Theorem of Calculus? That is, to what extent is $F(x) = c + \int_{[a,x]} f\, dm$ differentiable on $\mathbb{R}$? What functions arise in this way?

We restrict our attention to real valued functions. In fact, splitting a function f as a difference of two positive functions, we may suppose that $f \geq 0$; and so the integral will be monotone increasing.

4.1.1. DEFINITION. The *upper* and *lower derivative* from left and right of a real valued function f are defined as

$$\overline{D}_r f(x) = \limsup_{h\to 0^+} \frac{f(x+h)-f(x)}{h} \qquad \underline{D}_r f(x) = \liminf_{h\to 0^+} \frac{f(x+h)-f(x)}{h}$$
$$\overline{D}_l f(x) = \limsup_{h\to 0^+} \frac{f(x)-f(x-h)}{h} \qquad \underline{D}_l f(x) = \liminf_{h\to 0^+} \frac{f(x)-f(x-h)}{h}$$

Then f is *differentiable* at x if $\overline{D}_r f(x) = \underline{D}_r f(x) = \overline{D}_l f(x) = \underline{D}_l f(x) \in \mathbb{R}$.

This definition is clearly equivalent to the familiar definition from calculus: a function f is differentiable at x if and only if there is a finite limit

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h}.$$

We require a technical definition. A singleton $\{a\}$ and the empty set are considered to be *degenerate* intervals.

4.1.2. DEFINITION. If $E \subset \mathbb{R}$, a collection $\mathcal{J}$ of non-degenerate intervals is a *Vitali cover* if for every $x \in E$ and $\varepsilon > 0$, there is $I \in \mathcal{J}$ so that $x \in I$ and $m(I) < \varepsilon$.

4.1.3. VITALI COVERING LEMMA. *If $E \subset \mathbb{R}$ has finite outer measure ($m^*(E) < \infty$), $\mathcal{J}$ is a Vitali cover of E, and $\varepsilon > 0$, then there are disjoint intervals $I_1, \dots, I_N \in \mathcal{J}$ such that $m^*\bigl(E \setminus \bigcup_{j=1}^N I_j\bigr) < \varepsilon$.*

PROOF. Fix an open $U \supset E$ with $m(U) < \infty$. Let $\mathcal{J}' = \{I \in \mathcal{J} : I \subset U\}$. This is also a Vitali cover since if $x \in E$ and $\delta = \operatorname{dist}(x, U^c)$, then any interval $I \in \mathcal{J}$ containing x of length less than δ will belong to $\mathcal{J}'$.

Recursively choose disjoint $I_k \in \mathcal{J}'$ so that $m(I_k) > \alpha_k/2$ where

$$\alpha_k = \sup\{m(I) : I \in \mathcal{J}',\ I \text{ disjoint from } I_1, \dots, I_{k-1}\}.$$

Observe that $m\bigl(\bigcup_{k\ge 1} I_k\bigr) = \sum_{k\ge 1} m(I_k) \le m(U) < \infty$. Thus α_k is a summable sequence. So we may choose N so that $\sum_{k>N} m(I_k) < \varepsilon/5$. We claim that $I_1, \dots, I_N$ works.

Let $X = E \setminus \bigcup_{j=1}^N \overline{I_j}$. If $x \in X$, $\delta := \operatorname{dist}(x, \bigcup_{j=1}^N \overline{I_j}) > 0$. Hence there is some $I \in \mathcal{J}$ with $x \in I$ and $m(I) < \delta$. So I is disjoint from $\bigcup_{j=1}^N I_j$, and hence $m(I) \le \alpha_{N+1}$. Pick $K > N$ so that $\alpha_{K+1} < m(I) \le \alpha_K$. Then by construction, I cannot be disjoint from $\bigcup_{j=1}^K I_j$. Thus there is some k with $N < k \le K$ so that $I_k \cap I \ne \varnothing$. Now $m(I_k) \ge \alpha_k/2 \ge \alpha_K/2 \ge m(I)/2$. Hence

$$\operatorname{dist}(x, \text{midpoint of } I_k) \le \tfrac{1}{2} m(I_k) + m(I) \le \tfrac{5}{2} m(I_k).$$

Let J_k be the closed interval with the same midpoint as I_k but with $m(J_k) = 5m(I_k)$. Then $x \in J_k$. Therefore $X \subset \bigcup_{k>N} J_k$. If follows that

$$m^*(X) \le \sum_{k>N} m(J_k) = 5 \sum_{k>N} m(I_k) < \varepsilon.$$

This proves the claim. ■

4.1.4. THEOREM. *Let $f : [a,b] \to \mathbb{R}$ be monotone increasing. Then f is continuous except on a countable set, and is differentiable except on a set of measure 0. The derivative f' is integrable, and $\displaystyle\int_a^b f'\,dm \le f(b) - f(a)$.*

PROOF. Define $f(x) = f(a)$ for $x < a$ and $f(x) = f(b)$ for $x > b$. Since f is monotone, for $c \in [a,b]$, we have

$$f(c^-) = \lim_{x\to c^-} f(x) = \sup_{x<c} f(x) \le f(c) \le f(c^+) = \inf_{x>c} f(x) = \lim_{x\to c^+} f(x).$$

Thus f is continuous at c unless it is a jump discontinuity, in which case the jump has length $j(c) = f(c^+) - f(c^-)$. Clearly $\sum j(c) \le f(b) - f(a) < \infty$. In particular the number of points with jump at least $\delta > 0$ is at most $\delta^{-1}(f(b) - f(a)) < \infty$. It follows that the number of discontinuities is countable.

Clearly we have $\underline{D}_r f(x) \le \overline{D}_r f(x)$ and $\underline{D}_l f(x) \le \overline{D}_l f(x)$. We will show that $\overline{D}_l f(x) \le \underline{D}_r(x)$ a.e. and $\overline{D}_r f(x) \le \underline{D}_l f(x)$ a.e. For $u, v \in \mathbb{Q}$ with $u < v$, let

$$E_{u,v} = \{x : \underline{D}_r(x) < u < v < \overline{D}_l f(x)\} \quad \text{and} \quad E = \bigcup_{u<v\in\mathbb{Q}} E_{u,v}.$$

To show that $m^*(E) = 0$, it suffices to show that $m^*(E_{u,v}) = 0$ for all $u < v$.

Let $m^*(E_{u,v}) = s$ and fix an $\varepsilon > 0$. Choose an open set $U \supset E_{u,v}$ with $m(U) < s + \varepsilon$. Let $\mathcal{J} = \{I = [x, x+h] \subset U : f(x+h) - f(x) < uh\}$. By definition of $\underline{D}_r f(x)$, this contains arbitrarily small intervals $[x, x+h]$ for each $x \in E_{u,v}$. Thus $\mathcal{J}$ is a Vitali cover of E. By the Vitali Covering Lemma, we can find $I_1 = [x_1, x_1 + h_1], \ldots, I_N = [x_N, x_N + h_N]$ disjoint intervals in $\mathcal{J}$ so that $m^*(E_{u,v} \setminus \bigcup_{j=1}^N I_j) < \varepsilon$. Therefore

$$s - \varepsilon < \sum_{j=1}^N m(I_j) = \sum_{j=1}^N h_j < m(U) < s + \varepsilon,$$

and $m^*(E_{u,v} \cap \bigcup_{j=1}^N I_j) > s - \varepsilon$. Let

$$F = E_{u,v} \cap \bigcup_{j=1}^N (x_j, x_j + h_j) \subset \bigcup_{j=1}^N (x_j, x_j + h_j) =: V.$$

Consider $\mathcal{J}' = \{I = [x-k, x] \subset V : f(x) - f(x-k) > vk\}$. As for $\mathcal{J}$, one sees that $\mathcal{J}'$ is a Vitali cover of F. Choose disjoint intervals $J_i = [y_i - k_i, y_i] \in \mathcal{J}'$ for $1 \le i \le M$ so that $m^*(F \setminus \bigcup_{i=1}^M J_i) < \varepsilon$. Therefore

$$\sum_{i=1}^M k_i = \sum_{i=1}^M m(J_i) > m^*(F) - \varepsilon > s - 2\varepsilon.$$

Since the intervals J_i are disjoint and are contained in $\bigcup_{j=1}^N I_j$, we have that

$$\begin{aligned} v(s - 2\varepsilon) &< \sum_{i=1}^M k_i v < \sum_{i=1}^M f(y_i) - f(y_i - k_i) \\ &\le \sum_{j=1}^N f(x_j + h_j) - f(x_j) < \sum_{j=1}^N u h_j < u(s + \varepsilon). \end{aligned}$$

Letting $\varepsilon \to 0$ yields $vs \le us$, and thus $s = 0$. Hence $m^*(E_{u,v}) = 0$. Therefore, since m is complete, $m(E) = 0$. Consequently, $\overline{D}_l f(x) \le \underline{D}_r f(x)$ on E^c, that is a.e.

Similarly $\overline{D}_r f(x) \le \underline{D}_l f(x)$ a.e. Thus $f'(x)$ exists except on a set of measure 0. Note however that the derivative might be $+\infty$. We will show soon that this also happens only on a set of measure 0.

Define $g_n(x) = n\big(f(x + \frac{1}{n}) - f(x)\big)$ on $[a, b]$. Monotone functions are Borel, and thus are measurable. So g_n is measurable and positive. Moreover

$$\lim_{n\to\infty} g_n(x) = \lim_{n\to\infty} \frac{f(x + \frac{1}{n}) - f(x)}{1/n} = f'(x)$$

whenever $f'(x)$ is defined, which is almost everywhere. Therefore f' is measurable and $f' \geq 0$. We can apply Fatou's Lemma to this sequence to get

$$\begin{aligned}\int_a^b f'\,dm &\leq \liminf_{n\to\infty} \int_a^b g_n\,dm = \liminf_{n\to\infty}\ n\int_{a+\frac{1}{n}}^{b+\frac{1}{n}} f\,dm - n\int_a^b f\,dm \\ &= \liminf_{n\to\infty}\ n\int_b^{b+\frac{1}{n}} f(b)\,dm - n\int_a^{a+\frac{1}{n}} f\,dm \leq f(b) - f(a) < \infty.\end{aligned}$$

In particular, f' is integrable. This implies that $f'(x) < \infty$ a.e. ∎

4.1.5. EXAMPLE. Recall the *Cantor ternary function*. This is defined on $[0,1]$ by $f(0) = 0$, $f(1) = 1$. Then $f(x) = \frac{1}{2}$ for $x \in [\frac{1}{3}, \frac{2}{3}]$, the (closure of) the middle third removed in the construction of the Cantor set. Then we set $f(x) = \frac{1}{4}$ on $[\frac{1}{9}, \frac{2}{9}]$ and $f(x) = \frac{3}{4}$ on $[\frac{7}{9}, \frac{8}{9}]$. This repeats, on each middle third removed, f is defined as the midpoint between the values defined at the endpoints of the (larger) interval. Finally, for $x \in C$, where C is the Cantor set, we can define $f(x) = \sup\{f(y) : 0 \leq y \leq x,\ y \notin C\}$. This evidently yields a monotone increasing function. Moreover the values attained by f include all diadic rationals $2^{-n}k$ for $0 \leq k \leq 2^n$. So the range of f is dense in $[0,1]$. In particular, f cannot have any jump discontinuities. Therefore f is continuous.

Next observe that on each of the removed (open) intervals, $(\frac{1}{3}, \frac{2}{3})$, $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$, etc., the function f is constant. Therefore it is differentiable with $f'(x) = 0$. This occurs on $[0,1] \setminus C$, and since $m(C) = 0$, we see that f' is defined and finite almost everywhere. However

$$\int_0^1 f'\,dm = 0 < 1 = f(1) - f(0).$$

So it can certainly be the case that you cannot recover f from its derivative.

4.1.6. DEFINITION. A function $f : [a,b] \to \mathbb{R}$ is *bounded variation* (belongs to $\mathrm{BV}[a,b]$) if

$$V_a^b(f) = \sup\Big\{ \sum_{i=1}^n |f(x_i) - f(x_{i-1})| : n \geq 1,\ a = x_0 < x_1 < \ldots < x_n = b \Big\} < \infty.$$

4.1.7. THEOREM. *A function $f \in \mathrm{BV}[a,b]$ if and only if $f = g - h$ where g, h are monotone increasing.*

PROOF. If $f = g - h$ where g, h are monotone increasing, then for a partition $a = x_0 < x_1 < \ldots < x_n = b$,

$$\begin{aligned}\sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| &\le \sum_{i=1}^{n} |g(x_i) - g(x_{i-1})| + \sum_{i=1}^{n} |h(x_i) - h(x_{i-1})| \\ &= g(b) - g(a) + h(b) - h(a) =: L.\end{aligned}$$

So $V_a^b(f) \le L$.

Conversely, if $f \in BV$, define $g(x) = V_a^x(f)$. Clearly this is an increasing function of x since if $a \le x < y \le b$, $V_a^y(f) = V_a^x(f) + V_x^y(f)$. Let $h = g - f$, so that $f = g - h$. If $a \le x < y \le b$, then

$$h(y) - h(x) = V_x^y(f) - (f(y) - f(x)) \ge |f(y) - f(x)| - (f(y) - f(x)) \ge 0.$$

So h is also monotone increasing. ■

4.1.8. COROLLARY. *If $f \in BV[a,b]$, then f is continuous except on a countable set, $f'(x)$ exists* a.e.(m) *and f' is m-integrable.*

4.1.9. COROLLARY. *If $f \in L^1[a,b]$, then $F(x) = \displaystyle\int_a^x f\,dm$ is bounded variation.*

PROOF. Write $f = f_+ - f_-$ where $f_+ = f \vee 0$ and $f_- = (-f) \vee 0$ are positive integrable functions. Then $g(x) = \int_a^x f_+\,dm$ and $h(x) = \int_a^x f_-\,dm$ are monotone increasing. Hence $f = g - h$ is BV. ■

4.1.10. DEFINITION. A function $f : [a,b] \to \mathbb{R}$ is *absolutely continuous* (AC) if for all $\varepsilon > 0$, there is a $\delta > 0$ so that whenever (x_i, y_i) are disjoint intervals in $[a,b]$ and $\sum_{i=1}^n y_i - x_i < \delta$, then $\sum_{i=1}^n |f(y_i) - f(x_i)| < \varepsilon$.

4.1.11. EXAMPLE. The Cantor ternary function f is monotone, and thus BV. However it is not absolutely continuous. If x_i, y_i for $1 \le i \le 2^n$ are the endpoints of the intervals remaining after n stages of the construction of the Cantor set, then because f is constant on the complements of these intervals, we have

$$\sum_{i=1}^{n} |f(y_i) - f(x_i)| = 1 \quad \text{and} \quad \sum_{i=1}^{2^n} y_i - x_i = \left(\tfrac{2}{3}\right)^n.$$

Since $\left(\frac{2}{3}\right)^n \to 0$, f is not absolutely continuous.

4.1.12. PROPOSITION. *Let $f \in L^1(\mu)$ and $\varepsilon > 0$. Then there is a $\delta > 0$ so that whenever $\mu(A) < \delta$, $\displaystyle\int_A |f|\,d\mu < \varepsilon$.*

PROOF. Choose a simple function $0 \le \varphi \le |f|$ so that $\int \varphi\, d\mu > \int |f|\, d\mu - \frac{\varepsilon}{2}$. Since φ is simple, there is a constant M so that $\varphi \le M$. Set $\delta = \frac{\varepsilon}{2M}$. If $\mu(A) < \delta$, then

$$\int_A |f|\, d\mu < \int_A \varphi\, d\mu + \frac{\varepsilon}{2} \le M\mu(A) + \frac{\varepsilon}{2} < \varepsilon. \qquad \blacksquare$$

4.1.13. COROLLARY. *If $f \in L^1[a,b]$, then $F(x) = \int_a^x f\, dm$ is absolutely continuous.*

PROOF. Given $\varepsilon > 0$, let δ be provided by Proposition 4.1.12. Then if (x_i, y_i) are disjoint intervals in $[a,b]$ and $\sum_{i=1}^n y_i - x_i < \delta$, then

$$\sum_{i=1}^n |F(y_i) - F(x_i)| = \sum_{i=1}^n \Big| \int_{x_i}^{y_i} f\, dm \Big| \le \int_{\bigcup_{i=1}^n [x_i, y_i]} |f|\, dm < \varepsilon. \qquad \blacksquare$$

4.1.14. LEMMA. *If f is absolutely continuous on $[a,b]$, then it has bounded variation.*

PROOF. Take $\varepsilon = 1$, and find $\delta > 0$. Split $[a,b] = \bigcup_{j=1}^p [a_{j-1}, a_j]$ where $a_j - a_{j-1} < \delta$ for $1 \le j \le p$. If $[a_{j-1}, a_j]$ contains disjoint intervals (x_i, y_i) for $1 \le i \le n$, then $\sum_{i=1}^n y_i - x_i \le a_j - a_{j-1} < \delta$. Hence $\sum_{i=1}^n |f(y_i) - f(x_i)| < 1$. Taking the supremum yields $V_{a_{j-1}}^{a_j}(f) \le 1$. Thus $V_a^b(f) = \sum_{j=1}^p V_{a_{j-1}}^{a_j}(f) \le p$. $\blacksquare$

4.1.15. LEMMA. *If $f \in L^1[a,b]$ and $F(x) = \int_a^x f\, dm$ is monotone increasing, then $f \ge 0$* a.e.

PROOF. Let $E = \{x : f(x) < 0\}$ and $E_n = \{x : f(x) < -\frac{1}{n}\}$ for $n \ge 1$. Then if $m(E) > 0$, there is some n so that $m(E_n) > 0$. Let $\varepsilon = m(E_n)/2n$; and choose $\delta > 0$ using Proposition 4.1.12. Select an open set $U \supset E_n$ so that $m(U \setminus E_n) < \delta$. Write $U = \dot\bigcup_{i\ge1} (x_i, y_i)$. Then

$$\begin{aligned} 0 \le \sum_{i\ge1} F(y_i) - F(x_i) &= \int_U f\, dm \\ &= \int_{E_n} f\, dm + \int_{U\setminus E_n} f\, dm \le -\frac{m(E_n)}{n} + \varepsilon < 0. \end{aligned}$$

This contradiction shows that $m(E) = 0$, or that $f \ge 0$ a.e. $\blacksquare$

The following consequence is immediate.

4.1.16. Corollary. *Let $f \in L^1[a,b]$ and $F(x) = \int_a^x f\,dm$. If $F = 0$, then* $f = 0$ a.e.

The main result here is the following result of Lebesgue, his substitute for the Fundamental Theorem of Calculus, which answers the first part of the question posed at the beginning of this section.

4.1.17. Lebesgue Differentiation Theorem. *Let $f \in L^1[a,b]$ and $F(x) = c + \int_a^x f\,dm$. Then $F'(x) = f(x)$* a.e.

Proof. Since F is BV by Corollary 4.1.9, Corollary 4.1.8 shows that F' exists a.e. and the derivative is integrable. Extend f by setting $f(x) = 0$ for $x > b$; so that $F(x) = F(b)$ for $x > b$. Let $g_n(x) = n(F(x + \frac{1}{n}) - F(x))$. Then g_n converges pointwise a.e. to F'.

Case 1 $|f| \le M$. Then $g_n(x) = n \int_x^{x+\frac{1}{n}} f\,dm$. So

$$|g_n(x)| \le n \int_x^{x+\frac{1}{n}} M\,dx = M.$$

Thus $|g_n| \le M\chi_{[a,b]}$, which is integrable. By the LDCT and the fact that F is continuous, for $c \in [a,b]$,

$$\begin{aligned}\int_a^c F'\,dm &= \lim_{n\to\infty} \int_a^c g_n\,dm = \lim_{n\to\infty} n \int_a^c F(x + \tfrac{1}{n}) - F(x)\,dx \\ &= \lim_{n\to\infty} n \int_c^{c+\frac{1}{n}} F(x)\,dx - n \int_a^{a+\frac{1}{n}} F(x)\,dx \\ &= F(c) - F(a) = \int_a^c f\,dm.\end{aligned}$$

Therefore $\int_a^c (F' - f)\,dm = 0$ for all $c \in [a,b]$. By Corollary 4.1.16, $F' = f$ a.e.

Case 2 $f \ge 0$. Let $f_n = f \wedge n$ for $n \ge 1$. Then f_n is bounded and case 1 applies. Observe that

$$F(x) = \int_a^x f_n\,dm + \int_a^x f - f_n\,dm.$$

Because the second term is monotone increasing,

$$F'(x) = f_n(x) + \frac{d}{dx} \int_a^x f - f_n\,dm \ge f_n(x) \text{ a.e.}$$

Consequently, $F'(x) \geq \lim_{n\to\infty} f_n(x) = f(x)$ a.e. For $c \in [a,b]$,

$$\int_a^c F'(x)\,dm \leq F(c) - F(a) = \int_a^c f\,dm \leq \int_a^c F'(x)\,dm.$$

Therefore $\displaystyle\int_a^c F'(x) - f(x)\,dm = 0$ for all $c \in [a,b]$. Hence $F' = f$ a.e. by Corollary 4.1.16.

Case 3. Write $f = f_+ - f_-$ where $f_\pm \in L^+ \cap L^1(m)$. Then by case 2, we obtain that $F' = f_+ - f_- = f$ a.e. ∎

Next we wish to characterize which functions are integrals. We first need another lemma.

4.1.18. LEMMA. *If $f \in C[a,b]$ is absolutely continuous and $f' = 0$ a.e., then f is constant.*

PROOF. Fix $c \in (a,b]$. Let $\varepsilon > 0$. Obtain $\delta > 0$ from the definition of absolute continuity. Set $E = \{x \in (a,c) : f'(x) = 0\}$. So $m([a,c] \setminus E) = 0$. Define

$\mathcal{J} = \{[x, x+h] : x \in E,\ h > 0,\ [x, x+h] \subset (a,c)$ and $|f(x+h) - f(x)| < \varepsilon h\}$.

This collection is a Vitali cover of E because for each x, all sufficiently small h work. Therefore there are disjoint intervals $I_1, \dots, I_n$ in $\mathcal{J}$ so that

$$m\big([a,c] \setminus \bigcup_{j=1}^n I_j\big) = m\big(E \setminus \bigcup_{j=1}^n I_j\big) < \delta.$$

Write $I_j = (a_j, b_j)$ ordered so that

$$a < a_1 < b_1 < a_2 < b_2 < \cdots < a_n < b_n < c.$$

Then

$$\begin{aligned}
|f(c) - f(a)| \leq \sum_{j=1}^n &|f(b_j) - f(a_j)| \\
&+ \Big(|f(a_1) - f(a)| + \sum_{j=1}^{n-1} |f(a_{j+1}) - f(b_j)| + |f(c) - f(b_n)|\Big) \\
< \sum_{j=1}^n &\varepsilon |b_j - a_j| + \varepsilon < (c - a + 1)\varepsilon.
\end{aligned}$$

The second term in parentheses is at most ε by absolute continuity since the total length of the intervals is less than δ. Now $\varepsilon > 0$ was arbitrary, and therefore $f(c) = f(a)$; whence f is constant. ∎

4.1.19. Theorem. *Let $F : [a, b] \to \mathbb{R}$. The following are equivalent:*

(1) *There is an $f \in L^1(a, b)$ so that $F(x) = c + \int_a^x f\,dm$.*

(2) *F is absolutely continuous.*

(3) *F is differentiable* a.e.*, $F' \in L^1(a, b)$ and $F(x) = F(a) + \int_a^x F'\,dm$.*

Proof. Lebesgue's Differentiation Theorem shows that (1) implies (3). Clearly (3) implies (1). Lemma 4.1.9 shows that integrals are AC, so (1) implies (2). If (2) holds, then F is BV by Lemma 4.1.14; and thus by Corollary 4.1.8, F' exists a.e. and belongs to $L^1(m)$. Let

$$G(x) = F(a) + \int_a^x F'\,dm.$$

Then Lebesgue's Differentiation Theorem shows that $G' = F'$ a.e. Both F and G are absolutely continuous, and $(G - F)' = 0$ a.e. Therefore Lemma 4.1.18 shows that $G - F$ is constant; and $G(a) = F(a)$. So (3) holds. ■

4.2. Signed Measures

4.2.1. Definition. A *signed measure* on $(X, \mathcal{B})$ is a map $\nu : \mathcal{B} \to \mathbb{R} \cup \{\pm\infty\}$ such that $\nu(\varnothing) = 0$, ν takes at most one of the values $\pm\infty$, and ν is countably additive, meaning that if E_i are disjoint sets in $\mathcal{B}$, then $\nu\big(\dot{\bigcup}_{i\geq 1} E_i\big) = \sum_{i\geq 1} \nu(E_i)$.

4.2.2. Remark. The definition implies that if $\big|\nu\big(\dot{\bigcup}_{i\geq 1} E_i\big)\big| < \infty$, then the series $\sum_{i\geq 1} \nu(E_i)$ converges absolutely. One argument would be that the union is independent of order, and thus the sum should be independent of the order as well. But unconditionally convergent sequences of real numbers converge absolutely. An alternative argument, which is better, is that if the sum converges only conditionally, then the sum of the positive terms diverges, as does the sum of the negative terms. But then if $J = \{i : \nu(E_i) \geq 0\}$, we have that $\nu\big(\dot{\bigcup}_{i\in J} E_i\big) = \sum_{i\in J} \nu(E_i) = +\infty$ and $\nu\big(\dot{\bigcup}_{i\notin J} E_i\big) = \sum_{i\notin J} \nu(E_i) = -\infty$. This is not permitted for signed measures because we cannot make sense of $\nu(E \dot{\cup} F) = \nu(E) + \nu(F)$ is $\nu(E) = +\infty$ and $\nu(F) = -\infty$.

4.2.3. Examples.

(1) Let $f \in L^1_{\mathbb{R}}(\mu)$ and define $\nu(E) = \int_E f\,d\mu$. In this case, neither $\pm\infty$ arises as a value. The absolute convergence over countably many disjoint sets follows from the LDCT.

(2) Let $f \in L^+(\mu)$ and $g \in L^+ \cap L^1(\mu)$. Define $\nu(E) = \int_E f - g\,d\mu$. If f is not integrable, then $\nu(X) = \int f\,d\mu - \|g\|_1 = +\infty$. If $E = \dot{\bigcup}_{i\geq 1} E_i$ and $\int_E f\,d\mu < \infty$, then countable additivity follows as in (1).

4.2.4. DEFINITION. A *null set* for a signed measure ν is a measurable set E such that $\nu(F) = 0$ for all $F \subset E$, $F \in \mathcal{B}$. A *positive set* (or *negative set*) for ν is a measurable set E such that $\nu(F) \geq 0$ (or ≤ 0) for all $F \subset E$, $F \in \mathcal{B}$.

4.2.5. HAHN DECOMPOSITION THEOREM. *Let ν be a signed measure on $(X, \mathcal{B})$. Then there are sets $P, N \in \mathcal{B}$ so that $X = P \dot{\cup} N$, P is a positive set, and N is a negative set. If $X = P' \dot{\cup} N'$ is another decomposition into positive and negative sets, then $P \triangle P'$ is a null set.*

4.2.6. LEMMA. *If $0 < \nu(E) < \infty$, then there is a positive set $A \subset E$ with $\nu(A) > 0$.*

PROOF. If E contains a set of negative measure, choose a set $B_1 \subset E$ in $\mathcal{B}$ so that

$$\nu(B_1) \leq \max\big\{ -1, \tfrac{1}{2}\inf\{\nu(B) : B \subset E,\ B \in \mathcal{B}\}\big\}.$$

Recursively choose $B_n \subset E \setminus \dot{\bigcup}_{i=1}^{n-1} B_i$ with

$$\nu(B_n) \leq \max\big\{ -1, \tfrac{1}{2}\inf\{\nu(B) : B \subset E \setminus \dot{\bigcup_{i=1}^{n-1}} B_i,\ B \in \mathcal{B}\}\big\}.$$

Either this terminates because $A = E \setminus \dot{\bigcup}_{i=1}^{n} B_i$ is a positive set, or there is an infinite sequence. In this case, let $A = E \setminus \dot{\bigcup}_{i\geq 1} B_i$. Now by countable additivity,

$$\nu(E) = \nu(A) + \sum_{i\geq 1} \nu(B_i).$$

Since $|\nu(E)| < \infty$, this series converges absolutely. Therefore

$$\nu(A) = \nu(E) - \sum_{i\geq 1} \nu(B_i) < \infty.$$

Moreover this shows that $\nu(A) \geq \nu(E) > 0$.

The convergence also shows that $\nu(B_i) \to 0$. If $B \subset A$ had $\nu(B) < 0$, then $\nu(B) < 2\nu(B_n)$ for some n. But this contradicts the fact that

$$\inf\{\nu(B) : B \subset E \setminus \dot{\bigcup_{i=1}^{n-1}} B_i,\ B \in \mathcal{B}\}\big\} > 2\nu(B_n).$$

Hence if $B \subset A$, then $\nu(B) \geq 0$. Thus A is a positive set. ∎

4.2.7. LEMMA. *If A_n are positive sets, so is $A := \bigcup_{n\geq 1} A_n$.*

PROOF. If $B \subset A$ and $B \in \mathcal{B}$, let $B_n = B \cap \big(A_n \setminus \bigcup_{i=1}^{n-1} A_i\big)$ for $n \geq 1$. Then the B_n are pairwise disjoint sets in $\mathcal{B}$ such that $B = \dot{\bigcup}_{n\geq 1} B_n$. Since each A_n is positive, $\nu(B_n) \geq 0$. Therefore $\nu(B) = \sum_{n\geq 1} \nu(B_n) \geq 0$. ■

PROOF OF THE HAHN DECOMPOSITION. We may suppose that ν does not take the value $+\infty$ (by considering $-\nu$ instead if necessary). Let

$$m = \sup\{\nu(A) : A \text{ is positive}\}.$$

Choose positive sets A_n with $\nu(A_n) \to m$. Set $P = \bigcup_{n\geq 1} A_n$. By Lemma 4.2.7, P is a positive set. Moreover $\nu(P) = \nu(A_n) + \nu(P \setminus A_n) \geq \nu(A_n)$ for all $n \geq 1$, and hence $\nu(P) = m$. Since ν does not take the value $+\infty$, we see that $m < \infty$.

Let $N = P^c$. Claim: N is negative. If it isn't, there is a subset $E \subset N$ such that $\nu(E) > 0$. By Lemma 4.2.6, E contains a positive set A with $\nu(A) > 0$. Then $P \dot{\cup} A$ is positive with $\nu(P) + \nu(A) > m$. This contradicts the definition of m. Hence N must be negative.

Uniqueness. Suppose that $X = P' \dot{\cup} N'$ is a second decomposition of X into a positive and negative set. Then $A = P \setminus P' = N' \setminus N$ is both positive and negative, and hence is a null set. Similarly $B = P' \setminus P = N \setminus N'$ is a null set. Thus $P \triangle P' = A \cup B$ is a null set. ■

4.2.8. DEFINITION. Two signed measures μ, ν on $(X, \mathcal{B})$ are *mutually singular* ($\mu \perp \nu$) if there is a decomposition $X = A \dot{\cup} B$ such that A is ν-null and B is μ-null.

4.2.9. JORDAN DECOMPOSITION THEOREM. *If ν is a signed measure on $(X, \mathcal{B})$, there there is a unique pair of mutually singular (positive) measures ν_+ and ν_- such that $\nu = \nu_+ - \nu_-$.*

PROOF. Let $X = P \dot{\cup} N$ be the Hahn decomposition for ν. Define $\nu_+(A) = \nu(A \cap P)$ and $\nu_-(A) = -\nu(A \cap N)$. Then by construction, ν_+ and ν_- are positive measures supported on disjoint sets, so they are mutually singular, and $\nu = \nu_+ - \nu_-$.

For uniqueness, suppose that $\mu_+ \perp \mu_-$ and $\nu = \mu_+ - \mu_-$. Let $X = P' \dot{\cup} N'$ such that P' is μ_--null and N' is μ_+-null. Then P' is a positive set for ν and N' is a negative set. By the uniqueness of the Hahn decomposition, $P \triangle P'$ is a ν-null set. For any set $A \in \mathcal{B}$,

$$\mu_+(A) = \nu(A \cap P') = \nu(A \cap P) = \nu_+(A).$$

So $\mu_+ = \nu_+$. Similarly $\mu_- = \nu_-$. ■

4.2.10. DEFINITION. The *absolute value* of a signed measure with Jordan decomposition $\nu = \nu_+ - \nu_-$ is $|\nu| = \nu_+ + \nu_-$.

Note that a set $A \in \mathcal{B}$ is ν-null if and only if $|\nu|(A) = 0$.

4.3. Decomposing Measures

4.3.1. Definition. A signed measure ν is *absolutely continuous* with respect to a (positive) measure μ ($\nu \ll \mu$) if $A \in \mathcal{B}$ and $\mu(A) = 0$ implies that $\nu(A) = 0$.

Since any measurable subset of A is also a μ-null set, A must be ν-null. This is equivalent to saying that $|\nu|(A) = 0$. So $\nu \ll \mu$ if and only if $|\nu| \ll \mu$.

4.3.2. Radon-Nikodym Theorem. *Let μ and ν be σ-finite measures on $(X, \mathcal{B})$. Suppose that $\nu \ll \mu$. Then there is an $f \in L^+$ so that $\nu(E) = \displaystyle\int_E f\,d\mu$ for $E \in \mathcal{B}$. Also f is uniquely determined* a.e.(μ).

Proof. First we assume that $\mu(X) < \infty$ and $\nu(X) < \infty$. For each $r \in \mathbb{Q}^+$, $\nu - r\mu$ has a Hahn decomposition (P_r, N_r). Also, let $P_0 = X$ and $N_0 = \varnothing$. Define $f(x) = \sup\{r \geq 0 : x \in P_r\}$. Then for $t \geq 0$, $f^{-1}(t, \infty] = \bigcup_{r>t} P_r$ is measurable. So $f \in L^+$.

If $r < s$ in $\mathbb{Q}^+$, then P_s is a positive set for $\nu - s\mu$ and thus is positive for $\nu - r\mu = (\nu - s\mu) + (s - r)\mu$. Since N_r is a negative set for $\nu - r\mu$, we have $\mu(N_r \cap P_s) = 0$. Therefore $\mu\big(N_r \cap \bigcup_{s>r\in\mathbb{Q}^+} P_s\big) = 0$. Hence $f|_{N_r} \leq r$ a.e.(μ). Thus $\mu(f^{-1}(r, \infty]) \leq \mu(P_r)$.

Since P_r is positive for $\nu - r\mu$, $r\mu(P_r) \leq \nu(P_r)$. Hence

$$\mu(P_r) \leq \nu(P_r)/r \leq \nu(X)/r.$$

This goes to 0 as $r \to \infty$, and thus $\mu(f^{-1}(\infty)) \leq \lim_{r\to\infty} \mu(P_r) = 0$. Therefore $f < \infty$ a.e.(μ).

Let $E \in \mathcal{B}$. Fix N. Define $E_k = E \cap P_{\frac{k}{N}} \cap N_{\frac{k+1}{N}})$ for $k \geq 0$; and let $E_\infty = E \setminus \bigcup_{k\geq 0} E_k = E \cap \bigcap_r P_r$. Then $\mu(E_\infty) = 0$ and so $\nu(E_\infty) = 0$ as well. Observe that

$$(\nu - \tfrac{k}{N}\mu)(E_k) \geq 0 \geq (\nu - \tfrac{k+1}{N}\mu)(E_k).$$

Thus

$$\tfrac{k}{N}\mu(E_k) \leq \nu(E_k) \leq \tfrac{k+1}{N}\mu(E_k).$$

Also $\frac{k}{N} \leq f(x) \leq \frac{k+1}{N}$ for a.e. $x \in E_k$. So $\frac{k}{N}\chi_{E_k} \leq f\chi_{E_k} \leq \frac{k+1}{N}\chi_{E_k}$. Integrating yields

$$\tfrac{k}{N}\mu(E_k) \leq \int_{E_k} f\,d\nu \leq \tfrac{k+1}{N}\mu(E_k).$$

Summing over $k \in \mathbb{N}$, we obtain that

$$\sum_{k\geq 0} \tfrac{k}{N}\mu(E_k) \leq \sum_{k\geq 0} \nu(E_k) = \nu(E) \leq \sum_{k\geq 0} \tfrac{k+1}{N}\mu(E_k) = \sum_{k\geq 0} \tfrac{k}{N}\mu(E_k) + \frac{\mu(E)}{N}.$$

And similarly

$$\sum_{k\geq 0} \tfrac{k}{N}\mu(E_k) \leq \sum_{k\geq 0} \int_{E_k} f\,d\nu = \int_E f\,d\nu \leq \sum_{k\geq 0} \tfrac{k+1}{N}\mu(E_k) = \sum_{k\geq 0} \tfrac{k}{N}\mu(E_k) + \frac{\mu(E)}{N}.$$

Therefore

$$\Big|\nu(E) - \int_E f\,d\nu\Big| \leq \frac{\mu(E)}{N}.$$

However N is arbitrary, and hence $\nu(E) = \displaystyle\int_E f\,d\nu$.

Suppose that $f, g \in L^+$ such that $\nu(E) = \displaystyle\int_E f\,d\mu = \int_E g\,d\mu$ for $E \in \mathcal{B}$. Define $A = \{x : f(x) > g(x)\}$ and $A_n = \{x : f(x) \geq g(x) + \frac{1}{n}\}$. If $\mu(A) > 0$, then $\mu(A_n) > 0$ for some n. But then

$$\nu(A_n) = \int_{A_n} f\,d\mu \geq \int_{A_n} g + \tfrac{1}{n}\,d\mu = \int_{A_n} g\,d\mu + \tfrac{1}{n}\mu(A_n) > \nu(A_n).$$

Hence $f \leq g$ a.e.(μ). Similarly $g \leq f$ a.e.(μ), so that $f = g$ a.e.(μ).

For the general case, use the fact that μ and ν are σ-finite to chop X up into a disjoint union of countable many pieces X_n so that $\mu(X_n) < \infty$ and $\nu(X_n) < \infty$ for each $n \geq 1$. Then apply the previous argument on each piece, and combine. ■

4.3.3. EXAMPLE. On $([0,1], \mathrm{Bor}([0,1]))$, consider Lebesgue measure m and counting measure m_c. Then $m \ll m_c$ but there is no function $f \in L^+$ so that $m(E) = \displaystyle\int_E f\,dm_c$. But m_c is not σ-finite, and this shows the necessity of this condition in the Radon-Nikodym Theorem.

Next we deduce the corresponding results for signed measures.

4.3.4. COROLLARY. *If ν is a signed measure on $(X, \mathcal{B})$, there is a measurable function f with $|f| = 1$ so that $\nu(E) = \displaystyle\int_E f\,d|\nu|$.*

If in addition, $|\nu|$ and μ are σ-finite measures on $(X, \mathcal{B})$ and $\nu \ll \mu$, there is a measurable function $g = g_+ - g_-$ with at least one of $g_\pm$ integrable so that $\nu(E) = \displaystyle\int_E g\,d\mu$.

PROOF. Let $X = P \dot\cup N$ be the Hahn decomposition for ν. Define $f(x) = 1$ on P and $f(x) = -1$ on N. Then we have that $\nu = \nu_+ - \nu_-$ and $|\nu| = \nu_+ + \nu_-$

and $\nu_+(E) = \nu(E \cap P)$. So

$$\int_E f\, d|\nu| = \int_{E\cap P} 1\, d\nu_+ + \int_{E\cap N} -1\, d\nu_- = \nu_+(E) - \nu_-(E) = \nu(E).$$

If $|\nu| \ll \mu$ are σ-finite, then by the Radon-Nikodym theorem, there is a measurable function $h \in L^+$ so that $|\nu|(E) = \int_E h\, d\mu$. Then

$$\nu(E) = \int_{E\cap P} h\, d\mu - \int_{E\cap N} h\, d\mu = \int_E fh\, d\mu.$$

So let $g = fh = h\chi_P - h\chi_N =: g_+ - g_-$. Then in particular, one of $\nu(P)$ and $\nu(N)$ is finite, which means that one of $g_\pm$ is integrable. ■

The Radon-Nikodym Theorem is usually combined with the following decomposition result.

4.3.5. Lebesgue Decomposition Theorem. *Let ν, μ be two σ-finite measures on $(X, \mathcal{B})$. Then there is a unique decomposition $\nu = \nu_a + \nu_s$ so that $\nu_a \ll \mu$ and $\nu_s \perp \mu$.*

Proof. Let $\lambda = \mu + \nu$. This is also a σ-finite measure on $(X, \mathcal{B})$. Clearly $\mu \ll \lambda$ and $\nu \ll \lambda$. By the Radon-Nikodym Theorem, there are functions $f, g \in L^+$ so that

$$\mu(E) = \int_E f\, d\lambda \quad \text{and} \quad \nu = \int_E g\, d\lambda \quad \text{for} \quad E \in \mathcal{B}.$$

Let $A = f^{-1}(0, \infty]$ and $B = f^{-1}(0)$. Set $\nu_a(E) = \nu(E \cap A)$ and $\nu_s(E) = \nu(E \cap B)$. Clearly $\nu = \nu_a + \nu_s$. Also $\nu_s \perp \mu$ since it is supported on B, which is a μ-null set. If $\mu(E) = 0$, then $f\chi_E = 0$ a.e.(λ); and hence $\lambda(E \cap A) = 0$. Therefore $\nu_a(E) = \nu(E \cap A) = 0$. Thus $\nu_a \ll \mu$.

Uniqueness. Suppose that $\nu = \nu'_a + \nu'_s$ is another decomposition so that $\nu'_a \ll \mu$ and $\nu'_s \perp \mu$. Then there is a set $A' \in \mathcal{B}$ so that $\nu'_s(A') = 0$ and $\mu(B') = 0$, where $B' = A'^c$. Thus

$$\mu(B \cup B') = 0 \quad \text{and} \quad \nu_s(A \cap A') = \nu'_s(A \cap A') = 0.$$

Now if $E \subset B \cup B'$, then $\mu(E) = 0$ and so $\nu_a(E) = \nu'_a(E) = 0$; and so $\nu_s(E) = \nu'_s(E) = \nu(E)$. On the other hand, if $E \subset A \cap A'$, then $\nu_s(E) = \nu'_s(E) = 0$. Combining, we deduce that $\nu'_s = \nu_s$. Hence $\nu'_a = \nu_a$ as well. ■

We finish this section with a discussion of measures taking complex values. These measures are not allowed to take infinite values.

4.3.6. Definition. A *complex measure* on $(X, \mathcal{B})$ is a map $\nu : \mathcal{B} \to \mathbb{C}$ such that $\nu(\varnothing) = 0$ that is countably additive: If $E = \dot{\bigcup}_{i\geq 1} E_i$, then $\nu(E) = \sum_{i\geq 1} \nu(E_i)$ and this series converges absolutely.

4.3.7. REMARK. As in Remark 4.2.2, when $E = \dot{\bigcup}_{i\geq 1} E_i$, countable additivity together with the fact that every set has a finite measure forces absolute convergence of the series $\sum_{i\geq 1} \nu(E_i)$.

Note that $\operatorname{Re}\nu$ and $\operatorname{Im}\nu$ are finite signed measures. Thus there are positive *finite* measures ν_i for $1 \leq i \leq 4$ so that $\operatorname{Re}\nu = \nu_1 - \nu_2$ and $\operatorname{Im}\nu = \nu_3 - \nu_4$. Hence $\nu = \nu_1 - \nu_2 + i\nu_3 - i\nu_4$.

4.3.8. EXAMPLE. The main example (and essentially the only example) is obtained as follows: let μ be a measure and let $f \in L^1(\mu)$. Define $\nu(E) = \int_E f\,d\mu$. You can readily check that this is a complex measure. We will write $d\nu = f\,d\mu$ and $\dfrac{d\nu}{d\mu} = f$ in this case.

The second statement of the following result is also sometimes called the Radon-Nikodym Theorem.

4.3.9. THEOREM. *If ν is a complex measure on $(X, \mathcal{B})$, there is a unique finite measure $|\nu|$ and measurable function h with $|h| = 1$ a.e.$(|\nu|)$ so that $d\nu = h\,d|\nu|$. Moreover if μ is a σ-finite measure on $(X, \mathcal{B})$, then ν decomposes as $\nu_a + \nu_s$ where $\nu_a = f\,d\mu$ for $f \in L^1(\mu)$ and ν_s is supported on a μ-null set.*

PROOF. With the notation as in the discussion preceding the proposition, let $\mu = \nu_1 + \nu_2 + \nu_3 + \nu_4$. Then $\operatorname{Re}\nu \ll \mu$ and $\operatorname{Im}\nu \ll \mu$. By the Radon-Nikodym Theorem, there are $f, g \in L^1(\mu)$ so that $d\operatorname{Re}\nu = f\,d\mu$ and $d\operatorname{Im}\nu = g\,d\mu$. Hence $\nu(E) = \displaystyle\int_E f + ig\,d\mu$. Define $|\nu|(E) = \displaystyle\int_E |f + ig|\,d\mu$ and $h = \operatorname{sign}(f + ig)$. Then $|h| = 1$ a.e.$(|\nu|)$ and $\nu(E) = \displaystyle\int_E h\,d|\nu|$. Also $|\nu|(X) = \|f + ig\|_1 < \infty$; so $|\nu|$ is finite. Uniqueness is left as an exercise.

Now if μ is a σ-finite, the Lebesgue deomposition for $|\nu|$ yields $|\nu| = |\nu|_a + |\nu|_s$ where $d|\nu|_a = f\,d\mu$ for $f \in L^+ \cap L^1(\mu)$ and $|\nu|_s$ is supported on a μ-null set A. Then $d\nu = h\,d|\nu| = h\,d|\nu|_a + h\,d|\nu|_s = hf\,d\mu + h\,d|\nu|_s$. Thus $d\nu_a = hf\,d\mu \ll \mu$ and $hf \in L^1(\mu)$ and $d\nu_s = h\,d|\nu|_s$ is supported on the μ-null set A. ∎

Finally, we can integrate with respect to complex measures. If $d\nu = hd|\nu|$ is a complex measure and $f \in L^1(|\nu|)$, then we define

$$\int f\,d\nu := \int fh\,d|\nu|.$$

It is easy to check that this is linear. It is advisable to convert to integrals with respect to positive measures when taking limits.

CHAPTER 5

L^p spaces

5.1. L^p as a Banach space

5.1.1. DEFINITION. Let $(X, \mathcal{B}, \mu)$ be a measure space. For $1 \le p < \infty$, let

$$\text{“}L^p(\mu)\text{”} = \Big\{ f \text{ measurable, complex valued} : \|f\|_p^p = \int |f|^p \, d\mu < \infty \Big\}.$$

Set

$$\mathcal{N} = \{ f \text{ measurable, complex valued} : f = 0 \, \text{a.e.}(\mu) \}$$

and define $L^p(\mu) = \text{“}L^p(\mu)\text{”}/\mathcal{N}$ with the norm $\|[f]\|_p = \|f\|_p$.

“$L^\infty(\mu)$” $= \{ f$ measurable, complex valued : $\|f\|_\infty = \operatorname{ess\,sup} |f| < \infty \}$ where $\operatorname{ess\,sup} |f| = \sup\{ t \ge 0 : \mu(\{x : |f(x)| > t\} > 0 \}$. Set $L^\infty(\mu) = \text{“}L^\infty(\mu)\text{”}/\mathcal{N}$ with norm $\|[f]\|_\infty = \|f\|_\infty$.

By convention, we write elements of $L^p(\mu)$ as f, where the equivalence class is understood.

Note that “L^p” is a linear space because if $f, g \in L^p(\mu)$, then $\lambda f \in L^p(\mu)$ with $\|\lambda f\|_p = |\lambda| \, \|f\|_p$ for $\lambda \in \mathbb{C}$; and

$$|f + g|^p \le (2 \max\{|f|, |g|\})^p \le 2^p(|f|^p + |g|^p).$$

whence $\|f + g\|_p^p \le 2(\|f\|_p + \|g\|_p) < \infty$. Notice that $\mathcal{N}$ is a subspace of “$L^p(\mu)$”, so that $L^p(\mu)$ is also a vector space. Moreover, if f is measurable, then $\|f\|_p^p = \int |f|^p \, d\mu = 0$ if and only if $f = 0 \, \text{a.e.}(\mu)$ if and only if $f \in \mathcal{N}$. In particular, $\| \cdot \|_p$ is only a seminorm on “$L^p(\mu)$”. However, for $f \in L^p(\mu)$, we have $\|f\|_p = 0$ if and only if $f = 0$. To verify that $\| \cdot \|_p$ is a norm on $L^p(\mu)$, we need to verify the triangle inequality, which means eliminating the annoying 2 in the inequality above.

A function belongs to “$L^\infty(\mu)$” if it agrees with a bounded function a.e.(μ). The triangle inequality is very easy here. It is also very easy for $L^1(\mu)$ (see section 3.4).

5.1.2. MINKOWSKI'S INEQUALITY. *Let $(X, \mathcal{B}, \mu)$ be a measure space. For $1 < p < \infty$, the triangle inequality is valid for $L^p(\mu)$. Equality holds only when f and g lie in a 1-dimensional subspace.*

PROOF. Let $f, g \in L^p(\mu)$. We may suppose that neither is 0, as that case is trivial. Define $A = \|f\|_p$ and $B = \|g\|_p$, and set $f_0 = f/A$ and $g_0 = g/B$; so $\|f_0\|_p = 1 = \|g_0\|_p$.

Consider $\varphi(x) = x^p$ on $[0, \infty)$. Note that $\varphi''(x) = p(p-1)x^{p-2} > 0$ on $(0, \infty)$, and thus $\varphi(x)$ is a strictly convex function, meaning that for all $x_1, x_2 \in [0, \infty)$ and $0 \le t \le 1$,

$$\varphi(tx_1 + (1-t)x_2) \le t\varphi(x_1) + (1-t)\varphi(x_2),$$

with equality only when $x_1 = x_2$ or $t = 0$ or $t = 1$. That is every chord between distinct points on the curve $y = \varphi(x)$ lies strictly above the curve.

Hence for any $x \in X$,

$$\Big(\frac{A}{A+B}|f_0(x)| + \frac{B}{A+B}|g_0(x)|\Big)^p \le \frac{A}{A+B}|f_0(x)|^p + \frac{B}{A+B}|g_0(x)|^p$$

with equality only when $|f_0(x)| = |g_0(x)|$. Also

$$\frac{1}{A+B}|f(x) + g(x)| \le \frac{|f(x)| + |g(x)|}{A+B} = \frac{A}{A+B}|f_0(x)| + \frac{B}{A+B}|g_0(x)|,$$

with equality only when $\operatorname{sign}(f(x)) = \operatorname{sign}(g(x))$. Integrate the pth power:

$$\begin{aligned}\frac{1}{(A+B)^p}\int |f(x) + g(x)|^p\,d\mu &\le \frac{A}{A+B}\int |f_0(x)|^p\,dx + \frac{B}{A+B}\int_a^b |g_0(x)|^p\,d\mu \\ &= \frac{A}{A+B}\|f_0\|_p^p + \frac{B}{A+B}\|g_0\|_p^p = 1.\end{aligned}$$

Multiplying through by $(A+B)^p$ and take the pth root to get

$$\|f+g\|_p \le (A+B) = \|f\|_p + \|g\|_p.$$

For equality, we require that $|f_0(x)| = |g_0(x)|$ a.e. and $\operatorname{sign}(f(x)) = \operatorname{sign}(g(x))$ a.e. This implies that $f_0 = g_0$ a.e.(μ), so $g = Bf/A$; i.e., g is a scalar multiple of f. ■

Now that we have established that $L^p(\mu)$ is a normed space, we will show that it is complete, and so is a Banach space. This extends the result for $L^1(\mu)$, Theorem 3.4.2. This is another important piece of evidence that this is the right context for integration.

5.1.3. RIESZ-FISCHER THEOREM. *Let $(X, \mathcal{B}, \mu)$ be a measure space. For $1 \le p \le \infty$, $L^p(\mu)$ is complete.*

PROOF. Suppose that $(f_n)_{n\ge1}$ is a Cauchy sequence in $L^p(\mu)$. Select a subsequence $(n_j)_{j\ge1}$ so that $\|f_{n_j} - f_m\|_p < 2^{-j}$ for all $m > n_j$. Define

$$h_k = |f_{n_1}| + \sum_{j=1}^{k-1} |f_{n_{j+1}} - f_{n_j}| \quad \text{and} \quad h = \lim_{k\to\infty} h_k.$$

Note that $0 \le h_k \le h_{k+1}$. By the triangle inequality,

$$\|h_k\|_p \le \|f_{n_1}\|_p + \sum_{j=1}^{k-1} 2^{-j} < \|f_{n_1}\|_p + 1.$$

Since h_k^p are increasing to h^p, the MCT shows that $\|h\|_p \le \|f_{n_1}\|_p + 1$ when $p < \infty$. For $p = \infty$, $\|h\|_\infty \le \|f_{n_1}\|_p + 1$ follows directly. So $h \in L^p(\mu)$, and in particular, $h(x) < \infty$ a.e.(μ).

Note that $f_{n_k} = f_{n_1} + \sum_{j=1}^{k-1}(f_{n_{j+1}} - f_{n_j})$. Let

$$f = f_{n_1} + \sum_{j\ge 1}(f_{n_{j+1}} - f_{n_j}) = \lim_{k\to\infty} f_{n_k}.$$

This series converges absolutely whenever $h(x) < \infty$, and so is defined a.e.(μ). Moreover $|f| \le h$, so $f \in L^p(\mu)$. Also $|f - f_{n_k}|^p \le \big(\sum_{j\ge k} |f_{n_j} - f_{n_{j+1}}|\big)^p \le h^p$ when $p < \infty$. Thus by the LDCT,

$$\lim_{k\to\infty} \|f - f_{n_k}\|_p^p = \lim_{k\to\infty} \int |f - f_{n_k}|^p \, d\mu = \int \lim_{k\to\infty} |f - f_{n_k}|^p \, d\mu = 0.$$

When $p = \infty$, $\|f - f_{n_k}\|_\infty \le \sum_{j\ge k} \|f_{n_j} - f_{n_{j+1}}\|_\infty < 2^{1-k}$. Therefore f_{n_k} converges to f in $L^p(\mu)$. Since $(f_n)_{n\ge 1}$ is Cauchy, in fact $f_n \to f$ in $L^p(\mu)$. So $L^p(\mu)$ is complete. ■

5.1.4. Examples.

(1) $L^p(0,1) = L^p((0,1), m)$. Here Lebesgue measure, m, is a probability measure on $(0,1)$. If $1 \le p < r < \infty$ and $f \in L^r(0,1)$, then

$$\|f\|_p^p = \int |f|^p \, dm \le \int_{|f|\le 1} 1 \, dm + \int_{|f|>1} |f|^r \, dm \le 1 + \|f\|_r^r < \infty.$$

Thus $f \in L^p(0,1)$. So $L^1(0,1) \supset L^p(0,1) \supset L^r(0,1) \supset L^\infty(0,1)$. These subspaces are not closed in the larger spaces. In the next section, we will improve the norm inequality. We also write $L^p(\mathbb{R})$ for $L^p((\mathbb{R}, m))$.

(2) Consider $(\mathbb{N}, \mathcal{P}(\mathbb{N}), m_c)$. Then

$$L^p(\mathbb{N}, m_c) = l^p = \Big\{(a_i)_{i\ge 1} : \|(a_i)\|_p = \big(\sum_{i\ge 1} |a_i|^p\big)^{1/p} < \infty\Big\}$$

for $p < \infty$, and $L^\infty(\mathbb{N}, m_c) = l^\infty = \big\{(a_i)_{i\ge 1} : \|(a_i)\|_\infty = \sup_{i\ge 1} |a_i| < \infty\big\}$. If $1 \le p < r < \infty$ and $(a_i) \in l_p$, then

$$\begin{aligned} \|(a_i)\|_r^r &= \sum_{i\ge 1} |a_i|^r = \sum_{i\ge 1} |a_i|^p \, |a_i|^{r-p} \\ &\le \big(\sum_{i\ge 1} |a_i|^p\big) \sup |a_i|^{r-p} \le \|(a_i)\|_p^p \, \|(a_i)\|_p^{r-p} = \|(a_i)\|_p^r. \end{aligned}$$

Thus $\|(a_i)\|_r \le \|(a_i)\|_p$. In particular, $l^1 \subset l^p \subset l^r \subset l^\infty$.

(3) Let μ be the measure on $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ given by $\mu(A) = \infty$ is $A \neq \varnothing$ and $\mu(\varnothing) = 0$. Then $L^p(\mu) = \{0\}$ if $1 \leq p < \infty$ and $L^\infty(\mu) = l^\infty$. You need sets of non-zero finite measure to get interesting functions in $L^p(\mu)$.

5.1.5. PROPOSITION. *Let $(X, \mathcal{B}, \mu)$ be a measure and let $1 \leq p < \infty$. Then the simple functions of finite support ($\varphi = \sum_{i=1}^n a_i \chi_{A_i}$, where $\mu(A_i) < \infty$) are dense in $L^p(\mu)$. For $p = \infty$, the set of all simple functions is dense in $L^\infty(\mu)$.*

PROOF. Fix $f \in L^p(\mu)$ for $p < \infty$. Choose simple functions φ_n so that

$$|\varphi_n| \leq |\varphi_{n+1}| \leq |f| \quad \text{and} \quad f(x) = \lim_{n \to \infty} \varphi_n(x).$$

Then $|\varphi_n|^p \leq |f|^p$ is integrable, so $\varphi_n \in L^p(\mu)$. In particular, each set A_i in the definition of φ_n for which $a_i \neq 0$ must have finite measure. Also since $|f - \varphi_n| \leq |f| + |\varphi_n| \leq 2|f|$, we have $|f - \varphi_n|^p \leq 2^p |f|^p$. Therefore by the LDCT,

$$\lim_{n \to \infty} \|f - \varphi_n\|_p^p = \lim_{n \to \infty} \int |f - \varphi_n|^p \, d\mu = \int \lim_{n \to \infty} |f - \varphi_n|^p \, d\mu = 0.$$

So φ_n converge to f in $L^p(\mu)$.

When $p = \infty$, suppose that $\|f\|_\infty = N < \infty$. For $n \geq 1$, and $(j, k) \in [-nN, nN]^2$, let $A_{j,k} = \{x : \frac{j}{n} \leq \operatorname{Re} f(x) < \frac{j+1}{n}, \ \frac{k}{n} \leq \operatorname{Im} f(x) < \frac{k+1}{n}\}$. These are measurable sets, and $\varphi_n = \sum_{j,k=-nN}^{nN} \frac{j+ik}{n} \chi_{A_{j,k}}$ satisfies $\|f - \varphi_n\|_\infty \leq \frac{\sqrt{2}}{n}$. So the simple functions are dense. If $\mu(A) = \infty$, then χ_A is not the limit of simple functions with finite support. ■

Recall that if X is a topological space, then $C_c(X)$ denotes the space of continuous functions on X with compact support (i.e. $\overline{\{x : f(x) \neq 0\}}$ is compact).

5.1.6. COROLLARY. *If $1 \leq p < \infty$, then $C_c(\mathbb{R})$ is dense in $L^p(\mathbb{R})$. This is false for $p = \infty$.*

PROOF. Let $f \in L^p(\mathbb{R})$ and let $\varepsilon > 0$. Use Proposition 5.1.5 to find a simple function of finite support φ with $\|f - \varphi\|_p < \varepsilon/2$. Write $\varphi = \sum_{i=1}^n a_i \chi_{A_i}$, where $\mu(A_i) < \infty$. Let $L = \sup |a_i|$. By the regularity of Lebesgue measure (Theorem 1.5.6), there are compact sets K_i so that $K_i \subset A_i$ and $m(A_i \setminus K_i) < \frac{\varepsilon}{4nL}$. Choose disjoint open sets $U_i \supset K_i$ so that $m(U_i \setminus K_i) < \frac{\varepsilon}{4nL}$ and $\overline{U_i}$ is compact. For each $1 \leq i \leq n$, let $h_i = \frac{\operatorname{dist}(x, U_i^c)}{\operatorname{dist}(x, K_i) + \operatorname{dist}(x, U_i^c)}$. Note that this is a continuous function which vanishes off of U_i, has $h_i(x) = 1$ on K_i, and takes values in $[0, 1]$. Then $\chi_{A_i} - h_i$ vanishes on K_i and on $U_i^c \cap A_i^c$; so

$$\|\chi_{A_i} - h_i\|_p^p \leq m(A_i \setminus K_i) + m(U_i \setminus A_i) \leq \frac{\varepsilon}{4nL} + \frac{\varepsilon}{4nL} = \frac{\varepsilon}{2nL}.$$

Hence $h = \sum_{i=1}^n a_i h_i$ belongs to $C_c(\mathbb{R})$ and

$$\|\varphi - h\|_p \le \sum_{i=1}^n |a_i| \|\chi_{A_i} - h_i\|_p < nL\frac{\varepsilon}{2nL} = \frac{\varepsilon}{2}.$$

Thus $\|f - h\|_p < \varepsilon$.

Clearly the constant function 1 is not a limit in $L^\infty(\mathbb{R})$ of functions of compact support. So this fails for $p = \infty$. ∎

5.2. Duality for Normed Vector Spaces

In this section, we recall some basic facts about linear functionals on normed vector spaces.

5.2.1. Definition. Let $(V, \|\cdot\|)$ be a normed vector space over $\mathbb{F} \in \{\mathbb{R}, \mathbb{C}\}$. Let $\mathcal{L}(V, \mathbb{F})$ denote the vector space of linear maps of V into the scalars, *linear functionals*, and let V^* denote the *dual space of* V consisting of all continuous linear functionals.

5.2.2. Proposition. *Let $(V, \|\cdot\|)$ be a normed vector space over $\mathbb{F}$, and let $\varphi \in \mathcal{L}(V, \mathbb{F})$. The following are equivalent:*

(1) *φ is continuous.*

(2) $\|\varphi\| := \sup\{|\varphi(v)| : \|v\| \le 1\} < \infty$.

(3) *φ is continuous as $v = 0$.*

Proof. (2) $\Rightarrow$ (1). If $u \ne v \in V$, set $w = \frac{u-v}{\|u-v\|}$ and note that

$$|\varphi(u) - \varphi(v)| = |\varphi(u - v)| = |\varphi(w)|\,\|u - v\| \le \|\varphi\|\|u - v\|.$$

Hence φ is Lipschitz, and in particular is continuous.

(1) $\Rightarrow$ (3) is trivial.

(3) $\Rightarrow$ (2). Assume that (2) fails. Then there are vectors $v_n \in V$ with $\|v_n\| = 1$ and $|\varphi(v_n)| > n^2$. Thus $\frac{1}{n}v_n \to 0$ while $|\varphi(\frac{1}{n}v_n)| > n$ diverges. So φ is discontinuous at 0. The result follows. ∎

5.2.3. Theorem. *Let $(V, \|\cdot\|)$ be a normed vector space. Then $(V^*, \|\cdot\|)$ is a Banach space.*

Proof. First we show that $\|\cdot\|$ is a norm. Clearly $\|\varphi\| = 0$ if and only if $\varphi(v) = 0$ for all $v \in V$ with $\|v\| \le 1$. This forces $\varphi = 0$ by linearity. Also if

$\lambda \in \mathbb{F}$, then

$$\|\lambda\varphi\| = \sup_{\|v\|\le 1} |\lambda\varphi(v)| = |\lambda| \sup_{\|v\|\le 1} |\varphi(v)| = |\lambda|\,\|\varphi\|.$$

For the triangle inequality, take $\varphi, \psi \in V^*$.

$$\|\varphi+\psi\| = \sup_{\|v\|\le 1} |\varphi(v)+\psi(v)| \le \sup_{\|v\|\le 1} |\varphi(v)| + \sup_{\|v\|\le 1} |\psi(v)| = \|\varphi\| + \|\psi\|.$$

To establish completeness, let $(\varphi_n)_{n\ge 1}$ be a Cauchy sequence in V^*. For each $v \in V$, $|\varphi_m(v)-\varphi_n(v)| \le \|\varphi_m-\varphi_n\|\,\|v\|$. It follows that $(\varphi_n(v))_{n\ge 1}$ is a Cauchy sequence in $\mathbb{F}$. Thus we may define $\varphi(v) = \lim_{n\to\infty} \varphi_n(v)$. Then

$$\varphi(\lambda u+\mu v) = \lim_{n\to\infty} \varphi_n(\lambda u+\mu v) = \lim_{n\to\infty} \lambda\varphi_n(u)+\mu\varphi_n(v) = \lambda\varphi(u)+\mu\varphi(v).$$

Therefore φ is linear. Now let $\varepsilon > 0$ and select N so that if $m, n \ge N$, then $\|\varphi_m - \varphi_n\| < \varepsilon$. In particular, if $\|v\| \le 1$, we have $|\varphi_m(v) - \varphi_n(v)| < \varepsilon$. Holding m fixed and letting $n \to \infty$, we obtain that $|\varphi_m(v) - \varphi(v)| \le \varepsilon$. Taking the supremum over all v with $\|v\| \le 1$ yields $\|\varphi_m - \varphi\| \le \varepsilon$ when $m \ge N$. In particular, $\|\varphi\| \le \|\varphi_m\| + \|\varphi_m - \varphi\| < \infty$; so $\varphi \in V^*$. Moreover we have shown that $\lim_{m\to\infty} \varphi_m = \varphi$ in $(V^*, \|\cdot\|)$. So V^* is complete. ■

5.3. Duality for L^p

In this section, we determine the Banach space dual of the spaces $L^p(\mu)$ for $1 \le p < \infty$. We need another import inequality.

5.3.1. LEMMA. *If $a, b \in (0,\infty)$ and $0 \le t \le 1$, then*

$$a^t b^{1-t} \le ta + (1-t)b$$

with equality only when $a = b$ or $t = 0$ or 1.

PROOF. This is just the AMGM inequality. The function $f(x) = e^x$ is strictly convex. So $e^{t\alpha+(1-t)\beta} \le te^\alpha + (1-t)e^\beta$, with equality only when $\alpha = \beta$ or $t = 0$ or 1. Take $a = e^\alpha$ and $b = e^\beta$ and the result follows. ■

5.3.2. HÖLDER'S INEQUALITY. *Let $(X, \mathcal{B}, \mu)$ be a measure space. Let $1 < p < \infty$ and define q so that $\frac{1}{p} + \frac{1}{q} = 1$. If $f \in L^p(\mu)$ and $g \in L^q(\mu)$, then $fg \in L^1(\mu)$ and*

$$\|fg\|_1 \le \|f\|_p\,\|g\|_q.$$

Equality holds if and only if $|f|^p$ and $|g|^q$ are collinear.

PROOF. We may suppose that f, g are non-zero since the inequality is trivial if $fg = 0$ a.e.(μ). Let $f_0 = |f|/A$ where $A = \|f\|_p$ and $g_0 = |g|/B$ where

$B = \|g\|_q$. Apply Lemma 5.3.1 to $a = f_0(x)^p$ and $b = g_0(x)^q$ with $t = \frac{1}{p}$ and $1 - t = \frac{1}{q}$. Then

$$\frac{|f(x)g(x)|}{AB} = f_0(x)g_0(x) \le \frac{1}{p}f_0(x)^p + \frac{1}{q}g_0(x)^q = \frac{1}{pA^p}|f(x)|^p + \frac{1}{qB^q}|g(x)|^q.$$

Integrate to get

$$\frac{\|fg\|_1}{AB} \le \frac{1}{pA^p}\|f(x)\|_p^p + \frac{1}{qB^q}\|g(x)\|_q^q = \frac{1}{p} + \frac{1}{q} = 1.$$

Hence $\|fg\|_1 \le AB = \|f\|_p\,\|g\|_q$.

Equality holds in the first inequality only when $f_0(x)^p = g_0(x)^q$. For it to hold for the integral, this identity must hold a.e.(μ). Hence $|g|^q = \frac{B}{A}|f|^p$ a.e.(μ). ∎

For more elementary reasons, if $f \in L^1(\mu)$ and $g \in L^\infty(\mu)$, we have that $fg \in L^1(\mu)$ and $\|fg\|_1 \le \|f\|_1\,\|g\|_\infty$.

5.3.3. EXAMPLE. We return to Example 5.1.4(1). Suppose that μ is a probability measure. If $1 \le p < r < \infty$ and $f \in L^r(\mu)$, choose s so that $\frac{p}{r} + \frac{1}{s} = 1$. Note that $|f|^p \in L^{r/p}(\mu)$. Then

$$\|f\|_p^p = \big\||f|^p 1\big\|_1 \le \big\||f|^p\big\|_{r/p}\|1\|_s = \|f\|_r^p.$$

Hence $\|f\|_p \le \|f\|_r$.

More generally if $\mu(X) < \infty$, the same computation yields

$$\|f\|_p \le \|1\|_s^{1/p}\,\|f\|_r = \mu(X)^{1/sp}\,\|f\|_r = \mu(X)^{\frac{1}{p}-\frac{1}{r}}\,\|f\|_r.$$

When every point has measure 1, you get the l^p spaces, where the inequalities are reversed. See Example 5.1.4(2).

For the spaces $L^p(\mathbb{R})$, there is no containment between $L^p(\mathbb{R})$ and $L^r(\mathbb{R})$ when $p \ne r$.

We now get to the main result of this section.

5.3.4. THEOREM (Riesz). *Let $(X, \mathcal{B}, \mu)$ be a σ-finite measure. Suppose that $1 \le p < \infty$, and let q satisfy $\frac{1}{p} + \frac{1}{q} = 1$, where $q = \infty$ when $p = 1$. Then $L^p(\mu)^* = L^q(\mu)$ via the isometric pairing $L^q(\mu) \ni g \to \Phi_g$, where $\Phi_g(f) = \int fg\,d\mu$.*

5.3.5. LEMMA. *Suppose that μ is σ-finite, $1 \le p < \infty$ and $g \in L^q(\mu)$. Then $\|\Phi_g\| = \|g\|_q$.*

PROOF. First suppose that $1 < p < \infty$. By Hölder's inequality

$$|\Phi_g(f)| = \Big|\int fg\,d\mu\Big| \le \|fg\|_1 \le \|f\|_p\,\|g\|_q.$$

Hence $\|\Phi_g\| = \sup_{\|f\|_p \le 1} |\Phi_g(f)| \le \|g\|_q$. On the other hand, if $g \ne 0$, let $f = \dfrac{|g|^{q-1}\overline{\operatorname{sign}(g)}}{\|g\|_q^{q-1}}$. We will use that $p(q-1) = pq(1-\frac{1}{q}) = q$. Then $|f|^p = \dfrac{|g|^q}{\|g\|_q^q}$. So

$$\|f\|_p^p = \int \frac{|g|^q}{\|g\|_q^q}\, d\mu = 1.$$

Therefore

$$\|\Phi_g\| \ge \Phi_g(f) = \int \frac{|g|^{q-1}\overline{\operatorname{sign}(g)}}{\|g\|_q^{q-1}} g\, d\mu = \frac{1}{\|g\|_q^{q-1}} \int |g|^q\, d\mu = \|g\|_q.$$

For $p = 1$, take $g \in L^\infty(\mu)$. As above, $|\Phi_g(f)| \le \displaystyle\int |f|\, \|g\|_\infty\, d\mu = \|f\|_1\, \|g\|_\infty$. Conversely, given $\varepsilon > 0$, let $A = \{x : |g(x)| > \|g\|_\infty - \varepsilon\}$. Then $\mu(A) > 0$. Since μ is σ-finite, there is a measurable subset $E \subset A$ with $0 < \mu(E) < \infty$. Let $f = \dfrac{\overline{\operatorname{sign}(g)}}{\mu(E)} \chi_E$. Then $\|f\|_1 = 1$ and $\Phi_g(f) = \frac{1}{\mu(E)} \displaystyle\int_E |g|\, d\mu > \|g\|_\infty - \varepsilon$. Thus $\|\Phi_g\| = \|g\|_\infty$. ■

5.3.6. LEMMA. *Suppose that μ is σ-finite, $1 \le p < \infty$ and g is a measurable function such that*

$$\Big| \int \varphi g\, d\mu \Big| \le M \|\varphi\|_p \quad \textit{for all } \varphi \textit{ simple, finite support}.$$

Then $g \in L^q(\mu)$ and $\|g\|_q \le M$.

PROOF. First take $1 < p < \infty$, so that $q < \infty$. Suppose first that g is real valued. Choose simple functions ψ_n so that $|\psi_n| \le |\psi_{n+1}| \le |g|$ and $\psi_n \to g$. Since μ is σ-finite, we can write $X = \bigcup_{n\ge1} X_n$ where $X_n \subset X_{n+1}$ and $\mu(X_n) < \infty$ for all n. Then $\varphi_n = \psi_n \chi_{X_n}$ are simple functions with finite support such that $|\varphi_n| \le |\varphi_{n+1}| \le |g|$ and $\varphi_n \to g$. Analogous to the previous lemma, define

$$f_n = \frac{|\varphi_n|^{q-1} \operatorname{sign}(g)}{\|\varphi_n\|_q^{q-1}}.$$

Note that $\operatorname{sign}(g)$ take only the values $\pm1, 0$ and so f_n is a simple function of finite support. As in the previous lemma, $\|f_n\|_p = 1$. Therefore

$$\begin{aligned} M &\ge \sup_{n\ge1} \Big| \int f_n g\, d\mu \Big| = \sup_{n\ge1} \Big| \int \frac{|\varphi_n|^{q-1}|g|}{\|\varphi_n\|_q^{q-1}}\, d\mu \Big| \\ &\ge \sup_{n\ge1} \Big| \int \frac{|\varphi_n|^q}{\|\varphi_n\|_q^{q-1}}\, d\mu \Big| = \sup_{n\ge1} \|\varphi_n\|_q = \|g\|_q. \end{aligned}$$

The last equality follows from the MCT.

Now for complex valued g, note that $\operatorname{Re} g$ and $\operatorname{Im} g$ satisfy the hypotheses of the lemma. Thus they both belong to $L^q(\mu)$, and hence $g \in L^q(\mu)$. So by Lemma 5.3.5,

$\|\Phi_g\| = \|g\|_q$. Proposition 5.1.5 shows that simple functions of finite support are dense in $L^p(\mu)$, and hence the optimal constant M must be $\|\Phi_g\|$; so $\|g\|_q \le M$.

For $p = 1$, we need to show that $\|g\|_\infty \le M$. If this fails, then for some $N > M$, $\{x : |g(x)| > N\}$ has positive measure. Hence there is a $\theta \in \mathbb{R}$ so that $A = \{x : \operatorname{Re} e^{i\theta} g(x) > M\}$ has positive measure. Let $E \subset A$ have positive and finite measure. Define $f = e^{i\theta}\mu(E)^{-1}\chi_E$. Then $\|f\|_1 = 1$ and $M \ge \operatorname{Re} \int fg\,d\mu > M$. This is a contradiction, and hence $\|g\|_\infty \le M$. ■

PROOF OF THEOREM 5.3.4. First assume that $\mu(X) < \infty$. Let $\Phi \in L^p(\mu)^*$. Define $\nu(E) = \Phi(\chi_E)$ for $E \in \mathcal{B}$. Note that $\nu(\varnothing) = \Phi(0) = 0$. Also if $E = \dot{\bigcup}_{i\ge1} E_i$ is a disjoint union of sets in $\mathcal{B}$, then

$$\Big\|\chi_E - \sum_{i=1}^n \chi_{E_i}\Big\|_p^p = \mu\big(\dot{\bigcup}_{i>n} E_i\big) \to 0 \quad \text{as } n \to \infty.$$

Since Φ is continuous,

$$\nu(E) = \Phi(\chi_E) = \lim_{n\to\infty} \Phi\Big(\sum_{i=1}^n \chi_{E_i}\Big) = \sum_{i\ge1} \nu(E_i).$$

Therefore ν is countably additive, and thus is a complex measure. Moreover if $\mu(E) = 0$, then $\chi_E = 0$ a.e.(μ) and hence $\nu(E) = \Phi(0) = 0$. Thus $\nu \ll \mu$.

By the Radon-Nikodym Theorem 4.3.2, there is a measurable $g \in L^1(\mu)$ so that $\nu(E) = \displaystyle\int_E g\,d\mu$. If $\varphi = \sum_{i=1}^n a_i\chi_{E_i}$ is a simple function (with finite support), then

$$\Phi(\varphi) = \sum_{i=1}^n a_i \nu(E_i) = \int \varphi\,d\nu = \int \varphi g\,d\mu.$$

Thus

$$\Big|\int \varphi g\,d\mu\Big| = |\Phi(\varphi)| \le \|\Phi\|\,\|\varphi\|_p.$$

By Lemma 5.3.6, $g \in L^q(\mu)$, and so Φ_g is a continuous functional on $L^p(\mu)$ which agrees with Φ on all simple functions of finite support. Such functions are dense in $L^p(\mu)$ by Proposition 5.1.5. So by continuity, $\Phi = \Phi_g$. By Lemma 5.3.5, $\|\Phi\| = \|\Phi_g\| = \|g\|_q$. Moreover this shows that g is unique, because distinct $L^p(\mu)$ functions yield distinct functionals.

Now consider when $\mu(X) = \infty$. We can write $X = \bigcup_{n\ge1} X_n$ where $X_n \subset X_{n+1}$ and $\mu(X_n) < \infty$. We can restrict Φ to $L^p(X_n, \mu|_{X_n})$ (since this is a closed subspace of $L^p(\mu)$). Call it Φ_n. The first part of the proof shows that thee is a unique $g_n \in L^p(X_n, \mu|_{X_n})$ so that $\Phi_n = \Phi_{g_n}$ and $\|g_n\|_q \le \|\Phi\|$. Let $g = \bigcup_{n\ge1} g_n$ be the unique measurable function such that $g|_{X_n} = g_n$. The various functions must match up a.e.(μ) because of uniqueness. By the MCT,

$$\|g\|_q = \lim_{n\to\infty} \|g_n\|_q \le \|\Phi\|.$$

So Φ and Φ_g agree on the dense subspace $\bigcup_{n\geq 1} L^p(X_n, \mu|_{X_n})$ and they are continuous functionals; and so $\Phi = \Phi_g$. Again by Lemma 5.3.5, $\|\Phi\| = \|g\|_q$. ■

5.3.7. REMARK. In fact, σ-finiteness is not needed to obtain $L^p(\mu)^* = L^q(\mu)$ provided that $1 < p < \infty$. See Folland's book for the details.

However it is critical to establish that $L^1(\mu)^* = L^\infty(\mu)$. In Example 5.1.4 (3), each point in $\mathbb{N}$ has $\mu(\{n\}) = \infty$ and so $L^1(\mu) = \{0\}$ while $L^\infty(\mu) = l^\infty$. However $L^1(\mu)^* = \{0\}$.

A more subtle example is to take $(\mathbb{R}, \mathcal{P}(\mathbb{R}), m_c)$. Let $\mathcal{B}$ be the σ-algebra of countable and co-countable subsets of $\mathbb{R}$. Let $\mu = m_c|_{\mathcal{B}}$. Any function in $L^1(m_c)$ has countable support, and thus is $\mathcal{B}$-measurable. Therefore $L^1(\mu) = L^1(m_c)$. However changing the measure has a dramatic effect on the L^∞ spaces. The space $L^\infty(m_c) = l^\infty(\mathbb{R})$ is the space of all bounded functions on $\mathbb{R}$. However for a function to be $\mathcal{B}$-measurable, it must be constant on a co-countable set. In fact, $L^1(\mu)^* = L^1(m_c)^* = L^\infty(m_c) = l^\infty(\mathbb{R})$. One way to see this is that if $\Phi \in L^1(\mu)^*$, we can define $g(r) = \Phi(\chi_{\{r\}})$. Then $|g(r)| \leq \|\Phi\|$, so that g is a bounded function; and so $g \in L^\infty(m_c)$. Moreover, if $f \in L^1(\mu)$, then there is a countable (or finite) collection of real numbers $\{r_n : n \geq 1\}$ and scalars α_n so that $f = \sum_{n\geq 1} \alpha_n \chi_{\{r_n\}}$. Moreover $\|f\|_1 = \sum_{n\geq 1} |\alpha_n|$; whence this series converges absolutely. Therefore by continuity,

$$\Phi(f) = \sum_{n\geq 1} \alpha_n \Phi(\chi_{\{r_n\}}) = \sum_{n\geq 1} \alpha_n g(n).$$

Moreover this same computation shows that any element of $L^\infty(m_c)$ determines a distinct continuous functional on $L^1(\mu)$.

.

CHAPTER 6

Some Topology

This is a brief introduction skewed to certain things needed in the next section. This is not intended as a comprehensive introduction to point set topology. A good general reference is Willard's book [3].

6.1. Topological spaces

6.1.1. DEFINITION. A *topology* τ on a set X is a collection of subsets such that

(1) $\varnothing,\ X \in \tau$.

(2) If $\{U_\lambda : \lambda \in \Lambda\} \subset \tau$, then $\bigcup_{\lambda\in\Lambda} U_\lambda \in \tau$.

(3) If $U_1, \dots, U_n \in \tau$, then $\bigcap_{i=1}^n U_i \in \tau$.

The elements $U \in \tau$ are called *open sets*.

6.1.2. EXAMPLES.

(1) If (X, d) is a metric space, then U is open if for every $x \in U$, there is an $r > 0$ so that the open ball $b_r(x) \subset U$.

(2) If X is any set, the *discrete topology* has $\tau_d = \mathcal{P}(X)$, the collection of all subsets of X.

(3) If X is any set, the *trivial topology* has $\tau = \{\varnothing, X\}$.

(4) If $(X, \leq)$ is a totally ordered set, the intervals $(a, b) = \{x \in X : a < x < b\}$, $(-\infty, b) = \{x \in X : x < b\}$ and $(a, \infty) = \{x \in X : x > a\}$ are open, and the topology consists of arbitrary unions of such intervals.

(5) If (X, τ) is a topology and $Y \subset X$, the *induced topology* on Y is $\tau|_Y = \{U \cap Y : U \in \tau\}$.

6.1.3. DEFINITION. A set $F \subset X$ is *closed* if F^c is open. If $A \subset X$, the *closure* of A is $\overline{A} = \bigcap\{F : A \subset F,\ F \text{ closed}\}$. A point in $\overline{A}$ is called a *limit point* of A.

If $A \subset X$, then $a \in A$ is an *interior point* of A if there exists $U \in \tau$ with $a \in U \subset A$. If $A \subset X$, the *interior* of A is A^o or $\operatorname{int} A = \bigcup\{U \in \tau : U \subset A\}$.

If $x \in X$, a *neighbourhood* of x is a set N such that $x \in N^o$.

6.1.4. PROPOSITION.

(1) *Finite unions and arbitrary intersections of closed sets are closed.*
(2) $\overline{A}$ *is the smallest closed set containing* A.
(3) $x \in \overline{A}$ *if and only if every* $U \in \tau$ *with* $x \in U$ *has* $A \cap U \neq \varnothing$.
(4) $\overline{A} = A^{coc}$ *is the complement of the interior of* A^c.

PROOF. Since open sets are closed under arbitrary unions and finite intersections, the collection of closed sets is closed under arbitrary intersections and finite unions. Hence the intersection of all closed sets $F \supset A$ is closed, and is thus the smallest closed set containing A. Now $x \in \overline{A}$ if and only if $x \in F$ for every closed $F \supset A$ if and only if $x \notin U$ if U is open and disjoint from A. Finally

$$X \setminus \overline{A} = \bigcup \{U \in \tau : U \cap A = \varnothing\} = \bigcup \{U \in \tau : U \subset A^c\} = A^{co}. \qquad \blacksquare$$

6.1.5. DEFINITION. If σ and τ are two topologies on X, we say that σ is a *weaker topology* than τ, and τ is a *stronger topology* than σ, if $\sigma \subset \tau$.

6.1.6. PROPOSITION. *If* $S \subset \mathcal{P}(X)$, *then there is a weakest topology* τ *containing* S. *It consists of arbitrary unions of sets which are intersections of finitely many elements of* S.

PROOF. Clearly if $\tau \supset S$ is a topology, then it contains all intersections of finitely many elements of S, and arbitrary unions of these sets. The intersection of no sets is X by convention, and $\varnothing$ is the union of no sets, so they both belong to τ. This collection is clearly closed under arbitrary unions. To check that it is stable under intersection, observe that if $A_{\alpha,i}$ and $B_{\beta,j}$ are in S, then

$$\begin{aligned} &\bigcup_{\alpha \in A} A_{\alpha,1} \cap \cdots \cap A_{\alpha,n_\alpha} \;\cap\; \bigcup_{\beta \in B} B_{\beta,1} \cap \cdots \cap B_{\beta,m_\beta} \\ &= \bigcup_{\alpha \in A,\, \beta \in B} A_{\alpha,1} \cap \cdots \cap A_{\alpha,n_\alpha} \cap B_{\beta,1} \cap \cdots \cap B_{\beta,m_\beta}. \end{aligned}$$

Hence this collection is a topology. By construction, this is the weakest topology containing S. $\blacksquare$

6.1.7. DEFINITION. Say that $S \subset \mathcal{P}(X)$ is a *base* for a topology τ if every open set $U \in \tau$ is the union of elements of S. Also S is a *subbase* for a topology τ if the collection of finite intersections of elements of S is a base for τ.

6.1.8. EXAMPLES.

(1) If (X, d) is a metric space, then $\{b_{1/n}(x) : x \in X,\ n \geq 1\}$ is a base for the topology.

(2) $\{(r, s) : r < s \in \mathbb{Q}\}$ is a base for the topology of $\mathbb{R}$.

(3) Let $C[0, 1]$ denote the space of continuous functions on $[0, 1]$. For each $x \in [0, 1]$, $a \in \mathbb{C}$ and $r > 0$, let $U(x, a, r) = \{f \in C[0, 1] : f(x) \in b_r(a)\}$. Let τ be the topology generated by these sets. This is the topology of pointwise convergence. An open neighbourhood of f must contain a set of the form

$$\{g \in C[0, 1] : |g(x_i) - f(x_i)| < r \text{ for } 1 \leq i \leq n\}$$

for $x_1, \ldots, x_n \in [0, 1]$ and $r > 0$.

6.1.9. DEFINITION. A set A is *dense* in X if $X = \overline{A}$. X is *separable* if it has a countable dense subset. X is *first countable* if for each $x \in X$, there is a countable family $\{U_i\} \subset \tau$ with $x \in U_i$ which forms a *countable base of neighbourhoods of* x; i.e., if $x \in V$ is open, then there is some i so that $U_i \subset V$. X is *second countable* if there is a countable family of open sets which is a base for τ.

6.1.10. EXAMPLES.

(1) If (X, d) is a metric space and $x \in X$, then $\{b_{1/n}(x) : n \geq 1\}$ is a countable base of neighbourhoods of x. If X is separable, and $\{x_i : i \geq 1\}$ is dense in X, then $\{b_{1/n}(x_i) : i \geq 1,\ n \geq 1\}$ is a base for τ. Indeed, suppose that $x \in U$ is open. Pick $r > 0$ so that $b_r(x) \subset U$ and x_i so that $d(x, x_i) < 1/n < r/2$. Then $x \in b_{1/n}(x_i) \subset U$. So X is second countable. In particular, compact metric spaces are separable and so second countable.

(2) Consider the discrete topology τ_d on a set X. Since the topology is generated by $\{\{x\} : x \in X\}$, X is always first countable. However it is second countable if and only if X is countable if and only if X is separable.

6.2. Continuity

6.2.1. DEFINITION. A function $f : (X, \tau) \to (Y, \sigma)$ between topological spaces is *continuous* if for all $V \subset Y$ open, the set $f^{-1}(V)$ is open in X. Say that f is a *homeomorphism* if f is a bijection such that both f and f^{-1} are continuous.

6.2.2. EXAMPLES.

(1) The identity map $(X, \text{discrete}) \xrightarrow{\text{id}} (X, \tau) \xrightarrow{\text{id}} (X, \text{trivial})$ is a continuous bijection, however in both cases f^{-1} will be discontinuous provided that τ satisfies $\{\varnothing, X\} \subsetneq \tau \subsetneq \mathcal{P}(X)$.

(2) A function $f : (X, \text{trivial}) \to \mathbb{R}$ is continuous only if it is constant, while every function $f : (X, \text{discrete}) \to \mathbb{R}$ is continuous. On the other hand, a function $f : \mathbb{R} \to (X, \text{discrete})$ is continuous only if it is constant, while every function $f : \mathbb{R} \to (X, \text{trivial})$ is continuous.

(3) $f : (-1, 1) \to \mathbb{R}$ by $f(x) = \tan \frac{\pi x}{2}$ is a homeomorphism.

(4) Let $X = \{0, 1\}$. Let $\tau = \big\{\varnothing, \{0\}, X\big\}$. Then $\{1\}$ is closed, but $\{0\}$ is not, and $\overline{\{0\}} = X$. If $f : X \to \mathbb{R}$ is continuous, then f is constant.

(5) Let $X = [0, 1) \cup \{a, b\}$. Let the open sets in τ be $U \subset [0, 1)$ which are open in the usual metric on $[0, 1)$ together with sets $U \cup (r, 1) \cup \{a\}$, $U \cup (r, 1) \cup \{b\}$ and $U \cup (r, 1) \cup \{a, b\}$ for $r < 1$. Here the points $\{a\}$ and $\{b\}$ are closed because the complement is open. However if $a \in U$ and $b \in V$ are open sets, then $U \cap V \supset (r, 1)$ for some $r < 1$. That means that you cannot separate a and b from one another by open sets. If $f : X \to \mathbb{R}$ is continuous, then $f(a) = f(b)$.

6.2.3. Definition. Let $C^b(X)$ and $C^b_{\mathbb{R}}(X)$ or $C^b(X, \mathbb{R})$ denote the normed vector space of bounded continuous functions from X into $\mathbb{C}$ and $\mathbb{R}$, respectively, with norm $\|f\|_\infty = \sup_X |f(x)|$. Similarly, $C(X)$ and $C_{\mathbb{R}}(X)$ or $C(X, \mathbb{R})$ denote the vector space of continuous functions from X into $\mathbb{C}$ and $\mathbb{R}$, respectively.

6.2.4. Definition. A topological space is *Hausdorff* if for all $x, y \in X$ two distinct points, there are open sets $U \ni x$ and $V \ni y$ so that $U \cap V = \varnothing$.

6.2.5. Proposition. *If $C^b(X)$ separates points of X, i.e., for $x \neq y$ in X, there is a continuous function $f \in C^b(X)$ so that $f(x) \neq f(y)$, then X is Hausdorff.*

Proof. If $f(x) = \alpha$ and $f(y) = \beta$ and $r = |\alpha - \beta|/2 > 0$, then $x \in U = f^{-1}(b_r(\alpha))$ and $y \in V = f^{-1}(b_r(\beta))$ and $U \cap V = \varnothing$. ∎

Consider the Examples 6.2.2 (4) and (5) in light of this proposition.

Recall that $f_n \in C^b(X)$ *converge uniformly* to a function f if $\|f - f_n\|_\infty \to 0$. The following standard result for metric spaces extends easily.

6.2.6. Proposition. *The uniform limit f of a sequence $f_n \in C^b(X)$ is continuous.*

Proof. Let U be open in $\mathbb{C}$ and let $x \in f^{-1}(U)$. Then there is an $r > 0$ so that $b_r(f(x)) \subset U$. Choose n so large that $\|f - f_n\|_\infty < r/3$. Then $x \in V =$

$f_n^{-1}(b_{r/3}(f_n(x)))$ is open. If $y \in V$, then $|f_n(y) - f_n(x)| < r/3$, so

$$\begin{aligned} |f(y) - f(x)| &\le |f(y) - f_n(y)| + |f_n(y) - f_n(x)| + |f_n(x) - f(x)| \\ &< \|f - f_n\|_\infty + \frac{r}{3} + \|f - f_n\|_\infty < r. \end{aligned}$$

Hence $f(y) \in b_r(f(x)) \subset U$. Thus $V \subset f^{-1}(U)$. So f is continuous. ∎

The norm $\|f\|_\infty$ makes $C^b(X)$ into a normed vector space. In view of Proposition 6.2.5, the following is most interesting when X is Hausdorff.

6.2.7. THEOREM. *For any topological space, $C^b(X)$ is complete.*

PROOF. Let $(f_n)_{n\ge1}$ be a Cauchy sequence in $C^b(X)$. If $\varepsilon > 0$, there is an N so that if $N \le m < n$, then $\|f_n - f_m\|_\infty < \varepsilon$. In particular, for $x \in X$, the sequence $\big(f_n(x)\big)_{n\ge1}$ is Cauchy in $\mathbb{C}$. So we may define $f(x) = \lim_{n\to\infty} f_n(x)$ pointwise. However for $m \ge N$,

$$|f(x) - f_m(x)| = \lim_{n\to\infty} |f_n(x) - f_m(x)| \le \varepsilon.$$

Hence $\|f - f_m\|_\infty \le \varepsilon$. So convergence is uniform. By Proposition 6.2.6, f is continuous. Also $\|f\|_\infty = \lim_{n\to\infty} \|f_n\|_\infty < \infty$, and so f lies in $C^b(X)$. Therefore $C^b(X)$ is complete. ∎

6.3. Compactness

6.3.1. DEFINITION. An *open cover* of a set $A \subset X$ is a collection of open sets $\{U_\lambda : \lambda \in \Lambda\}$ such that $A \subset \bigcup_\Lambda U_\lambda$. A set A is *compact* if every open cover has a finite subcover, i.e., a finite subset $U_{\lambda_1}, \dots, U_{\lambda_n}$ such that $A \subset \bigcup_{i=1}^n U_{\lambda_i}$.

6.3.2. EXAMPLE. In 6.2.2 (4), the point $\{0\}$ is compact but not closed.

6.3.3. PROPOSITION. *If X is compact and $A \subset X$ is closed, then A is compact.*

If X is Hausdorff and $A \subset X$ is compact, then A is closed. Moreover, if $x \notin A$, there are disjoint open sets $U \supset A$ and $V \ni x$.

PROOF. If $\mathcal{U} = \{U_\lambda : \lambda \in \Lambda\}$ is an open cover of A, then $\mathcal{U} \cup \{A^c\}$ is an open cover of X. By compactness, it has a finite subcover $U_{\lambda_1}, \dots, U_{\lambda_n}, A^c$. Hence $U_{\lambda_1}, \dots, U_{\lambda_n}$ covers A; whence A is compact.

Suppose that X is Hausdorff and $A \subset X$ is compact, and let $x \in A^c$. For each $a \in A$, there are open sets $a \in U_a$ and $x \in V_a$ so that $U_a \cap V_a = \varnothing$. Clearly

$\{U_a : a \in A\}$ is an open cover of A. By compactness, there is a finite subcover $U_{a_1}, \dots, U_{a_n}$. Let $V = \bigcap_{i=1}^n V_{a_i}$. Then $x \in V$ is open, and

$$V \cap A \subset \bigcup_{i=1}^{n} V \cap U_{a_i} = \varnothing.$$

Hence $x \notin \overline{A}$. So A is closed. Moreover $A \subset U = \bigcup_{i=1}^n U_{a_i}$, and $U \cap V = \varnothing$. ∎

6.3.4. DEFINITION. A family $\{A_\lambda : \lambda \in \Lambda\}$ of subsets of X has the *finite intersection property* (FIP) if whenever $\lambda_1, \dots, \lambda_n$ are finitely many elements of Λ, then $\bigcap_{i=1}^n A_{\lambda_i} \neq \varnothing$.

6.3.5. PROPOSITION. *A topological space X is compact if and only if every family $\mathcal{F} = \{A_\lambda : \lambda \in \Lambda\}$ of closed sets with FIP has non-empty intersection $\bigcap \mathcal{F} := \bigcap_\Lambda A_\lambda \neq \varnothing$.*

PROOF. Suppose that X is compact and $\mathcal{F}$ has FIP. Define open sets $U_\lambda = A_\lambda^c$. If $\bigcap \mathcal{F} = \varnothing$, then $\bigcup_\Lambda U_\lambda = \left(\bigcap \mathcal{F}\right)^c = X$. So $\mathcal{U} = \{U_\lambda : \lambda \in \Lambda\}$ is an open cover of X. By compactness, there is a finite subcover $U_{\lambda_1}, \dots, U_{\lambda_n}$. Hence $\bigcap_{i=1}^n A_{\lambda_i} = \left(\bigcup_{i=1}^n U_{\lambda_i}\right)^c = \varnothing$, contradicting FIP. Therefore $\bigcap \mathcal{F} \neq \varnothing$.

Conversely, suppose that $\mathcal{U} = \{U_\lambda : \lambda \in \Lambda\}$ is an open cover of X. Define closed sets $A_\lambda = U_\lambda^c$. If there is no finite subcover, then $\bigcap_{i=1}^n A_{\lambda_i} = \left(\bigcup_{i=1}^n U_{\lambda_i}\right)^c \neq \varnothing$; and thus $\mathcal{F} = \{A_\lambda : \lambda \in \Lambda\}$ has FIP. But then $\bigcap \mathcal{F} \neq \varnothing$. Therefore $\bigcup_\Lambda U_\lambda = \left(\bigcap \mathcal{F}\right)^c \neq X$, contradicting the fact that $\mathcal{U}$ is an open cover. Hence $\mathcal{F}$ does not have FIP, so there is a finite set $A_{\lambda_1}, \dots, A_{\lambda_n}$ such that $\bigcap_{i=1}^n A_{\lambda_i} = \varnothing$. Hence $\bigcup_{i=1}^n U_{\lambda_i} = \left(\bigcap_{i=1}^n A_{\lambda_i}\right)^c = X$. Therefore X is compact. ∎

6.3.6. PROPOSITION. *If $f : (X, \tau) \to (Y, \sigma)$ is continuous and $A \subset X$ is compact, then $f(A)$ is compact.*

PROOF. Let $\{V_\lambda : \lambda \in \Lambda\}$ is an open cover of $f(A)$ in Y. Define $U_\lambda = f^{-1}(V_\lambda)$. These are open sets by continuity, and they cover A. Thus there is a finite subcover $U_{\lambda_1}, \dots, U_{\lambda_n}$. Then since $V_\lambda \supset f(U_\lambda)$, it follows that $V_{\lambda_1}, \dots, V_{\lambda_n}$ covers $f(A)$. Hence $f(A)$ is compact. ∎

The following important consequence follows directly.

6.3.7. EXTREME VALUE THEOREM. *If (X, τ) is compact and $f \in C(X)$, then $|f|$ attains it maximum. In particular, $\|f\|_\infty < \infty$.*

6.3.8. DEFINITION. If $(X_\lambda, \tau_\lambda)$ are topological spaces for $\lambda \in \Lambda$, we define the *product space* to be $X = \prod_\Lambda X_\lambda = \{(x_\lambda) : x_\lambda \in X_\lambda\}$ with the weakest topology τ which makes the coordinate projections $\pi_\lambda : X \to X_\lambda$ by $\pi_\lambda(x) = x_\lambda$

continuous. That is, the sets $\pi_\lambda^{-1}(U) = \prod_{\mu\in\Lambda\setminus\{\lambda\}} X_\mu \times U$ are open and form a subbase for the topology.

6.3.9. Remarks.
(1) The product topology τ consist of arbitrary unions of finite intersections of the subbase. So if $\lambda_1, \dots, \lambda_n \in \Lambda$ and $U_i \in \tau_{\lambda_i}$, then the sets of the form

$$U_1 \times \cdots \times U_n \times \prod_{\mu\in\Lambda\setminus\{\lambda_i, 1\le i\le n\}} X_\mu$$

form a base for the topology.

(2) If Λ is finite, this is a familiar construction in the metric space case. Indeed, if (X_i, d_i) are metric spaces for $1 \le i \le n$, then

$$D\big((x_1, \dots, x_n), (y_1, \dots, y_n)\big) = \max\big\{d_i(x_i, y_i) : 1 \le i \le n\big\}$$

is a metric on the product, and the metric topology coincides with the product topology.

(3) When Λ is infinite, it often requires the Axiom of Choice to be able to say that X is non-empty.

6.3.10. Theorem. *If X_i are compact for $1 \le i \le n$, then $X = \prod_{i=1}^n X_i$ is compact.*

Proof. It suffices to show that $X\times Y$ is compact if both X and Y are compact, as the result follows by finitely many repetitions. Let $\mathcal{W} = \{W_\lambda : \lambda \in \Lambda\}$ be an open cover of $X \times Y$. For each $(x, y) \in X \times Y$, there is a $\lambda(x, y)$ so that $(x, y) \in W_{\lambda(x,y)}$. Hence this set contains a basic open set $W_{\lambda(x,y)} \supset U_{x,y} \times V_{x,y} \ni (x, y)$ for open sets $U_{x,y}$ in X and $V_{x,y}$ in Y. Fix $y \in Y$. The sets $\{U_{x,y} : x \in X\}$ is an open cover of X. Select a finite subcover $U_{x_1^y,y}, \dots, U_{x_{n_y}^y,y}$. Then $U_{x_i^y,y} \times V_{x_i^y,y}$ for $1 \le i \le n_y$ covers $X \times \{y\}$. Let $V_y = \bigcap_{i=1}^{n_y} V_{x_i^y,y}$. This is open, contains y, and $U_{x_i^y,y} \times V_{x_i^y,y}$ for $1 \le i \le n_y$ covers $X \times V_y$. Now $\{V_y : y \in Y\}$ is an open cover of Y. Let $V_{y_1}, \dots, V_{y_m}$ be a finite subcover. Then the finite collection $\{U_{x_i^{y_j}, y_j} \times V_{x_i^{y_j}, y_j} : 1 \le i \le n_{y_j},\ 1 \le j \le m\}$ covers $X \times Y$. Therefore $\{W_{\lambda(x_i^{y_j}, y_j)} : 1 \le i \le n_{y_j},\ 1 \le j \le m\}$ covers $X \times Y$. ∎

6.3.11. Remark. An infinite product of compact spaces is also compact. This is known as Tychonoff's Theorem. It turns out to be equivalent to the Axiom of Choice. We usually prove it in the functional analysis course.

6.4. Separation Properties

There is a whole hierarchy of *separation properties* to classify how nice a topological space is.

6.4.1. DEFINITION. A topological space is T_0 if $x \neq y \in X$, then there is an open set containing one of these points, but not the other.
A topological space is T_1 if points are closed.
A topological space is T_2 if it is Hausdorff.
A topological space is T_3 if it is T_1 and *regular*: given a closed set A and a point $x \notin A$, there are disjoint open sets $U \supset A$ and $V \ni x$.
A topological space is $T_{3.5}$ or *Tychonoff* if it is T_1 and *completely regular*: given a closed set A and a point $x \notin A$, there is a continuous function $f : X \to [0,1]$ so that $f(x) = 1$ and $f|A = 0$.
A topological space is T_4 if it is T_1 and *normal*: given disjoint closed sets A, B, there are disjoint open sets $U \supset A$ and $V \supset B$.

6.4.2. REMARKS.
(1) The T_0 property is very weak. Example 6.2.2(4) is T_0 but not T_1.

(2) T_1 is also a very weak property. It implies T_0 since if $x \neq y$, the set $\{x\}^c$ is an open set containing y but not x. To be T_1 it is enough to find open sets $U \ni x$ with $y \notin U$ and $V \ni y$ with $x \notin V$. (Exercise.) Example 6.2.2(5) is T_1 but not Hausdorff. Points are closed in Hausdorff spaces, so T_2 implies T_1.

(3) The trivial topology is regular because there are no points which are disjoint from a non-empty closed set. But throwing in the T_1 condition makes a T_3 space Hausdorff because you can take your closed set to be $\{y\}$.

(4) Completely regular spaces are regular, because if $x \notin A$ and A is closed, let $f : X \to [0,1]$ be a continuous function with $f(x) = 1$ and $f|_A = 0$. Then $U = \{y : f(y) > 1/2\}$ and $V = \{y : f(x) < 1/2\}$ are disjoint open sets with $x \in U$ and $A \subset V$. Again we need to add the T_1 property to exclude examples like the trivial topology.

(5) The T_1 property ensures that T_4 spaces are Hausdorff. A metric space (X, d) is normal. If A and B are disjoint closed sets, let $U = \{x : d(x, A) < d(x, B)\}$ and $V = \{x : d(x, A) > d(x, B)\}$. It is easy to check that these are disjoint open sets.

6.4.3. PROPOSITION. *Compact Hausdorff spaces are normal.*

PROOF. Let A, B be disjoint closed subsets of a compact Hausdorff space X. Then A and B are compact by Proposition 6.3.3. Moreover since X is Hausdorff, that same Proposition shows that for each point $x \in B$, there are disjoint open sets

$U_x \supset A$ and $V_x \ni x$. The collection $\{V_x : x \in B\}$ is an open cover of B. Let $V_{x_1}, \dots, V_{x_n}$ be a finite subcover. Set $V = \bigcup_{i=1}^n V_{x_i}$ and $U = \bigcap_{i=1}^n U_{x_i}$. These are disjoint open sets with $A \subset U$ and $B \subset V$. ■

Now we prove that T_4 spaces have lots of continuous functions.

6.4.4. URYSOHN'S LEMMA. *Let X be a normal topological space, and let A and B be disjoint closed sets. Then there is a continuous function $f : X \to [0,1]$ such that $f|_A = 0$ and $f|_B = 1$.*

PROOF. Normality implies the following property: if A is closed and W is open and $A \subset W$, then there is an open set U such that $A \subset U \subset \overline{U} \subset W$. To see this, take $B = W^c$. Use normality to find disjoint open sets $U \supset A$ and $V \supset B$. Then $\overline{U} \subset V^c \subset W$.

Start with $U_1 = B^c$. Find an open $U_{1/2}$ so that $A \subset U_{1/2} \subset \overline{U_{1/2}} \subset U_1$. Repeating this procedure recursively, we find open sets $U_{k/2^n}$ for $1 \le k \le 2^n$ and $n \ge 1$ so that

$$A \subset U_{k/2^n} \subset \overline{U_{k/2^n}} \subset U_{(k+1)/2^n} \quad \text{for} \quad 1 \le k < 2^n.$$

Let $D = \{k/2^n : 1 \le k \le 2^n,\ n \ge 1\}$. Define $f(x) = \inf\{r \in D : x \in U_r\}$ if $x \in U_1$ and $f|_B = 1$. Clearly $0 \le f \le 1$ and $f|_A = 0$.

<u>Claim:</u> f is continuous. Note that

$$f^{-1}\big([0,t)\big) = \bigcup_{r<t,\, r\in D} U_r \quad \text{is open for } t \in [0,1].$$

Also for $0 \le t < 1$, since $t < r < s$ for $r, s \in D$ implies that $\overline{U_r} \subset U_s$,

$$f^{-1}\big([0,t]\big) = \bigcap_{r>t,\, r\in D} f^{-1}\big([0,r)\big) = \bigcap_{r>t,\, r\in D} U_r = \bigcap_{r>t,\, r\in D} \overline{U_r}.$$

This is closed, and therefore $f^{-1}\big((t,1]\big) = \big(\bigcap_{r>t,\, r\in D} \overline{U_r}\big)^c$ is open. Hence, $f^{-1}\big((s,t)\big)$ is open for $s < t$, and so f is continuous. ■

6.4.5. REMARK. If (X,d) is a metric space and A and B are disjoint closed sets, define

$$f(x) = \frac{d(x,A)}{d(x,A) + d(x,B)}.$$

This satisfies the conclusion of Urysohn's Lemma.

6.4.6. COROLLARY. *If X is a T_4 space, then it is Tychonoff ($T_{3.5}$).*

6.4.7. COROLLARY. *If X is a compact Hausdorff space, $C(X)$ separates points.*

Urysohn's Lemma implies the following significant strengthening.

6.4.8. TIETZE'S EXTENSION THEOREM. *Let X be a normal topological space, and let $A \subset X$ be a closed set. If $f : A \to [a,b]$ is continuous, there is a continuous function $F : X \to [a,b]$ such that $F|_A = f$.*

PROOF. After scaling, we may assume that the range is $[-1,1]$. Let $A_1 = f^{-1}([-1,-\frac{1}{3}])$ and $B_1 = f^{-1}([\frac{1}{3},1])$. By Urysohn's Lemma, there is a function $g_1 : X \to [-\frac{1}{3},\frac{1}{3}]$ so that $g_1|_{A_1} = -\frac{1}{3}$ and $g_1|_{B_1} = \frac{1}{3}$. Then $f_1 = f - g_1|_A$ has range in $[-\frac{2}{3},\frac{2}{3}]$. Repeat the process, setting $A_2 = f_1^{-1}([-\frac{2}{3},-\frac{2}{9}])$ and $B_2 = f_1^{-1}([\frac{2}{9},\frac{2}{3}])$, and finding $g_2 : X \to [-\frac{2}{9},\frac{2}{9}]$ with $g_2|_{A_2} = -\frac{2}{9}$ and $g_2|_{B_2} = \frac{2}{9}$.

Then $f_2 = f_1 - g_2|_A$ has range in $[-\left(\frac{2}{3}\right)^2, \left(\frac{2}{3}\right)^2]$. Recursively we obtain functions $g_n : X \to [-2\cdot 3^{-n}, 2\cdot 3^{-n}]$ so that $f_n = f - \sum_{i=1}^n g_n|_A$ has range in $[-\left(\frac{2}{3}\right)^n, \left(\frac{2}{3}\right)^n]$. Let $g = \sum_{n\ge 1} g_n$. Then $g|_A = f$ and

$$\|g\|_\infty \le \sum_{n\ge 1} \|g_n\|_\infty = \sum_{n\ge 1} 2\cdot 3^{-n} = 1. \qquad \blacksquare$$

6.4.9. DEFINITION. A Hausdorff space X is *locally compact* (LCH) if every point has a compact neighbourhood; i.e. for $x \in X$, there is a compact set K with $x \in \operatorname{int} K$.

6.4.10. PROPOSITION. *Locally compact Hausdorff spaces are regular.*

PROOF. Let A be a closed set in a LCH space X, and $x \notin A$. Let K be a compact neighbourhood of x. Then $A \cap K$ is compact. So by Proposition 6.3.3, there are disjoint open sets $U \ni x$ and $V \supset A \cap K$. Then $W = U \cap \operatorname{int} K \subset K \setminus V$ is an open neighbourhood of x and $\overline{W} \subset K \setminus V \subset A^c$. Hence $\overline{W}^c \supset A$ is open and disjoint from W. Therefore X is regular. $\blacksquare$

6.4.11. DEFINITION. If $f : X \to \mathbb{C}$, the *support* of f is

$$\operatorname{supp}(f) = \overline{\{x : f(x) \ne 0\}}.$$

If X is LCH, $C_c(X)$ denotes the space of continuous $\mathbb{C}$-valued functions with compact support. Let $C_0(X)$ be the closure of $C_c(X)$ in $(C^b(X), \|\cdot\|_\infty)$.

There is a weaker version of Urysohn's Lemma valid for LCH spaces.

6.4.12. PROPOSITION. *Locally compact Hausdorff spaces are completely regular. Moreover, if X is LCH, $A \subset X$ is closed, $B \subset X$ is compact and $A \cap B = \varnothing$, then there is $f \in C_c(X)$ with $0 \le f \le 1$, $f|_A = 0$ and $f|_B = 1$.*

PROOF. Arguing as in the previous proof, for each $x \in B$, there is a compact neighbourhood $\overline{W_x} \ni x$ which is disjoint from A. We can take $W_x = \operatorname{int} \overline{W_x}$. Then $\{W_x : x \in B\}$ is an open cover. Let $W_{x_1}, \dots, W_{x_n}$ be a finite subcover. Then $W = \bigcup_{i=1}^n W_{x_i}$ is open containing B and $\overline{W} = \bigcup_{i=1}^n \overline{W_{x_i}}$ is compact and disjoint from A.

Let $C = \overline{W} \setminus W$. Then B and C are disjoint closed sets in the compact Hausdorff space $\overline{W}$. By Urysohn's Lemma, there is a continuous function $f : \overline{W} \to [0,1]$ with $f|_B = 1$ and $f|_C = 0$. Extend f to a function on X by setting $f|_{\overline{W}^c} = 0$. Now $A \subset W^c$, and thus $f|_A = 0$. To see that f is continuous, note that if $a \geq 0$, $f^{-1}(a,b)$ is open in W, and hence open in X. If $a < 0 < b$, then $D = f^{-1}([b,\infty)$ is closed in $\overline{W}$, and thus $f^{-1}(a,b) = D^c$ is open in X. Since $\operatorname{supp}(f) \subset \overline{W}$, we have $f \in C_c(X)$. ■

There is also a weaker version of Tzietze's Theorem valid for LCH spaces.

6.4.13. COROLLARY. *Suppose that X is LCH, $A \subset X$ is closed, $B \subset X$ is compact, $A \cap B = \varnothing$ and $g \in C(B)$. Then there is an $f \in C_c(X)$ with $f|_A = 0$ and $f|_B = g$, and $\|f\|_\infty = \|g\|_\infty$.*

PROOF. The setup is the same as the previous proof. Define $f|B = g$ and $f|_C = 0$, and extend this to a continuous function on $\overline{W}$ using Tietze's Theorem. Then extend f to all of X by setting $f|_{\overline{W}^c} = 0$. This is continuous with compact support as in the previous proof. ■

6.4.14. DEFINITION. If U is open, say $f \prec U$ for a continuous function f if $0 \leq f \leq 1$ and $\operatorname{supp} f \subset U$. Suppose that $\mathcal{U} = \{U_\lambda : \lambda \in \Lambda\}$ is an open cover of $K \subset X$. Then a *partition of unity* for K *relative to* $\mathcal{U}$ is a collection of functions $g_\lambda \prec U_\lambda$ which is *locally finite*, i.e., for each $x \in X$, $\{\lambda : g_\lambda(x) \neq 0\}$ is finite, and $\sum_\Lambda g_\lambda(x) = 1$ for all $x \in K$.

6.4.15. PROPOSITION. *Let X be LCH, and let $K \subset X$ be compact. Suppose that $\mathcal{U} = \{U_1, \dots, U_n\}$ is an open cover of K. Then there are $g_i \prec U_i$ in $C_c(X)$ which is a partition of unity for K relative to $\mathcal{U}$.*

PROOF. For each $x \in K$, there is some $U_i \ni x$. Pick a compact neighbourhood with $x \in C_x \subset U_i$. Now $\{\operatorname{int} C_x : x \in K\}$ is an open cover with finite subcover $C_{x_1}, \dots, C_{x_m}$. Let $K_i = \bigcup\{C_{x_j} : C_{x_j} \subset U_i\}$. Then $K_i \subset U_i$ and $K \subset \bigcup_{i=1}^n K_i =: C$ is compact . Let W be an open neighbourhood of C with $\overline{W}$ compact. By Urysohn's lemma, there are functions h_i so that $\chi_{K_i} \leq h_i \prec U_i \cap W$. Let $g_i = h_i / \sum_{i=1}^n h_i$. ■

6.5. Nets*

Sequences are not sufficient for dealing with convergence in general topological spaces, including many that arise in normal contexts. The replacement is the notion of a net, which you can think of as a very wide and very long generalized sequence. This optional section is not required in the next chapter.

We will need to call on the Axiom of Choice frequently. Recall that this is the assumption that whenever $\{A_\lambda : \lambda \in \Lambda\}$ is a collection of non-empty sets, there is a selection (choice function) $\varphi : \Lambda \to \bigcup_\Lambda A_\lambda$ so that $\varphi(\lambda) \in A_\lambda$ for all $\lambda \in \Lambda$.

6.5.1. Definition. A *partial order* on a set Λ is a relation $\leq$ satisfying

(1) $\lambda \leq \lambda$ for $\lambda \in \Lambda$ (reflexive)

(2) $\lambda \leq \mu$ and $\mu \leq \lambda$ implies that $\lambda = \mu$ (antisymmetric)

(3) $\lambda \leq \mu$ and $\mu \leq \nu$ implies that $\lambda \leq \nu$ (transitive).

Then $(\Lambda, \leq)$ is called a *poset*. A poset is *upward directed* if for $\lambda_1, \lambda_2 \in \Lambda$, there is a $\mu \in \Lambda$ so that $\lambda_1 \leq \mu$ and $\lambda_2 \leq \mu$.

6.5.2. Definition. A *net* in X is an upward directed poset Λ with a function $j : \Lambda \to X$, say $x_\lambda = j(\lambda)$. We usually write the net as $(x_\lambda)_\Lambda$. A net $(x_\lambda)_\Lambda$ *converges to* x in (X, τ) if for every open set $U \ni x$, there is $\lambda_0 \in \Lambda$ so that $x_\lambda \in U$ for every $\lambda \geq \lambda_0$. We write $\lim_\Lambda x_\lambda = x$.

A *subnet* $(y_\gamma)_\Gamma$ of $(x_\lambda)_\Lambda$ is given by a cofinal function $\varphi : \Gamma \to \Lambda$ so that $y_\gamma = x_{\varphi(\gamma)}$, where we say that φ is *cofinal* if for all $\lambda \in \Lambda$, there is a $\gamma_0 \in \Gamma$ so that $\varphi(\gamma) \geq \lambda$ for all $\gamma \geq \gamma_0$. It is convenient if φ is *monotone*, meaning that $\gamma_1 \leq \gamma_2$ implies that $\varphi(\gamma_1) \leq \varphi(\gamma_2)$. But this is not necessary.

We present a detailed example to explain why nets are needed, and how to use them.

6.5.3. Example. Let $X = \mathbb{N}_0 \times \mathbb{N}_0$. Declare that $U \subset X$ is open if $(0,0) \notin U$; and that a set $U \ni (0,0)$ is open if $\{m : \pi_1^{-1}(m) \cap U \text{ is cofinite in } \mathbb{N}_0\}$ is cofinite in $\mathbb{N}_0$. It is easy to verify that this defines a topology.

(a) X is Hausdorff because $\{(m,n)\}$ is open if $m+n \geq 1$ and $\{(m,n)\}^c$ is an open neighbourhood of $(0,0)$.

(b) $(0,0) \in \overline{X \setminus \{(0,0)\}}$ because every open set $U \ni (0,0)$ intersects $X \setminus \{(0,0)\}$.

(c) However no sequence $x_k = (m_k, n_k)$ in $X \setminus \{(0,0)\}$ converges to $(0,0)$. There are two cases. If $\{m_k : k \geq 1\}$ is bounded, pick m_0 so that $m_k = m_0$ infinitely often. The set $U = \{(m,n) : m \neq m_0\} \cup \{(0,0)\}$ is an open neighbourhood of $(0,0)$, and the sequence is not eventually in U. Otherwise, there is a sequence

$k_i \to \infty$ so that $m_{k_i} < m_{k_{i+1}}$ for $i \geq 1$. Then $U = X \setminus \{x_{k_i} : i \geq 1\}$ is an open neighbourhood of $(0,0)$, and the sequence is not eventually in U.

(d) There is a net in $X \setminus \{(0,0)\}$ converging to $(0,0)$. Let $\Lambda = \{U \in \tau : (0,0) \in U\}$ where $U \leq V$ if $U \supset V$. (We say that Λ is ordered by containment.) This is directed because $U, V \leq U \cap V$. Order $X \setminus \{(0,0)\}$ by

$$(0,1), (1,0), (0,2), (1,1), (2,0), (0,3), (1,2), (2,1), (3,0), \ldots.$$

Define x_U to be the least element in this list which belongs to U. (This avoids any issues with the Axiom of Choice.) Then (x_U) converges to $(0,0)$ because given an open neighbourhood $U \ni (0,0)$, we have $x_V \in V \subset U$ whenever $U \leq V$.

(e) The sequence $(0,1), (1,0), (0,2), (1,1), (2,0), (0,3), (1,2), (2,1), \ldots$ has a subnet converging to $(0,0)$. Let Λ be the net just constructed. Define $\varphi(U) = x_U$ considered as an element in this sequence. To see that this map is cofinal, let (m_0, n_0) be in this sequence, and set $N_0 = m_0 + n_0$. Let

$$U_0 = X \setminus \{(m,n) : 1 \leq m+n \leq N_0\}.$$

Then if $U_0 \leq U$, it follows that $x_U = (m,n)$ with $m+n > N_0$ and thus $\varphi(U)$ follows (m_0, n_0) in the sequence. Therefore this is a subnet of the sequence which converges to $(0,0)$.

Now we show that nets replace sequences in some familiar results.

6.5.4. PROPOSITION. *Let $A \subset X$. Then $x \in \overline{A}$ if and only if there is a net $(a_\lambda)_\Lambda$ in A such that $\lim_\Lambda a_\lambda = x$.*

PROOF. Suppose that $x \in \overline{A}$. By Proposition 6.1.4(3), every open neighbourhood U of x intersects A. Let $\mathcal{O}(x)$ be the open neighbourhoods of x ordered by containment. By the Axiom of Choice, we can pick a point $a_U \in A \cap U$ for each $U \in \mathcal{O}(x)$. The net $(a_U)_{\mathcal{O}(x)}$ converges to x by construction.

Conversely, suppose that $(a_\lambda)_\Lambda$ is a net in A such that $\lim_\Lambda a_\lambda = x$. Then for any neighbourhood U of x, there is a λ so that $a_\lambda \in U$. In particular, $A \cap U \neq \varnothing$. Hence by Proposition 6.1.4(3), $x \in \overline{A}$. ∎

6.5.5. THEOREM. *Let $f : (X,\tau) \to (Y,\sigma)$. Then f is continuous if and only if whenever $(x_\lambda)_\Lambda$ is a net in X converging to x, it follows that $f(x) = \lim_\Lambda f(x_\lambda)$.*

PROOF. Suppose that f is continuous, and let $(x_\lambda)_\Lambda$ be a net in X converging to x. Let V be an open neighbourhood of $f(x)$. Then $U = f^{-1}(V)$ is an open neighbourhood of x. By convergence, there is a $\lambda_0 \in \Lambda$ so that $x_\lambda \in U$ for all $\lambda \geq \lambda_0$. Hence $f(x_\lambda) \in f(U) \subset V$ for all $\lambda \geq \lambda_0$. That means that $f(x) = \lim_\Lambda f(x_\lambda)$.

Conversely, suppose f is not continuous. Thus there is an open set $V \subset Y$ such that $U = f^{-1}(V)$ is not open. Then $\overline{U^c} \cap U$ contains a point x. By Proposition 6.5.4, there is a net $(x_\lambda)_\Lambda$ in U^c with limit x. Therefore $f(x_\lambda) \in f(U^c) \subset V^c$. Since V^c is closed, any limit point of this net must remain in V^c by Proposition 6.5.4. Therefore it cannot converge to $f(x)$ which lies in V. So $f(x) \neq \lim_\Lambda f(x_\lambda)$. ■

6.5.6. THEOREM. *A topological space X is compact if and only if every net in X has a convergent subnet.*

PROOF. Suppose that every net in X has a convergent subnet. Consider a collection $\mathcal{F} = \{C_\alpha : \alpha \in A\}$ of closed sets with the FIP. Let

$$\Lambda = \{F \subset A : F \text{ is finite, non-empty}\}$$

ordered by inclusion, i.e., $F \le G$ if $F \subset G$. This is an upward directed poset: $F_1, F_2 \le F_1 \cup F_2$. For each $F \in \Lambda$, use the Axiom of Choice to select a point $x_F \in \bigcap_{\alpha \in F} C_\alpha$. This is possible since the finite intersection is non-empty. Then $(x_F)_\Lambda$ is a net in X. Let $(y_\gamma)_\Gamma$ be a subnet with limit x; where $\varphi : \Gamma \to \Lambda$ and $y_\gamma = x_{\varphi(\gamma)}$. For any $\alpha \in A$, there is a $\gamma_\alpha \in \Gamma$ so that $\gamma \ge \gamma_\alpha$ implies that $\varphi(\gamma) \ge \{\alpha\}$. Hence $y_\gamma \in C_\alpha$ for all $\gamma \ge \gamma_\alpha$. Since C_α is closed, the limit point $x \in C_\alpha$. This holds for all $\alpha \in A$. Therefore $x \in \bigcap \mathcal{F}$. By Proposition 6.3.5, it follows that X is compact.

Conversely suppose that X is compact. Let $(x_\lambda)_\Lambda$ be a net in X. For each $\lambda \in \Lambda$, define $C_\lambda = \overline{\{x_\mu : \mu \ge \lambda\}}$. Then $\mathcal{F} = \{C_\lambda : \lambda \in \Lambda\}$ is a collection of non-empty closed sets. It has FIP because if $\lambda_1, \dots, \lambda_n \in \Lambda$, the upward directed property ensures that there is some $\lambda_0 \in \Lambda$ so that $\lambda_i \le \lambda_0$ for $1 \le i \le n$. Hence $\bigcap_{i=1}^n C_{\lambda_i} \supset C_{\lambda_0} \neq \varnothing$. By Proposition 6.3.5, there is a point $x \in \bigcap_\Lambda C_\lambda$.

Now we build a subnet with limit x. Let $\mathcal{O}(x)$ be the set of all open neighbourhoods of x. Let $\Gamma = \Lambda \times \mathcal{O}(x)$ with order $(\lambda, U) \le (\mu, V)$ if $\lambda \le \mu$ and $U \supset V$. Let $S_{\lambda,U} = \{\mu \in \Lambda : \mu \ge \lambda \text{ and } x_\mu \in U\}$. This set is non-empty because $x \in C_\lambda \cap U = \overline{\{x_\mu : \mu \ge \lambda\}} \cap U$; and thus by Proposition 6.5.4, $x_\mu \in U$ for some $\mu \ge \lambda$. Use the Axiom of Choice to select $\mu = \varphi(\lambda, U) \in S_{\lambda,U}$ for each $(\lambda, U) \in \Gamma$. The map $\varphi : \Gamma \to \Lambda$ is cofinal because if $\lambda_0 \in \Lambda$, then every $(\lambda, U) \ge (\lambda_0, X)$ will have $\varphi(\lambda, U) = \mu \ge \lambda \ge \lambda_0$. So $y_{\lambda,U} = x_{\varphi(\lambda,U)}$ defines a subnet $(y_{\lambda,U})_\Gamma$ of $(x_\lambda)_\Lambda$.

Finally we claim that $\lim_\Gamma y_{\lambda,U)} = x$. Indeed, let $U \in \mathcal{O}(x)$ be any open neighbourhood of x. Fix some $\lambda_0 \in \Lambda$. Whenever $(\lambda, V) \ge (\lambda_0, U)$, we have $y_{\lambda,V} \in V \subset U$. Thus this net converges to x. ■

CHAPTER 7

Functionals on $C_c(X)$ and $C_0(X)$

7.1. Radon measures

Let X be a locally compact Hausdorff space. Let $C_c(X)$ denote the space of continuous functions with compact support (i.e. $\overline{\{x : f(x) \neq 0\}}$ is compact). Let $C_0(X)$ be the space of all continuous functions *vanishing at infinity*, meaning that $\{x : |f(x)| \geq \varepsilon\}$ is compact for all $\varepsilon > 0$. Both spaces are endowed with the sup norm $\|f\|_\infty = \sup |f(x)|$. It is easy to see that $C_0(X) = \overline{C_c(X)}^{\|\cdot\|_\infty}$. When X is compact, $C_c(X) = C_0(X) = C(X)$.

7.1.1. DEFINITION. A Borel measure μ is *outer regular* on $E \in \mathrm{Bor}(X)$ if

$$\mu(E) = \inf\{\mu(U) : E \subset U \text{ open}\}$$

and *inner regular* on E if

$$\mu(E) = \sup\{\mu(K) : E \supset K \text{ compact}\}.$$

We call μ a *regular measure* if it is both inner and outer regular on all Borel sets. A Borel measure μ is a *Radon measure* if $\mu(K) < \infty$ for all compact sets K, it is outer regular on all Borel sets and inner regular on open sets.

7.1.2. REMARKS.
(1) Some books define a Radon measure to be a regular measure which takes finite values on all compact sets. This is somewhat less general (so we would require an additional hypothesis such as σ-compactness), and does not simplify the proof of the main result.

(2) If $\mu(E) < \infty$, then μ is regular on E if and only if for $\varepsilon > 0$, there is an open set U and a compact set K so that $K \subset E \subset U$ and $\mu(U \setminus K) < \varepsilon$.

(3) There are compact Hausdorff spaces which support finite Borel measures which are not regular. See the exercises in [2].

7.1.3. PROPOSITION. *If X is LCH, μ is a Radon measure on X and $E \in$* $\mathrm{Bor}(X)$ *is σ-finite, then μ is regular on E.*

PROOF. Let $E \in \mathrm{Bor}(X)$ such that $\mu(E) < \infty$, and let $\varepsilon > 0$. By outer regularity, there is an open $U \supset E$ with $\mu(U \setminus E) < \varepsilon/2$. Then there is an open set $V \supset U \setminus E$ with $\mu(V) < \varepsilon/2$. Since μ is inner regular on U, there is a compact $L \subset U$ with $\mu(U \setminus L) < \varepsilon/2$. Then $K = L \setminus V$ is compact and $K \subset U \setminus V \subset E$. Finally,

$$\mu(E \setminus K) < \mu(U \setminus K) \leq \mu(U \setminus L) + \mu(V) < \varepsilon.$$

Thus μ is regular on E.

If E is a σ-finite set, write $E = \dot{\bigcup}_{i\geq 1} E_i$ where $\mu(E_i) < \infty$. Find compact sets $K_n \subset E_n$ with $\mu(E_n \setminus K_n) < 2^{-n}\varepsilon$. Then $C_p = \bigcup_{n=1}^{p} K_n$ are compact. Set $K = \bigcup_{n\geq 1} K_n$. Then $K \subset E$ and $\mu(E \setminus K) \leq \sum_{n\geq 1} \mu(E_n \setminus K_n) < \varepsilon$. By continuity from below, we get $\lim_{p\to\infty} \mu(C_p) = \mu(K) > \mu(E) - \varepsilon$. Therefore μ is regular on E. ■

The following immediate corollary covers many cases of interest.

7.1.4. COROLLARY. *Every σ-finite Radon measure is regular.*

In particular, if X is σ-compact (i.e. a countable union of compact sets) and μ is a Radon measure on X, then μ is regular.

7.1.5. COROLLARY. *If X is a separable, locally compact metric space and μ is a Radon measure on X, then μ is regular.*

PROOF. We show that X is σ-compact. A separable metric space is second countable (see Example 6.1.10(1)). Indeed, if $\{x_i : i \geq 1\}$ is a dense subset, then

$$\mathcal{T} = \{b_r(x_i) : i \geq 1,\ r \in \mathbb{Q}^+\}$$

is a countable base for the topology. For each x_i, local compactness implies that there is some $r_i > 0$ so that $\overline{b_{r_i}(x_i)}$ is compact. Then

$$\mathcal{T}_0 = \{b_r(x_i) : i \geq 1,\ r \in \mathbb{Q}^+,\ r \leq r_i\}$$

is also a base for the topology. Hence

$$X = \bigcup_{\mathcal{T}_0} b_r(x_i) = \bigcup_{\mathcal{T}_0} \overline{b_r(x_i)}$$

is the union of countably many compact sets. The result now follows from Corollary 7.1.4. ■

7.1.6. PROPOSITION. *If X is a separable, locally compact metric space and μ is a finite Borel measure on X, then μ is regular and hence Radon.*

PROOF. Let $\mathcal{S}$ be the collection of all Borel sets on which μ is regular. The proof of Corollary 7.1.5 shows that X is σ-compact. Write $X = \bigcup_{n\geq 1} K_n$ where K_n are compact. If C is closed, then the sets $C_n = C \cap \bigcup_{i=1}^{n} K_i$ are compact, and

$C_n \subset C_{n+1} \subset \bigcup_{n\geq 1} C_n = C$. By continuity from below, $\mu(C) = \lim_{n\to\infty} \mu(C_n)$. So μ is inner regular on closed sets.

As X is a metric space, a closed set C is a G_δ (let $U_n = \{x : d(x, C) < \frac{1}{n}\}$). Since μ is finite, continuity from above shows that if $C = \bigcap_{n\geq 1} U_n$ and $U_n \supset U_{n+1}$, then $\mu(C) = \lim_{n\to\infty} \mu(U_n)$. Thus μ is regular on C.

Next we claim $\mathcal{S}$ is closed under complements. Let $A \in \mathcal{S}$. By Remark 7.1.2(2), for $\varepsilon > 0$, there is a compact set K and open set U with $K \subset A \subset U$ and $\mu(U \setminus K) < \varepsilon$. Therefore $U^c \subset A^c \subset K^c$, and $\mu(K^c \setminus U^c) = \mu(U \setminus K) < \varepsilon$. The set U^c is closed but not necessarily compact. Since $U^c \in \mathcal{S}$, there is a compact set $L \subset U^c$ so that $\mu(U^c \setminus L) < \varepsilon - \mu(K^c \setminus U^c)$. Thus $\mu(K^c \setminus L) < \varepsilon$. Therefore μ is regular on A^c.

Now we show that $\mathcal{S}$ is closed under countable unions. Let $A_i \in \mathcal{S}$ for $i \geq 1$, and let $A = \bigcup_{i\geq 1} A_i$. Given $\varepsilon > 0$, find compact sets K_i and open sets U_i so that $K_i \subset A_i \subset U_i$ and $\mu(U_i \setminus K_i) < 2^{-i}\varepsilon$. Set

$$C_n = \bigcup_{i=1}^{n} K_i, \quad C = \bigcup_{i\geq 1} K_i \quad \text{and} \quad U = \bigcup_{i\geq 1} U_i.$$

Then $C_n \subset A \subset U$, each C_n is compact, and U is open. Therefore

$$\lim_{n\to\infty} \mu(U \setminus C_n) = \mu(U \setminus C) \leq \sum_{i\geq 1} \mu(U_i \setminus K_i) < \sum_{i\geq 1} 2^{-i}\varepsilon = \varepsilon.$$

The first step uses continuity from above, which requires the finiteness of μ.

Hence $\mathcal{S}$ is a σ-algebra. As $\mathcal{S}$ contains all closed sets, it contains all Borel sets. So μ is regular and finite, whence Radon. ■

7.2. Positive functionals on $C_c(X)$

7.2.1. DEFINITION. A *positive linear functional* φ on $C_c(X)$ or $C_0(X)$ is a linear functional such that $\varphi(f) \geq 0$ if $f \geq 0$.

7.2.2. PROPOSITION. *Let X be a locally compact Hausdorff space; and let μ be a Borel measure on X which is finite on compact sets.*

(1) $\Phi_\mu(f) = \int f\, d\mu$ *determines a positive linear functional on $C_c(X)$.*

(2) Φ_μ *is continuous with respect to the sup norm (i.e. $|\Phi_\mu(f)| \leq C\|f\|_\infty$ for $f \in C_c(X)$) if and only if $M = \sup\{\mu(K) : K \text{ is compact}\} < \infty$. Moreover $\|\Phi_\mu\| = M$.*

(3) *In particular, if μ is Radon, then Φ_μ is continuous if and only if $\|\mu\| := \mu(X) < \infty$.*

PROOF. If $f \in C_c(X)$ with $\|f\|_\infty \le 1$ and $\operatorname{supp}(f) \subset K$, where K is compact, then

$$|\Phi_\mu(f)| = \Big|\int f\,d\mu\Big| \le \int_K |f|\,d\mu \le \mu(K).$$

So this is a linear functional which is norm continuous on the subspace of functions supported on K. Clearly it takes positive functions to positive values.

By Proposition 6.4.12, given a compact set K, there is a function $h \in C_c(X)$ so that $\chi_K \le h \le 1$. Hence $\|\Phi_\mu\| \ge \Phi_\mu(h) \ge \mu(K)$. Taking the supremum over all compact K shows that $\|\Phi_\mu\| \ge M = \sup\{\mu(K) : K \text{ is compact}\}$. Therefore if Φ_μ is continuous with respect to the sup norm, then $M < \infty$. Conversely, if M is finite and $f \in C_c(X)$ with $\|f\|_\infty \le 1$, then $K = \operatorname{supp} f$ is compact. Thus

$$|\Phi_\mu(f)| \le \int |f|\,d\mu \le \int_K d\mu = \mu(K) \le M.$$

Hence $\|\Phi_\mu\| = M$.

If μ is Radon, then it is inner regular on X; so $\mu(X) = M$. Thus Φ_μ is continuous if and only if $\|\mu\| := \mu(X) < \infty$. ■

7.2.3. PROPOSITION. *Let X be a LCH space. Suppose that φ is a positive linear functional on $C_c(X)$. If $K \subset X$ is compact, then there is a constant C_K so that $|\varphi(f)| \le C_K\|f\|_\infty$ for all $f \in C_c(X)$ with $\operatorname{supp}(f) \subset K$.*

A positive linear functional on $C_0(X)$ is continuous, i.e. there is a $C < \infty$ so that $|\varphi(f)| \le C\|f\|_\infty$ for all $f \in C_0(X)$.

PROOF. As in the previous proof, there is a function $h \in C_c(X)$ such that $\chi_K \le h \le \chi_L$ for some compact set L. Let $f \in C_c(X)$ with $\operatorname{supp}(f) \subset K$ with $\|f\|_\infty \le 1$. First suppose that f is real valued. Then $-h \le f \le h$, so $0 \le f+h \le 2h$. By positivity, $0 \le \varphi(f)+\varphi(h) \le 2\varphi(h)$; whence $|\varphi(f)| \le \varphi(h)$. If f takes arbitrary complex values, let $\varphi(f) = e^{i\theta}|\varphi(f)|$. Replace f by $e^{-i\theta}f$ so that we may suppose that $\varphi(f) > 0$. Write $f = \operatorname{Re} f + i \operatorname{Im} f$. Since real functions are sent to real values, $0 \le \varphi(f) = \varphi(\operatorname{Re} f) + i\varphi(\operatorname{Im} f)$ implies that $\varphi(f) = \varphi(\operatorname{Re} f)$. Hence $|\varphi(f)| \le \varphi(h)$. So $C_K = \varphi(h)$ is the desired constant.

Now suppose that φ is defined on $C_0(X)$. Let

$$M = \sup\{\varphi(f) : 0 \le f \le 1,\ f \in C_0(X)\}.$$

If $M = \infty$, we can choose $f_n \in C_0(X)$ with $0 \le f_n \le 1$ and $\varphi(f_n) > 2^n$. Define $f = \sum_{k\ge1} 2^{-k} f_{n_k}$. This converges uniformly, and so lies $C_0(X)$. By positivity,

$$\varphi(f) \ge \sup_{p\ge1} \varphi\Big(\sum_{n=1}^{p} 2^{-n} f_n\Big) = \sup_{p\ge1} \sum_{n=1}^{p} 2^{-n}\varphi(f_n) \ge \sup_{p\ge1} p = \infty.$$

This is impossible, and thus $M < \infty$. We now argue as in the first paragraph to deduce that $|\varphi(f)| \le \varphi(|f|) \le M\|f\|_\infty$ for f in $C_0(X)$. ■

For convenience of notation, given $U \subset X$ is open, we will write $f \prec U$ if $f \in C_c(X)$, $0 \le f \le 1$ and $\operatorname{supp}(f) \subset U$. Also recall that $C_c(X, \mathbb{R})$ is a lattice, and we write $f \vee g = \max\{f, g\}$ and $f \wedge g = \min\{f, g\}$.

7.2.4. RIESZ-MARKOV THEOREM. *Let φ be a positive linear functional on $C_c(X)$ for a LCH space X. Then there is a unique Radon measure μ on X so that $\varphi = \Phi_\mu$.*

PROOF. For $U \subset X$ open, define

$$\rho(U) = \sup\{\varphi(f) : f \prec U\}.$$

Use this to define an outer measure μ^* given by

$$\mu^*(E) = \inf\Big\{ \sum_{i\ge1} \rho(U_i) : E \subset \bigcup_{i\ge1} U_i,\ U_i \text{ open}\Big\}.$$

Let $\bar\mu$ denote the complete measure on the σ-algebra of μ^*-measurable sets.

Claim 1: $\mu^*(U) = \rho(U)$ for U open. Since $U \subset U$, we have $\mu^*(U) \le \rho(U)$. Suppose that $U \subset \bigcup_{i\ge1} U_i$ for U_i open. Take $f \prec U$. Then $K = \operatorname{supp}(f)$ is a compact subset of U, and hence $K \subset U_1 \cup \dots \cup U_n$ for some n. By Proposition 6.4.15, there is a partition of unity $g_i \in C_c(X)$ with $g_i \prec U_i \cap U$ such that $\chi_K \le \sum_{i=1}^n g_i \prec U$. Then $f = \sum_{i=1}^n f g_i$ and $f g_i \prec U_i \cap U$. Therefore

$$\varphi(f) = \sum_{i=1}^n \varphi(f g_i) \le \sum_{i=1}^n \rho(U_i).$$

Taking the supremum over $f \prec U$, we see that $\rho(U) \le \sum_{i=1}^\infty \rho(U_i)$.

This means that there is no advantage to using more than one open set to cover E. That is, if $E \subset \bigcup_{i\ge1} U_i = U$ for U_i open, then $\rho(U) \le \sum_{i\ge1} \rho(U_i)$. Therefore $\mu^*(E) = \inf\{\rho(U) : E \subset U,\ U \text{ open}\}$.

Claim 2: an open set U is μ^*-measurable, i.e., $\mu^*(E) = \mu^*(E\cap U) + \mu^*(E \setminus U)$ for all sets $E \subset X$. Recall that $\mu^*(E) \le \mu^*(E \cap U) + \mu^*(E \setminus U)$ is always valid. Thus is is enough to prove the reverse inequality when $\mu^*(E) < \infty$.

First let E be open with $\mu^*(E) < \infty$. Let $\varepsilon > 0$. Then for some $f \prec E \cap U$,

$$\mu^*(E \cap U) = \rho(E \cap U) < \varphi(f) + \varepsilon$$

Let $K = \operatorname{supp}(f)$. Then $\mu^*(E \setminus U) \le \mu^*(E \setminus K) = \rho(E \setminus K) < \varphi(g) + \varepsilon$ for some $g \prec E \setminus K$. Since f, g have disjoint supports, $f + g \le \chi_E$. Therefore

$$\mu^*(E \cap U) + \mu^*(E \setminus U) < \varphi(f+g) + 2\varepsilon \le \rho(E) + 2\varepsilon.$$

Letting $\varepsilon \to 0$, we get $\mu^*(E \cap U) + \mu^*(E \setminus U) \le \mu^*(U)$, and thus equality holds.

Now let E be arbitrary with $\mu^*(E) < \infty$. For $\varepsilon > 0$, pick $E \subset V$ with V open so that $\rho(V) < \mu^*(E) + \varepsilon$. Then

$$\mu^*(E \cap U) + \mu^*(E \setminus U) \le \mu^*(V \cap U) + \mu^*(V \setminus U) = \mu^*(V) < \mu^*(E) + \varepsilon.$$

Let $\varepsilon \to 0$ to get the non-trivial inequality. Thus U is μ^*-measurable.

Therefore $\bar{\mu}$ is defined on the σ-algebra generated by all open sets, namely $\mathrm{Bor}(X)$. Let μ be the Borel measure obtained by restricting $\bar{\mu}$ to $\mathrm{Bor}(X)$.

Claim 3: μ is Radon. First μ is outer regular since for $E \in \mathrm{Bor}(X)$,

$$\mu(E) = \mu^*(E) = \inf\{\mu(U) : E \subset U,\ U \text{ open}\}.$$

If K is compact, there is an open set $W \supset K$ with $L = \overline{W}$ compact. By Urysohn's Lemma, there is a function f with $\chi_K \le f \prec W$. Thus

$$\mu(K) \le \mu(W) \le C_L < \infty,$$

where C_L is the constant from Proposition 7.2.3. By outer regularity, we can choose an open $W \supset K$ so that $\mu(W) < \mu(K) + \varepsilon$. If f is chosen with $\chi_K \le f \prec W$, then $\varphi(f) < \mu(W) < \mu(K) + \varepsilon$.

On the other hand, suppose $f \in C_c(X)$ such that $\chi_K \le f$. Let $\varepsilon > 0$, and set $V = \{x : f(x) > 1 - \varepsilon\}$. Then $K \subset V$, and V is open. Let $g \prec V$ so that $\varphi(g) > \mu(V) - \varepsilon \ge \mu(K) - \varepsilon$. Observe that $(1-\varepsilon)g \le f$; and hence $\varphi(f) \ge (1-\varepsilon)(\mu(K) - \varepsilon)$. Letting $\varepsilon \to 0$ yields $\varphi(f) \ge \mu(K)$. Combining these two results shows that

$$\mu(K) = \inf\{\varphi(f) : \chi_K \le f,\ f \in C_c(X)\}.$$

Now if U is open, we have that

$$\mu(U) = \sup\{\varphi(f) : f \prec U\}.$$

Suppose that $r < \mu(U)$. Choose $g \prec U$ with $\varphi(g) > r$. Set $K = \mathrm{supp}(g)$, which is a compact subset of U. By Urysohn's Lemma, there is an $h \in C_c(X)$ with $\chi_K \le h \prec U$. By the previous paragraph, there is an $f \in C_c(X)$ with $\chi_K \le f$ and $\varphi(f) < \mu(K) + \varepsilon$. So $\chi_K \le h \wedge f \prec U$. Therefore

$$r < \varphi(g) \le \varphi(h \wedge f) \le \varphi(f) < \mu(K) - \varepsilon.$$

Hence $\mu(K) > r - \varepsilon$. It follows that

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

Thus μ is inner regular on open sets. Consequently μ is Radon.

Claim 4: $\varphi = \Phi_\mu$. By linearity, it suffices to verify this for $0 \le f \le 1$. Given $N \in \mathbb{N}$, let $t_i = \frac{i}{N}$ for $0 \le i \le N$. Define

$$f_i = \big((f \vee t_{i-1}) \wedge t_i\big) - t_{i-1} \quad \text{and} \quad K_i = \{x : f(x) \ge t_i\} \quad \text{for} \quad 1 \le i \le N,$$

and $K_0 = \mathrm{supp}(f)$. Then $\frac{1}{N}\chi_{K_i} \le f_i \le \frac{1}{N}\chi_{K_{i-1}}$ for $1 \le i \le N$ and $f = \sum_{i=1}^N f_i$. Integrating, we obtain

$$\frac{1}{N}\mu(K_i) \le \Phi_\mu(f_i) \le \frac{1}{N}\mu(K_{i-1}).$$

We also obtain $\frac{1}{N}\mu(K_i) \le \varphi(f_i)$. If $K_{i-1} \subset U$ for U open, then $Nf_i \prec U$ and so $\varphi(f_i) \le \frac{1}{N}\mu(U)$. But μ is outer regular, and hence $\varphi(f_i) \le \frac{1}{N}\mu(K_{i-1})$. Therefore

$$\frac{1}{N}\mu(K_i) \le \varphi(f_i) \le \frac{1}{N}\mu(K_{i-1}).$$

Now sum both of these inequalities from 1 to N, we obtain

$$\frac{1}{N}\sum_{i=1}^{N}\mu(K_i) \le \Phi_\mu(f_i),\ \varphi(f) \le \frac{1}{N}\sum_{i=0}^{N-1}\mu(K_i).$$

Hence

$$\Big|\Phi_\mu(f) - \varphi(f)\Big| \le \frac{\mu(K_0) - \mu(K_N)}{N} \le \frac{\mu(K_0)}{N}.$$

Since N was arbitrary, we obtain $\Phi_\mu(f) = \varphi(f)$.

Claim 5: uniqueness. Suppose that ν is a Radon measure such that $\Phi_\nu = \varphi$. Then if U is open and $f \prec U$, then $\varphi(f) = \Phi_\nu(f) \le \nu(U)$. On the other hand, since ν is inner regular on U, if $r < \nu(U)$, there is a compact $K \subset U$ with $\nu(K) > r$. Take some $f \in C_c(X)$ with $\chi_K \le f \prec U$ and observe that $\varphi(f) = \Phi_\nu(f) \ge \nu(K) > r$. Therefore

$$\nu(U) = \sup\{\varphi(f) : f \prec U\} = \mu(U).$$

Both measures are outer regular, and therefore they agree on all Borel sets. ■

7.2.5. COROLLARY. *Let X be a LCH space. Suppose that φ is a positive linear functional on $C_0(X)$. Then there is a unique finite Radon measure on X so that $\varphi = \Phi_\mu$.*

PROOF. The restriction of φ to $C_c(X)$ is positive. Thus by the Riesz-Markov Theorem, there is a unique Radon measure μ on X so that $\varphi = \Phi_\mu$ on $C_c(X)$. By Proposition 7.2.3, φ is continuous and thus $\varphi = \Phi_\mu$ on $C_0(X)$, and moreover

$$\mu(X) = \sup\{\varphi(f) : 0 \le f \le 1,\ f \in C_c(X)\} \le C < \infty.$$

So μ is a finite measure. ■

7.3. Linear functionals on $C_0(X)$

To discuss the continuous linear functionals on $C_0(X)$, we need complex Borel measures. Since they are finite measures, things simplify somewhat. A complex measure μ is called *regular* if $|\mu|$ is regular.

7.3.1. PROPOSITION. *Let X be a LCH space. Let μ be a complex Borel measure on X. Then $\Phi_\mu(f) = \int f\,d\mu$ defines a continuous linear functional on $C_0(X)$ such that $\|\Phi_\mu\| \le \|\mu\| := |\mu|(X)$. If μ is regular, then $\|\Phi_\mu\| = \|\mu\|$.*

PROOF. If $f \in C_0(X)$,

$$|\Phi_\mu(f)| = \Big|\int f\,d\mu\Big| \le \int |f|\,d|\mu| \le \|f\|_\infty\,|\mu|(X).$$

Thus $\|\Phi_\mu\| \le \|\mu\|$.

Suppose that μ is regular. By the Radon-Nikodym Theorem, there is a Borel function h so that $|h| = 1$ and $\mu = h|\mu|$. Given $\varepsilon > 0$, chop the unit circle $\mathbb{T}$ into disjoint arcs $I_1, \ldots, I_n$ of length at most ε and midpoints ζ_j. Set $E_j = h^{-1}(I_j)$. By regularity, find compact $K_j \subset E_j$ so that $|\mu|(E_j \setminus K_j) < \varepsilon/n$. Find disjoint open sets $U_j \supset K_j$. By Urysohn's Lemma, there are functions $f_j \in C_c(X)$ with $\chi_{K_j} \le f_j \prec U_j$. Let $f = \sum_{j=1}^n \zeta_j f_j$. Then $\|f\|_\infty = 1$ and

$$\begin{aligned}\big|\|\mu\| - \Phi_\mu(f)\big| &= \Big|\int \bar{h} - f\,d\mu\Big| \le \int \big|\bar{h} - f\big|\,d|\mu| \\ &\le \sum_{j=1}^n \int_{K_j} \big|\bar{h} - \zeta_j\big|\,d|\mu| + \sum_{j=1}^n 2|\mu|(E_j \setminus K_j) \\ &\le \frac{\varepsilon}{2}\|\mu\| + 2\varepsilon.\end{aligned}$$

Letting $\varepsilon \to 0$ yields $\|\Phi_\mu\| = \|\mu\|$. ∎

7.3.2. COROLLARY. *If μ, ν are regular complex Borel measures on a LCH space X, then $\Phi_\mu = \Phi_\nu$ if and only if $\mu = \nu$.*

7.3.3. DEFINITION. If X is a LCH space, let $M(X)$ be the vector space of all complex regular Borel measures on X with norm $\|\mu\| = |\mu|(X)$.

7.3.4. RIESZ REPRESENTATION THEOREM. *Let X be a LCH space. Then $C_0(X)^*$ is isometrically isomorphic to $M(X)$ via the pairing $\mu \to \Phi_\mu$.*

PROOF. Proposition 7.3.1 shows that the map from $M(X)$ into $C_0(X)^*$ is an isometric linear map. On the other hand, Corollary 7.2.5 of the Riesz-Markov Theorem shows that every positive linear functional on $C_0(X)$ is Φ_μ for a finite positive regular Borel measure μ. The result will follow if the linear span of the positive linear functional on $C_0(X)$ is all of $C_0(X)^*$.

We call a linear functional *self-adjoint* if $\varphi(f) \in \mathbb{R}$ for $f \in C_0(X, \mathbb{R})$. Let $\varphi \in C_0(X)^*$. Define $\tilde{\varphi}(f) = \overline{\varphi(\bar{f})}$. Set $\operatorname{Re}\varphi = \frac{1}{2}(\varphi + \tilde{\varphi})$ and $\operatorname{Im}\varphi = \frac{1}{2i}(\varphi - \tilde{\varphi})$.

Then if $f \in C_0(X, \mathbb{R})$ is real valued,

$$(\operatorname{Re}\varphi)(f) = \frac{\varphi(f) + \overline{\varphi(f)}}{2} = \operatorname{Re}\varphi(f) \in \mathbb{R}.$$

Similarly $\operatorname{Im}\varphi$ takes real values on $C_0(X, \mathbb{R})$. So $\operatorname{Re}\varphi$ and $\operatorname{Im}\varphi$ are self-adjoint. Note that these functionals are not real valued on $C_0(X)$. Observe that $\varphi = \operatorname{Re}\varphi + i\operatorname{Im}\varphi$. So $C_0^X)^*$ is spanned by the self-adjoint linear functionals.

Claim: if φ is self-adjoint, there are positive linear functionals φ_1, φ_2 on $C_0(X)$ such that $\varphi = \varphi_1 - \varphi_2$. This is an analogue of the Jordan decomposition theorem. If $f \geq 0$, define

$$\varphi_1(f) = \sup\{\varphi(g) : 0 \leq g \leq f\}.$$

Then $0 \leq \varphi_1(f) \leq \sup_{0 \leq g \leq f} \|\varphi\| \, \|g\| = \|\varphi\| \, \|f\|$. Also for $t > 0$, $\varphi_1(tf) = t\varphi_1(f)$.

Additivity: suppose that $f_1 \geq 0$ and $f_2 \geq 0$ in $C_0(X)$. If $0 \leq g_i \leq f_i$, then $0 \leq g_1 + g_2 \leq f_1 + f_2$. Hence $\varphi_1(f_1 + f_2) \geq \varphi(g_1) + \varphi(g_2)$. Taking the supremum over all choices of g_i yields $\varphi_1(f_1 + f_2) \geq \varphi_1(f_1) + \varphi_1(f_2)$. On the other hand, if $0 \leq g \leq f_1 + f_2$, let $g_1 = g \wedge f_1$ and

$$g_2 = g - g_1 = \begin{cases} 0 & \text{if } g(x) \leq f_1(x) \\ g(x) - f_1(x) \leq f_2(x) & \text{if } g(x) > f_1(x). \end{cases}$$

So $0 \leq g_2 \leq f_2$. Thus $\varphi(g) = \varphi(g_1) + \varphi(g_2) \leq \varphi_1(f_1) + \varphi_1(f_2)$. Taking the supremum over $0 \leq g \leq f_1 + f_2$ yields the reverse inequality : $\varphi_1(f_1 + f_2) \leq \varphi_1(f_1) + \varphi_1(f_2)$ Therefore $\varphi_1(f_1 + f_2) = \varphi_1(f_1) + \varphi_1(f_2)$.

Now we extend φ_1 to $C_0(X)$. If $f \in C_0(X)$, we write $f = \operatorname{Re} f + i \operatorname{Im} f = f_1 - f_2 + if_3 - if_4$ where $f_1 = \operatorname{Re} f \vee 0$, $f_2 = -\operatorname{Re} f \vee 0$, $f_3 = \operatorname{Im} f \vee 0$ and $f_4 = -\operatorname{Im} f \vee 0$. Define $\varphi_1(f) = \varphi_1(f_1) - \varphi_1(f_2) + i\varphi_1(f_3) - i\varphi_1(f_4)$. It is left to the reader to verify that φ_1 is linear. Once that it checked, it is clear that φ_1 is a continuous positive linear functional. Set $\varphi_2 = \varphi_1 - \varphi$. Then if $f \geq 0$,

$$\varphi_2(f) = \sup\{\varphi(g) - \varphi(f) : 0 \leq g \leq f\} \geq \varphi(f) - \varphi(f) = 0.$$

So φ_2 is also a positive linear functional; and $\varphi = \varphi_1 - \varphi_2$.

We have shown that the linear span of the positive linear functionals is all of $C_0(X)^*$. Therefore by the Riesz-Markov theorem for $C_0(X)$, each is given as integration against a finite Radon measure, hence regular. Thus the whole functional is given by integration against a regular complex Borel measure. ■

CHAPTER 8

A Taste of Probability*

This chapter is purely for interest. I expect to give one lecture on some of this, but it won't be tested. This chapter closely follows Folland's book [2].

8.1. The language of probability

While the basic theorems of probability theory often coincide with the analysts view of measure theory, the vocabulary is completely different. This chart explains the correspondence for some familiar notions.

Analyst's term	Probabalist's term
measure space $(X, \mathcal{B}, \mu)$	sample space $(\Omega, \mathcal{B}, P)$ and $P(\Omega) = 1$.
σ-algebra	σ-field
measurable set	event
real measurable function	*random variable* X
integral $\int f\,d\mu$	*expectation* $E(X)$
$\mu(\{x : f(x) < a\})$	$P(X < a)$
$\|f\|_p < \infty$	X has a finite pth moment
almost everywhere (a.e.)	*almost surely* (a.s.)
Borel probability measure on $\mathbb{R}$	distribution
charcteristic function χ_A	*indicator function* $\mathbf{1}_A$

In analysis, we say that a sequence f_n of measurable functions on $(X, \mathcal{B}, \mu)$ *converges in measure* to f if for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \mu\big(\{x : |f(x) - f_n(x)| > \varepsilon\}\big) = 0.$$

In probability theory, this is called *convergence in probability*.

If X is a random variable, $E(X)$ is the *expected value* or mean. The *standard deviation* (which is finite if $X \in L^2(P)$) is

$$\sigma(X) = \sqrt{E\big((X - E(X))^2\big)} = \|X - E(X)\|_2.$$

If $\varphi : (\Omega, \mathcal{B}) \to (\Omega', \mathcal{B}')$ is a measurable map, define P_φ on $(\Omega', \mathcal{B}')$ by $P_\varphi(B) = P(\varphi^{-1}(B))$. For example, if $X : \Omega \to \mathbb{R}$ is a random variable, the *distribution function of* X is

$$F(t) = P(X \le t) = P_X\big((-\infty, t]\big).$$

Moreover $P_X = dF = \mu_F$ is the Lebesgue-Stieltjes measure of F. A family of random variables $\{X_\alpha\}_{\alpha\in A}$ are *identically distributed* if $P_{X_\alpha} = P_{X_\beta}$ for all $\alpha, \beta \in A$. The *joint distribution* of $(X_1, \dots, X_n)$ is $P_{(X_1,\dots,X_n)}$ where we consider $(X_1, \dots, X_n) : \Omega \to \mathbb{R}^n$.

Many properties of random variables can be recovered from their distributions.

8.1.1. PROPOSITION. *If* $\varphi : (\Omega, \mathcal{B}) \to (\Omega', \mathcal{B}')$ *is measurable and* $f : \Omega' \to \mathbb{R}$ *is measurable with respect to* P_φ*, then*

$$\int_{\Omega'} f \, dP_\varphi = \int_\Omega f \circ \varphi \, dP.$$

PROOF. Let $f = \chi_A$. Then

$$\begin{aligned} \int_{\Omega'} \chi_A \, dP_\varphi &= P_\varphi(A) = P(\varphi^{-1}(A)) \\ &= \int_\Omega \chi_{\varphi^{-1}(A)} \, dP = \int_\Omega \chi_A \circ \varphi \, dP. \end{aligned}$$

Thus the result is true for simple functions. It extends to all positive measurable functions by the MCT. Then it extends to all integrable functions by the LDCT. ■

8.1.2. EXAMPLES.

(1) $E(X) = \displaystyle\int X \, dP = \int_{\mathbb{R}} t \, dP_X.$

(2) $\sigma^2(X) = E\big((X - E(X))^2\big) = \displaystyle\int_{\mathbb{R}} \big(t - E(X)\big)^2 \, dP_X.$

(3) $E(X + Y) = \displaystyle\int_{\mathbb{R}^2} s + t \, dP_{(X,Y)}(s, t).$

8.2. Independence

8.2.1. DEFINITION. Events $\{A_i : i \in I\}$ are *independent* if whenever $i_1, \dots, i_k$ in I are distinct, then

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{s=1}^{k} P(A_{i_s}).$$

Random variables $\{X_i : i \in I\}$ are *independent* if for every choice of Borel sets B_i, the collection $\{X_i^{-1}(B_i) : i \in I\}$ is independent. That is, whenever $i_1, \dots, i_k \in I$

are distinct,

$$P_{(X_{i_1},\dots,X_{i_k})}\big(\prod_{s=1}^{k} B_{i_s}\big) = P\big(\bigcap_{s=1}^{k} X_{i_s}^{-1}(B_i)\big) = \prod_{s=1}^{k} P(A_{i_s}) = \prod_{s=1}^{k} P_{X_{i_s}}\big(\prod_{s=1}^{k} B_{i_s}\big).$$

In other words, $P_{(X_{i_1},\dots,X_{i_k})} = \prod_{s=1}^{k} P_{X_{i_s}}$.

8.2.2. EXAMPLE. It is not sufficient to check independence for pairs of variables. Let $X = \{1,2,3,4\}$, $P(i) = \frac{1}{4}$, and $A_1 = \{2,3\}$, $A_2 = \{1,3\}$ and $A_3 = \{1,2\}$. You can check that $P(A_i) = \frac{1}{2}$ and $P(A_i \cap A_j) = \frac{1}{4}$ for $i \neq j$, but $P(A_1 \cap A_2 \cap A_3) = 0$.

8.2.3. PROPOSITION. *Let X, Y be independent integrable random variables. Then $E(XY) = E(X)\,E(Y)$.*

PROOF. Compute

$$\begin{aligned} E(XY) &= \int st\, dP_{(X,Y)}(s,t) = \int\!\!\int st\, dP_X\, dP_Y \\ &= \int s\, dP_X(s) \int t\, dP_Y(t) = E(X)\,E(Y). \end{aligned}$$

■

8.2.4. PROPOSITION. *Let $\{X_{i,j} : 1 \le j \le J_i,\ 1 \le i \le n\}$ be independent random variables. If $f_i : \mathbb{R}^{J_i} \to \mathbb{R}$ are Borel for $1 \le i \le n$, and $Y_i = f_i(X_{i,1},\dots,X_{i,J_i})$, then $\{Y_1,\dots,Y_n\}$ are independent.*

PROOF. Let $\bar{X}_i = (X_{i,1},\dots,X_{i,J_i})$. Then $Y_i^{-1}(B) = \bar{X}_i^{-1}(f_i^{-1}(B))$. Thus if $Y = (Y_1,\dots,Y_n)$ and $X = (\bar{X}_1,\dots,\bar{X}_n)$,

$$Y^{-1}(B_1 \times \dots \times B_n) = \bigcap_{i=1}^{n} \bar{X}_i^{-1}(f_i^{-1}(B_i)) = X^{-1}\big(\prod_{i=1}^{n} f_i^{-1}(B_i)\big).$$

Therefore

$$\begin{aligned} P\big(Y^{-1}(B_1 \times \dots \times B_n)\big) &= P_X\big(\prod_{i=1}^{n} f_i^{-1}(B_i)\big) \\ &= \prod_{i=1}^{n}\big(\prod_{j=1}^{J_i} P_{X_{i,j}}\big)\big(\prod_{i=1}^{n} f_i^{-1}(B_i)\big) \\ &= \prod_{i=1}^{n}\big(\prod_{j=1}^{J_i} P_{X_{i,j}}\big(f_i^{-1}(B_i)\big)\big) \end{aligned}$$

$$= \prod_{i=1}^{n} P_{\bar{X}_i}(f_i^{-1}(B_i)) = \prod_{i=1}^{n} P_{y_i}(B_i).$$

So $\{Y_1, \dots, Y_n\}$ are independent. ■

8.2.5. COROLLARY. *If $\{X_i\}$ are independent variables in $L^2(P)$, then*

$$\sigma^2(X_1 + \cdots + X_n) = \sum_{i=1}^{n} \sigma^2(X_i).$$

PROOF. Let $Y_i = X_i - E(X_i)$; so $E(Y_i) = 0$. Then $\{Y_i\}$ are independent, so $E(Y_iY_j) = E(Y_i)\,E(Y_j) = 0$ for $i \neq j$. Therefore

$$\sigma^2(X_1 + \cdots + X_n) = E\Big(\Big(\sum_{i=1}^{n} X_i - E(X_i)\Big)^2\Big) = E\Big(\Big(\sum_{i=1}^{n} Y_i\Big)^2\Big)$$
$$= \sum_{i=1}^{n} E(Y_i^2) + \sum_{i \neq j} E(Y_i)\,E(Y_j) = \sum_{i=1}^{n} \sigma^2(X_i).$$ ■

8.3. The Law of Large Numbers

The following easy lemma is surprisingly useful.

8.3.1. LEMMA (Chebychev's Inequality). *If $X \geq 0$ is a random variable and $a > 0$, then $P(X \geq a) \leq E(X)/a$.*

PROOF. $E(X) = \int_0^\infty t\,dP_X(t) \geq \int_a^\infty a\,dP_X(t) = aP(X \geq a).$ ■

8.3.2. COROLLARY. *Let X be a random variable such that $\sigma(X) < \infty$ (i.e. $X \in L^2(P)$). Then for $b > 0$, $P(|X - E(X)| \geq b) \leq \sigma^2(X)/b^2$.*

PROOF. Let $Y = (X - E(X))^2$; so $E(Y) = \sigma^2(X)$. Hence by Chebychev's inequality, $P(|X - E(X)| \geq b) = P(Y \geq b^2) \leq E(Y)/b^2 = \sigma^2(X)/b^2$. ■

8.3.3. Weak Law of Large Numbers. *Let $\{X_i\}_{i\geq 1}$ be independent L^2 random variables. Denote $E(X_i) = \mu_i$ and $\sigma^2(X_i) = \sigma_i^2$ for $i \geq 1$. If $\lim\limits_{n\to\infty} \frac{1}{n^2} \sum\limits_{i=1}^{n} \sigma_i^2 = 0$, then for any $\varepsilon > 0$,*

$$\lim_{n\to\infty} P\Big(\big|\tfrac{1}{n} \sum_{i=1}^{n} X_i - \mu_i\big| > \varepsilon\Big) = 0.$$

That is, $\lim\limits_{n\to\infty} \frac{1}{n} \sum\limits_{i=1}^{n} X_i - \mu_i = 0$ in probability.

Proof. Let $S_n = \frac{1}{n} \sum\limits_{i=1}^{n} (X_i - \mu_i)$; so $E(S_n) = 0$ and by Corollary 8.2.5, $\sigma^2(S_n) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2(X_i - \mu_i) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma_i^2$. Hence by Corollary 8.3.2,

$$P(|S_n| > \varepsilon) \leq \sigma^2(S_n)/\varepsilon^2 \to 0.$$

That is, S_n converges to 0 in probability. ∎

8.3.4. Borel-Cantelli Lemma. *Let $\{A_n\}_{n\geq 1}$ be events in a probability space. Define $B = \limsup A_n := \bigcap_{k\geq 1} \big(\bigcup_{n\geq k} A_n\big)$.*

(a) *If $\sum_{n\geq 1} P(A_n) < \infty$, then $P(B) = 0$.*

(b) *If $\{A_n\}$ are independent and $\sum_{n\geq 1} P(A_n) = \infty$, then $P(B) = 1$.*

Proof. (a) $P(B) \leq P\big(\bigcup_{n\geq k} A_n\big) \leq \sum_{n\geq k} P(A_n) \to 0$.

(b)
$$\begin{aligned} 1 - P\Big(\bigcup_{n\geq k} A_n\Big) &= P\Big(\bigcap_{n\geq k} A_n^c\Big) = \prod_{n\geq k} P(A_n^c) \\ &= \prod_{n\geq k} 1 - P(A_n) \leq \prod_{n\geq k} e^{-P(A_n)} = e^{-\sum_{n\geq k} P(A_n)} = 0. \end{aligned}$$

Thus $P(B) = 1$. ∎

8.3.5. Kolmogorov's Inequality. *Let $X_1, \dots, X_n$ be independent random variables with $E(X_i) = 0$ and $\sigma^2(X_i) = \sigma_i^2$. Then for $a > 0$,*

$$P\big(\max_{1\leq k\leq n} |X_1 + \dots + X_k| \geq a\big) \leq \frac{1}{a^2} \sum_{i=1}^{n} \sigma_i^2.$$

Proof. Let $S_k = X_1 + \dots + X_k$. Let $A = \big\{\max\{|S_1|, \dots, |S_n|\} \geq a\big\}$ and $A_k = \big\{|S_j| < a,\ 1 \leq j < k,\ |S_k| \geq a\big\}$. So $A = \dot\bigcup_{k=1}^{n} A_k$.

Since X_i are independent and mean 0,

$$\sum_{i=1}^{n} \sigma_i^2 = \sigma^2(X_1 + \cdots + X_n) = E(S_n^2) \geq E(S_n^2 \chi_A) = \sum_{k=1}^{n} E(S_n^2 \chi_{A_k}).$$

Also note that $S_k \chi_{A_k}$ depends only on $X_1, \ldots, X_k$ and $S_n - S_k$ depends only on $X_{k+1}, \ldots, X_n$. Hence these are independent quantities, so that

$$E\big((S_k \chi_{A_k})(S_n - S_k)\big) = E(S_k \chi_{A_k}) E(X_{k+1} + \cdots + X_n) = 0.$$

Therefore

$$\begin{aligned} E(S_n^2 \chi_{A_k}) &= E\big((S_k \chi_{A_k} + (S_n - S_k)\chi_{A_k})^2\big) \\ &= E(S_k^2 \chi_{A_k}) + 2E\big((S_k \chi_{A_k})(S_n - S_k)\big) + E\big((S_n - S_k)^2 \chi_{A_k})\big) \\ &\geq E(S_k^2 \chi_{A_k}) \geq a^2 P(A_k). \end{aligned}$$

Consequently,

$$a^2 P(A) = \sum_{k=1}^{n} a^2 P(A_k) \leq \sum_{k=1}^{n} E(S_n^2 \chi_{A_k}) \leq \sum_{i=1}^{n} \sigma_i^2.$$

Thus $P(A) \leq \frac{1}{a^2} \sum_{i=1}^{n} \sigma_i^2$. ■

8.3.6. Lemma. *Let* $\{X_i : i \geq 1\}$ *be independent random variables with* $E(X_i) = 0$ *and* $\sigma^2(X_i) = \sigma_i^2$. *If* $\sum_{i \geq 1} \sigma_i^2 < \infty$, *then* $P\big(\sum_{i \geq 1} X_i$ *converges*$\big) = 1$.

Proof. Let $S_n = X_1 + \cdots + X_n$. So $S_{n+k} - S_n = X_{n+1} + \cdots + X_{n+k}$. Given $\delta > 0$, choose N so that $\sum_{i \geq N} \sigma_i^2 < \delta$. Then if $\varepsilon > 0$ and $n \geq N$, Kolmogorov's inequality yields

$$P\big(\max_{1 \leq k \leq m} |S_{n+k} - S_n| \geq \varepsilon\big) \leq \frac{1}{\varepsilon^2} \sum_{i=n+1}^{n+m} \sigma_i^2.$$

Letting $m \to \infty$, we get

$$P\big(\max_{k \geq 1} |S_{n+k} - S_n| \geq \varepsilon\big) \leq \frac{1}{\varepsilon^2} \sum_{i > n} \sigma_i^2 < \frac{\delta}{\varepsilon^2}.$$

Choose $\delta = 2^{-3j}$ and $\varepsilon = 2^{-j}$ and obtain an integer n_j so that

$$A_j = \big\{ \max_{k \geq 1} |S_{n_j + k} - S_{n_j}| \geq 2^{-j} \big\}$$

has $P(A_j) < 2^{-j}$. By the Borel-Cantelli Lemma, since $\sum_{j \geq 1} P(A_j) < \infty$, we have $P\big(\bigcap_{k \geq 1} \bigcup_{j \geq k} A_j\big) = 0$. Hence $P\big(\bigcup_{i \geq 1} \bigcap_{j \geq i} A_j^c\big) = 1$. Therefore a.s., x belongs to $\bigcap_{j \geq i} A_j^c$ for some i, meaning that $\max_{k \geq 1} |S_{n_j + k} - S_{n_j}| < 2^{-j}$ for $j \geq i$. This means that $S_n(x)$ is Cauchy and thus converges with probability 1. ■

8.3.7. LEMMA. *If $(a_i)_{i\geq 1}$ is a sequence of real numbers and $\sum_{i\geq 1} \frac{a_i}{i}$ converges, then $\lim\limits_{n\to\infty} \frac{1}{n} \sum_{i=1}^n a_i = 0$.*

PROOF. Let $b_n = \frac{a_n}{n}$ and $S_n = \sum_{i=1}^n b_i$. Let $L = \lim_{n\to\infty} S_n$. Then $L = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} S_i$. Compute

$$\begin{aligned} \frac{1}{n}\sum_{i=1}^n a_i &= \frac{1}{n}\sum_{i=1}^n ib_i = \frac{1}{n}\sum_{i=1}^n \sum_{k=i}^n b_k \\ &= \frac{1}{n}\sum_{i=1}^n S_n - S_{i-1} = S_n - \frac{1}{n}\sum_{k=0}^{n-1} S_k \to 0. \end{aligned}$$

■

8.3.8. THEOREM (Kolmogorov). *Let $\{X_i : i \geq 1\}$ be independent random variables with $E(X_i) = \mu_i$ and $\sigma^2(X_i) = \sigma_i^2$. If $\sum_{i\geq 1} \frac{1}{i^2}\sigma_i^2 < \infty$, then*

$$\lim_{n\to\infty} \tfrac{1}{n}\Big(\sum_{i=1}^n X_i - \mu_i\Big) = 0 \text{ a.s.}$$

PROOF. Let $Y_i = \frac{1}{i}(X_i - a_i)$. Then $E(Y_i) = 0$, $\sigma^2(Y_i) = \frac{1}{i^2}\sigma_i^2$ and $\{Y_i\}$ are independent. By hypothesis, $\sum_{i\geq 1} \sigma^2(Y_i) < \infty$. Hence by Lemma 8.3.6, $P(\sum_{i\geq 1} Y_i \text{ converges}\} = 1$. Therefore by Lemma 8.3.7, with $a_i = P(X_i)$,

$$P\Big(\lim_{n\to\infty} \tfrac{1}{n} \sum_{i=1}^n (X_i - \mu_i) = 0\Big) = 1.$$

■

8.3.9. STRONG LAW OF LARGE NUMBERS. *Let $\{X_i\}_{i\geq 1}$ be independent, identically distributed random variables.*

(a) *If X_i are in L^1 with $E(X_i) = \mu$, then*

$$P\big(\lim_{n\to\infty} \tfrac{1}{n}(X_1 + \cdots + X_n) = \mu\big) = 1.$$

(b) *If X_i are not in L^1, then*

$$P\big(\limsup_{n\to\infty} \tfrac{1}{n}|X_1 + \cdots + X_n| = \infty\big) = 1.$$

PROOF. (a) By replacing X_n by $X_n - \mu$, we may suppose that $E(X_n) = 0$. Let $\lambda = P_{X_n}$, which is independent of n. Then $\|X_n\|_1 = E(|X_n|) = \displaystyle\int_{\mathbb{R}} |t|\, d\lambda$ and

$\int_{\mathbb{R}} t\,d\lambda = E(X_n) = 0$. Let $Y_n = X_n\chi_{\{|X_n|\le n\}}$ and $Z_n = X_n - Y_n$. Then

$$\sum_{n\ge 1} P(Z_n \ne 0) = \sum_{n\ge 1} P(|X_n| > n) = \sum_{n\ge 1} \int_{-\infty}^{-n} + \int_n^{\infty} d\lambda$$
$$= \int_{-\infty}^{\infty} \lfloor |t| \rfloor\, d\lambda \le \int_{-\infty}^{\infty} |t|\, d\lambda = \|X_1\|_1 < \infty.$$

By the Borel-Cantelli Lemma, $P(Z_n \ne 0 \text{ infinitely often}) = 0$. Therefore

$$P\big(\tfrac{1}{n}\sum_{i=1}^{n} Z_i \to 0\big) = 1.$$

Now $\sigma^2(Y_n) = E(Y_n^2) = \int_{-n}^{n} t^2\,d\lambda$. Hence by the MCT,

$$\sum_{n\ge 1} \tfrac{1}{n^2}\sigma^2(Y_n) = \sum_{n\ge 1} \int_{-n}^{n} \frac{t^2}{n^2}\,d\lambda = \int_{-\infty}^{\infty} \Big(\sum_{n\ge 1} \tfrac{1}{n^2}\chi_{[-n,n]}\Big) t^2\,d\lambda$$

for $n-1 < |t| \le n$, we have $\sum_{n\ge 1} \frac{1}{n^2}\chi_{[-n,n]}(t) = \sum_{k\ge n} \frac{1}{k^2} < \frac{2}{n}$ and thus

$$\le \int_{-\infty}^{\infty} 2|t|\,d\lambda < \infty.$$

Therefore by Kolmogorov's Theorem, $P\big(\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} Y_i = 0\big) = 1$. Consequently,

$$P\big(\lim_{n\to\infty} \tfrac{1}{n}\sum_{i=1}^{n} X_i = 0\big) \ge P\big(\lim_{n\to\infty} \tfrac{1}{n}\sum_{i=1}^{n} Y_i = 0\big) \cdot P\big(\lim_{n\to\infty} \tfrac{1}{n}\sum_{i=1}^{n} Z_i = 0\big) = 1.$$

(b) In this case, for any $C > 0$,

$$\sum_{n\ge 1} P(|X_n| \ge Cn) = \sum_{n\ge 1} \int \chi_{\{|t|\ge Cn\}}\,d\lambda \ge \int_{\mathbb{R}} \frac{|t|}{C} - 1\,d\lambda = +\infty.$$

By the Borel-Cantelli Lemma, $P(|X_n| \ge Cn \text{ infinitely often}) = 1$. But then

$$\max\Big\{\frac{|X_1 + \cdots + X_{n-1}|}{n-1}, \frac{|X_1 + \cdots + X_n|}{n}\Big\} \ge \frac{C}{2} \quad \text{infinitely often a.s.}$$

Therefore $\limsup_{n\to\infty} \frac{1}{n}|X_1 + \cdots + X_n| = \infty$ a.s. ■

BIBLIOGRAPHY

[1] S. Friedberg, A. Insel and L. Spence, *Linear Algebra. 5th ed.*, Prentice Hall, 2018, Englewood Cliffs, NJ

[2] G.B. Folland, *Real Analysis: modern techniques and their applications*, 2nd ed., John Wiley & Sons, New York, 1999.

[3] S. Willard, *General Topology*, Addison-Wesley Pub. Co, Reading, 1970.

INDEX