SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

The Lebesgue Integral

av

John Möller

2026 - No K6

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

The Lebesgue Integral

John Möller

Självständigt arbete i matematik 15 högskolepoäng, grundnivå

Handledare: Matthew de Courcy-Ireland

2026

Abstract

The aim of this thesis is to develop the Lebesgue integral and to study its fundamental convergence properties. Beginning with measure theory and measurable functions, the Lebesgue integral is constructed and its basic properties are established. The Monotone Convergence Theorem, Fatou's Lemma, and the Dominated Convergence Theorem are proved and used to explore key structural features of the space L^1. These results are then applied to integrals with respect to measures defined by densities and to the formulation of probability theory in a measure-theoretic framework. As a final application, the Central Limit Theorem is proved using methods from Fourier analysis.

Abstract

Målet med denna uppsats är att utveckla Lebesgueintegralen och studera dess grundläggande konvergensegenskaper. Med utgångspunkt i måttteori och mätbara funktioner konstrueras Lebesgueintegralen och dess grundläggande egenskaper fastställs. Den monotona konvergensteoremet, Fatous lemma och det dominerade konvergensteoremet bevisas och används för att undersöka centrala strukturella egenskaper hos rummet $L1$. Dessa resultat tillämpas sedan på integraler med avseende på mått definierade via tätheter samt på formuleringen av sannolikhetsteori inom en måttteoretisk ram. Som en avslutande tillämpning bevisas centrala gränsvärdessatsen med hjälp av metoder från Fourieranalys.

Contents

1 Introduction

The classical theory of integration, as developed through the Riemann integral, is sufficient for many elementary applications but shows significant limitations when dealing with limiting processes. In particular, convergence of sequences of functions often fails to interact well with Riemann integration, requiring strong assumptions such as uniform convergence.

The Lebesgue integral provides a framework in which integration is compatible with much weaker forms of convergence. By shifting focus from partitions of the domain to measuring sets on which functions take given values, the theory naturally incorporates limits, approximation, and convergence arguments that arise throughout analysis and probability theory.

The aim of this thesis is to develop the Lebesgue integral and its main convergence theorems in a concrete and systematic way, starting from basic measure-theoretic concepts. Rather than pursuing maximal generality, the emphasis is placed on clarity and on results that are directly useful for applications.

After introducing σ-algebras, measures, and measurable functions, the Lebesgue integral is defined and its basic properties are established. The Monotone Convergence Theorem, Fatou's Lemma, and the Dominated Convergence Theorem are then proved and form the core technical tools of the thesis. These results are used to study fundamental properties of the space L^1, as well as integrals with respect to measures defined by densities.

In the final chapters, the measure-theoretic framework is applied to probability theory. Random variables are treated as measurable functions, expectations are expressed as Lebesgue integrals, and independence is formulated in measure-theoretic terms. The thesis concludes with a proof of the Central Limit Theorem using Fourier analytic methods, illustrating how Lebesgue integration and L^1 techniques naturally arise in probability theory.

2 Motivation

In 1902, Henri Lebesgue published *Intégrale, longueur, aire* [Leb02], introducing what is now known as the Lebesgue integral. Rather than defining integration by partitioning the domain, as in the Riemann theory, Lebesgue's approach begins by partitioning the codomain and weighting values according to the measure of their preimages in the domain.

This perspective naturally leads to the notion of measuring subsets of the domain and, inevitably, to the question of which sets are measurable at all. The motivation for the resulting definitions is often developed through the theory of inner and outer measures, as in early texts like *Vorlesungen über reelle Funktionen* [Car13] or in later textbooks such as *Measure, Lebesgue Integrals, and Hilbert Spaces* [KF60].

Since inner/outer measures lies outside the scope of this text, an alternative motivation is given. The class of continuous functions will be attempted to be extended in a natural way so that it becomes closed under Cauchy sequences and algebraic operations. As the focus of this text will mainly be on the limiting theorems yielded from measure theory.

A function f being continuous means that for all open sets U in the codomain, $f^{-1}(U)$ is also open. Let us now instead naively define measurable functions and sets as supersets of continuous functions and open sets respectively.

Example 2.1 (unit function). Consider the continuous function

$$u_n(x) = \begin{cases} 1 - e^{-nx} & \text{if } x > 0 \\ 0 & \text{otherwise .} \end{cases}$$

One can see that the limit $u(x) = \lim_{n\to\infty} u_n(x)$ will become

$$u(x) = \begin{cases} 1 & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

which is not continuous.

Definition 2.2 (unit function)**.** The function $u(x)$ in definition 2.1 will be called the **unit function**.

If we want measurable functions to be closed under Cauchy sequences, we must include the unit function.

Definition 2.3 (characteristic function)**.** Let A be a set, then the characteristic function χ is defined as

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise.} \end{cases}$$

Proposition 2.4 (Characteristic function on an open interval)**.** *Given* (a,b) *is an open interval on* $\mathbb{R}$ *then* $\chi_{(a,b)}(x) = u(x-a) - u(x-b)$.

Proof. Follows from cases on $x \leq a$, $a < x < b$ and $b \leq x$. □

Since the unit function is measurable and we want measurable functions to be closed under subtraction we have that the characteristic function on an open interval must be measurable.

Proposition 2.5 (Basic properties of the characteristic function)**.** *Let* $A, A_1, A_2, \ldots$ *be sets. The following properties hold:*

a) $\chi_{A^c}(x) = 1 - \chi_A(x)$,

b) $\chi_{A_1 \cap A_2 \cap \ldots}(x) = \chi_{A_1}(x) \cdot \chi_{A_2}(x) \cdot \ldots$,

c) $\chi_{A_1 \cup A_2 \cup \ldots}(x) = 1 - \chi_{A_1^c}(x) \cdot \chi_{A_2^c}(x) \cdot \ldots \cdot$.

We can also of course get finite versions of b) and c) by letting sets after a certain point be all null sets.

Proof. Statement a) follows directly from cases on $x \in A$ and $x \in A^c$. Statement b) follows also from cases, as if $x \in A_1 \cap A_2 \cap \ldots$ then $x \in A_i$ for all i and thus $\chi_{A_i}(x) = 1$ for all i and all partial products are 1. If $x \notin A_1 \cap A_2 \cap \ldots$ then there exists an n where $x \notin A_n$, thus $\chi_{A_n}(x) = 0$ and all partial products after n will contain a zero and the product becomes 0. Statement c) follows from a) and b) together with De Morgan's law. □

Proposition 2.6 (Characteristic function on open set)**.** *If* A *is an open set on* $\mathbb{R}$ *then there exists a sequence of open intervals* $A_1, A_2, \ldots$ *such that*

$$\chi_A(x) = 1 - \chi_{A_1^c}(x) \cdot \chi_{A_2^c}(x) \cdot \ldots$$

Proof. Since $\mathbb{R}$ is a second countable topological space [Mun00, p. 190] we have a countable open cover of $\mathbb{R}$. If we take the intersection of each set with A we then get a countable union that equals A. The formula then follows from statement c in proposition 2.5. □

This means that if A is measurable then χ_A must be measurable. Which implies by proposition 2.6 that if $A, A_1, A_2, \ldots$ are open sets then the functions

- $\chi_{A^c}(x)$,
- $\chi_{A_1 \cap A_2 \cap \ldots}(x)$,

- $\chi_{A_1 \cup A_2 \cup \ldots}(x)$

must be measurable. Which implies the sets

- A^c,
- $A_1 \cap A_2 \cap \ldots$,
- $A_1 \cup A_2 \cup \ldots$,

must be measurable.

3 Measure Theory

Definition 3.1 (sigma-algebra)**.** Let X be a nonempty set. A nonempty subset $\mathcal{A} \subseteq \mathcal{P}(X)$ is called a σ-algebra of X if the sets of $\mathcal{A}$ are closed under

- complements
- countable intersections
- countable unions (including the empty union).

Definition 3.2 (Measurable space)**.** A tuple $(X, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra of X, is called a **measurable space**. If $\mathcal{A}$ is clear by context, then X might be called a **measurable space**.

Theorem 3.3 (Intersection of σ-algebras)**.** *Let $\{\mathcal{A}_i\}_{i\in I}$ be a collection of σ-algebras on X. Then $\cap_{i\in I}\mathcal{A}_i$ is a σ-algebra on X.*

Proof. Let $A, A_1, A_2, A_3, \ldots \in \cap_{i\in I}\mathcal{A}_i$. Then $A \in \mathcal{A}_i$ for all i which means that $A^c \in \mathcal{A}_i$ for all i since they are all σ-algebras and thus $A^c \in \cap_{i\in I}\mathcal{A}_i$. Likewise $A_1 \cap A_2 \cap \ldots$ and $A_1 \cup A_2 \cup \ldots$ are in all $\mathcal{A}_i$ since they are all σ-algebra, and thus they are also in $\cap_{i\in I}\mathcal{A}_i$. □

Definition 3.4 (Generated σ-algebra)**.** Let X be a set and $\mathcal{E} \subseteq \mathcal{P}(X)$. The intersection of all σ-algebras containing $\mathcal{E}$ is a σ-algebra due to theorem 3.3 and will be denoted $\mathcal{M}(\mathcal{E})$ and called the σ-algebra **generated** by $\mathcal{E}$.

Definition 3.5 (Borel σ-algebra)**.** Let X be a topological space, that is to say the open sets of X are defined. Then the σ-algebra generated by the open sets of X is called the **Borel σ-algebra** on X and denoted $\mathcal{B}_X$.

Unless anything else is mentioned, $\mathbb{C}$ and $\mathbb{R}$ will be assumed to be measurable spaces with their respective borel σ-algebra.

Definition 3.6 (Measure)**.** Let $(X, \mathcal{M})$ be a measurable space. A measure on that space is a function $\mu : \mathcal{M} \to [0, \infty]$ such that

1. $\mu(\emptyset) = 0$,
2. For a disjoint sequence $\{E_j\}_1^\infty$ in $\mathcal{M}$, then $\mu(\cup E_j) = \sum \mu(E_j)$.

The triple $(X, \mathcal{M}, \mu)$ (or just X if $\mathcal{M}$ and μ are clear by context) is called a **measure space**.

If $\mathcal{M}$ is the borel σ-algebra on X then it is called a **Borel measure space** and μ is specifically called a **borel measure**.

If $\mu(X) < \infty$ then the space is called finite.

If there exists a countable sequence of subsets of finite measure that has as their union the space X, then the space is called σ-finite.

Remark 3.7. In the definition of a measure we let its codomain be $[0, \infty]$. It is to be understood as $[0, \infty) \cup \{\infty\}$ where ∞ is a symbol such that any sum or product that contains it as a term or factor somewhere is ∞ except when it is a product with 0 in which case it is 0.

Definition 3.8 (Null sets, almost everywhere and completeness)**.** Consider a measure space $(X, \mathcal{M}, \mu)$. If $\mu(E) = 0$, then E is called a **null set**. If a statement is true for all $x \in A$ where A^c is a null set, then that statement is true **almost everywhere**. If $\mathcal{M}$ contains all subsets of null sets then the measure space is **complete**.

Completeness is a good property to have in a measure space, because it would be counterintuitive if something being true at more points than what is almost everywhere stops it from being true almost everywhere. Conveniently given a measure space one can always create a superset of the σ-algebra that is complete and a unique measure that agrees on the original σ-algebra.

Theorem 3.9 (Completion)**.** *Let $(X, \mathcal{M}, \mu)$ be a measure space and let $\overline{\mathcal{M}}$ be the set of all $E \cup F$ where $E \in \mathcal{M}$ and F is some subset of a null set. Then $\overline{\mathcal{M}}$ is a σ-algebra on X and there exists a unique $\hat{\mu}$ such that $(X, \overline{\mathcal{M}}, \hat{\mu})$ is a complete measure space.*

Proof. Let $\mathcal{N}$ be the set of all null sets. Both $\mathcal{M}$ and $\mathcal{N}$ are closed under countable intersections.

For each $E \cup F \in \overline{\mathcal{M}}$ there exists a $N \in \mathcal{N}$ such that $F \subseteq N$ by definition. Let's replace all $E \cup F \in \overline{\mathcal{M}}$ where $E \cap F \neq \emptyset$ with the representation $E \cup (F - E)$ as it's the same set. Now we can assume for $E \cup F \in \overline{\mathcal{M}}$ that $E \cap F = \emptyset$. Also when talking about a null set N that F is a subset of, replace it with $N - E$. This is still in $\overline{\mathcal{M}}$ as $N - E = N \cap E^c$ and it is a null set as

$$\begin{aligned} 0 &= \mu(N) \\ &= \mu((E \cap N) \cup (E^c \cap N)) \\ &= \mu(E \cap N) + \mu(E^c \cap N) \\ &\geq \mu(E^c \cap N) \\ &\geq 0. \end{aligned}$$

Now we can assume that for $E \cup F \in \overline{\mathcal{M}}$ that there exists a $N \in \mathcal{N}$ so that $F \subseteq N$ and $N \cap E = \emptyset$. We have that

$$\begin{aligned} (E \cup N) \cap (N^c \cup F) &= (E \cap N^c) \cup (E \cap F) \cup (N \cap N^c) \cup (N \cap F) \\ &= E \cup (E \cap F) \cup \emptyset \cup F \\ &= E \cup F. \end{aligned}$$

Which implies that $(E \cup F)^c = (E \cup N)^c \cup (N - F)$, and $(E \cup N)^c \in \overline{\mathcal{M}}$ and $(N - F) \subseteq N$ which means that $(E \cup F)^c \subseteq \overline{\mathcal{M}}$. We can conclude that $\overline{\mathcal{M}}$ is a σ-algebra.

We define $\bar{\mu}$ with $\bar{\mu}(E \cup F) = \mu(E)$. To see that it is well defined, assume $E_1 \cup F_1 = E_2 \cup F_2$ where $F_1 \subseteq N_1$ and $F_2 \subseteq N_2$ for null sets N_1, N_2. This means $E_1 \subseteq E_2 \cup F_2 \subseteq E_2 \cup N_2$ which implies $\mu(E_1) \leq \mu(E_2) + \mu(N_2) = \mu(E_2)$, and by switching indexes this argument holds for yielding $\mu(E_2) \leq \mu(E_1)$.

To see $\bar{\mu}$ is unique let $E \cup F$ be in the completion. As mentioned earlier we can assume $E \cap F \neq \emptyset$. Let $\bar{\mu}'$ be any complete measure on the completion that agrees with μ on the original sigma algebra.

$$\bar{\mu}'(E \cup F) = \bar{\mu}'(E) + \bar{\mu}'(F) = \bar{\mu}'(E) = \mu(E) = \bar{\mu}(E \cup F).$$

Thus, for any measurable set E in $\mathcal{M}$, $\mu'(E)$ equals $\mu(E)$. ☐

Definition 3.10 (Completion)**.** The complete measure space in theorem (3.9) is called the **completion** of $(X, \mathcal{M}, \mu)$.

Lemma 3.11. *Given a complete measure space if f_n is measurable for all $n \in \mathbb{N}$ and $f_n \to f$ μ-a.e., then f is measurable.*

Proof. See Proposition 2.11 in [Fol99]. ☐

Lemma 3.12. *Consider a measure space $(X, \mathcal{M}, \mu)$. If f is measurable in the completion, then there exists a $\mathcal{M}$-measurable function such that it equals f $\bar{\mu}$-almost everywhere.*

Proof. See proposition 2.12 in [Fol99]. ☐

Lemma 3.13. *Consider a measure space. If f_n is a sequence of measurable functions that converges almost everywhere to f then one can redefine f almost everywhere to make it measurable*

Proof. Since measurable functions remain measurable in the completion, f is measurable in the completion by lemma 3.11. Lemma 3.12 then ensures there exists a measurable function almost everywhere equal to f in the original space. ☐

Theorem 3.14 (Distribution determines measure)**.** *Let $F : \mathbb{R} \to \mathbb{R}$ be increasing and right continuous. There exists a unique measure μ_F on $\mathbb{R}$ such that $\mu_F((a, b]) = F(b) - F(a)$ for all a, b.*

Proof. See Theorem 1.16 in [Fol99]. ☐

Definition 3.15 (Lebesgue-Stieltjes Measure)**.** A measure yielded from a function F as described in theorem 3.14 will be called a Lebesgue-Stieltjes Measure.

Definition 3.16 (Lebesgue measure)**.** Let $F(x) = x$ where $F : \mathbb{R} \to \mathbb{R}$. The completion of $(\mathbb{R}, \mathcal{M}, \mu_F)$ is called the **Lebesuge measure space**, and the measure associated with it is called the **Lebesgue measure** and denoted m.

4 Integration

Definition 4.1 (Measurable function)**.** Let $(X, \mathcal{M})$, $(Y, \mathcal{N})$ be two measurable spaces. A function $f : X \to Y$ **measurable** if $f^{-1}(N) \in \mathcal{M}$ for all $N \in \mathcal{N}$.

Definition 4.2 (Simple function)**.** Let f be a complex valued measurable function on a measurable space $(X, \mathcal{M})$. If the range of f range $(f) = \{z_1, z_2, \dots, z_n\}$ is finite then f is called a **simple function**. Let $E_j = f^{-1}(\{z_j\})$, it is clear that

$$f(x) = \sum_{j=1}^{n} z_j \cdot \chi_{E_j}(x).$$

This is called the **standard representation** of f.

Definition 4.3 (Integral of simple function)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and let $\phi : X \to \mathbb{R}$ be a positive measurable simple function. Given a standard representation

$$\phi(x) = \sum_{j=1}^{n} x_j \cdot \chi_{E_j}(x)$$

the integral of a simple function is defined as

$$\int \phi := \sum_{j=1}^{n} x_j \cdot \mu(E_j).$$

Theorem 4.4. *Let $(X, \mathcal{M}, \mu)$ be a measure space and let $a, b \in \mathbb{C}$ and let A, B be measurable sets. Then*

$$\int (a\chi_A + b\chi_b) = \int a\chi_A + \int b\chi_b.$$

Proof.

$$\begin{aligned}
\int (a\chi_A + b\chi_B) &= \int (a\chi_{A\setminus B} + (a+b)\chi_{A\cap B} + b\chi_{B\setminus A} + 0 \cdot \chi_{(A\cup B)^c}) \\
&= a\mu(A \setminus B) + (a+b)\mu(A \cap B) + b\mu(B \setminus A) \\
&= a(\mu(A \setminus B) + \mu(A \cap B)) + b(\mu(B \setminus A) + \mu(A \cap B)) \\
&= a\mu(A) + b\mu(B) \\
&= \int (a\chi_A + 0 \cdot \chi_{A^c}) + \int (b\chi_B + 0 \cdot \chi_{B^c}) \\
&= \int a\chi_A + \int b\chi_B.
\end{aligned}$$

□

Corollary 4.5. *For simple functions ϕ, θ*

$$\int (\phi + \theta) = \int \phi + \int \theta.$$

Definition 4.6 (Integral of positive function)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f : X \to \mathbb{R}$ be a positive valued measurable function.

Then the integral is defined as

$$\int f := \sup \left\{ \int \phi : \phi \text{ positive simple function with } \phi \leq f \right\},$$

where the integrals inside the set use definition 4.3.

Definition 4.7 (Integral of real valued function)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f : X \to \mathbb{R}$ be a measurable function.

Let $f^+ = \max(f, 0)$ and $f^- = \max(-f, 0)$ be the positive respectively negative parts of the function f such that $f = f^+ - f^-$. These functions are measurable by corollary 2.8 in [Fol99].

Then the integral of f is defined as

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

where the integrals on the right hand side uses definition 4.6.

Definition 4.8 (Lebesgue integral)**.** Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f : X \to \mathbb{C}$ be a measurable function. The functions $\text{Re}(f)$ and $\text{Im}(f)$ are measurable functions by corollary 2.5 in [Fol99].

Then the **Lebesgue integral** of f is defined as

$$\int f := \int \text{Re}(f) + i \int \text{Im}(f),$$

If we want to make what measure is being used explicit the notation

$$\int f d\mu$$

can be used. If f is written as dependent on a variable x, then the notation

$$\int f(x) d\mu(x)$$

can be used. Furthermore if we understand what measure to use from context we can use the familiar notation of

$$\int f(x)dx.$$

The Lebesgue integral will be the assumed integral used unless otherwise stated.

Lemma 4.9. *Let f and g be two real measurable functions so that $f \leq g$. Then*

$$\int f\, d\mu \leq \int g\, d\mu.$$

Proof. First assume that f and g are positive functions. Since $f \leq g$ we have that

$$\left\{\int \phi : \phi \text{ is simple and } \phi \leq f\right\} \subseteq \left\{\int \phi : \phi \text{ is simple and } \phi \leq g\right\}.$$

Thus it follows that

$$\sup\left\{\int \phi : \phi \text{ is simple and } \phi \leq f\right\} \leq \sup\left\{\int \phi : \phi \text{ is simple and } \phi \leq g\right\},$$

which per definition mean

$$\int f \leq \int g.$$

Now for any two real functions we can observe that if $f \leq g$ then $f^+ \leq g^+$ and $-f^- \leq -g^-$. If we expand the definition of the lebesgue integral on real functions

$$\int f = \int f^+ - \int f^-$$
$$\int g = \int g^+ - \int g^-$$

we see that term by term it implies

$$\int f \leq \int g.$$

□

Definition 4.10. Let $(X, \mathcal{M}, \mu)$ be a measure space and let $f : X \to \mathbb{C}$ be a measurable function. If

$$\int |f| d\mu < \infty$$

then f is called integrable.

Since $\text{Re}f \leq |f|$ and $\text{Im}f \leq |f|$ then lemma 4.9 implies

$$\int \text{Re}f \leq \int |f| \leq \infty$$
$$\int \text{Im}f \leq \int |f| \leq \infty$$

which implies

$$\int f = \int \text{Re}f + i \int \text{Im}f \leq \infty$$

Lemma 4.11. *Let $(X, \mathcal{M}, \mu)$ be a measure space, let $f, g : X \to \mathbb{C}$ be integrable functions and let $a \in \mathbb{C}$. Then*

$$\int (af + g) = a \int f + \int g.$$

Proof. For any constant $c \in \mathbb{C}$ it follows directly that $\int cf = c \int f$ from the definition. Since $|af + g| \leq |a||f| + |g|$ and $|f|$ and $|g|$ are integrable it follows that $af + g$ is integrable. It suffices to prove now that $\int (f + g) = \int f + \int g$. It suffices further to prove for when $f, g \in \mathbb{R}^+$ since one can match up imaginary and real parts from the definition and likewise positive and negative parts. We will prove it by showing inequalities in both directions apply. We begin with

$$\int f + \int g \leq \int (f + g).$$

We can observe there exists simple ϕ, θ such that $\phi \leq f$, $\theta \leq g$, $\int f - \int \phi < \epsilon$ and $\int g - \int \theta < \epsilon$ for all $\epsilon > 0$. This implies that $\phi + \theta \leq f + g$ which further implies by lemma 4.9 that $\int (\phi + \theta) \leq \int (f + g)$ as well as $\int \phi \leq \int f$ and $\int \theta \leq \int g$. By corollary 4.5,

$$\int f + \int g \leq \int \phi + \int \theta + 2\epsilon = \int (\phi + \theta) + 2\epsilon \leq \int (f + g) + 2\epsilon,$$

for all $\epsilon > 0$. This implies that

$$\int f + \int g \leq \int (f+g).$$

Now for the other direction let φ be a simple function such that $\varphi \leq f + g$ such that $\int(f+g) - \int \varphi \leq \epsilon$ for all $\epsilon > 0$. Let $\varphi = \sum z_n \chi_{E_n}$ be the standard representation. Now define

$$\phi := \sum_n \left(\inf_{x \in E_n} f(x) \right) \chi_{E_j},$$
$$\theta := \sum_n \left(\inf_{x \in E_n} g(x) \right) \chi_{E_j}.$$

It follows directly that $\phi \leq f$ and $\theta \leq g$ but also $\varphi \leq \phi$ and $\varphi \leq \theta$ since if any factor to a term was bigger in φ it would imply either $\varphi(x) > f(x)$ or $\varphi(x) > g(x)$ for some x. This in turn trivially implies $\varphi \leq \phi + \theta$. Using lemma 4.9 again we get the following inequalities

$$\int \phi \leq \int f,$$
$$\int \theta \leq \int g,$$
$$\int \varphi \leq \int (\phi + \theta).$$

Now we can reason

$$\begin{aligned}
\int (f+g) &\leq \int \varphi + \epsilon \\
&\leq \int (\phi + \theta) + \epsilon \\
&= \int \phi + \int \theta + \epsilon \\
&\leq \int f + \int g + \epsilon,
\end{aligned}$$

for all $\epsilon > 0$ which implies

$$\int (f+g) \leq \int f + \int g.$$

□

Lemma 4.12. *For integrable f and g,*

$$f = g \ a.e. \quad \iff \quad \int |f-g| = 0.$$

Proof. If $f = g$ almost everywhere, then $|f - g| = 0$ almost everywhere. For any term $z_j\mu(E_j)$ in the integral of a simple function ϕ with $0 \le \phi \le |f - g|$, either $z_j = 0$ or E_j is a measurable subset of a null set. In either case, $\mu(E_j) = 0$, so the integral of any such simple function is zero. Hence, the supremum over all such simple functions is zero, and thus

$$\int |f-g| = 0.$$

Conversely, suppose $\int |f - g| = 0$. Let ϕ be a simple function such that $0 \le \phi \le |f - g|$. Then

$$\int \phi \le \int |f-g| = 0,$$

which implies $\int \phi = 0$.

Now assume, for contradiction, that $|f - g|$ is not zero almost everywhere. Then there exists $c > 0$ and a measurable set E with $\mu(E) > 0$ such that $|f - g| \ge c$ on E. Consider the simple function $\phi = c\chi_E$. Then

$$\int |f-g| \ge \int \phi = c\mu(E) > 0,$$

which contradicts the assumption that $\int |f - g| = 0$.

Therefore, $|f - g| = 0$ almost everywhere, completing the proof. □

Theorem 4.13. *Let f be a bounded real-valued function on $[a, b]$. If f is Riemann integrable, then f is Lebesgue integrable and*

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

where m is the lebesgue measure.

Proof. See theorem 2.28 in [Fol99]. □

5 Limiting Theorems

The reward of having developed this theory is that it yields powerful theorems for interchanging the limit to be either inside or outside the integral in question. This will prove to be very useful in the remaining chapters.

Lemma 5.1. *Let E_n be a sequence of sets such that $E_1 \subset E_2 \subset E_3 \subset \ldots$ and the union of all equals E. Then*

$$\lim_{n\to\infty} \int \chi_{E_n} f = \int \chi_E f.$$

Proof. It follows from proposition 2.13d and theorem 1.8c in [Fol99]. □

The characteristic function here basically specifies which subset to integrate over, thus the following notation will also be used in text:

$$\lim_{n\to\infty} \int_{E_n} f = \int_E f.$$

Theorem 5.2 (Monotone Convergence Theorem)**.** *Let $f_n : X \to \mathbb{R}^+ \cup \{\infty\}$ be measurable functions for all integers $n \geq 1$ (sets in $\mathbb{R}^+ \cup \{\infty\}$ are defined as measurable if their intersection with $\mathbb{R}$ is measurable). If $\{f_n\}$ is a sequence in L^1 and $f_j \leq f_{j+1}$ for all j, then*

$$\int \lim_{n\to\infty} f_n = \lim_{n\to\infty} \int f_n.$$

Proof. Since the sequence $\{f_n\}$ is increasing, it follows that $f := \lim_{n\to\infty} f_n$ exists (possibly taking the value $+\infty$ at some points). Moreover, we have $f_n \leq f$ for all n. Lemma 4.9 implies $\lim_{n\to\infty} \int f_n \leq \int f$. Hence, it suffices to show that $\lim_{n\to\infty} \int f_n \geq \int f$.

Let $\alpha \in (0,1)$, and let ϕ be a simple function such that $0 \leq \phi \leq f$. Define $h_n(x) := f_n(x) - \alpha\,\phi(x)$. Since h_n is measurable, the sets $E_n := h_n^{-1}([0,\infty)) = \{x \in X : f_n(x) \geq \alpha\,\phi(x)\}$ are measurable. One can see that $\bigcup_{n=1}^{\infty} E_n = X$.

We now have $\int f_n \geq \int_{E_n} f_n \geq \alpha \int_{E_n} \phi$. Taking the limit as $n \to \infty$ together with lemma 5.1 we get $\lim_{n\to\infty} \int f_n \geq \alpha \int \phi$. Since this is true for all $\alpha \in (0,1)$ it is also true for $\alpha = 1$. Taking the supremum over all simple $\phi \leq f$ one obtains $\lim_{n\to\infty} \int f_n \geq \int f$. □

Theorem 5.3 (Fatou's Lemma). *If $\{f_n\}$ is any sequence positively integrable functions, then*

$$\int (\liminf f_n) \leq \liminf \int f_n.$$

Proof. Let $k \geq 1$ and $j \geq k$. Then

$$\inf_{n \geq k} f_n \leq f_j,$$

which implies

$$\int \inf_{n \geq k} f_n \leq \int f_j.$$

Since it was true for all $j \geq k$ we can choose the j so the right hand side is as small as possible:

$$\int \inf_{n \geq k} f_n \leq \inf_{j \geq k} \int f_j.$$

Now, from the definition of $\liminf$ and the monotone convergence theorem we can reason as follows:

$$\int \liminf_{n\to\infty} f_n \;=\; \int \lim_{k\to\infty} \inf_{n\geq k} f_n \;=\; \lim_{k\to\infty} \int \inf_{n\geq k} f_n \;\leq\; \lim_{n\to\infty} \inf_{j\geq n} \int f_n \;\leq\; \liminf_{n\to\infty} \int f_n.$$

□

Theorem 5.4 (The Dominated Convergence Theorem). *Let $\{f_n\}$ be a sequence of integrable functions such that (a) $f_n \to f$ a.e. and (b) there exists a nonnegative integrable g such that $|f_n| \leq g$ a.e. for all n. Then f is integrable and*

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

Proof. The function f is measurable (possibly after redefinition on a null set) by lemma 3.13. Since $|f| \leq g$ a.e., and g is integrable it means f is integrable. By taking real and imaginary parts, it suffices to assume that f_n and f are real-valued. In this case, we have $g + f_n \geq 0$ a.e. and $g - f_n \geq 0$ a.e.

Thus, by Fatou's lemma and lemma 4.11,

$$\int g + \int f = \int (g + f_n) \leq \liminf \int (g + f_n) = \int g + \liminf \int f_n,$$

and

$$\int g - \int f \leq \liminf \int (g - f_n) = \int g - \limsup \int f_n.$$

Therefore,

$$\liminf \int f_n \geq \int f \geq \limsup \int f_n,$$

and the result follows. □

6 The space L^1

In a utopia, one would always have

$$\lim_{n\to\infty} \int f_n = \int \lim_{n\to\infty} f_n.$$

For Riemann integration, the main way to ensure this equality is to restrict sequences to uniformly convergent ones, which is a relatively strong requirement. However, with the Monotone Convergence Theorem and the Dominated Convergence Theorem introduced in the previous chapter, much weaker conditions suffice to interchange the limit and the integral.

The strength of the Monotone Convergence Theorem lies in its simplicity: requiring the sequence to be positive and monotonically increasing is often natural and straightforward. The strength of the Dominated Convergence Theorem is that it imposes no restrictions on the mode of convergence; it only requires that the sequence be dominated by an integrable function, which is usually easy to verify if the limit can indeed be interchanged. On the other hand, it does restrict the domain of the functions: they must belong to the space of integrable functions. This space, therefore, comes very close to the utopia in which limits and integrals can always be interchanged. Understanding the structure of this space is thus of particular interest.

Throughout this chapter, we will assume a fixed measure space $(X, \mathcal{M}, \mu)$.

Definition 6.1 (Banach space)**.** A **Banach space** is a normed vector space $(V, \|\cdot\|)$ that is *complete*, meaning that every Cauchy sequence in V converges to an element in V. That is, for any sequence $(v_n)_{n\geq 1}$ in V such that

$$\forall \epsilon > 0, \exists N : m, n \geq N \implies \|v_m - v_n\| < \epsilon,$$

there exists $v \in V$ with $\|v_n - v\| \to 0$ as $n \to \infty$.

Theorem 6.2 (Equivalence classes form a normed vector space)**.** *Define an equivalence relation $f \sim g$ iff $f = g$ a.e., and let $[f]$ denote the equivalence class of f.*

Define

$$\|[f]\|_{L^1} := \int |f|.$$

Then $\|\cdot\|_{L^1}$ *is a well-defined norm, and the set of equivalence classes under* $\sim$ *forms a normed vector space.*

Proof. Follows directly from lemmas 4.11, 4.9 and 4.12.

□

Definition 6.3 (L^1 space)**.** Following Theorem 6.2, we define

$$L^1 := \{[f] : f \text{ is integrable}\},$$

the normed vector space of equivalence classes of integrable functions. The statement $f \in L^1$ will both refer to the statement that f is in itself an equivalence class or a specific measurable function in one of the equivalence classes in L^1 depending on context.

Theorem 6.4 (L^1 Banach)**.** *The space* L^1 *is a Banach space.*

Proof. In light of theorem 6.2 what remains to be proved is that L^1 is complete.

Let (f_n) be a Cauchy sequence in L^1. For each $k \geq 1$, set $\varepsilon_k = 2^{-k}$. Thus there exists N_k such that

$$m, n \geq N_k \implies \|f_m - f_n\|_1 < 2^{-k}.$$

Construct a subsequence (f_{n_k}) inductively by

$$n_1 := N_1, \qquad n_{k+1} := \max(1 + n_k, N_{k+1}).$$

Then $n_{k+1} > n_k$ and

$$\|f_{n_{k+1}} - f_{n_k}\|_1 < 2^{-k}.$$

Define $g_k := |f_{n_{k+1}} - f_{n_k}|$ and $G_m := \sum_{k=1}^{m} g_k$. Then G_m is monotone increasing and

$$\int G_m = \sum_{k=1}^{m} \|f_{n_{k+1}} - f_{n_k}\|_1 \leq \sum_{k=1}^{\infty} 2^{-k} = 1.$$

By the Monotone Convergence Theorem, $G := \sum_{k=1}^{\infty} g_k \in L^1$. For almost every x, the series

$$\sum_{k=1}^{\infty} (f_{n_{k+1}}(x) - f_{n_k}(x))$$

converges absolutely (because otherwise $\int G = \infty$). Define

$$f(x) := f_{n_1}(x) + \sum_{k=1}^{\infty}(f_{n_{k+1}}(x) - f_{n_k}(x)).$$

Then $f_{n_k} \to f$ pointwise a.e. and

$$|f_{n_k} - f| \leq \sum_{j=k}^{\infty} |f_{n_{j+1}} - f_{n_j}| \leq G \in L^1.$$

By the Dominated Convergence Theorem, $\|f_{n_k} - f\|_1 \to 0$. Since (f_n) is Cauchy, the full sequence converges to the same limit f in L^1. □

Lemma 6.5 (Signature is Simple)**.** *If $f : X \to \mathbb{R}$ is measurable, then* $\operatorname{sgn}(f)$ *is a simple function.*

Proof. Since $\{x \in \mathbb{R} : x < 0\}$, $\{x \in \mathbb{R} : x > 0\}$ and $\{0\}$ are measurable sets and f is measurable we have that $\operatorname{sgn}(f)^{-1}(1)$, $\operatorname{sgn}(f)^{-1}(-1)$ and $\operatorname{sgn}(f)^{-1}(0)$ are all measurable and thus

$$\operatorname{sgn}(f) = -1 \cdot \chi_{\operatorname{sgn}(f)^{-1}(-1)} + 1 \cdot \chi_{\operatorname{sgn}(f)^{-1}(1)} + 0 \cdot \chi_{\operatorname{sgn}(f)^{-1}(0)}.$$

□

Lemma 6.6 (Linear Combinations of Simple Functions)**.** *If $\phi_1, \ldots, \phi_n$ are simple functions and $\alpha_1, \ldots, \alpha_n \in \mathbb{R}$, then*

$$\sum_{i=1}^{n} \alpha_i \phi_i$$

is a simple function.

Proof. Follows directly from 4.5 and 4.11. □

Theorem 6.7. *Let (X, μ) be a finite measure space, and $f \in L^1(X)$ with bounded range of Y. Suppose Y is partitioned into measurable sets $Y_1, \ldots, Y_n$ such that*

$$\sup_{y \in Y_i} |y| - \inf_{y \in Y_i} |y| \leq c.$$

Then there exists a simple function ϕ such that

$$\|f - \phi\|_{L^1} \leq c\, \mu(X)$$

and

$$|\phi| < |f|.$$

Proof. Define the partition of X, $X_i := f^{-1}(Y_i)$ and the function

$$\phi := \sum_{i=1}^{n} \Big(\inf_{y \in Y_i} |y| \Big) \operatorname{sgn}(f)\, \chi_{X_i}.$$

For $x \in X_i$, if we factor out the signature we get

$$|f(x) - \phi(x)| = \Big| |f(x)| - \inf_{y \in Y_i} |y| \Big| \leq \sup_{y \in Y_i} |y| - \inf_{y \in Y_i} |y| \leq c.$$

Integrating over X gives

$$\|f - \phi\|_{L^1} = \sum_{i=1}^{n} \int_{X_i} |f - \phi| \, d\mu \leq \sum_{i=1}^{n} \int_{X_i} c \, d\mu = c\, \mu(X).$$

That ϕ is bounded by $|f|$ is clear from construction. □

Theorem 6.8. *Let (X, μ) be a σ-finite measure space and let $f \in L^1(X)$. For every $\varepsilon > 0$ there exists a simple function ϕ such that*

$$\|f - \phi\|_{L^1} < \varepsilon \quad \textit{and} \quad |\phi| \leq |f| \ \textit{a.e.}$$

Proof. We first assume $\mu(X) < \infty$.

Define the truncation

$$f_M := f\, \chi_{\{|f| \leq M\}}.$$

Then $|f_M| \leq |f|$ and

$$\|f - f_M\|_1 = \int |f - f\chi_{\{|f|\leq M\}}| = \int |f||1 - \chi_{\{|f|>M\}}| = \int |f|\chi_{\{|f|>M\}}$$

which goes to 0 by monotone convergence. Hence there exists $M > 0$ such that

$$\|f - f_M\|_1 < \varepsilon/2.$$

Since f_M has bounded range, Theorem 6.7 yields a simple function ϕ satisfying

$$\|f_M - \phi\|_{L^1} < \varepsilon/2 \quad \text{and} \quad |\phi| \leq |f_M|.$$

Because $|f_M| \leq |f|$, we obtain $|\phi| \leq |f|$. By the triangle inequality,

$$\|f - \phi\|_{L^1} \leq \|f - f_M\|_{L^1} + \|f_M - \phi\|_{L^1} < \varepsilon.$$

Now let (X, μ) be a σ-finite measure space. Let $A_1, A_2, \ldots$ be a sequence of sets that are finite and has union X. Now define $X_n := A_1 \cup A_2 \cup \ldots \cup A_n$ then $\mu(X_n) < \infty$ and

$$\lim_{n\to\infty} \int_{X\setminus X_n} |f| = 0.$$

Define $f^{(n)} := f\,\chi_{X_n}$. Then $f^{(n)} \in L^1(X_n)$, $f^{(n)} \in L^1(X)$ (when domains are appropriately adjusted), and we also have $|f^{(n)}| \leq |f|$, and $\|f - f^{(n)}\|_{L^1(X)} \to 0$.

Given $\varepsilon > 0$, choose n so that

$$\|f - f^{(n)}\|_{L^1(X)} < \varepsilon/2.$$

Applying the finite-measure case to $f^{(n)}$ on X_n, there exists a simple function ϕ such that

$$\|f^{(n)} - \phi\|_{L^1(X_n)} < \varepsilon/2 \quad \text{and} \quad |\phi| \leq |f^{(n)}| \leq |f|.$$

Again,

$$\begin{aligned}\|f - \phi\|_{L^1(X)} &\leq \|f - f^{(n)}\|_{L^1(X)} + \|f^{(n)} - \phi\|_{L^1(X)} \\ &\leq \|f - f^{(n)}\|_{L^1(X)} + \|f^{(n)} - \phi\|_{L^1(X_n)} \\ &< \varepsilon.\end{aligned}$$

This completes the proof. □

Corollary 6.9. *Let (X, μ) be a σ-finite measure space. The set of simple functions is dense in $L^1(X)$.*

This result is actually true for any measure space X. But from this point on general spaces will be assumed to be σ-finite if the result requires so.

At first glance, convergence in L^1 may appear to be a very strong notion, perhaps even stronger than almost everywhere convergence. However, neither mode of convergence implies the other.

Indeed, almost everywhere convergence does not imply L^1 convergence, as shown by the sequence

$$f_n(x) = n\,\chi_{(0,1/n)},$$

for which $f_n \to 0$ almost everywhere, while

$$\|f_n\|_{L^1} = 1 \quad \text{for all } n.$$

Conversely, L^1 convergence does not imply almost everywhere convergence. Consider defining

$$f_1 = \chi_{[0,1]}, \quad f_2 = \chi_{[0,1/2]}, \quad f_3 = \chi_{[1/2,1]}, \quad f_4 = \chi_{[0,1/4]},$$
$$f_5 = \chi_{[1/4,1/2]}, \quad f_6 = \chi_{[1/2,3/4]}, \quad f_7 = \chi_{[3/4,1]},$$

and in general,

$$f_n = \chi_{[j/2^k,\,(j+1)/2^k]} \quad \text{where } n = 2^k + j \text{ with } 0 \le j < 2^k.$$

Since the width always decreases, the L^1 norm tends to zero but since for each point in the domain it will revisit 1 infinitely many times , we can't say any point in the domain converges to 0 and thus we can't conclude almost everywhere convergence.

This example suggest the following phenomenon: although L^1 convergence does not force almost everywhere convergence of the whole sequence, the obstruction comes from values that keep reappearing at different points. This leads one to suspect that L^1 convergence should still guarantee the existence of a subsequence that converges almost everywhere. This is indeed the case.

Lemma 6.10. *If $f_n \to f$ in $L^1(X)$, then there exists a subsequence (f_{n_k}) such that*

$$f_{n_k}(x) \to f(x) \quad \textit{for almost every } x \in X.$$

Proof. Corollary 2.32 in [Fol99]. □

Corollary 6.11. *Let (X, μ) be a measure space and let $f \in L^1(X)$. Then there exists a sequence of simple functions that converges almost everywhere to f and is bounded by $|f|$.*

The last corollary will be very useful when combined with the Dominated Convergence Theorem. So far, the only direct way to evaluate integrals has been via the original supremum definition, but as seen previously suprema can be awkward to work with. Using these two results, if $f \in L^1$ we can now do the powerful move of reasoning

$$\int f d\mu = \int \lim_{n\to\infty} \sum_{j=1}^{N_n} z_j^n \chi_{E_j^n} = \lim_{n\to\infty} \int \sum_{j=1}^{N_n} z_j^n \chi_{E_j^n} = \lim_{n\to\infty} \sum_{j=1}^{N_n} z_j^n \mu(E_j^n).$$

7 Change of Measure via Densities

As a quick demonstration of the use of corollary 6.11 with the dominated convergence theorem we look at how to evaluate integrals with respect to measures that are themselves defined by an integral.

Definition 7.1 (Measure defined by a density)**.** Let μ be a measure and $f : \Omega \to [0, \infty]$ be in L^1. The measure ν defined by

$$\nu(E) := \int_E f \, d\mu, \quad \text{for all measurable } E,$$

is called a (finite) measure with density f. We write $d\nu = f \, d\mu$.

Theorem 7.2. *Let ν be a finite measure with density f with respect to μ, and let $g \in L^1(\nu)$ and assume $gf \in L^1(\mu)$, then*

$$\int g \, d\nu = \int gf \, d\mu.$$

Proof. Using dominated convergence and corollary 6.11 we have

$$\begin{aligned}
\int g \, d\nu &= \int \lim_{n\to\infty} \phi_n \, d\nu \\
&= \lim_{n\to\infty} \int \phi_n \, d\nu \\
&= \lim_{n\to\infty} \sum z_j^n \nu(E_j^n) \\
&= \lim_{n\to\infty} \sum z_j^n \int \chi_{E_j^n} f \, d\mu \\
&= \lim_{n\to\infty} \int \phi_n f \, d\mu \\
&= \int \lim_{n\to\infty} \phi_n f \, d\mu \\
&= \int gf \, d\mu.
\end{aligned}$$

The final application of the dominated convergence theorem is justified as follows. Let

$$N = \{x : |\phi_n(x)| > |g(x)|\}.$$

By construction from Corollary 6.11, we have $|\phi_n| \leq |g|$ ν-almost everywhere, and hence $\nu(N) = 0$. By definition of ν, this implies

$$\int_N f \, d\mu = 0,$$

and therefore $f = 0$ μ-almost everywhere on N.

On the complement N^c, we have $|\phi_n| \leq |g|$, which yields

$$|\phi_n f| \leq |gf|.$$

On N, since $f = 0$ μ-almost everywhere, the same inequality holds μ-almost everywhere as $0 \leq 0$. Consequently,

$$|\phi_n f| \leq |gf| \quad \mu\text{-almost everywhere on the whole space.}$$

□

8 Probability Theory expressed in measure theory

A large part of the framework of probability theory is based on events existing in a space and measuring the size of different sets of events. Thus it is no suprise that measure theory provides a convenient framework for probability theory.

Also the limiting theorems will be very useful in proving theorems such as the Central Limit Theorem as will be shown in the next chapter.

Definition 8.1 (Probability space)**.** Given a space (Ω, P), the measure P is called a probability measure if $P(\Omega) = 1$ and if that is the case the space is called a probability space. Given a probability space, P is called the distribution of the space.

Definition 8.2 (Random Variable)**.** A **random variable** X is a measurable function on a probability space to $(\mathbb{R}, m)$.

Definition 8.3 (Discrete Random Variable)**.** A **discrete random variable** is a random variable that is simple.

Definition 8.4 (Distribution of Random Variable)**.** Given a random variable, we define a measure P_X called the distribution of X defined as

$$P_X(E) := P(X^{-1}(E)) = \int \chi_E \, dP.$$

Corollary 8.5. *Given a random variable X on (Ω, P) to the measure space (Ω', P_X). Then (Ω', P_X) is also a probability space.*

Definition 8.6 (Distribution Function)**.** Given a probability space (Ω, P) where $\Omega = \mathbb{R}$, we define the distribution function

$$F_P(x) := P((-\infty, x]).$$

If the distribution is the distribution of a random variable X we also denote it as F_X.

Definition 8.7 (Probability Density Function)**.** Given a distribution has a differentiable distribution function, then $\frac{dF(x)}{dx}$ is called the probability density function and is denoted $f_X(x)$.

Definition 8.8 (Expectation)**.** Let X be a random variable on (Ω, P). We say X has an expected value if X is in $L^1(\Omega)$. We define the expectation of X, to be $\int X dP$.

Lemma 8.9 (Expectation with distribution function)**.** *Let X be a random variable with expectation and distribution P_X. Given $h \in L^1(\mathbb{R}, P_X)$ then*

$$E[h(X)] = \int h(X) dP = \int h(x) dP_X(x).$$

Proof. Since h exists in L^1 then there exists a sequence of simple functions ϕ_n converging to it and dominated by $|h|$. And thus we can use dominated convergence to reason as follows.

$$\begin{aligned}\int h(x)dP_X(x) &= \lim_{n\to\infty}\int \phi_n dP_X(x)\\ &= \lim_{n\to\infty}\sum_{j=1}^{n} a_j^n P_X(E_j^n)\\ &= \lim_{n\to\infty}\sum a_j^n P(X^{-1}(E_j^n))\\ &= \lim_{n\to\infty}\sum a_j^n \int \chi_{E_j^n}(X(x))dP(x)\\ &= \int \left(\lim_{n\to\infty}\sum a_j^n \chi_{E_j^n}(X(x))\right) dP(x)\\ &= \int \left(\lim_{n\to\infty}\phi_n(X(x))\right) dP(x)\\ &= \int h(X(x))dP(x).\end{aligned}$$

□

Definition 8.10 (Independence)**.** Random variables $X_1, \ldots, X_n$ are independent if for all measurable functions $h : \mathbb{R}^n \to \mathbb{R}$ such that the right-hand side integral exists and is finite,

$$\int h(X_1, \ldots, X_n)\, dP = \int \cdots \int h(x_1, \ldots, x_n)\, dP_{X_1}(x_1) \cdots dP_{X_n}(x_n).$$

Theorem 8.11 (Calculation of expectation)**.** *If X is a discrete random variable then*

$$E[X] = \sum x \cdot P_X(\{x\}),$$

where x ranges over the image of X.

If X has a probability density function f_X and $x \mapsto x$ has expectation in (Ω', P_X) then

$$E[X] = \int x \cdot f_X(x)dP(x).$$

Proof. Let X be discrete. Then the definition of the lebesgue integral gives

$$\begin{aligned}\int XdP &= \sum x \cdot P(X^{-1}(\{x\})\\ &= \sum x \cdot P_X(\{x\}).\end{aligned}$$

Let now X have a probability density function f_X and let $x \mapsto x$ be in $L^1(\Omega', P_X)$. By the fundamental theorem of calculus we have

$$\begin{aligned}\int_{-\infty}^{x} f_X(t)dt &= \int_{-\infty}^{x} \frac{dF_X}{dt}(t)dt \\ &= F_X(x) - \lim_{t\to-\infty} F(t) \\ &= F_X(x) - \lim_{t\to-\infty} P_X((-\infty, t]) \\ &= F_X(x).\end{aligned}$$

By 3.14 there exists a unique measure μ with the property that

$$\mu((a,b]) := \int_a^b f_X(t)dt.$$

And we can thus see that P_X must be that unique measure μ.

Thus by 4.13 $P_X(E) = \int \chi_E f_X dP$. Since $x \mapsto x$ is assumed to be in L^1 we get from dominated convergence and lemma 8.9 that

$$\begin{aligned}\int X dP &= \int x dP_X \\ &= \lim_{n\to\infty} \int \phi_n \, dP_X \\ &= \lim_{n\to\infty} \sum_j a_{jn} P_X(E_{jn}) \\ &= \lim_{n\to\infty} \sum_j a_{jn} \int \chi_{E_{jn}} f_X \, dP \\ &= \int \left(\lim_{n\to\infty} \sum_j a_{jn} \chi_{E_{jn}} \right) f_X \, dP \\ &= \int x f_X \, dP.\end{aligned}$$

□

9 The Central Limit Theorem

As a final application of the theory developed, the Central Limit Theorem will be proven using a bit of fourier analysis. The proof will closely follow *Real analysis: modern techniques and their applications* [Fol99] but adapted to fit the terminology and results already proven in this text.

Definition 9.1 (The Fourier Transform)**.** Let $f \in L^1(\mathbb{R})$, the Fourier transform of f is

$$\hat{f}(\xi) := \int f(x)e^{-2\pi i\xi x}dx,$$

and the inverse fourier transform of f as

$$f^{\vee}(x) := \hat{f}(-x) = \int f(\xi)e^{2\pi i\xi x}d\xi.$$

In the context of probability theory such as in *Probability: A Graduate Course* [Gut06], the fourier transform is called the characteristic function.

Definition 9.2 (Convolution)**.** Let f, g be two measurable functions on $\mathbb{R}$. The convolution $f * g$ is defined as

$$f * g(x) = \int f(x-y)g(y)dy.$$

Theorem 9.3 (Properties of convolution)**.** *Let $f, g, h \in L^1(\mathbb{R})$. Then:*

1. $f * g \in L^1(\mathbb{R})$.
2. $f * g = g * f$.
3. $(f * g) * h = f * (g * h)$.
4. $\|f * g\|_{L^1} \leq \|f\|_{L^1} \|g\|_{L^1}$.

Proof. Follows from proposition 8.6 and 8.7 in [Fol99]. □

Lemma 9.4 (Change of Variables for Linear Maps)**.** *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear map, and let f be in L^1. Then $f \circ T$ is in L^1 and*

$$\int f(x)\, dx = |\det T| \int f(Tx)\, dx.$$

Proof. See Theorem 2.44(a) in [Fol99]. □

Corollary 9.5. *Let $\phi \in L^1(\mathbb{R})$, and for $t > 0$ define*

$$\phi_t(x) := \frac{1}{t}\, \phi\Big(\frac{x}{t}\Big).$$

Then for all $t > 0$

$$\int \phi_t(x)\, dx = \int \phi(x)\, dx.$$

Lemma 9.6. *Let $f \in L^1(\mathbb{R})$. Then*

$$\lim_{t\to 0} \|f(x+t) - f(x)\|_{L^1} = 0.$$

Proof. See proposition 8.5 in [Fol99]. □

Theorem 9.7. *Suppose $\varphi, f \in L^1$ and $\int \varphi(x)\,dx = a$. Then $f * \varphi_t \to af$ in the L^1 norm as $t \to 0$.*

Proof. Let $t > 0$, then

$$\begin{aligned} f * \varphi_t(x) - af(x) &= \int [f(x-y) - f(x)]\varphi_t(y)\,dy \\ &= |\frac{1}{t}| \int [f(x - \frac{yt}{t}) - f(x)]\varphi(\frac{yt}{t})\,dy. \end{aligned}$$

By theorem 9.4 we get that

$$f * \varphi_t(x) - af(x) = \int [f(x - ty) - f(x)]\varphi(y)\,dy.$$

Apply absolute value and then integrate:

$$\|f * \varphi_t - af\|_{L^1} \le \int \|f(x-ty) - f(x)\|_{L^1} |\varphi(y)| dy.$$

The expression $\|f(x-ty) - f(x)\|_{L^1}$ is bounded by $2\|f\|_{L^1}$ by the triangle inequality and the fact that integration is invariant under transposition. Thus by dominated convergence and lemma 9.6 it follows that as t goes to zero, $\|f * \varphi_t - af\|_1$ also goes to zero.

□

Lemma 9.8. *If $f(x) = e^{-\pi a|x|^2}$ where $a > 0$, then*

$$\hat{f}(\xi) = a^{-1/2} e^{-\pi|\xi|^2/a}.$$

Proof. See proposition 8.24 in [Fol99]. □

Lemma 9.9. *Let $f \in L^1$. If $h(x) := e^{2\pi i\eta x} f(x)$ then $\hat{h}(\xi) = \hat{f}(\xi - \eta)$.*

Proof.

$$\hat{h}(\xi) = \int e^{-2\pi i\xi x} h(x) dx = \int e^{-2\pi i\xi x} e^{2\pi i\eta x} f(x) dx = \int e^{-2\pi i(\xi-\eta)x} f(x) dx = \hat{f}(\xi - \eta).$$

□

Theorem 9.10 (Fubini)**.** *Let $f : X \times Y \to \mathbb{C}$ be measurable. If*

$$\int \left(\int |f(x,y)|\, dy \right) dx < \infty,$$

then the following integrals exist and are equal:

$$\int \left(\int f(x,y)\, dy \right) dx = \int \left(\int f(x,y)\, dx \right) dy.$$

Proof. See Theorem 2.37 in [Fol99]. □

Theorem 9.11 (Fourier Inversion Theorem)**.** *If $f \in L^1$ and $\hat{f} \in L^1$, then $(\hat{f})^\vee = (\hat{f}^\vee) = f$ almost everywhere.*

Proof. Given $t > 0$ and $x \in \mathbb{R}$, set

$$\varphi(\xi) = \exp(2\pi i \xi \cdot x - \pi t^2 |\xi|^2).$$

By lemma 9.9 and lemma 9.8,

$$\hat{\varphi}(y) = t^{-1} \exp(-\pi |x-y|^2/t^2) = \phi_t(x-y),$$

where $\phi(x) = e^{-\pi|x|^2}$ and the subscript t has the meaning in corollary 9.5. Consider the following integral, and what it equals:

$$\begin{aligned}
\int e^{-\pi t^2 |\xi|^2} e^{2\pi i \xi \cdot x} \hat{f}(\xi)\, d\xi &= \int \hat{f}(\xi)\, \varphi(\xi)\, d\xi \\
&= \int \int f(y) e^{-2\pi i \xi y} dy\, \varphi(\xi)\, d\xi \\
&= \int \int f(y)\, \varphi(\xi) e^{-2\pi i \xi y} dy\, d\xi.
\end{aligned}$$

Since,

$$\int \int |f(y)\, \varphi(\xi) e^{-2\pi i \xi y}| dy\, d\xi = \int \int |f(y)|\, |\varphi(\xi)| dy\, d\xi = \|f\|_1 \|\varphi\|_1,$$

we can switch the order of integration with Fubini's theorem.

$$\begin{aligned}\int e^{-\pi t^2|\xi|^2} e^{2\pi i\xi\cdot x}\hat{f}(\xi)\,d\xi &= \int\int f(y)\,\varphi(\xi)e^{-2\pi i\xi y}dy\,d\xi \\ &= \int\int f(y)\,\varphi(\xi)e^{-2\pi i\xi y}d\xi dy \\ &= \int f(y)\int \varphi(\xi)e^{-2\pi i\xi y}d\xi dy \\ &= \int f(y)\,\hat{\varphi}(y)\,dy \\ &= \int f(y)\,\phi_t(x-y)\,dy \\ &= (f*\phi_t)(x).\end{aligned}$$

Since $\int e^{-\pi|x|^2}\,dx = 1$, by theorem 9.7 we have $f*\phi_t \to f$ in the L^1 norm as $t \to 0$. On the other hand, since $\hat{f} \in L^1$ and the exponential terms have absolute value less than 1 the dominated convergence theorem yields

$$\lim_{t\to 0}\int e^{-\pi t^2|\xi|^2} e^{2\pi i\xi\cdot x}\hat{f}(\xi)\,d\xi = \int e^{2\pi i\xi\cdot x}\hat{f}(\xi)\,d\xi = \hat{f}^\vee(x).$$

And we also have from lemma 9.4 that

$$\begin{aligned}\int e^{2\pi i\xi\cdot x}\hat{f}(\xi)\,d\xi &= \int e^{-2\pi i(-\xi)\cdot x}\hat{f}(-(-\xi))\,d\xi \\ &= \int e^{-2\pi i(-\xi)\cdot x} f^\vee(-\xi)\,d\xi \\ &= \int e^{-2\pi i\xi x} f^\vee(\xi)\,d\xi \\ &= (f^\vee)^\wedge(x).\end{aligned}$$

Thus we can conclude that $f = \hat{f}^\vee = (f^\vee)^\wedge$ almost everywhere. ☐

Definition 9.12 (Some spaces)**.** The set $C^\infty(\mathbb{R})$ will denote all smooth functions on $\mathbb{R}$ and $C_0(\mathbb{R})$ will denote the set of continuous vanishing functions on $\mathbb{R}$.

The *Schwartz space* $\mathcal{S}(\mathbb{R})$ will denote the set of all functions $f \in C^\infty(\mathbb{R})$ such that for all integers $k, m \geq 0$,

$$\sup_{x\in\mathbb{R}}\left|x^k\,\partial^m f(x)\right| < \infty.$$

Lemma 9.13. *If $f \in \mathcal{S}(\mathbb{R})$ then $\hat{f} \in \mathcal{S}(\mathbb{R})$*

Proof. See Corollary 8.23 in [Fol99]. ☐

Corollary 9.14. *If $f \in \mathcal{S}(\mathbb{R})$ then $f^\vee \in \mathcal{S}(\mathbb{R})$.*

Proof. If $f \in \mathcal{S}(\mathbb{R})$ then by the chain rule $g(x) = f(-x)$ is also in $\mathcal{S}(\mathbb{R})$ which means because of the previous result that $\hat{g}(\xi)$ is in $\mathcal{S}(\mathbb{R})$. But using lemma 9.4 we have that

$$\begin{aligned}\hat{g}(\xi) &= \int g(x)\, e^{-2\pi i x\xi} dx \\ &= \int f(-x) e^{2\pi i(-x)\xi} dx \\ &= \int f(x) e^{2\pi i x\xi} dx \\ &= f^\vee(\xi).\end{aligned}$$

Thus $f^\vee(\xi)$ is in $\mathcal{S}(\mathbb{R})$. □

Definition 9.15 (Vague convergence)**.** Let $\mu_1, \mu_2, \ldots$ and μ be measures on $\mathbb{R}$. We say $\mu_1, \mu_2, \ldots$ converges vaguely to μ if for all $f \in C_0(\mathbb{R})$ $\int f d\mu_1, \int f d\mu_2, \ldots$ converges to $\int f d\mu$.

Lemma 9.16. *Let $\mu_1, \mu_2, \ldots$ and μ be measures on $\mathbb{R}$ such that*

$$\sup \mu_n(\mathbb{R}) < \infty.$$

Let $D \subset C_0(\mathbb{R})$ be dense. If

$$\int f\, d\mu_n \to \int f\, d\mu \quad \textit{for all } f \in D,$$

then $\mu_n \to \mu$ vaguely.

Proof. See Proposition 5.17 in [Fol99]. □

Lemma 9.17. *The Schwartz space $\mathcal{S}(\mathbb{R})$ is dense in $C_0(\mathbb{R})$ and a subset of $L^1(\mathbb{R}, m)$.*

Proof. For density in $C_0(\mathbb{R})$ see proposition 8.17 in [Fol99].

To see it's a subset of $L^1(\mathbb{R}, m)$ let $f \in \mathcal{S}(\mathbb{R})$. Then the definition implies it is continuous, bounded and that there exists an M such that

$$|f(x)| \leq \frac{M}{x^2}.$$

Using the fact f is continuous and bounded as well as theorem 4.13 this implies $f \in L^1$ by

$$\begin{aligned}
\int |f(x)|dm &\leq \int_{|x|>1} |f(x)|dm + \int_{|x|\leq 1} |f(x)|dm \\
&\leq \int_{|x|>1} \frac{M}{x^2} dm + \int_{|x|\leq 1} |f(x)|dm \\
&= 2M \int_1^\infty \frac{1}{x^2} dx + \int_{-1}^1 |f(x)|dx \\
&= 2M + \int_{-1}^1 |f(x)|dx.
\end{aligned}$$

□

Corollary 9.18. *Let $P_1, P_2, \ldots$ and P be probability measures on $\mathbb{R}$. If*

$$\int f\, dP_n \to \int f\, dP \quad \textit{for all } f \in \mathcal{S}(\mathbb{R}),$$

then $P_n \to P$ vaguely.

Proof. Follows from lemma 9.16 and 9.17. □

Lemma 9.19. *Let $\mu_1, \mu_2, \ldots$ and μ be positive Lebesgue-Stieltjes measures on $\mathbb{R}$ with their respective distribution functions $F_1, F_2, \ldots$ and F. If $\mu_1, \mu_2, \ldots$ converges vaguely to μ then $F_1(x), F_2(x), \ldots$ converges to $F(x)$ for all x at which F is continuous.*

Proof. See proposition 7.19 in [Fol99]. □

Lemma 9.20. *Let $f_1, f_2, \ldots$ and f be in $L^1(\mathbb{R}, m)$. Define $F_k(x) = \int \chi_{(-\infty,x]} f_k$ and $F(x) = \int \chi_{(-\infty,x]} f$. Further assume that $\lim_{x\to\infty} F_k(x) = lim_{x\to\infty} F(x) = 1$ (that is they are distribution functions to probability measures).*

If $\hat{f}_1, \hat{f}_2, \ldots$ converges pointwise to $\hat{f}$ then $F_1(x), F_2(x), \ldots$ converges to $F(x)$ for all x at which F is continuous.

Proof. Define $\mu_k(E) = \int \chi_E f_k dm$ and $\mu(E) = \int \chi_E f dm$. Let $g \in \mathcal{S}(\mathbb{R})$, that means $|g|$ is bounded and thus in $L^1(\mu_k)$ for all k since

$$\int |g| d\mu_k \leq \int \sup_{x\in\mathbb{R}} |g(x)| d\mu_k = \mu_k(\mathbb{R}) \sup_{x\in\mathbb{R}} |g(x)| = \sup_{x\in\mathbb{R}} |g(x)|.$$

By the fourier inversion theorem we have that

$$\int g\,d\mu_k = \int\int g^\vee(y)e^{-2\pi iy\cdot x}\,dy\,d\mu_k(x).$$

Further by corollary 9.14 we have that $g^\vee \in \mathcal{S}(\mathbb{R})$ and thus also $g^\vee \in L^1(\mathbb{R}, m)$ by lemma 9.17. We can now see that the integrand is absolutely integrable by

$$\begin{aligned}\int\int |g^\vee(y)e^{-2\pi iy\cdot x}|\,dm(y)\,d\mu_k(x) &\leq \int\int |g^\vee(y)||e^{-2\pi iy\cdot x}|\,dm(y)\,d\mu_k(x)\\ &= \int\int |g^\vee(y)|\,dm(y)\,d\mu_k(x)\\ &= \int |g^\vee(y)|\,dm(y)\int\,d\mu_k(x)\\ &= \int |g^\vee(y)|\,dm(y).\end{aligned}$$

Thus we can use Fubini's theorem and get

$$\begin{aligned}\int g\,d\mu_k &= \int\int g^\vee(y)e^{-2\pi iy\cdot x}\,dy\,d\mu_k(x)\\ &= \int\int g^\vee(y)e^{-2\pi iy\cdot x}\,d\mu_k(x)\,dy\\ &= \int g^\vee(y)\int e^{-2\pi iy\cdot x}\,d\mu_k(x)\,dy.\end{aligned}$$

For all y and r the function $x \mapsto \chi_{(-r,r)e^{-2\pi iyx}}$ is clearly in $L^1(\mathbb{R}, m)$. By corollary 6.11 we can find a.e. bounded sequence of simple functions converging a.e. to it. Therefore we can express $e^{-2\pi iyx} = \lim_{r\to\infty}\lim_{n\to\infty}\sum a_j^n\chi_{E_j^n}$. Using dominated convergence we can continue reasoning that

$$\begin{aligned}\int g\,d\mu_k &= \int g^\vee(y)\int e^{-2\pi iy\cdot x}\,d\mu_k(x)\,dy\\ &= \int g^\vee(y)\int\left[\lim_{r\to\infty}\lim_{n\to\infty}\sum a_j^n\chi_{E_j^n}\right]\,d\mu_k(x)\,dy\\ &= \int g^\vee(y)\left[\lim_{r\to\infty}\lim_{n\to\infty}\sum a_j^n\mu_k(E_j^n)\right]\,dy\\ &= \int g^\vee(y)\left[\lim_{r\to\infty}\lim_{n\to\infty}\sum\int a_j^n\chi_{E_j^n}f_k\,dm\right]\,dy\\ &= \int g^\vee(y)\int e^{-2\pi iy\cdot x}f_k\,dm\,dy\\ &= \int g^\vee(y)\hat{f}_k(y)\,dy.\end{aligned}$$

By lemma 9.13 and 9.14 $g^\vee$ and $\hat{f}_k$ are in $\mathcal{S}(\mathbb{R})$ and thus bounded, which means that $g^\vee\hat{f}_k$ is bounded which implies by the dominated convergence theorem and our assumption that $\hat{f}_k \to \hat{f}$ pointwise that

$$\int g\,d\mu_k \to \int g^\vee(y)\hat{f}(y)dy.$$

Since our argument for

$$\int g\,d\mu_k = \int g^\vee(y)\hat{f}_k(y)\,dy$$

is still valid when replacing μ_k with μ and f_k with f we also have that

$$\int g\,d\mu = \int g^\vee(y)\hat{f}(y)\,dy.$$

We can now conclude that

$$\int g\,d\mu_k \to \int g\,d\mu.$$

Since we assumed g could be any function in $\mathcal{S}(\mathbb{R})$ it follows from corollary 9.18 that $\mu_k \to \mu$ vaguely. Which implies by lemma 9.19 that $F_k(x) \to F(x)$ for all x where F is continuous. □

Lemma 9.21. *Let $f \in L^1(\mathbb{R})$ and suppose that $x^n f(x) \in L^1(\mathbb{R})$ for positive integers $n \leq k$ where $k \geq 0$. Then the Fourier transform $\hat{f}$ is k times continuously differentiable, and*

$$\frac{d^k}{d\xi^k}\hat{f}(\xi) = [(-2\pi i x)^k f\,\widehat{]}.$$

Proof. See theorem 8.22d in [Fol99]. □

Lemma 9.22. *Let*

$$h(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}.$$

Then

$$\widehat{h}(\xi) = e^{-2\pi^2\xi^2}.$$

Proof. See Proposition 8.24 in [Fol99].

□

Theorem 9.23. *Let (Ω, P) be a probability space. For $i \geq 1$ let X_i be independent random variables on (Ω, P) such that there exists a $g \in L^1(m)$ so that for all i*

$$P_{X_i}(E) = \int \chi_E g\,dm.$$

Further assume for all $i \geq 1$ that

$$E[X_i^2] = 1, \quad E[X_i] = 0.$$

Define $Y_n = \frac{1}{\sqrt{n}}(X_1 + X_2 + \ldots + X_n)$. Assume for each Y_n there exists a function f_n such that

$$P_{Y_i}(E) = \int \chi_E f_n \, dm.$$

Then

$$\lim_{n\to\infty} \int_{-\infty}^{x} f_n dt = \int_{-\infty}^{x} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt.$$

Proof. Let P_X denote the distribution all P_{X_i} are equal to. Because of independence and theorem 7.2 we get

$$\begin{aligned}
\hat{f}_n(\xi) &= \int e^{-2\pi i \xi x} g \, dx \\
&= \int e^{-2\pi i \xi x} \, dP_{\frac{1}{\sqrt{n}}(X_1+X_2+\ldots+X_n)}(x) \\
&= \int e^{-2\pi i \xi \frac{1}{\sqrt{n}}(X_1+X_2+\ldots+X_n)} \, dP \\
&= \int e^{-2\pi i \xi \frac{1}{\sqrt{n}}(x_1+\ldots+x_n)} \, dP_{X_1}(x_1) \ldots dP_{X_n}(x_n) \\
&= \int e^{-2\pi i \xi \frac{1}{\sqrt{n}} x_1} \, dP_{X_1}(x_1) \cdot \ldots \cdot \int e^{-2\pi i \xi \frac{1}{\sqrt{n}} x_1} \, dP_{X_n}(x_n) \\
&= \left(\int e^{-2\pi i \frac{\xi}{\sqrt{n}} x} \, dP_X(x) \right)^n \\
&= \left(\int e^{-2\pi i \frac{\xi}{\sqrt{n}} x} g(x) \, dm(x) \right)^n \\
&= \left(\hat{g}(\frac{\xi}{\sqrt{n}}) \right)^n .
\end{aligned}$$

Because $E[X^2] = 1$ and $E[X] = 0$ we have that

$$\begin{aligned}
\int x g dm &= 0 \\
\int x^2 g dm &= 1,
\end{aligned}$$

which means by lemma 9.21 $\hat{f} \in C^2$. Further it implies that

$$\begin{aligned}
\hat{g}(0) &= \int f(x) e^{-2\pi i x \cdot 0} dm(x) = \int f dm = 1, \\
\hat{g}'(0) &= -2\pi i \int x f e^{-2\pi i x \cdot 0} dm(x) = -2\pi i \int x f dm = 0, \\
&\text{and} \\
\hat{g}''(0) &= -4\pi^2 \int x^2 f e^{-2\pi i x \cdot 0} dm(x) = -4\pi^2 \int x^2 f dm = -4\pi^2.
\end{aligned}$$

Thus by Taylor's theorem (see [PB10] for proof), we have

$$\hat{g}(\xi) = 1 - 2\pi^2 \xi^2 + o(\xi^2),$$

where $o(\alpha)$ denotes a quantity that satisfies $\alpha^{-1} o(\alpha) \to 0$ as $\alpha \to 0$. We thus get

$$\hat{f}_n(\xi) = \left[\hat{g}(n^{-1/2}\xi)\right]^n = \left[1 - 2\pi^2 \frac{\xi^2}{n} + o\left(\frac{\xi^2}{n}\right)\right]^n.$$

Since $\log(1+z) = z + o(z)$, we get

$$\log \hat{f}_n(\xi) = n \log \left[1 - 2\pi^2 \frac{\xi^2}{n} + o\left(\frac{\xi^2}{n}\right)\right] = -2\pi^2 \xi^2 + n \cdot o\left(\frac{\xi^2}{n}\right),$$

which tends to $-2\pi^2\xi^2$ as $n \to \infty$. In other words, $\hat{f}_n(\xi) \to e^{-2\pi^2\xi^2}$ as $n \to \infty$ for all ξ. From lemma 9.22 and lemma 9.20 it follows that

$$\lim_{n\to\infty} \int_{-\infty}^{x} f_n dt = \int_{-\infty}^{x} \frac{e^{t^2/2}}{\sqrt{2\pi}} dt.$$

□

References

[Car13] Constantin Carathéodory. *Vorlesungen über reelle Funktionen.* Springer-Verlag, 2013.

[Fol99] Gerald B Folland. *Real analysis: modern techniques and their applications*, volume 40. John Wiley & Sons, 1999.

[Gut06] Allan Gut. *Probability: a graduate course*, volume 200. Springer, 2006.

[KF60] AndreĭNikolaevich Kolmogorov and Sergei Vasil'evich Fomin. *Elementy teorii funktsii i funktsional'nogo analiza.* Moskovskii universitet, 1960.

[Leb02] Henri Lebesgue. Intégrale, longueur, aire. *Annali di Matematica Pura ed Applicata (1898-1922)*, 7(1):231–359, 1902.

[Mun00] James Munkres. *Topology James Munkres Second Edition.* Prentice Hall, 2000.

[PB10] Arne Persson and Lars-Christer Böiers. *Analys i en variabel.* Studentlitteratur AB, 2010.