MAT 425 Notes

Max Chien

Spring 2025

Contents

1 Introductory Measure Theory **3**
1.1 Motivations . . . 3
1.2 Abstract Measure Theory . . . 5
1.3 Measures and Integration . . . 11
1.4 Limit Theorems . . . 15

2 The Lebesgue Measure **20**
2.1 Premeasures and Outer Measures . . . 20
2.2 The Lebesgue Premeasure . . . 27
2.3 Regularity of Borel Measures . . . 30
2.4 Product Measures . . . 33
2.5 Fubini-Tonelli . . . 35

3 Change of Variables and the Radon-Nikodym Derivative **39**
3.1 The Change of Variables Formula . . . 39
3.2 L^p Spaces . . . 41
3.3 Lebesgue Decomposition and The Radon-Nikodym Derivative . . . 45
3.4 Complex Measures . . . 50

4 Lebesgue Differentiation **57**
4.1 Change of Variables in $\mathbb{R}^n$ . . . 61

5 Probability Theory **64**
5.1 Preliminaries . . . 64
5.2 Independence . . . 67
5.3 Asymptotic Analysis . . . 68
5.4 Large Deviations . . . 78
5.5 Kolmogorov Extension Theorem . . . 79

Definitions **83**

Introduction

This document contains notes taken for the class MAT 425: Integration Theory and Hilbert Spaces at Princeton University, taken in the Spring 2025 semester. These notes are primarily based on lectures by Professor Jacob Shapiro. Other references used in these notes include *Real Analysis* by Elias Stein and Rami Shakarchi, *Real and Complex Analysis* by Walter Rudin, *Real Analysis (2nd Edition)* by Halsey Royden, *The Elements of Integration and Lebesgue Measure* by Robert Bartle, *Measure Theory* by Paul Halmos, and *Real Analysis: Modern Techinques and Their Applications* by Gerald Folland. Since these notes were primarily taken live, they may contains typos or errors.

Chapter 1

Introductory Measure Theory

1.1 Motivations

The formal study of measure theory is motivated historically by the insufficiency of the Riemann integral as a complete tool for describing integration. Considering some bounded function $f : [a, b] \to \mathbb{R}$, there are many desirable properties that we might expect from an integral.

1. We might ask that the integral produces the average value of the function f on $[a, b]$, as
$$\overline{f} = \frac{1}{b-a} \int_a^b f$$
2. Geometrically, we can interpret the integral as the signed area between the graph of f and the x-axis:
$$A = \int_a^b f$$
3. We also think of integrals as the continuous generalization of summation.

Recall that the Riemann integral of f over $[a, b]$ is defined by considering, for fixed $N \in \mathbb{N}$, the upper and lower sums L_N, U_N defined by

$$L_N(f) = \frac{b-a}{N} \sum_{j=0}^{N-1} \inf \left\{ f(x) : x \in a + [n, n+1] \frac{b-a}{N} \right\}$$
$$U_N(f) = \frac{b-a}{N} \sum_{j=0}^{N-1} \sup \left\{ f(x) : x \in a + [n, n+1] \frac{b-a}{N} \right\}$$

We say that f is Riemann integrable with integral $I = \int_a^b f \in \mathbb{R}$ if $\lim L_N, \lim U_N$ both exist and are equal to I.

From our previous studies, Lebesgue's criterion gave a convenient characterization of Riemann integrable functions.

Definition 1.1

A set $S \subseteq \mathbb{R}$ has **measure zero** if for any $\varepsilon > 0$ there exists a collection $\{U_n\}_{n\in\mathbb{N}}$ of open intervals such that $S \subseteq \bigcup U_n$ and $\sum |U_n| < \varepsilon$, where $|U_n|$ is the length of U_n.

Example 1.1

The Cantor set $\mathcal{C}$ has measure zero. This is a consequence of the fact that at each iterative step in the construction of the Cantor set, we have a collection of open intervals covering the Cantor set, and the total length at step k is given by $\left(\frac{2}{3}\right)^k \to 0$.

Theorem 1.1: Lebesgue's Theorem

A bounded function $f : [a, b] \to \mathbb{R}$ is Riemann integrable if and only if the set of discontinuities of f has measure zero.

In particular, continuous functions are always Riemann integrable. The indicator function $\chi_{\mathcal{C}}$ of the Cantor set is Riemann integrable, since its discontinuities are of measure zero. However, $\chi_{\mathbb{Q}}$ (restricted to some compact interval) is not, since it is discontinuous at *every* point (this is precisely Dirichlet's function).

One can define a Riemann integral for unbounded functions or on unbounded domains by considering appropriate limits of Riemann integrals on compact intervals.

Example 1.2

The improper integral $\int_0^1 \frac{1}{\sqrt{x}}\,\mathrm{d}x$ is computed as

$$\int_{[0,1]} \frac{1}{\sqrt{x}}\,\mathrm{d}x = \lim_{n\to\infty} \int_{[\frac{1}{n},1]} \frac{1}{\sqrt{x}}\,\mathrm{d}x = \lim_{n\to\infty} 2\sqrt{x}|_{\frac{1}{n}}^{1} = \lim_{n\to\infty} \left[2 - \frac{2}{\sqrt{n}}\right] = 2$$

This method may be naturally extended to functions with a finite number of "integrable" discontinuities, or sometimes countable discontinuities. However, the following example shows that it fails in the general case.

Example 1.3

Let $\{\eta_n\}_{n\in\mathbb{N}}$ be an enumeration of the set $(0, 1) \cap \mathbb{Q}$. Define $f_n : [0, 1] \to \mathbb{R}$ by

$$f_n : x \mapsto \begin{cases} \frac{1}{\sqrt{x-\eta_n}} : & x > \eta_n \\ 0 : & x \leq \eta_n \end{cases}$$

Then define

$$f(x) := \sum_{n=1}^{\infty} 2^{-n} f_n(x)$$

By density, f is unbounded in every open subinterval of $[0,1]$. As a result, there is no limit of intervals increasing to $[0,1]$ which we could use to define the integral of f over $[0,1]$, in the sense used in the previous example.

To try to figure out a way around this, note that our work in the previous example shows that

$$\int_{[0,1]} f_n = 2\sqrt{1-\eta_n}$$

Now, consider the (unjustified) interchange of the integral and sum:

$$\int_{[0,1]} f = \int_{[0,1]} \sum_{n=1}^{\infty} 2^{-n} f_n \longrightarrow \sum_{n=1}^{\infty} 2^{-n} \int_{[0,1]} f_n = \sum_{n=1}^{\infty} 2^{-n} 2\sqrt{1-\eta_n} < \infty$$

As the above example demonstrates, an important question in analysis is which operations respect the limiting process. In particular, we know that uniform convergence respects the limit:

Theorem 1.2

Let $f_n : [a,b] \to \mathbb{R}$ be a sequence of bounded Riemann integrable functions which converge uniformly to f. Then f is Riemann integrable and $\lim \int_{[a,b]} f_n = \int_{[a,b]} f$.

However, it is desirable to us to apply this interchange under weaker hypotheses than uniform convergence, so that we can develop a more powerful and general theory of integration.

Example 1.4

Consider again the enumeration $\{\eta_n\}_{n\in\mathbb{N}}$ of $(0,1)\cap\mathbb{Q}$. Define

$$f_n \coloneqq \chi_{\{\eta_j : j\in[1,n]\}}$$

In words, $f_n(x) = 1$ if $x = \eta_j$ for some $j \leq n$ and 0 otherwise. $\int_{[0,1]} f_n = 0$ for all n, so we would like to assign the value 0 to $\int_{[0,1]} \lim f$. However, observe that f_n converges pointwise to Dirichlet's function, which is not Riemann integrable.

The development of the Lebesgue integral, which solves many issues with the Riemann integral, will be accomplished by first discussing the general theory of measure and integration, and following the construction of the Lebesgue measure and integral.

1.2 Abstract Measure Theory

The development of a measure space structure on a set is accomplished by defining a collection of "measurable" subsets, not unlike a topology, which satisfies particular structural constraints.

Definition 1.2

Let X be a set, and consider a collection of subsets $\mathcal{M} \subseteq \mathcal{P}(X)$. We say that $\mathcal{M}$ is a **σ-algebra** on X if

1. $X \in \mathcal{M}$,
2. If $A \in \mathcal{M}$ then $X \setminus A \in \mathcal{M}$,
3. If $\{A_n\}_{n\in\mathbb{N}}$ is a countable collection of elements of $\mathcal{M}$, then $\bigcup A_n \in \mathcal{M}$.

If $\mathcal{M}$ is a σ-algebra on X, then $(X, \mathcal{M})$ is called a **measurable space**. An element of $\mathcal{M}$ is called a **measurable set**. If the σ-algebra on X is understood by context, then $\operatorname{Meas}(X)$ denotes the set of measurable subsets of X (that it, it denotes the implied σ-algebra).

Notice that while a topology is required to be closed under arbitrary unions, a σ-algebra is only required to be closed under countable unions. Moreover, the following follows directly from the axioms of σ-algebras:

Proposition 1.3

$\varnothing \in \operatorname{Meas}(X)$ and $\operatorname{Meas}(X)$ is clsoed under countable intersections.

For comparison, recall the following definition of a topology:

Definition 1.3

Let X be a set, and consider a collection of subsets $\mathcal{T} \subseteq \mathcal{P}(X)$. We say that $\mathcal{T}$ is a **topology** on X if

1. $X, \varnothing \in \mathcal{T}$,
2. $\bigcap_{n=1}^{N} V_n \in \mathcal{T}$ whenever each $V_n \in \mathcal{T}$,
3. $\bigcup_{\alpha\in A} V_\alpha \in \mathcal{T}$ whenever $V_\alpha \in \mathcal{T}$ for an arbitrary indexing set A.

By direct comparison, a topology is not automatically a σ-algebra, since it may not be closed under complements.

Again in analogy to topology, recall that continuous functions are the morphisms of topological spaces. Thus, we can ask which functions can be considered to be the morphisms of measurable spaces. Indeed, just as continuous functions are topologically characterized by preserving open sets under preimages, we define measurable space morphisms similarly:

Definition 1.4

A function $f : X \to Y$ for measurable spaces X, Y is said to be a **measurable function** if $f^{-1}(A) \in \operatorname{Meas}(X)$ whenever $A \in \operatorname{Meas}(Y)$.

It follows immediately that the composition of measurable functions is measurable.

As with topologies, any set automatically comes equipped with two σ-algebras: the power set $\mathcal{P}(X)$ and $\{\varnothing, X\}$. These are the largest and smallest σ-algebras on X, respectively.

Example 1.5

Let $X = \{1, 2, 3, 4\}$. Then the following is a nontrivial σ-algebra:

$$\mathcal{M} = \{\varnothing, X, \{1, 2\}, \{3, 4\}\}$$

Generalizing the above, for any $A \subseteq X$, the σ-algebra $\{\varnothing, X, A, X \setminus A\}$ is the smallest σ-algebra containing A.

Remark 1.1

The arbitrary intersection of σ-algebras on a common set is again a σ-algebra, but not necessarily unions.

Definition 1.5

Let $f : X \to Y$, where X is an arbitrary set and Y is a measurable space. Then the σ-algebra $\sigma(f)$ **generated** by f is

$$\sigma(f) \coloneqq \left\{ f^{-1}(A) : A \in \operatorname{Meas}(Y) \right\}$$

Essentially, $\sigma(f)$ is generated by pulling back the measurable structure of Y through f. It is straightforward to verify that $\sigma(f)$ is actually a σ-algebra, and it follows that $\sigma(f)$ is the smallest σ-algebra on X such that f is measurable. In other words, if $\mathcal{M}$ is a σ-algebra on X, then f is measurable with respect to $(X, \mathcal{M}), Y$ if and only if $\sigma(f) \subseteq \mathcal{M}$.

We can generalize the construction of "smallest σ-algebra" type constructions to find the smallest σ-algebra containing a certain collection of subsets. It is somewhat nonobvious that such an algebra exists or is unique.

Theorem 1.4

Let $\mathcal{F} \subseteq \mathcal{P}(X)$. Then there exists a unique minimal σ-algebra $\sigma(\mathcal{F})$ on X such that $\mathcal{F} \subseteq \sigma(\mathcal{F})$.

Proof. Let Ω be the set of collection of all σ-algebras on X which contain $\mathcal{F}$. Ω is nonempty since $\mathcal{P}(X) \subseteq \Omega$. Define

$$\sigma(\mathcal{F}) = \bigcap_{\mathcal{M} \in \Omega} \mathcal{M}$$

Since the arbitrary intersection of σ-algebras is a σ-algebra, $\sigma(\mathcal{F})$ is indeed a σ-algebra. Moreover, by construction $\sigma(\mathcal{F})$ is contained in any element of Ω, and it is thus minimal. $\square$

As we remarked above, a topology is not in general a σ-algebra. The two notions are linked by considering the Borel σ-algebra, which is generated by the open sets on a space.

Definition 1.6

Let X be a topological space with topology $\mathcal{T}$. Then the **Borel σ-algebra** on X is given by $\mathscr{B}(X) = \sigma(\mathcal{T})$.

Note that since σ-algebras are closed under complements, by definition the closed sets on X are in $\mathscr{B}(X)$. It is also the case that countable intersections of open sets and countable unions of closed sets are contained in $\mathscr{B}(X)$, when this is not necessarily true in $\mathcal{T}$. Elements of a Borel σ-algebra are called **Borel sets**. In general, when we refer to topological spaces without specifying a σ-algebra, the Borel algebra is implicitly taken.

Under Hausdorff's terminology, sets which are the countable union of closed sets are denoted F_σ sets. Analogously, sets which are the countable intersection of open sets are denoted G_δ sets.

To make more precise the connection between topologies and measurable spaces through the Borel σ-algebra, we make the following claim:

Proposition 1.5

Let $f : X \to Y$ be a mapping between topological spaces such that $f^{-1}(V) \in \mathscr{B}(X)$ for any open set $V \subseteq Y$. Then f is measurable with respect to $\mathscr{B}(X), \mathscr{B}(Y)$.

Proof. Define the collection

$$\mathcal{M} = \left\{ A \in \mathcal{P}(Y) : f^{-1}(A) \in \mathscr{B}(X) \right\}$$

It can be verified that $\mathcal{M}$ is a σ-algebra on Y. Then, by assumption the open sets in Y are contained in $\mathcal{M}$. Moreover, by definition $\mathscr{B}(Y)$ is the smallest σ-algebra containing the open sets. Therefore we have $\mathrm{Open}(Y) \subseteq \mathscr{B}(Y) \subseteq \mathcal{M}$. Since $\mathscr{B}(Y)$ is contained in $\mathcal{M}$ it follows by definition that f is measurable with respect to $\mathscr{B}(X), \mathscr{B}(Y)$. $\square$

Note that the above proposition implies that any continuous mapping between topological spaces is measurable with respect to their Borel algebras. We prove the following statement, which will aid our understanding of complex measurable functions:

Proposition 1.6

Let X be a measurable space and Y a topological space. Let $u, v : X \to \mathbb{R}$ be measurable and $\varphi : \mathbb{R}^2 \to Y$ be continuous. Then $h : X \to Y$ defined by

$$h(x) = \varphi(u(x), v(x))$$

is measurable with respect to $\mathrm{Meas}(X), \mathscr{B}(Y)$.

Proof. From the previous proposition, φ is measurable with respect to $\mathscr{B}(\mathbb{R})$ and $\mathscr{B}(Y)$. Let $f : X \to \mathbb{R}^2$ be $x \mapsto (u(x), v(x))$. Then $h = \varphi \circ f$, and the composition of measurable functions is measurable. So it suffices to show f is measurable with respect to $\text{Meas}(X), \mathscr{B}(\mathbb{R})$.

Take some rectangle $R = I_1 \times I_2$ for intervals I_1, I_2. Then $f^{-1}(\mathbb{R}) = u^{-1}(I_1) \cap v^1(I_2)$. $f^{-1}(\mathbb{R})$ is then a measurable set since both u, v are measurable functions. Now, let $V \in \text{Open}(\mathbb{R}^2)$. Then V can be written as the countable union of rectangles. So we have

$$f^{-1}(V) = f^{-1}\left(\bigcup_{n=1}^{\infty} R_n\right) = \bigcup_{n=1}^{\infty} f^{-1}(R_n) \in \text{Meas}(X)$$

From the previous proposition it follows that f is measurable. $\square$

We can now use this fact to produce measurable functions from other measurable functions.

Theorem 1.7

Let X be a measurable space. Then:

1. If $u, v : X \to \mathbb{R}$ are measurable, then so is $u + iv : X \to \mathbb{C}$.
2. If $f : X \to \mathbb{C}$ is measurable, then so are $\text{Re}(f), \text{Im}(f), |f|$.
3. If $f, g : X \to \mathbb{C}$ are measurable then $f + g$ and fg are both measurable.
4. If $A \in \text{Meas}(X)$ then $\chi_A : X \to \mathbb{R}$ is measurable as well.
5. If $f : X \to \mathbb{C}$ is measurable then there exists $\alpha : X \to \mathbb{C}$ measurable such that $f = \alpha|f|$.

It is often of interest to us to work in the extended real line, so that we can consider functions or measures which assign infinite values to some points or sets. This is also helpful since the extended real line is compact.

Definition 1.7

The **extended real line** is denoted $[-\infty, \infty]$ or $\overline{\mathbb{R}}$, and consists of the set $\mathbb{R} \cup \{\pm\infty\}$, together with the topology that contains open sets in $\mathbb{R}$ and countable unions of sets of the form $(a, \infty]$ and $[-\infty, a)$.

Theorem 1.8

Let $f : X \to \overline{\mathbb{R}}$ with X a measurable space. If

$$f^{-1}((a, \infty]) \in \text{Meas}(X)$$

for all $a \in \mathbb{R}$, then f is measurable.

Proof. The point is to show that any open set in $\overline{\mathbb{R}}$ may be constructed countably from sets of the form $(a, \infty]$.

First we consider sets of the form $[-\infty, a)$. Let $\{a_n\} \to a$ be a sequence of points with $a_n < a$ for all a_n. Then

$$[-\infty, a) = \bigcup_{n=1}^{\infty} [-\infty, a_n] = \bigcup_{n=1}^{\infty} (a_n, \infty]^c$$

so $f^{-1}([-\infty, a)) \in \mathrm{Meas}(X)$. We can similarly write

$$(a, b) = [-\infty, b) \cap (a, \infty]$$

so that $f^{-1}((a, b)) \in \mathrm{Meas}(X)$ as well. Now it follows that any open set in the topology on $\overline{\mathbb{R}}$ has a preimage in $\mathrm{Meas}(X)$, so it follows that f is measurable with respect to the Borel algebra on $\overline{\mathbb{R}}$. □

Theorem 1.9

Let $f_n : X \to \overline{\mathbb{R}}$ be a sequence of measurable functions. Then the functions $\sup f_n, \limsup f_n, \inf f_n, \liminf f_n$, which are defined pointwise, are all measurable.

Proof. By the previous theorem, it suffices to check that $(\sup f_n)^{-1}((a, \infty])$ is measurable for all $a \in \mathbb{R}$, which we will do by expressing these sets as countable unions of preimages through the individual f_n.

We claim that

$$(\sup f_n)^{-1}((a, \infty]) = \bigcup_{n=1}^{\infty} f_n^{-1}((a, \infty])$$

While this is not true in general, it holds for the half-open infinite intervals. We show double inclusion.

($\subseteq$) Let $x \in (\sup f_n)^{-1}((a, \infty])$. Then $\sup f_n(x) > a$. Thus there exists n such that $f_n(x) > \sup f_n - \varepsilon$ for ε sufficiently small that $\sup f_n - \varepsilon > a$. So $x \in f_n^{-1}(a, \infty]$.

($\supseteq$) Similarly, if $x \in \bigcup_{n=1}^{\infty} f_n^{-1}((a, \infty])$, then there exists n with $f_n(x) > a$, which then implies that $\sup f_n(x) > a$ as well.

By hypothesis, $f_n^{-1}((a, \infty]) \in \mathrm{Meas}(X)$ for all n. Thus $\sup f_n$ is measurable. Of course this is true for inf as well.

To show that lim sup is measurable as well, we simply use the representation of lim sup as

$$\limsup a_n = \inf_{n \geq 1} \left(\sup_{m \geq n} a_m \right)$$

Thus $\limsup f_n$ and $\liminf f_n$ are both measurable as well. □

Corollary 1.10

If $\lim f_n$ exists and each $f_n : X \to \overline{\mathbb{R}}$ is measurable, then so is $\lim f_n$.

Proof. If the limit exists then it is equal to both the lim sup and lim inf. □

Corollary 1.11

If $f, g : X \to \overline{\mathbb{R}}$ are measurable then so is $\max\{f, g\}$ and $\min\{f, g\}$.

Proof. Define $f_1 = f$ and $f_n = g$ for all $n \geq 2$. □

The following theorem, which is a direct result of the above, is useful for considering an arbitrary function in terms of two nonnegative functions, which are easier to work with.

Proposition 1.12

For any $f : X \to \overline{\mathbb{R}}$, we can decompose it into positive and negative parts as $f = f^+ - f^-$, with

$$f^+ := \max\{f, 0\}$$
$$f^- := -\min\{f, 0\}$$

If f is measurable then so are f^+, f^-.

Proof. Based on the previous theorems, we just need to show that the constant zero function is measurable. But this is clear since the preimage of any subset of $\overline{\mathbb{R}}$ will be all of X if the subset contains 0, and $\varnothing$ otherwise. □

1.3 Measures and Integration

Our next goal is to define integration of measurable functions. To do so, we will first consider simple functions, which will be the smallest building blocks that we define an integral on.

Definition 1.8

A function $s : X \to \mathbb{C}$ is a **simple function** if it has finite image. A simple nonnegative function is a simple function $s : X \to [0, \infty)$.

Because a simple function s assumes only finitely many values, we can always express it as the weighted sum of characteristic functions:

$$s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$$

where the α_i are the values in the image, and the A_i are their preimages.

Proposition 1.13

A simple function expressed as

$$s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$$

is measurable if and only if each A_i is measurable.

Proposition 1.14

Products and sums of simple functions are simple.

Proof. Clearly there are only finitely many values in the image. □

The utility of simple functions is that we may use them to approximate arbitrary measurable functions. Thus, so long as our integral operator interchanges with limits, we will be free to define integrals solely over simple functions.

Theorem 1.15

Let $f : X \to [0, \infty]$ be measurable. Then there exists a sequence of simple nonnegative measurable functions $s_n : X \to [0, \infty)$ such that:

- $0 \leq s_1 \leq s_2 \leq \ldots \leq f$.
- $s_n \to f$ pointwise.

Proof. We first provide an approximation for the identity, and then compose this with our function f. This approximation is made easier since we only need a pointwise limit. Thus we can consider a step function which both has finer steps (in order to approach the identity), and approximates the identity on a larger range (so that at there are always finite points in the range). Thus, we define

$$\varphi_n(t) = \begin{cases} 2^{-n}\lfloor 2^n t \rfloor, & 0 \leq t < n \\ n, & t \geq n \end{cases}$$

φ_n is simple since it has $\sim 2^{-n}$ values in its image. Additionally its preimages are half-open intervals so φ_n is measurable.

Now, we need to show that $\varphi_n \leq \varphi_{n+1}$ and φ_n converges to the identity pointwise. To show this, we prove that $t - 2^{-n} < \varphi_n(t) \leq t$ for all t. Then it is clear that as $n \to \infty$, φ_n approaches the identity.

Now, the conclusion to the proof is to set $s_n \coloneqq \varphi_n \circ f$. s_n is simple since we factor through the simple function φ_n, and it is measurable as the composition of measurable functions. □

Now, we have established the technical background to define integration of simple functions. To do this, we essentially just assign each possible preimage of the simple functions a weight (which must be additive). Such a weight is a generalization of the notions of area, volume, mass, and so on, and is called a measure. We make two slightly different definitions for real and complex measures.

Definition 1.9

A **complex measure** on X is a function $\mu : \mathrm{Meas}(X) \to \mathbb{C}$ which is countably additive, meaning that whenever $\{A_n\}$ is a countable sequence of pairwise disjoint measurable sets, we have

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

Definition 1.10

A **nonnegative measure** on X is a map $\mu : \mathrm{Meas}(X) \to [0, \infty]$ which is countably additive, and such that there is at least one set $A \in \mathrm{Meas}(X)$ with finite measure. A **measure space** is a triple $(X, \mathcal{M}, \mu)$ where $(X, \mathcal{M})$ is a mesaurable space and μ is a measure on $(X, \mathcal{M})$.

Note it follows that $\mu(\varnothing) = 0$, which would not be true in the nonnegative case if we did not require the existence of a finite measure set.

Proposition 1.16

If μ is a nonnegative meausre on X, then:

1. $\mu(\varnothing) = 0$.
2. μ is finitely additive.
3. If $A \subseteq B$ then $\mu(A) \leq \mu(B)$.
4. If $A_1 \subseteq A_2 \subseteq \ldots$ is a countable sequence of measurable sets, then
$$\lim \mu(A_n) = \mu\left(\bigcup A_n\right)$$
5. If $A_1 \supseteq A_2 \supseteq \ldots$ is a countable sequence of measurable sets and at least one A_n has finite measure, then
$$\lim \mu(A_n) = \mu\left(\bigcap A_n\right)$$

Roughly speaking, (4) and (5) tell us that measures may be approximated from either inside or outside.

Proof. 1. Take A measurable with finite measure, and consider the sequence $A_1 = A$, $A_n = \varnothing$ For $n \geq 2$. Then

$$\infty > \mu(A) = \mu\left(\bigcup A_n\right) = \sum \mu(A_n) = \mu(A) + \sum \mu(\varnothing)$$

which implies that we must have $\mu(\varnothing) = 0$.

2. Follows from countable additivity now that we know $\mu(\varnothing) = 0$.

3. We write $B = A \sqcup (B \setminus A)$ and apply (2).

4. Define $B_1 = A_1$, $B_2 = A_2 \setminus A_1$, and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. Then apply countable additivity.

5. EXERCISE □

The most important example of a nonnegative measure is the Lebesgue measure. Because it is harder to define, we start by defining a few simpler measures.

Definition 1.11

Let X be a measurable space with $\operatorname{Meas}(X) = \mathcal{P}(X)$. The **counting measure** is defined as $c : \operatorname{Meas}(X) \to [0, \infty]$ such that $c(A)$ is the cardinality of A (possibly infinite).

Definition 1.12

Let X be a measurable space with $\operatorname{Meas}(X) = \{\varnothing, X, \{x_0\}, X \setminus \{x_0\}\}$ for some distinguished point x_0. Then the **Dirac delta measure** at x_0 is defined by

$$S \mapsto \begin{cases} 1, & x_0 \in S \\ 0, & x_0 \notin S \end{cases}$$

We can now define the integral of a positive function against a measure. We will do so by first defining the integral of simple functions, then passing to the limit.

Definition 1.13

Let $\mu : \operatorname{Meas}(X) \to [0, \infty]$ be a nonnegative measure. Let $s = \sum_{i=1}^{n} \alpha_i \chi_{A_i}$ be a simple measurable function, and let $E \in \operatorname{Meas}(X)$. Then we define the **Lebesgue integral** of s over E with respect to μ to be

$$\int_E s \, \mathrm{d}\mu := \sum_{i=1}^{n} \alpha_i \mu(A_i \cap E)$$

By convention, if $\alpha_i = 0$ on a set of infinte measure, the entire term is considered to be zero.

Definition 1.14

Let $f : X \to [0, \infty]$ be measurable, and let $\mu : \mathrm{Meas}(X) \to [0, \infty]$ be a nonnegative measure. Let $E \in \mathrm{Meas}(X)$. Then the **Lebesgue integral** of f over E with respect to μ is

$$\int_E f \, \mathrm{d}\mu := \sup_{0 \le s \le f} \int_E s \, \mathrm{d}\mu$$

where the supremum is taken over all nonnegative simple measurable functions which satisfy $0 \le s \le f$.

Note that the second definition agrees with the first since the supremum is attained by f.

Example 1.6

Set $X = \mathbb{N}$, $\mathrm{Meas}(X) = \mathcal{P}(X)$, and c to be the counting measure on X. Then

$$\int_A f \, \mathrm{d}c = \sum_{x \in A} f(x)$$

when $A \subseteq \mathbb{N}$. This is clear for finite A but requires limit theorems for countable A. Thus we have represented the sum as an integral against the counting measurable, meaning that our integral theorems will apply to sums as well.

1.4 Limit Theorems

We now turn to the question of interchanging the limit operator and integral, which is a major motivation for the definition of the integral in this way. We begin first with a few elementary properties.

Proposition 1.17

Let $0 \le f \le g$ be nonnegative measurable functions. Then:

1. $\int f \le \int g$.
2. If $A \subseteq B$ then $\int_A f \le \int_B f$.
3. If $0 \le c < \infty$, then $\int cf = c \int f$.
4. If $f \equiv 0$ then $\int_E f = 0$ for any measurable E, even if E has infinite measure.
5. If E is measurable with $\mu(E) = 0$, then $\int_E f = 0$.
6. For E measurable, $\int_E f = \int \chi_E f$.

Theorem 1.18

Let $s, t \geq 0$ be nonnegative simple functions and μ a measure. Define

$$\varphi_s(E) = \int_E s \, \mathrm{d}\mu$$

Then φ_s is a measure, and $\varphi_{s+t} = \varphi_s + \varphi_t$.

Proof. Let $E = \bigsqcup E_i$ be the disjoint countable union of some E_i. By definition,

$$\varphi_s(E) = \sum_{i=1}^{n} \alpha_i \mu(E \cap A_i) = \sum_{i=1}^{n} \alpha_i \sum_{j=1}^{\infty} \mu(E_j \cap A_i)$$

Because s is simple we can interchange the finite sum:

$$\sum_{i=1}^{n} \alpha_i \sum_{j=1}^{\infty} \mu(E_j \cap A_i) = \sum_{j=1}^{\infty} \sum_{i=1}^{n} \alpha_i \mu\left(E_j \cap E_i\right) = \sum_{j=1}^{\infty} \varphi_s(E_j)$$

Thus φ_s is a measure. Linearity follows since we are only adding two simple functions, and so there are at most finitely many sets to work with. $\square$

Example 1.7

To give an example of a sequence where the limit and integral cannot be interchanged, define $f_n = n\chi_{(0,1/n)}$. Then $\int f_n = 1$ for all n, but the pointwise limit is 0 everywhere.

We now prove our first limit theorem:

Theorem 1.19: Monotone Convergence Theorem

Let $0 \leq f_n \nearrow f \leq \infty$ be a sequence of nonnegative measurable functions. Then f is measurable and

$$\int f_n \to \int f$$

Proof. First note that the sequence $\int f_n$ is monotone increasing, so it has a limit (in the extended reals). Thus we have

$$L = \lim \int f_n \leq \int f$$

Pick a simple function $s \leq f$ and $\varepsilon < 1$. We want to show that $L \geq \varepsilon \int s$, which will then prove the result by taking $\varepsilon \to 1$ and $s \to f$.

For each n, define

$$E_n = \{x : f_n(x) \geq \varepsilon s(x)\}$$

For any point $x \in X$, we have $f_n(x) \to f(x) > \varepsilon s(x)$, so

$$\bigcup E_n = X$$

Then for each n we have

$$\int_{E_n} \varepsilon s \leq \int_{E_n} f_n \leq \int_X f_n \to L$$

We also have

$$\int_{E_n} \varepsilon s \to \int_X \varepsilon s$$

so

$$\int \varepsilon s \leq L$$

for all $\varepsilon < 1, s \leq f$. Thus

$$\int f \leq L$$

so we have both inequalities and thus

$$\int f = L = \lim \int f_n$$

□

Corollary 1.20

If f, g are nonnegative and measurable then $\int f + g = \int f + \int g$.

Proof. Take two sequences of simple functions $s_i \nearrow f$ and $t_i \nearrow g$. The monotone convergence theorem gives the result. □

Corollary 1.21

If $f_n \geq 0$ is a sequence of nonnegative measurable functions then

$$\int \sum_{n=1}^{\infty} f_n(x) = \sum_{n=1}^{\infty} \int f_n(x)$$

Proof. Combine the monotone convergence theorem with the previous corollary. □

Corollary 1.22

If a_{ij} is a sequence of nonnegative numbers then

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} a_{ij} = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} a_{ij}$$

Proof. We write one of the sums as an integral with the counting measure. □

Lemma 1.23: Fatou's Lemma

Let $f_n \geq 0$ be a sequence of nonnegative measurable functions. Then

$$\int \liminf f_n \leq \liminf \int f_n$$

Proof. Define $g_n(x) = \inf_{m\geq n} f_m(x)$. Then by definition, $g_n \nearrow \liminf f_n$. Also $\int g_n \leq \int f_n$ for each n. So by monotone convergence we have

$$\int \liminf f_n = \lim \int g_n = \liminf \int g_n \leq \liminf \int f_n \qquad \square$$

Having established limit theorems for nonnegative functions, we now make our definition of arbitrary integrals.

Definition 1.15

Let $f : X \to \overline{\mathbb{R}}$ be a measurable function. Writing $f = f^+ - f^-$, we define

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

For a complex measurable function $F = u + iv : X \to \mathbb{C}$, we define

$$\int F = \int u + i \int v$$

Clearly this definition agrees with our previous one. However, there is a slight subtlety, which is that our definition may end up with an expression like $\infty - i\infty$. As such, we restrict this definition to those f which make the integral absolutely convergent (meaning $\int |f| < \infty$).

Proposition 1.24

For f measurable,

$$\left| \int f \right| \leq \int |f|$$

Proof. For f real valued, we write

$$\left| \int f \right| = \left| \int f^+ - \int f^- \right| \leq \left| \int f^+ \right| + \left| \int f^- \right| = \int f^+ + \int f^- = \int |f| \qquad \square$$

A similar proof shows the result for complex functions.

Our new integral inherits the properties we have shown for integrals of nonnegative functions, assuming the limits are finite. To capture this we make the following classification:

Definition 1.16

Let μ be a measure on X. Then we define the L^1 **space** to be

$$L^1(\mu) = \left\{ f : X \to \mathbb{C} : \int |f| \, \mathrm{d}\mu < \infty \right\}$$

Theorem 1.25: Dominated Convergence Theorem

If $f_n \to f$ and there exists $g \in L^1$ such that $|f_n| \le g$, then:

- $f_n \in L^1$,
- $\lim \int |f - f_n| = 0$ (equivalently, $f_n \to f$ in L^1),
- $\lim \int f_n = \int f$ (weak convergence)

Proof. First note that we have

$$|f_n| \le g \longrightarrow |f| \le g$$

so $f_n, f \in L^1$. Moreover, we have

$$|f_n - f| \le 2g$$

so the differences are in L^1 as well. Moreover, we have $2g - |f_n - f| \ge 0$. Thus we can apply Fatou's lemma:

$$\begin{aligned} \int 2g = \int \lim \left(2g - |f - f_n|\right) = \int \liminf \left(2g - |f - f_n|\right) \\ \le \liminf \int \left(2g - |f - f_n|\right) = \int 2g + \liminf \int -|f - f_n| \end{aligned}$$

Because $\int 2g < \infty$, we can subtract it from both sides to see that

$$0 \le \liminf \left(- \int |f - f_n| \right) \implies \limsup \int |f - f_n| \le 0$$

Since the RHS is nonnegative we conclude that $\lim \int |f - f_n|$ exists and is equal to zero. To demonstrate weak convergence, we have

$$\left| \int f_n - \int f \right| = \left| \int f_n - f \right| \le \int |f_n - f| \to 0$$

□

Example 1.8

Consider $f_n = n\chi_{(0,1/n^2)}$. These functions are bounded by $g(x) = \frac{1}{\sqrt{x}} \in L^1$. Moreover, we have

$$\lim \int f_n = \lim \frac{1}{n} = 0 = \int 0 = \int \lim f_n$$

Chapter 2

The Lebesgue Measure

To this point we have defined integrals in a way that allows us to interchange them with limit operators in various settings. We have also defined an appropriate σ-algebra, $\mathscr{B}(\mathbb{R})$, on $\mathbb{R}$, which we can use to work with this integral. Now we have to define a measure on $\mathscr{B}(\mathbb{R})$ that extends the Riemann integral. To make this definition we will essentially present an existence and uniqueness proof.

More precisely we show that there exists a unique positive, *translation invariant* meausre $\lambda : \mathscr{B}(\mathbb{R}) \to [0, \infty]$ such that $\lambda([0, 1]) = 1$.

In this search it will also turn out that the measurable sets under λ is larger than the Borel algebra.

Definition 2.1

For a set $S \subseteq \mathbb{R}$ and $x \in \mathbb{R}$, we define **translation** by

$$S + x := \{s + x : s \in S\}$$

A measure μ is called **translation invariant** if $\mu(S) = \mu(S + x)$ for all x, S.

Our work will involve first developing theorems about how to construct measures out of more primitive objects. Applying this to $\mathbb{R}$ with some geometric intuition will give us the Lebesgue measure.

2.1 Premeasures and Outer Measures

Consider some nonempty set X, and let $\rho : E \to [0, \infty]$ be a map which is initially defined on some subset E of $\mathcal{P}(X)$, with $\rho(\varnothing) = 0$. We do not assume that E is a σ-algebra; however it will generate the σ-algebra that is used by the final measure.

Definition 2.2

If X is nonempty, an **outer measure** on X is a map $\varphi : \mathcal{P}(X) \to [0, \infty]$ such that

1. $\varphi(\varnothing) = 0$,
2. $\varphi(A) \leq \varphi(B)$ whenever $A \subseteq B$,
3. $\varphi\left(\bigcup A_n\right) \leq \sum \varphi(A_n)$ for any countable collection of sets A_n.

Note that an outer measure is not a measure.

We now define an outer measure $\varphi_\rho : \mathcal{P}(X) \to [0, \infty]$ using the data from ρ.

Proposition 2.1

Let X be nonempty, $\rho : E \to [0, \infty]$ for $E \subseteq \mathcal{P}(X)$ containing $\{\varnothing, X\}$, and $\rho(\varnothing) = 0$. Then the function $\varphi_\rho : \mathcal{P}(X) \to [0, \infty]$ defined by

$$\varphi_\rho(A) \coloneqq \inf\left\{\sum \rho(E_n) : \{E_n\}_{n\in\mathbb{N}} \subseteq E, A \subseteq \bigcup E_n\right\}$$

is an outer measure. Here the infimum is over all countable covers of A with elements of E. If no such cover exists then by definition the infimum is ∞.

Proof. It is clear that $\varphi_\rho(\varnothing) = 0$ since we can take the cover to be $E_n = \varnothing$. To show monotonicity, if $A \subseteq B$ then any cover of B covers A, so $\varphi_\rho(A) \leq \varphi_\rho(B)$ (this still holds when one or both sets admit no covers).

If $A = \bigcup A_n$, then for any $\varepsilon > 0$ we can pick covers $\{E_{n,i}\}_i$ for each n such that

$$\sum_{i=1}^{\infty} \rho(E_{n,i}) \geq \varphi_\rho(A_n) - \frac{\varepsilon}{2^n}$$

Then the collection $\{E_{n,i}\}_{n,i}$ is a countable cover of A, and we have

$$\sum_{n,i} \rho(E_{n,i}) = \sum_n \sum_i \rho(E_{n,i}) = \sum_n \left(\varphi(A_n) - \frac{\varepsilon}{2^n}\right) = \sum_n \varphi_\rho(A_n) - \varepsilon$$

Taking $\varepsilon \to 0$ and taking the infimum, we conclude that

$$\varphi_\rho\left(\bigcup_n A_n\right) \leq \sum_n \varphi_\rho(A_n) \qquad \square$$

Example 2.1

Taking E to be the set of intervals and letting $\rho((a, b)) = b - a$, we generate the Lebesgue outer measure.

So far we have placed no assumptions on ρ. In order to get outer measures and measures which we can work with nicely, it is helpful to impose a few conditions. To see this, we examine some possible difficulties with pathological ρ. For instance, if ρ itself is not countably additive, then φ_ρ could fail to coincide with ρ on E.

Definition 2.3

If φ is an outer measure on X, a set $A \subseteq X$ is called **φ-measurable** if for all $Q \in \mathcal{P}(X)$,

$$\varphi(Q) = \varphi(Q \cap A) + \varphi(Q \cap A^c)$$

The set of φ-measurable sets is denoted $\mathscr{A}_\varphi$.

Essentially, a φ-measurable set splits with respect to measure. It is not a priori obvious that nonmeasurable sets should exist under this definition, but we will see later that they do. Note that we always have

$$\varphi(Q) \leq \varphi(Q \cap A) + \varphi(Q \cap A^c)$$

by countable subadditivity of φ. Thus in general we can check φ-measurability just by verifying that

$$\varphi(Q) \geq \varphi(Q \cap A) + \varphi(Q \cap A^c)$$

Moreover, when $\varphi(Q) = \infty$ this is automatically true.

A natural question is then to ask whether φ_ρ-measurable sets form a σ-algebra. The answer to this question is yes; moreover the restriction theorem that we prove shows that φ_ρ is also a measure when restricted to these sets.

Theorem 2.2: Caratheodory's Restriction Theorem

Let X be a nonempty set and φ an outer measure on X. Then $\mathscr{A}_\varphi$ is a σ-algebra, and $\mu_\varphi := \varphi|_{\mathscr{A}_\varphi}$ is a measure.

Proof. Take $\varnothing$. By the remark above, it suffices to show that for any $Q \in X$,

$$\varphi(Q) \geq \varphi(Q \cap \varnothing) + \varphi(Q \cap \varnothing^c)$$

But this is clear since the right hand side is just

$$\varphi(\varnothing) + \varnothing(Q) = \varnothing(Q)$$

It is also obvious that $\mathscr{A}_\varphi$ is closed under complements since the definition treats A, A^c symmetrically.

To show closure under countable unions, we first show finite unions. For $A, B \in \mathscr{A}_\varphi$, and pick $Q \in \mathcal{P}(X)$ with $\varphi(Q) < \infty$ (recall from above that we can assume finite outer measure). Then

$$\begin{aligned}\varphi(Q) &= \varphi(Q \cap A) + \varphi(Q \cap A^c)\\ &= \varphi(Q \cap A \cap B) + \varphi(Q \cap A \cap B^c) + \varphi(Q \cap A^c \cap B) + \varphi(Q \cap A^c \cap B^c)\end{aligned}$$

We have the identity
$$A \cup B = (A \cap B) \cup (A \cap B^c) \cup (A^c \cap B)$$
Since φ is an outer measure, it follows that
$$\varphi(A \cup B) \leq \varphi(A \cap B) + \varphi(A \cap B^c) + \varphi(A^c \cap B)$$
take a sequence $\{A_n\} \subseteq \mathscr{A}_\varphi$. Intersecting with Q on both sides, we have
$$\varphi(Q) \geq \varphi(Q \cap (A \cup B)) + \varphi(Q \cap A^c \cap B^c) = \varphi(Q \cap (A \cup B)) + \varphi(Q \cap (A \cup B)^c)$$
Now we extend this to countable unions $\bigcup A_n$. It suffices to assume that the A_n are pairwise disjoint by picking
$$A_n' = A_n \setminus \left(\bigcup_{m=1}^{n-1} A_m \right)$$
Then A_n' are in $\mathscr{A}_\varphi$ by our work showing that complements and finite unions were closed. Now take Q with $\varphi(Q) < \infty$. Then for any N, $\bigcup^N A_n \in \mathscr{A}_\varphi$. Therefore we can write
$$\begin{aligned}
\varphi(Q) &= \varphi\left(Q \cap \left(\bigcup_{n=1}^N A_n\right)\right) + \varphi\left(Q \cap \left(\bigcup_{n=1}^N A_n\right)^c\right) \\
&\geq \varphi\left(Q \cap \left(\bigcup^N A_n\right) \cap A_n\right) + \varphi\left(Q \cap \left(\bigcup^N A_n\right) \cap A_N^c\right) \\
&= \varphi\left(Q \cap (A)_N\right) + \varphi\left(Q \cap \left(\bigcup^{N-1} A_n\right)\right) \\
&\vdots \\
&\geq \sum_{n=1}^N \varphi(Q \cap A_n) + \varphi\left(Q \cap \left(\bigcup^N A_n\right)^c\right)
\end{aligned}$$
Now, we have $\bigcup^\infty A_n \supseteq \bigcup^N A_n$, so we have
$$\varphi\left(Q \cap \left(\bigcup^N A_n\right)^c\right) \geq \varphi\left(Q \cap \left(\bigcup^\infty A_n\right)^c\right)$$
Taking $N \to \infty$, we have
$$\varphi(Q) \geq \sum_{n=1}^\infty \varphi(Q \cap A_n) + \varphi\left(Q \cap \left(\bigcup^\infty \varphi(A_n)\right)^c\right)$$
Since φ is countably subadditive,
$$\sum_{n=1}^\infty \varphi(Q \cap A_n) \geq \varphi\left(Q \cap \left(\bigcup^\infty A_n\right)\right)$$

Thus

$$\varphi(Q) \geq \varphi\left(Q \cap \left(\bigcup^{\infty} A_n\right)\right) + \varphi\left(Q \cap \left(\bigcup^{\infty} \varphi(A_n)\right)^c\right)$$

showing that $\bigcup^{\infty} A_n \in \mathscr{A}_\varphi$. Thus $\mathscr{A}_\varphi$ is a σ-algebra.

We know there is a set with finite measure since $\mu_\varphi(\varnothing) = 0$. To demonstrate finite additivity, pick $A, B \in \mathscr{A}_\varphi$ disjoint. Then

$$\mu_\varphi(A \cup B) = \varphi(A \cup B) = \varphi((A \cup B) \cap A) + \varphi((A \cup B) \cap A^c) = \varphi(A) + \varphi(B) = \mu_\varphi(A) + \mu_\varphi(B)$$

The proof for countable additivity is the same. □

It is also worth noting that the measure produced by the restriction theorem has the property of being a "complete" measure.

Definition 2.4

A measure $\mu : \text{Meas}(X) \to [0, \infty]$ is said to be **complete** if for any $M \subseteq N$ with N measurable and $\mu(N) = 0$, M is measurable.

In short, any subset of a measure zero set is measurable. (We already know that any such measurable set has measure zero, but a priori it is not clear that such sets are measurable in the first place).

Proposition 2.3

Given X and φ an outer measure on X, the measure μ_φ as defined in Caratheodory's Restriction Theorem is complete.

Proof. Pick $B \in \mathscr{A}_\varphi$ with $\mu_\varphi(B) = 0$, and take $A \subseteq B$. Take some $Q \subseteq X$ with $\varphi(Q) < \infty$. Then we have

$$Q \cap A \subseteq Q \cap B \subseteq B \implies \varphi(Q \cap A) \leq \varphi(B) = \mu_\varphi(B) = 0$$

Also we have

$$Q \cap A^c \subseteq Q \implies \varphi(Q \cap A^c) \leq \varphi(Q)$$

so

$$\varphi(Q) \geq \varphi(Q \cap A) + \varphi(Q \cap A^c)$$

and thus $A \in \mathscr{A}_\varphi$. □

Proposition 2.4

Given any measure $\mu : \mathcal{M} \to [0, \infty]$, there exists a unique complete measure $\overline{\mu} : \overline{\mathcal{M}} \to [0, \infty]$, wehre $\overline{\mathcal{M}} \supseteq \mathcal{M}$ is another σ-algebra on X and $\overline{\mu}|_{\mathcal{M}} = \mu$.

We have thus illustrated a method to pass from a primitive function $\rho : E \to [0, \infty]$ to a full measure μ_{φ_ρ} on $\mathscr{A}_{\varphi_\rho}$. To that end it is worth investigating the relationship between $\sigma(E)$ and $\mathscr{A}_{\varphi_\rho}$. In order to properly do this it is best to impose additional conditions on ρ.

Definition 2.5

An **algebra** on a set X is a collection $\mathscr{A} \subseteq \mathcal{P}(X)$ which contains X, is closed under complements, and closed under finite unions.

Note the only difference between a σ-algebra and an algebra is we require σ-algebras to be closed under countable unions as well.

Definition 2.6

Let $\mathscr{A} \subseteq \mathcal{P}(X)$ be an algebra. Then a function $\rho : \mathscr{A} \to [0, \infty]$ is called a **premeasure** if:

1. $\rho(\varnothing) = 0$,
2. If $\{A_n\} \subseteq \mathscr{A}$ is a countable collection of pairwise disjoint sets, and if in addition $\bigcup A_n \in \mathscr{A}$, then
$$\rho\left(\bigcup A_n\right) = \sum \rho(A_n).$$

Proposition 2.5

If $\rho : \mathscr{A} \to [0, \infty]$ is a premeasure, then for $A \subseteq B$ with $A, B \in \mathscr{A}$, $\rho(A) \leq \rho(B)$.

Proof. $B \setminus A \in \mathscr{A}$, so by countable additivity

$$\rho(B) = \rho(A) + \rho(B \setminus A) \geq \rho(A) \qquad \square$$

Adding the condition that ρ is a premeasure ensures that our extended constructions are proper extensions, in the sense that they are consistent with our original data.

Proposition 2.6

Suppose $\rho : \mathscr{A} \to [0, \infty]$ is a premeasure. Then $\varphi_\rho|_{\mathscr{A}} = \rho$. Moreover, $\mathscr{A} \subseteq \mathscr{A}_{\varphi_\rho}$.

Proof. Note that clearly $\varphi_\rho|_{\mathscr{A}} \leq \rho$, since any set in $\mathscr{A}$ is a cover for itself.

For the reverse inequality, pick $Q \in \mathscr{A}$ and a countable cover $\{E_n\} \subseteq \mathscr{A}$. Note that by monotonicity we can assume the E_n are disjoint (call the disjoint parts $\{F_n\}$), and we can also assume that $Q = \sqcup F_n$; that is the F_n do not overcover Q. Then we have

$$\rho(Q) = \rho\left(\sqcup_{n=1}^{\infty} F_n\right) = \sum_{n=1}^{\infty} \rho(F_n) \leq \sum_{n=1}^{\infty} \rho(E_n)$$

So $\rho \leq \varphi_\rho|_{\mathscr{A}}$ as well and thus $\varphi_\rho|_{\mathscr{A}} = \rho$.

To show that $\mathscr{A} \subseteq \mathscr{A}_{\varphi_\rho}$, pick $A \in \mathscr{A}$ and $Q \subseteq X$. Then by the definition of φ_ρ, we can pick a countable cover $\{E_n\} \subseteq \mathscr{A}$ with

$$\varphi_\rho(Q) \geq \sum_{n=1}^{\infty} \rho(E_n) - \varepsilon$$

Since each $E_n \in \mathscr{A}$ and $A \in \mathscr{A}$, we have $\rho(E_n) = \rho(E_n \cap A) + \rho(E_n \cap A^c)$. Thus we have

$$\varphi_\rho(Q) + \varepsilon \geq \sum_{n-1}^{\infty} \rho(E_n) = \sum_{n=1}^{\infty} \rho(E_n \cap A) + \rho(E_n \cap A^c) = \varphi_\rho(E_n \cap A) + \varphi_\rho(E_n \cap A^c)$$

Since $Q \cap A \subseteq \bigcup E_n \cap A$, by countable subadditivity of φ we have

$$\varphi_\rho(Q \cap A) \leq \sum_{n=1}^{\infty} \varphi_\rho(E_n \cap A)$$

and similarly for the A^c term. Thus we have

$$\varphi_\rho(Q) + \varepsilon \geq \varphi_\rho(Q \cap A) + \varphi_\rho(Q \cap A^c)$$

Taking $\varepsilon \to 0$, we see that A is φ_ρ-measurable and thus $\mathscr{A} \subseteq \mathscr{A}_{\varphi_\rho}$. □

Thus, since we can now properly use premeasures to build outer measures, we can apply the restriction theorem to actually extend premeasures.

Theorem 2.7: Caratheodory's Extension Theorem

Let $\rho : \mathscr{A} \to [0, \infty]$ be a premeasure and φ_ρ, $\mathscr{A}_{\varphi_\rho}$, μ_{φ_ρ} be as defined above. Then:

1. $\sigma(\mathscr{A}) \subseteq \mathscr{A}_{\varphi_\rho}$;
2. If $\nu : \sigma(\mathscr{A}) \to [0, \infty]$ is any other measure such that $\nu|_{\mathscr{A}} = \rho$, then $\nu \leq \mu_{\varphi_\rho}$ on $\sigma(\mathscr{A})$, and moreover for any $E \in \sigma(\mathscr{A})$ with $\mu_{\varphi_\rho}(E) < \infty$, $\nu(E) = \mu_{\varphi_\rho}(E)$;
3. If X is **σ-finite** with respect to ρ, meaning that there is a countable collection $\{A_n\} \subseteq \mathscr{A}$ with $\rho(A_n) < \infty$ and $X = \bigcup A_n$, then μ_{φ_ρ} is the unique extension of ρ to $\sigma(\mathscr{A})$.

Proof.

1. This is clear since $\mathscr{A} \subseteq \mathscr{A}_{\varphi_\rho}$ with $\mathscr{A}_{\varphi_\rho}$ a σ-algebra. Then by definition $\sigma(\mathscr{A})$ is the smallest σ-algebra containing $\mathscr{A}$ so $\sigma(\mathscr{A}) \subseteq \mathscr{A}_{\varphi_\rho}$.
2. Let $E \in \sigma(\mathscr{A})$ and pick a cover $\{E_n\} \subseteq \mathscr{A}$. Then we have subadditivity (note the E_n are not necessarily disjoint):

$$\nu(E) \leq \sum_{n=1}^{\infty} \nu(E_n) = \sum_{n=1}^{\infty} \rho(E_n)$$

Taking the infimum, we have $\nu(E) \leq \mu_{\varphi_\rho}(E)$.

Note that in general, by approximation from inside, if we pick a sequence of sets $\{E_n\} \subseteq \mathscr{A}$, then

$$\nu\left(\bigcup_{n=1}^{\infty} E_n\right) = \lim_{n\to\infty} \nu(E_n) = \lim_{n\to\infty} \mu_{\varphi_\rho}(E_n)$$

Suppose $\mu_{\varphi_\rho}(E) < \infty$. By the infimum property, for $\varepsilon > 0$ we can pick a cover $\{E_n\} \subseteq \mathscr{A}$ such that

$$\mu_{\varphi_\rho}\left(\bigcup_{n=1}^{\infty} E_n\right) < \mu_{\varphi_\rho}(E) + \varepsilon \implies \mu_{\varphi_\rho}\left(\left(\bigcup_{n=1}^{\infty} E_n\right) \setminus E\right) < \varepsilon$$

Then we have

$$\mu_{\varphi_\rho}(E) \leq \mu_{\varphi_\rho}\left(\bigcup_{n=1}^{\infty} E_n\right) = \nu\left(\bigcup_{n=1}^{\infty} E_n\right)$$

$$= \nu\left(\left(\bigcup_{n=1}^{\infty} E_n\right) \cap E\right) + \nu\left(\left(\bigcup_{n=1}^{\infty}\right) \cap E^c\right) = \nu(E) + \nu\left(\left(\bigcup_{n=1}^{\infty} E_n\right) \setminus E\right)$$

$$\leq \nu(E) + \mu_{\varphi_\rho}\left(\left(\bigcup_{n=1}^{\infty} E_n\right) \setminus E\right) \leq \nu(E) + \varepsilon$$

3. If X is σ-finite with respect to ρ, then pick a pairwise disjoint countable collection $\{A_n\} \subseteq \mathscr{A}$ with $\rho(A_n) < \infty$ and $X = \bigcup A_n$. Then we have

$$\mu(E) = \sum_{n=1}^{\infty} \mu(E \cap A_n) = \sum_{n=1}^{\infty} \nu(E \cap A_n) = \nu(E) \qquad \square$$

2.2 The Lebesgue Premeasure

Having now developed a theory of how to define measures on simpler sets, we apply this to construct the Lebesgue measure.

To do this, we first need to consider the simpler sets on which we will define a premeasure. For the case of the Levesgue measure, we will take the collection of half open intervals:

$$\mathscr{A}_0 \coloneqq \{\varnothing\} \cup \{(a,b] : a \in [-\infty,\infty), a < b\} \cup \{(a,\infty) : a \in [-\infty,\infty)\} \subseteq \mathcal{P}(\mathbb{R})$$

This is the set of all intervals which are open on the left and closed on the right, with appropriate consideration of infinite endpoints.

Definition 2.7

An **elementary family** on X is a collection $\mathscr{F} \subseteq \mathcal{P}(X)$ such that

1. $\varnothing \in \mathscr{F}$,
2. $\mathscr{F}$ is closed under finite intersections,

3. For any $E \in \mathscr{F}$, $X \setminus E$ is a finite disjoint union of elements of $\mathscr{F}$.

Proposition 2.8

$\mathscr{A}_0$ is an elementary family on $\mathbb{R}$.

Proof. 1 is true by definition.

For intersections, in general we have

$$(a,b] \cap (a',b'] = \begin{cases} \varnothing \\ (a',b'], \quad a < a' < b' < b \\ \vdots \end{cases}$$

Here we do not show all of the cases but it is true in general. Complements are similar:

$$\begin{aligned} \varnothing^c &= \mathbb{R} \\ (a,b]^c &= (-\infty, a] \cup (b, \infty) \\ &\vdots \end{aligned}$$

$\square$

Proposition 2.9

If $\mathcal{E}$ is an elementary family then the collection $\mathscr{A}$ of finite disjoint unions of elements of $\mathcal{E}$ is an algebra.

Proposition 2.10

$\sigma(\mathscr{A}) = \mathscr{B}(\mathbb{R})$, where $\mathscr{A}$ is the algebra given by the collection of finite disjoint unions of elements of the half open intervals $\mathscr{A}_0$.

Proof. We can write half open intervals as countable intersections of open intervals:

$$(a,b] = \bigcap_{n=1}^{\infty} \left(a, b + \frac{1}{n}\right)$$

so $\sigma(\mathscr{A}_0) \subseteq \mathscr{B}(\mathbb{R})$. In the other direction, we have

$$(a,b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right)$$

Here we have handwaved some of the cases away but it is nevertheless true that $\mathscr{B}(\mathbb{R}) \subseteq \sigma(\mathscr{A})$ as well. $\square$

Now, we have developed an suitable algebra to define a premeasure on. (Note that we would have liked to simply define our premeasure on intervals. However this is not an algebra and so not sufficient to define a premeasure on. Nevertheless we can define essentially the same premeasure on disjoint unions thereof; with extra work to verify that it is well-defined.)

Definition 2.8

We define the **Lebesgue premeasure** $\rho : \mathscr{A} \to [0, \infty]$ by

$$\rho\left(\bigsqcup_{j=1}^{n} (a_j, b_j]\right) := \sum_{j=1}^{n} b_j - a_j$$
$$\rho(\varnothing) = 0$$

To check that it is well defined, note that alternate representations like $(0, 1] = (0, 1/2] \cup (1/2, 1]$ must "telescope" and hence since they are disjoint unions this is well defined.

Theorem 2.11

ρ is well defined, and it is a premeasure on $\mathscr{A}$.

Note that $\mathbb{R}$ is σ-finite with respect to ρ since it is the union of $[n, n+1]$ for $n \in \mathbb{Z}$.

Definition 2.9

The **Lebesgue measure** on $\mathbb{R}$ is the unique measure $\lambda = \mu_{\varphi_\rho}$ produced by applying Caratheodory's extension theorem to the Lebesgue premeasure ρ. Elements of the associated σ-algebra $\mathscr{A}_{\varphi_\rho}$ are called **Lebesgue measurable** subsets of $\mathbb{R}$.

Caratheodory's extension theorem assures us that λ is the unique extension of ρ to $\mathscr{B}(\mathbb{R})$. However, we need to remove the arbitrary choice of ρ for full generality of λ. This is accomplished by showing that λ satisfies another uniqueness condition.

Theorem 2.12

There exists a unique measure $\lambda : \mathscr{B}(\mathbb{R}) \to [0, \infty]$ such that $\lambda([0, 1]) = 1$ and λ is translation invariant.

Proof. Existence is satisfied by the Lebesgue measure we just showed. It is translation invariant since the premeasure it is defined on is, and it is not hard to verify that $[0, 1]$ has Lebesgue measure 1.

For uniqueness, pick another measure $\tilde{\lambda} : \mathscr{B}(\mathbb{R}) \to [0, \infty]$ obeying the hypotheses of the theorem. To show that $\lambda = \tilde{\lambda}$, we will need some regularity properties of Borel measures. We will delay the proofs of these properties, but essentially the statement is that measures on sufficiently nice Borel algebras may be approximated from the outside by open sets and from the inside from compact sets. More formally, for any $\mu : \mathscr{B}(\mathbb{R}) \to [0,\infty]$ and $S \in \mathscr{B}(\mathbb{R})$,

$$\mu(S) = \inf\{\mu(U) : S \subseteq U \in \text{Open}(\mathbb{R})\}$$
$$\mu(S) = \sup\{\mu(K) : S \supseteq K \in \text{Comapct}(\mathbb{R})\}$$

In particular it suffices to show that $\lambda, \tilde{\lambda}$ agree on open sets. Moreover since any open set in $\mathbb{R}$ is the countable disjoint union of open intervals, we can just show this for open intervals, and by translation invaraicne we just need to show that for $a > 0$,

$$\lambda((0,b)) = \tilde{\lambda}((0,b))$$

In essence, since we already know $\lambda([0,1]) = 1$, we need to prove a scaling property of translation invariant measures.

To show that the endpoints of intervals will not be a problem, we show that singletons have measure zero in both $\lambda, \tilde{\lambda}$. If we pick N distinct points $x_1, \dots, x_N$ in $[0,1]$, then by translation invariance we have

$$N\lambda(\{0\}) = \lambda\left(\bigcup_{j=1}^{N}\{x_j\}\right) = \lambda([0,1]) = 1$$

Taking $N \to \infty$, we conclude $\lambda(\{0\}) = 0$, and similarly for $\tilde{\lambda}$.

Claim: $\tilde{\lambda}\left(\left[0, \frac{1}{n}\right]\right) = \frac{1}{n}$. This is clear since $[0,1]$ is formed of n translated copies:

$$1 = \lambda([0,1]) = \lambda\left(\bigsqcup_{k=0}^{n-2}\left[\frac{k}{n}, \frac{k+1}{n}\right) \cup \left[\frac{n-1}{n}, 1\right]\right) = \sum_{k=0}^{n-1}\lambda\left(\left[\frac{k}{n}, \frac{k+1}{n}\right]\right) = n\lambda\left(\left[0, \frac{1}{n}\right]\right)$$

Similar work shows that $\lambda, \tilde{\lambda}$ obey a rational scaling factor:

$$\tilde{\lambda}\left(\left[0, \frac{m}{n}\right]\right) = \frac{m}{n}$$

To show that this is the case for real endpoints, we can use approximation from inside or outside for arbitrary measures to conclude that for $b > 0$ and $r_n \nearrow b$,

$$\lambda([0,b]) = \lim \lambda([0, r_n]) = \lim r_n = b$$

Thus we conclude that $\lambda = \tilde{\lambda}$ on the open intervals and hence all open sets, so that $\lambda = \tilde{\lambda}$ on $\mathscr{B}(\mathbb{R})$ (pending the regularity conditions for Borel measures). □

2.3 Regularity of Borel Measures

Remark

The argument outlined by Prof. Shapiro in this chapter during lecture used a flawed argument. The statements which are correct have been left here, but a proper proof of Borel regularity uses the Kakutani-Markov-Riesz theorem (here see Folland 7.7).

In this section we discuss regularity properties connecting topological and measure spaces over arbitrary sets with sufficiently nice properties. These proofs are taken from Bogachev's *Measure Theory*.

Definition 2.10

Given a topological space X, a **Borel measure** is a measure on $\mathscr{B}(X)$.

Definition 2.11

Let $\mu : \mathscr{B}(X) \to [0, \infty]$ be a Borel measure on X. A set $A \in \mathscr{B}(X)$ is called **μ-outer regular** if

$$\mu(A) = \inf \{\mu(U) : A \subseteq U \in \text{Open}(X)\}$$

and similarly it is **μ-inner regular** if either A is open or $\mu(A) < \infty$, and also

$$\mu(A) = \sup \{\mu(K) : A \supseteq K \in \text{Compact}(X) \cap \mathscr{B}(X)\}$$

We say that μ is **regular** if every Borel set is μ-inner and μ-outer regular.

Definition 2.12

A topological space X is a **Hausdorff space** if for any $x, y \in X$ there exist open sets $x \in U$ and $y \in V$ with U, V disjoint.

Note that in a Hausdorff space, every compact set is closed and hence Borel, so there is no need to check measurability.

This leads us to the natural question of what assumptions must be placed on our topological space X such that every Borel measure is regular. The first is a requirement that essentially ensures the topological operations are compatible with the Borel algebra operations:

Definition 2.13

A topological space X is called **second countable** if there exists a countable basis for its topology.

Definition 2.14

A topological space X is called **σ-compact** if there exists a countable collection of compact sets $\{K_n\}_{n=1}^{\infty} \subseteq \text{Compact}(X)$ such that $X = \bigcup K_n$.

Definition 2.15

A measure $\mu : \mathscr{B}(X) \to [0, \infty]$ is **locally finite** if for any $x \in X$ there exists $U \ni x$ open with $\mu(U) < \infty$.

Proposition 2.13

If $\mu : \mathscr{B}(X) \to [0, \infty]$ is locally finite and X is Hausdorff, then $\mu(K) < \infty$ for any K compact.

Proof. (Note that since X is Hausdorff, K is closed and hence Borel, so it is actually measurable). For each $x \in K$ take $U_x \ni x$ of finite measure. Then $K \subseteq \bigcup_{x \in K} U_x$. Picking a finite subcover, we establish that $\mu(K) < \infty$. □

We state the following without proof, which connects the criterion we just established with μ-regular sets. This is proved with monotonicity and approximation.

Theorem 2.14

Let X be a topological space and $\mu : \mathscr{B}(X) \to [0, \infty]$ a Borel measure. A set $A \in \mathscr{B}(X)$ of finite measure is μ-regular if and only if for all $\varepsilon > 0$ there exists U_ε open and $K_\varepsilon \in \mathscr{B}(X)$ compact with $K_\varepsilon \subseteq A \subseteq U_\varepsilon$ and $\mu(U_\varepsilon \setminus K_\varepsilon) < \varepsilon$.

Definition 2.16

A topological space X is called **locally compact** if for any $x \in X$ there exists $U \ni x$ open with $\overline{U}$ compact.

Theorem 2.15

Let X be locally compact, σ-compact, and Hausdorff with $\mu : \mathscr{B}(X) \to [0, \infty]$ a σ-finite and locally finite Borel measure. The:

1. For any $\varepsilon > 0$, $A \in \mathscr{B}(X)$, there exists U_ε open and F_ε closed with $\mu(U_\varepsilon \setminus F_\varepsilon) < \varepsilon$ with $F_\varepsilon \subseteq A \subseteq U_\varepsilon$.
2. μ is regular.
3. For any $A \in \mathscr{B}(X)$ there exists $F \in F_\sigma$ and $G \in G_\delta$ with $F \subseteq A \subseteq G$ and $\mu(G \setminus F) = 0$.

In particular, any Borel set may be written as the countable union of closed sets ($F \in F_\sigma$) and a measure zero set.

2.4 Product Measures

So far we have only defined a Lebesgue measure on $\mathbb{R}$, but we would like to extend this naturally to $\mathbb{R}^n$. Under sufficient assumptions we can ensure uniqueness of a Borel measure on $\mathbb{R}^n$; however showing existence may be done in two ways. One way to do so is to rerun the Lebesgue construction on $\mathbb{R}^n$, with rectangles in place of intervals. To do this more abstractly, we can define what it means to construct measures on product spaces more generally.

Definition 2.17

Let $\{X_\alpha\}_{\alpha \in A}$ be a collection of measurable spaces with associated σ-algebras $\{\mathcal{M}_\alpha\}_{\alpha\in A}$. Then the **product space** is the Cartesian product of the X_α:

$$X = \prod_{\alpha \in A} X_\alpha = \left\{ f : A \to \bigcup_{\alpha \in A} X_\alpha : f(\alpha) \in X_\alpha \right\}$$

For $\beta \in A$, let $\pi_\beta : X \to X_\beta$ denote the canonical projection map. We endow X with the **product σ-algebra**

$$\mathcal{M} := \sigma\left(\left\{\pi_\alpha^{-1}(E_\alpha) : \alpha \in A, E_\alpha \in \mathcal{M}_\alpha\right\}\right) = \sigma\left(\{\pi_\alpha\}_{\alpha\in A}\right)$$

We notate this product as $\mathcal{M} = \bigotimes_{\alpha\in A} \mathcal{M}_\alpha$.

Proposition 2.16

If the index set is countable then $\mathcal{M}$ is generated by the "rectangular sets":

$$\mathcal{M} = \sigma\left(\left\{\prod_{\alpha\in A} E_\alpha : E_\alpha \in \mathcal{M}_\alpha\right\}\right)$$

Proof. Homework. $\square$

Having defined product spaces, we can now define the natural way to define a product measure. Here we will restrict ourselves to the case of finite products.

Definition 2.18

A **rectangular subset** of $\prod_{\alpha\in A} X_\alpha$ is a set of the form $\prod_{\alpha \in A} E_\alpha$ where $E_\beta \subseteq X_\beta$.

Proposition 2.17

Let $X = \prod_{j=1}^n X_j$ be a finite product space with associated nonnegative measures $\{\mu_j : \mathcal{M}_j \to [0,\infty]\}_{j=1}^n$. We write $\mathscr{A}_0$ to denote the measurable rectangular sets in X. Then $\mathscr{A}_0$ is an elementary family.

Proof. Clearly $\varnothing, X \in \mathscr{A}_0$. Intersections are closed since

$$\prod_{j=1}^{n} E_j \cap \prod_{k=1}^{n} F_k = \prod_{j=1}^{n} (E_j \cap F_j)$$

Also since our index set is finite, the complement of rectangular sets are finite disjoint unions of rectangular sets. □

As a result it follows then that the collection $\mathscr{A}$ of finite disjoint unions of elements in $\mathscr{A}_0$ is an algebra. Note that because A is finite, $\mathcal{M} = \bigotimes_{j=1}^{n} \mathcal{M}_j = \sigma(\mathscr{A}_0)$. Since $\mathcal{M}$ is by definition the smallest σ-algebra containing $\mathscr{A}_0$ and $\mathscr{A}_0 \subseteq \mathscr{A} \subseteq \sigma(\mathscr{A})$, we conclude that $\sigma(\mathscr{A}) = \mathcal{M}$.

As in the case of the Lebesgue measure, we now define a premeasure on $\mathscr{A}$ by

$$\rho\left(\bigsqcup_{j=1}^{m} \prod_{k=1}^{n} E_{k,j}\right) = \sum_{j=1}^{m} \prod_{k=1}^{n} \mu_k(E_{k,j})$$

The fact that this is well defined essentially follows from additivity of each μ_j.

Proposition 2.18

ρ is a premeasure.

Proof. $\rho(\varnothing) = 0$ since each μ_j is a measure. Also it is essentially clear by definition that ρ is finitely additive. If a countable union $\bigcup_{n=1}^{\infty} A_n \in \mathscr{A}$, then since $\mathscr{A}$ is composed of finite unions, we must have $\bigcup_{n=1}^{\infty} A_n = \bigsqcup_{j=1}^{k} B_j$. Then we have

$$\bigcup_{n=1}^{\infty} A_n = \bigsqcup_{j=1}^{k} \prod_{\ell=1}^{n} F_{\ell,j} = \sum_{j=1}^{k} \prod_{\ell=1}^{n} \mu_\ell(F_{\ell,j})$$

Prof. Shapiro did not make it clear how to finish this argument. □

Now that we have a premeasure, we can run the Caratheodory construction to generate a complete measure $\mu : \mathscr{A}_{\varphi_\rho} \to [0, \infty]$ where $\mu|_{\mathscr{A}} = \rho$ and $\mathscr{A}_{\varphi_\rho}$ is a σ-algebra containing $\sigma(\mathscr{A}) = \mathcal{M}$.

Thus we have to note that in general, the product σ-algebra is not the algebra on which the product measure is complete.

In particular we define the Lebesgue measure on $\mathbb{R}^n$ to be the n-fold product measure of the Lebesgue measure on $\mathbb{R}$. The domain of this completed measure is the set of Lebesgue measurable sets on $\mathbb{R}^n$. Note that this is strictly larger than simply the product σ-algebra:

$$\mathscr{L}_n \supsetneq \mathscr{L} \otimes \ldots \otimes \mathscr{L}$$

For instance $A \times B \subseteq \mathbb{R}^2$ is measurable when A is not measurable but B has measure zero.

2.5 Fubini-Tonelli

Here we develop results on iterated integrals with respect to measures. In multivariable analysis we noted that for sufficiently nice functions, the relation

$$\int_{X\times Y} f = \int_X \left(\int_Y f_x\right) = \int_Y \left(\int_X f_y\right)$$

holds. Proving this for arbitrary measures is particularly helpful since it allows us to extend these results to, say, double sums and combinations of sums and integrals. In the language of measure theory we would like to prove that given a sufficiently nice measurable function $f : X \times Y \to \mathbb{C}$, we have

$$\int_{X\times Y} f \,\mathrm{d}(\mu\times\nu) = \int_X \left(\int_Y f_x \,\mathrm{d}\nu\right) \mathrm{d}\mu = \int_Y \left(\int_X f_y \,\mathrm{d}\mu\right) \mathrm{d}\nu$$

The Fubini and Tonelli theorems provide two conditions for this to hold. In short the Tonelli theorem applies to nonnegative product measurable functions while Fubini's applies to L^1 functions. It is often helpful to use Tonelli's theorem on $|f|$ to satisfy the L^1 hypothesis for Fubini's theorem.

Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be two measure spaces, and as before equip $X \times Y$ with the product σ-algebra $\mathcal{M}\otimes\mathcal{N}$, with the product measure $\mu\times\nu$ defined on $\overline{\mathcal{M}\otimes\mathcal{N}}$. Let $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$ be the canonical projections. For any $M \otimes N$-measurable function $f : X \times Y \to \mathbb{C}$ and $x \in X, y \in Y$, we define the **sections** of f by

$$f_x : \begin{cases} Y \to \mathbb{C} \\ y \mapsto f(x, y) \end{cases}$$
$$f^y : \begin{cases} X \to \mathbb{C} \\ x \mapsto f(x, y) \end{cases}$$

Similarly, for a subset $A \in \mathcal{M}\otimes\mathcal{N}$ we consider the **sections** by

$$A_2(x) = \{y : (x, y) \in A\} \subseteq Y$$
$$A_1(y) = \{x : (x, y) \in A\} \subseteq X$$

Of course in order to consider iterated integrals we will need the sections f_x, f^y to be measurable as well. (Note that this is equivalent to measurability of set sections).

Proposition 2.19

1. Let $f : X \times Y \to \mathbb{C}$ be $\mathcal{M}\otimes\mathcal{N}$ measurable, and fix $x \in X, y \in Y$. Then f_x, f^y are both measurable with respect to $\mathcal{N}, \mathcal{M}$, respectively.

2. Let $A \in \mathcal{M}\otimes\mathcal{N}$, and fix $x \in X, y \in Y$. Then $A_1(y), A_2(x)$ are $\mathcal{M}, \mathcal{N}$ measurable, respectively.

Proof. 1. This follows from point 2 since $(f_x)^{-1}(B) = (f^{-1}(B))_2(x)$ whenever $B \in \mathscr{B}(\mathbb{C})$.

2. Define the collection of sets whose sections are measurable:

$$\mathscr{R} = \{E \subseteq X \times Y : E_1(y) \in \mathcal{M}, E_2(x) \in \mathcal{N} \forall (x, y) \in X \times Y\}$$

Note that all measurable rectangular sets are in $\mathscr{R}$, since the sections of $U \times V$ are U or V (or $\varnothing$). Since the rectangles generate $\mathcal{M} \otimes \mathcal{N}$, we just need to show that $\mathscr{R}$ is a σ-algebra.

Clearly $X \times Y \in \mathscr{R}$. Also taking complements passes nicely through set sections since we are just complementing relative to the section:

$$(E^c)_2(x) = (E_2(x))^c$$

Similarly

$$\left(\bigcup_{j=1}^{\infty} E_j\right)_2 (x) = \bigcup_{j=1}^{\infty} (E_j)_2(x)$$

□

Definition 2.19

A **monotone class** on a nonempty set X is a subset $\mathscr{C} \subseteq \mathcal{P}(X)$ which is closed under:

1. countable increasing unions (if $E_n \nearrow E$ and $E_n \in \mathscr{C}$, then $E \in \mathscr{C}$);
2. countable decreasing intersections (if $E_n \searrow E$ and $E_n \in \mathscr{C}$, then $E \in \mathscr{C}$).

Obviously any σ-algebra must be a monotone class. Also the arbitrary intersection of monotone classes is a monotone class, as is common with any construction of this type.

Proposition 2.20

For any $E \subseteq \mathcal{P}(X)$, there exists a unique smallest monotone class containing E, denoted $\mathscr{C}(E)$. This is called the monotone class **generated by** E.

Proof. Take the intersection over all monotone classes containing E. This is nonempty since $\mathcal{P}(X)$ is a monotone class. □

Lemma 2.21: Monotone Class Lemma

If $\mathscr{A} \subseteq \mathcal{P}(X)$ is a algebra of sets then

$$\sigma(\mathscr{A}) = \mathscr{C}(\mathscr{A})$$

Proof. Clearly we have $\sigma(\mathscr{A}) \supseteq \mathscr{C}(\mathscr{A})$ since $\sigma(\mathscr{A})$ is itself a monotone class.

To see that $\mathscr{C}(\mathscr{A})$ is a σ-algebra, consider for any $E \in \mathscr{C}(\mathscr{A})$ the collection

$$\mathscr{D}_E(\mathscr{A}) \coloneqq \{F \in \mathscr{C}(\mathscr{A}) : E \setminus F, F \setminus E, E \cap F \in \mathscr{C}(\mathscr{A})\}$$

Clearly $\varnothing, E \in \mathscr{D}_E(\mathscr{A})$. Also $F \in \mathscr{D}_E(\mathscr{A}) \iff E \in \mathscr{D}_F(\mathscr{A})$. Also because relative complements and intersections commute with increasing unions and decreasing intersections, $\mathscr{D}_E$ is itself a monotone class. Clearly it must contain $\mathscr{A}$ since $\mathscr{A}$ is an algebra, so $\mathscr{D}_E(\mathscr{A}) = \mathscr{C}(\mathscr{A})$. This shows that $\mathscr{C}(\mathscr{A})$ contains the whole set and is closed under complements.

TODO: show closure under countable intersections. □

The following theorem says that the functions assigning measures of sections are themselves measurable.

Theorem 2.22

Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces. If $E \in \mathcal{M} \otimes \mathcal{N}$ then $x \mapsto \nu(E_2(x))$ and $y \mapsto \mu(E_1(y))$ are both measurable, and

$$(\mu \times \nu)(E) = \int_X \nu(E_2) \, \mathrm{d}\mu = \int_Y \mu(E_1) \, \mathrm{d}\nu$$

Proof. Let us first assume that $\mu(X), \nu(Y)$ are both finite. Define the class

$$\mathscr{C} = \{E \in \mathcal{M} \otimes \mathcal{N} \text{ s.t. the theorem holds}\}$$

It is clear by simple multiplication that rectangular sets are in $\mathscr{C}$. Also it is true for finite disjoint unions thereof, by additivity. Then since the algebra generating $\mathcal{M} \otimes \mathcal{N}$ is contained in $\mathscr{C}$, we merely need to show that $\mathscr{C}$ is a monotone class.

Consider some increasing sequence $E_n \nearrow E$ with $E_n \in \mathscr{C}$. To show that $E \in \mathscr{C}$, define for each $n \in \mathbb{N}, y \in Y$, $f_n(y) = \mu((E_n)_1(y))$. Then $f_n(y) \nearrow f(y) \coloneqq \mu(E_1(y))$, which proves that f is measurable since it is the pointwise limit of measurable functions. Then by the monotone convergence theorem,

$$\begin{aligned}
\int \mu(E_1) \, \mathrm{d}\nu &= \lim_n \int_\mu ((E_n)_1) \, \mathrm{d}\nu \\
&= \lim_n (\mu \times \nu)(E_n) \\
&= (\mu \times \nu)(E)
\end{aligned}$$

For the intersection we apply the same strategy, and use the assumption that μ, ν are finite for approximating from the outside. In general this can be relaxed to σ-finiteness. □

Theorem 2.23: Tonelli's Theorem

Let $(X, \mathcal{M}, \mu), (Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces. Let $f : X \times Y \to [0, \infty]$ be measurable with respect to $\mathcal{M} \otimes \mathcal{N}$. Then

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

Proof. The previous theorem shows that this holds for indicator functions of measurable sets, and hence also for simple functions. Then taking a sequence $f_n \nearrow f$ □

Theorem 2.24: Fubini's Theorem

Let X, Y be σ-finite measure spaces as in Tonelli's Theorem. Suppose $f : X \times Y \to \mathbb{C}$ is $\mathcal{M} \otimes \mathcal{N}$ measurable and also is L^1. Then

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

A result of the combination of Tonelli's theorem and Fubini's theorem is that for any measurable $f : X \times Y \to \mathbb{C}$,

$$\int_{X\times Y} |f| \,\mathrm{d}(\mu \times \nu) = \int_X \left(x \mapsto \int_Y |f|_x \,\mathrm{d}\nu \right) \mathrm{d}\mu = \int_Y \left(y \mapsto \int_X |f|^y \,\mathrm{d}\mu \right) \mathrm{d}\nu$$

and if any of them are finite, then the same equality holds for f as well.

An important technical point is that these theorems assume that f is measurable with respect to $\mathcal{M} \otimes \mathcal{N}$, not to the completion of the product space. This is in general a more restrictive condition, and the theorem does not hold in general for functions which are only measurable with respect to $\overline{\mathcal{M} \otimes \mathcal{N}}$.

Chapter 3

Change of Variables and the Radon-Nikodym Derivative

In this chapter we investigate ways that measures can be related to each other, particularly in ways remniscient of the classical change of variables formula for functions in $\mathbb{R}^d$.

3.1 The Change of Variables Formula

In the next section we work to reprove the change of variables formula from the Riemann integral for the Lebesgue integral.

Definition 3.1

Let $(X, \mathcal{M}, \mu)$ be a measure space and $(Y, \mathcal{N})$ a measurable space. Let $\varphi : X \to Y$ be measurable. Then the **pushforward measure** μ_φ on Y is given by

$$\mu_\varphi(A) \coloneqq \mu(\varphi^{-1}(A))$$

Proposition 3.1

The pushforward measure is in fact a measure.

Proof. $\varnothing$ has finite measure since $\mu_\varphi(\varnothing) = \mu(\varnothing) = 0$. Also since we factor through φ^{-1}, which preserves disjoint unions, we have for $\{A_i\} \subseteq \mathcal{N}$ disjoint

$$\mu_\varphi\left(\bigcup_{i=1}^{\infty} A_i\right) = \mu\left(\varphi^{-1}\left(\bigcup_{i=1}^{\infty} A_i\right)\right) = \mu\left(\bigcup_{i=1}^{\infty} \varphi^{-1}(A_i)\right) = \sum_{i=1}^{\infty} \mu(\varphi^{-1}(A_i)) = \sum_{i=1}^{\infty} \mu_\varphi(A_i)$$

□

Theorem 3.2: Change of Variables

Let $(X, \mathcal{M}, \mu)$ be a measure space and $(Y, \mathcal{N})$ a measurable space, with $\varphi : X \to Y$ measurable. If $f \in L^1(Y, \mu_\varphi)$, then $f \circ \varphi \in L^1(X, \mu)$ and

$$\int_Y f \, \mathrm{d}\mu_\varphi = \int_X f \circ \varphi \, \mathrm{d}\mu$$

Proof. We begin by proving the theorem for nonnegative simple functions $f = \sum \alpha_i \chi_{A_i}$. In this case we have

$$\begin{aligned} \int_Y f \, \mathrm{d}\mu_\varphi = \int_Y \left(\sum_{i=1}^n \alpha_i \chi_{A_i} \right) \mathrm{d}\mu_\varphi &= \sum \alpha_i \mu_\varphi(A_i) = \sum \alpha_i \mu(\varphi^{-1}(A_i)) \\ &= \sum \alpha_i \int_X \chi_{\varphi^{-1}(A_i)} \, \mathrm{d}\mu = \int_X \sum \alpha_i \chi_{\varphi^{-1}(A_i)} \, \mathrm{d}\mu \end{aligned}$$

Observe that $\chi_{A_i} \circ \varphi = \chi_{\varphi^{-1}(A_i)}$ so that this becomes

$$\int_X \sum \alpha_i \chi_{A_i} \circ \varphi \, \mathrm{d}\mu = \int_X f \circ \varphi \, \mathrm{d}\mu$$

Now if $f : Y \to [0, \infty]$ is a general nonnegative measurable function, then we can take a sequence of nonnegative simple functions $f_n \nearrow f$ pointwise. By the monotone convergence theorem this follows then:

$$\int_Y f \, \mathrm{d}\mu_\varphi = \int_Y (\lim f_n) \, \mathrm{d}\mu_\varphi = \lim \int_Y f_n \, \mathrm{d}\mu_\varphi = \lim \int_X f_n \circ \varphi \, \mathrm{d}\mu$$

Now if $f_n \nearrow f$, then also $f_n \circ \varphi \nearrow f \circ \varphi$, so we can apply the monotone convergence theorem again:

$$\lim \int_X f_n \circ \varphi \, \mathrm{d}\mu = \int_X \lim f_n \circ \varphi \, \mathrm{d}\mu = \int_X f \circ \varphi \, \mathrm{d}\mu$$

This proves the statement for nonnegative functions and hence for complex valued functions as well. $\square$

In the above case we only required φ to be measurable. In the case that φ is injective, then we have stronger results. Specifically, if φ is injective then

$$\chi_A = \chi_{\varphi(A)} \circ \varphi$$

Therefore for $A \in \mathcal{M}$, we can localize the change of variables using characteristic functions:

$$\int_A f \circ \varphi \, \mathrm{d}\mu = \int_X \chi_A f \circ \varphi \, \mathrm{d}\mu = \int_X (\chi_{\varphi(A)} \circ \varphi)(f \circ \varphi) \, \mathrm{d}\mu = \int_Y \chi_{\varphi(A)} f \, \mathrm{d}\mu_\varphi = \int_{\varphi(A)} f \, \mathrm{d}\mu_\varphi$$

Note that while we need to assume the existence of a left inverse to write $\chi_A = \chi_{\varphi(A)} \circ \varphi$, we do not actually need to assume that the inverse is measurable.

Example 3.1

Consider $\varphi(t) = t^2$ on $[-2, 2]$, which is not injective. As a result if we only look at the interval $[0, 2]$,

$$\chi_{[0,2]}(t) \neq \chi_{[0,4]}(t^2)$$

In general we have the inequality $\chi_A(t) \leq \chi_{\varphi(A)}(\varphi(t))$.

Now let λ be the Lebesgue measure on $\mathbb{R}^n$ and consider $\varphi : \mathbb{R}^n \to \mathbb{R}^n$ invertible. If f is measurable then by our work immediately above,

$$\int_A f \circ \varphi \, \mathrm{d}\lambda = \int_{\varphi(A)} f \, \mathrm{d}\lambda_\varphi$$

However the measure $\mathrm{d}\lambda_\varphi$ is in general difficult to work with, so we need to derive easier expressions for it. In particular the notation $\mathrm{d}\lambda_\varphi$ suggests that it might be somehow related to derivatives in the rough sense of a formula

$$\mathrm{d}\lambda_\varphi = \frac{\mathrm{d}\lambda_\varphi}{\mathrm{d}\lambda} \, \mathrm{d}\lambda$$

The quantity $\frac{\mathrm{d}\lambda_\varphi}{\mathrm{d}\lambda}$ is currently undefined, but as we will show in the Radon-Nikodym theorem, this quantity can be seen to exist.

3.2 L^p Spaces

Recall that for a measure space $(X, \mathcal{M}, \mu)$, we defined the L^1 space (with the alternate notations $L^1(X)$, $L^1(\mu)$, $L^1(X \to \mathbb{C}, \mu)$) to be the set[1]

$$L^1(X) = \left\{ f : X \to \mathbb{C} : f \text{ measurable}, \int_X |f| \, \mathrm{d}\mu < \infty \right\}$$

By the linearity of the integral, this is a $\mathbb{C}$-vector space. However, its dimension as such a vector space is infinite (when X is). When studying finite dimensional vector spaces, isomorphisms with $\mathbb{C}^n$ induce the Euclidean norm on any vector space. The topology induced by such a norm is independent of the isomorphism chosen, and all the topologies between spaces are therefore equivalent.

In contrast, for infinite dimensional vector spaces the choice of norm or metric is specific to the space itself.

Definition 3.2

A **norm** on a $\mathbb{C}$-vector space V is a function $\|\cdot\| : V \to [0, \infty)$ which is:

1. Homogeneous: $\|\alpha v\| = |\alpha| \|v\|$.
2. Satisfies the triangle inequality: $\|u + v\| \leq \|u\| + \|v\|$.

[1] Actually, L^1 is the space of equivalence classes of functions defined up to a set of measure zero

3. Positive definite: $\|v\| > 0$ if $\|v\| \neq 0$.

In the case of L^1, the norm is defined by

$$\|f\|_{L^1(X)} = \int_X |f| \,\mathrm{d}\mu$$

It turns out that $(L^1(X), \|f\|_{L^1(X)})$ is a completed normed $\mathbb{C}$-vector space, or a **Banach space**.

Theorem 3.3

If $f : (a, b) \to \mathbb{R}$ is bounded and Riemann integrable, then it is Lebesgue measurable and the integrals agree:

$$\int_a^b f \,\mathrm{d}x = \int_{[a,b]} f \,\mathrm{d}\lambda$$

Definition 3.3

Let $p \in [1, \infty]$ and let $f \in L^p$. Then the equivalence class of f in L^p is the set of functions which agree except possibly on a set of measure zero:

$$[f] = \{g \in L^p : \mu(\{f \neq g\}) = 0\}$$

Strictly speaking, we can distinguish L^p from the collection of equivalence classes in L^p, denoted as $\tilde{L^p}$. The L^p-norm is inherited by $\tilde{L^p}$ in the natural way, and for $\tilde{L^p}$ is in fact a complete normed space, or a Banach space. For convenience we will ignore this distinction. We can extend this definition to consider spaces $L^p(X)$ for any $p \in [1, \infty)$. The L^p **space** is defined as the set (of equivalence classes) of functions such that

$$L^p(X) = \left\{ f : X \to \mathbb{C} \text{ measurable}, \int_X |f|^p \,\mathrm{d}\mu < \infty \right\}$$

with the norm given by

$$\|f\|_p = \sqrt[p]{\int_X |f|^p \,\mathrm{d}\mu}$$

Homogeneity is essentially immediate due to the pth root. For L^1 the triangle inequality follows from the inequality for $\mathbb{C}$. To show positive definiteness (say for L^1), we need to show that if $\int_X |f| \,\mathrm{d}\mu = 0$ then $f = 0$. In general this is not strictly true; however it can be shown that f is nonzero only on a set of measure zero. As a result we identify those functions which differ on a set of measure zero, and in this case any f with $\int_X |f| \,\mathrm{d}\mu = 0$ is in the equivalence class of the constant zero function. This can be equivalently phrased by saying that f is equal to zero μ-almost everywhere.

In order to show that $\|\cdot\|_p$ is a norm for $p \neq 1$, we will need to establish some other inequalities.

Proposition 3.4: Jensen's Inequality

Let $(X, \mathcal{M}, \mu)$ be a finite measure space; that is $\mu(X) < \infty$ (where we can assume $\mu(X) = 1$ by scaling). Let $f \in L^1(X \to (a,b), \mu)$ and let $\varphi : (a,b) \to \mathbb{R}$ be convex[a]. Then

$$\varphi\left(\int_X f \,\mathrm{d}\mu\right) \le \int_X \varphi \circ f \,\mathrm{d}\mu$$

[a]A convex function satisfies $\varphi(tx + (1-t)y) \le t\varphi(x) + (1-t)\varphi(y)$ for $x, y \in (a,b)$, $t \in [0,1]$. This is the same as saying φ is concave up.

Definition 3.4

Let $p, q \in [1, \infty]$. If $\frac{1}{p} + \frac{1}{q} = 1$ then p, q are called **conjugate pairs**.

Proposition 3.5: Holder's Inequality

Let $p \in (1, \infty)$ and $(X, \mathcal{M}, \mu)$ be a measure space. Let q be the conjugate pair of p, and let $f, g : X \to \mathbb{C}$ be measurable. Then

$$\left|\int_X \overline{f} g \,\mathrm{d}\mu\right| \le \|f\|_{L^p} \|g\|_{L^q}$$

Proposition 3.6: Minkowski's Inequality

Let $p \in (1, \infty)$ and $(X, \mathcal{M}, \mu)$ be a measure space. If $f, g : X \to [0, \infty]$ be measurable. Then

$$\|f + g\|_p \le \|f\|_p + \|g\|_p$$

Definition 3.5

A **Banach space** is a normed vector space that is complete with respect to its norm.

Definition 3.6

A **Hilbert space** is a Banach space $\mathscr{H}$ equipped with an inner product $\langle \cdot, \cdot \rangle : \mathscr{H} \times \mathscr{H} \to \mathbb{C}$ which is sesquilinear (meaning it is conjugate linear in the first slot and linear in the second), and such that $\|\cdot\| = \langle \cdot, \cdot \rangle$.

A priori, both the norm and inner product are simply given to use. However, it is worth trying to find out if there is a way that the inner product may be derived from a given norm. This derivation is known as the polarization identity, However, it only is an inner product subject to the *parallelogram law*.

Definition 3.7

A norm $\|\cdot\|$ on a vector space V is said to satisfy the **parallelogram law** if for any $\psi, \varphi \in V$,

$$\|\psi + \varphi\|^2 + \|\psi - \varphi\|^2 \leq 2\|\psi\|^2 + 2\|\varphi\|^2$$

Proposition 3.7

If a Banach space V is equipped with a norm $\|\cdot\|$ that satisfies the parallelogram law, then the function $\langle\cdot,\cdot\rangle : V \times V \to \mathbb{C}$ given by the **polarization identity**

$$\langle \varphi, \psi \rangle = \frac{1}{4}\left[\|\varphi + \psi\|^2 - \|\varphi - \psi\|^2 + i\|i\psi - \varphi\|^2 - i\|i\psi + \varphi\|^2\right]$$

is an inner product on V.

In particular, any Banach space which satisfies the parallelogram law can be converted into a Hilbert space with the inner product given by the polarization identity. In fact, a stronger statement holds:

Claim 3.1

If a Banach space V does not satisfy the parallelogram law, then there is no inner product on V such that $\|\cdot\| = \sqrt{\langle\cdot,\cdot\rangle}$.

It can be shown that the parallelogram law is violated for L^p, for any $p \in [1, \infty]$ *except* $p = 2$. Thus L^2 is the only such space which is a Hilbert space (note that this holds for all $L^2(X \to \mathbb{C}, \mu)$, regardless of the underlying space X). The inner product on L^2 is given by

$$\langle f, g \rangle_{L^2} := \int_X \overline{f} g \, \mathrm{d}\mu$$

Definition 3.8

Two functions $f, g \in L^2$ are said to be **orthogonal** if $\langle f, g \rangle_{L^2} = 0$.

Definition 3.9

A **bounded operator** on a Hilbert space $(\mathscr{H}, \|\cdot\|_{\mathscr{H}})$ is a $\mathbb{C}$-linear functional $\Lambda : \mathscr{H} \to \mathbb{C}$ with finite **operator norm**

$$\|\Lambda\|_{\mathrm{op}} := \sup\{\|\Lambda\psi\|_{\mathbb{C}} : \psi \in \mathscr{H}, \|\psi\|_{\mathscr{H}} \leq 1\} < \infty$$

The following is an important result from functional analysis that we may use to study L^2:

Theorem 3.8: Riesz Representation Theorem

Let $\mathscr{H}$ be a Hilbert space. Then for any bounded operator $\Lambda : \mathscr{H} \to \mathbb{C}$, there exists a unique $\psi_\Lambda \in \mathscr{H}$ such that
$$\Lambda(\cdot) = \langle \psi_\Lambda, \cdot \rangle$$

In the context of L^2, we can use this result to express different measures as integrals against particular functions. In particular, let μ, λ be two measures on X, and consider the functional
$$L^2(\mu) \ni f \mapsto \int f \, \mathrm{d}\lambda$$
While this map is $\mathbb{C}$-linear, we need to find the conditions under which this operator is bounded. In such a case, the Riesz representation theorem tells us that we will be able to represent integration against λ as integration against μ with a specific function:
$$\int_X f \, \mathrm{d}\lambda = \int_X \overline{g_\lambda} f \, \mathrm{d}\mu$$
for $g_\lambda \in L^2(\mu)$.

3.3 Lebesgue Decomposition and The Radon-Nikodym Derivative

Given a measure space $(X, \mathcal{M}, \mu)$ and a function $f : X \to [0, \infty]$, we can define a new measure $\varphi_{\mu,f} : \mathcal{M} \to [0, \infty]$ by
$$\varphi_{\mu,f}(A) := \int_A f \, \mathrm{d}\mu$$
We can then ask the question, given a measurable space $(X, \mathcal{M})$ with a fixed measure μ, is it necessarily the case that all other measures on $(X, \mathcal{M})$ may be expressed using this construction?

It turns out that the answer is no. In particular, if $\mu(A) = 0$ for some A, then for any choice of f, $\varphi_{\mu,f}(A) = 0$ as well. Thus any measure which assigns A nonzero measure cannot be represented in this way. However, we will show that this is in fact the only barrier to such a representation.

Definition 3.10

Let $(X, \mathcal{M})$ be a measurable space with two measures μ, ν. We say that ν is **absolutely continuous** with respect to μ (denoted $\nu \ll \mu$) if whenever $\mu(E) = 0$, $\nu(E) = 0$.

Example 3.2

If $F : X \to [0, \infty]$, then $\varphi_{\mu,F} \ll \mu$.

Example 3.3

Let $\varphi : X \to X$ be measurable and moreover measure-preserving. Then the pushforward measure $\mu_\varphi : \cdot \mapsto \mu(\varphi^{-1}(\cdot))$ is absolutely continuous with respect to μ.

Definition 3.11

We say that a measure $\nu : \mathcal{M} \to \mathbb{C}$ is **concentrated** on $A \in \mathcal{M}$ if $\nu = \nu(A \cap \cdot)$.

Intuitively, the above says that all of the measure of a subset of X is contained in A. Equivalently, $\nu(E) = 0$ whenever $A \cap E = \varnothing$.

Definition 3.12

Two measures μ, ν are said to be **mutually singular** (denoted $\mu \perp \nu$) if there exist disjoint sets $A, B \in \mathcal{M}$ such that μ is concentrated on A and ν is concentrated on B.

Proposition 3.9

If $\nu \ll \mu$ and $\mu \perp \lambda$ then $\nu \perp \lambda$.

Proposition 3.10

If $\nu \ll \mu$ and $\nu \perp \mu$ then $\nu = 0$.

Theorem 3.11: Lebesgue Decomposition Theorem

Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space with μ a nonnegative measure. Let $\lambda : \mathcal{M} \to \mathbb{C}$ be a complex measure. Then there exist two unique measures $\lambda_{\text{ac}}, \lambda_{\text{s}} : \mathcal{M} \to \mathbb{C}$ such that:

- $\lambda = \lambda_{\text{ac}} + \lambda_{\text{s}}$,
- $\lambda_{\text{ac}} \ll \mu$,
- $\lambda_{\text{s}} \perp \mu$

Proof. This theorem will be proved in conjugation with Theorem 3.14 □

Note that here λ_{ac} and λ_{s} stand for the absolutely continuous and singular parts of λ, respectively.

Lemma 3.12

Let $\mu : \mathcal{M} \to [0, \infty]$ be σ-finite. Then there exists $w : X \to (0, 1)$ with $w \in L^1(\mu)$.

Proof. Since μ is σ-finite we may pick $\{E_n\} \subseteq \mathcal{M}$ such that $X = \bigcup_n E_n$ and $\mu(E_n) < \infty$. Then we can define

$$w_n(x) = \begin{cases} 0, & x \notin E_n \\ \frac{1}{2^n(1+\mu(E_n))}, & x \in E_n \end{cases}$$

Then define $w = \sum w_n$. Each w_n is measurable, and also bounded by $\frac{1}{2^n}$, so that w is bounded by 1 (it is strictly less than 1 as long as any E_n has positive measure). Also

$$\int_X w \,\mathrm{d}\mu = \sum_n \int_X w_n \,\mathrm{d}\mu = \sum_n \frac{\mu(E_n)}{2^n(1+\mu(E_n))} < \infty$$

□

Lemma 3.13

Let $\mu : \mathcal{M} \to [0, \infty)$ be a positive finite measure and $f \in L^1(X \to \mathbb{C}, \mu)$, with $F \in \mathrm{Closed}(\mathbb{C})$ such that

$$\frac{\int_E f \,\mathrm{d}\mu}{\mu(E)} \in F$$

for all E measurable with positive measure. Then $f(x) \in F$ for almost all $x \in X$.

Proof. If $F = \mathbb{C}$ we are done. Otherwise F^c is the union of countably many open balls $\{B_{\varepsilon_n}(z_n)\}_n$. We want to show that f lies outside of each ball μ-almost everywhere; that is that

$$\mu(f^{-1}(B_{\varepsilon_n}(z_n))) = 0$$

Suppose there is some n such that this is not true. Then applying the assumption to $E = f^{-1}(B_{\varepsilon_n}(z_n))$,

$$\frac{\int_{f^{-1}(B_{\varepsilon_n}(z_n))} f \,\mathrm{d}\mu}{\mu(f^{-1}(B_{\varepsilon_n}(z_n)))} \in F$$

Since $B_{\varepsilon_n}(z_n) \subseteq F^c$, we know that

$$\left| \frac{\int_{f^{-1}(B_{\varepsilon_n}(z_n))} f \,\mathrm{d}\mu}{\mu(f^{-1}(B_{\varepsilon_n}(z_n)))} - z_n \right| \geq \varepsilon_n$$

But also

$$\begin{aligned} \left| \frac{\int_{f^{-1}(B_{\varepsilon_n}(z_n))} f \,\mathrm{d}\mu}{\mu(f^{-1}(B_{\varepsilon_n}(z_n)))} - z_n \right| &= \frac{1}{\mu(f^{-1}(B_{\varepsilon_n}(z_n)))} \left| \int_{f^{-1}(B_{\varepsilon_n}(z_n))} (f - z_n) \,\mathrm{d}\mu \right| \\ &\leq \frac{1}{\mu(f^{-1}(B_{\varepsilon_n}(z_n)))} \int_{f^{-1}(B_{\varepsilon_n}(z_n))} |f - z_n| \,\mathrm{d}\mu \end{aligned}$$

But on $f^{-1}(B_{\varepsilon_n}(z_n))$ we have $|f - z_n| < \varepsilon_n$ by definition, so that

$$\frac{1}{\mu(f^{-1}(B_{\varepsilon_n}(z_n)))} \int_{f^{-1}(B_{\varepsilon_n}(z_n))} |f - z_n| \,\mathrm{d}\mu < \varepsilon_n$$

contradiction. Thus $\mu(f^{-1}(B_{\varepsilon_n}(z_n))) = 0$ for all n, and it follows that $\mu(f^{-1}(F^c)) = 0$. □

Theorem 3.14: Radon-Nikodym

Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space with μ a nonnegative measure. Let $\lambda : \mathcal{M} \to \mathbb{C}$ be a finite complex measure such that $\lambda \ll \mu$. Then there exists a unique function $h \in L^1(X \to \mathbb{C}, \mu)$ such that $\lambda = \varphi_{\mu,h}$. This function, known as the **Radon-Nikodym derivative** of λ with respect to μ, is denoted

$$h = \frac{\mathrm{d}\lambda}{\mathrm{d}\mu}$$

which satisfies

$$\int_X g \,\mathrm{d}\lambda = \int_X g \frac{\mathrm{d}\lambda}{\mathrm{d}\mu} \,\mathrm{d}\mu$$

We will first prove the theorem for those λ which are finite and positive. After doing so we will develop some results on complex measures that will allow us to prove the general case.

Proof of Lebesgue Decomposition and Radon-Nikodym for λ Finite, Positive. From Lemma 3.12, pick $w : X \to (0,1)$ with $w \in L^1(\mu)$. Define

$$\tilde{\mu}(E) := \varphi_{\mu,w}(E) = \int_E w \,\mathrm{d}\mu$$

$\tilde{\mu}(X) < \infty$ since w is L^1. Also $N \in \mathcal{M}$ has zero μ-measure if and only if it has zero $\tilde{\mu}$-measure.

Since we assume that $\lambda : \mathcal{M} \to [0, \infty)$ is positive and finite, $\tilde{\mu}$ is also positive and finite, so we can define a new positive and finite measure:

$$\varphi = \lambda + \tilde{\mu}$$

Note that this also implies that for simple functions s,

$$\int_X s \,\mathrm{d}\varphi = \int_X s \,\mathrm{d}\lambda + \int_X sw \,\mathrm{d}\mu$$

and by the monotone convergence theorem it follows that for $f : X \to \mathbb{C}$ measurable,

$$\int_X f \,\mathrm{d}\varphi = \int_X \mathrm{d}\lambda + \int_X fw \,\mathrm{d}\mu$$

Claim: The map which takes $f \in L^2(\varphi)$ to $\int_X f \,\mathrm{d}\lambda$ is a bounded $\mathbb{C}$-linear operator.

This map is $\mathbb{C}$-linear since the integral is. To show that it is bounded, if $\|f\|_{L^2(\varphi)} \leq 1$ then

$$\begin{aligned}\left|\int_X f \,\mathrm{d}\lambda\right| &\leq \int_X |f| \,\mathrm{d}\lambda - \int_X |f| \,\mathrm{d}\varphi - \int_X |f| w \,\mathrm{d}\mu \leq \int_X |f| \,\mathrm{d}\varphi = \langle |f|, 1 \rangle_{L^2(\varphi)} \\ &\leq \|f\|_{L^2(\varphi)} \|1\|_{L^2(\varphi)} = \|f\|_{L^2(\varphi)} \sqrt{\varphi(X)} \leq \sqrt{\varphi(X)}\end{aligned}$$

Thus we have shown that this map is $\mathbb{C}$-linear and bounded. In particular, the Riesz representation theorem tells us that there exists a unique $g \in L^2(\varphi)$ such that this operator is equivalently given by

$$\int_X f \,\mathrm{d}\lambda \hookleftarrow f \mapsto \langle g, f\rangle_{L^2(\varphi)} = \int_X \overline{g} f \,\mathrm{d}\varphi$$

Note that strictly speaking g is only unique φ-almost-everywhere, since Riesz really only gives us an equivalence class. We want to show that g may be chosen such that $g \geq 0$.

Since φ is positive (we can assume that the measures are not the zero measure, since the theorem holds in that case), there exists $E \in \mathcal{M}$ such that $\varphi(E) > 0$. Picking $f = \chi_E$, we then have

$$\lambda(E) = \int_X \chi_E \,\mathrm{d}\lambda = \int_E \overline{g} \,\mathrm{d}\varphi$$

As measures, we have $0 \leq \lambda \leq \varphi$, so for E we know $0 \leq \lambda(E) \leq \varphi(E)$, or equivalently

$$0 \leq \frac{\int_E \overline{g} \,\mathrm{d}\varphi}{\varphi(E)} \leq 1$$

Now by Lemma 3.13, g takes values in $[0,1]$ φ-almost everywhere, so we can pick a representative of the equivalence class which is in $[0,1]$ everywhere. In particular we may ignore the distinction between g and $\overline{g}$. This means that we have

$$\begin{aligned} \int_X f \,\mathrm{d}\lambda = \int_X fg \,\mathrm{d}\varphi &= \int_X fg \,\mathrm{d}\lambda + \int_X fgw \,\mathrm{d}\mu \\ \implies \int_X (1-g) f \,\mathrm{d}\lambda &= \int_X fgw \,\mathrm{d}\mu \end{aligned} \tag{$*$}$$

Now define $A = g^{-1}([0,1))$ and $B = g^{-1}(\{1\})$. A, B are measurable and disjoint so we may define two mutually singular measures by

$$\begin{aligned} \lambda_{\mathrm{ac}} &:= \lambda(A \cap \cdot) \\ \lambda_{\mathrm{s}} &:= \lambda(B \cap \cdot) \end{aligned}$$

Applying $(*)$ to $f = \chi_B$, we have

$$0 = \int_X \underbrace{(1-g)}_{=0} f \,\mathrm{d}\lambda = \int_X w \,\mathrm{d}\mu = \tilde{\mu}(B)$$

This tells us that $\tilde{\mu}$ is concentrated on B and hence $\lambda_s \perp \tilde{\mu} \implies \lambda_s \perp \mu$.

If we do the same for $f = \chi_E \sum_{j=0}^n g^j$ for E measurable and $n \in \mathbb{N}$, we get

$$\int_E gw \sum_{j=0}^n g^j \,\mathrm{d}\mu = \int_E (1-g) \sum_{j=0}^n g^j \,\mathrm{d}\lambda = \int_E (1 - g^{n+1}) \,\mathrm{d}\lambda$$

On B, $g = 1$ so $1-g^{n+1} = 0$. On A, $g^{n+1} \to 0$ monotonically. The sequence $\{gw \sum_{j=0}^n g^j\}$ is monotonically increasing, so it converges pointwise to some limiting function $h : X \to [0, \infty]$. Then by monotone convergence we have

$$\int_E h \,\mathrm{d}\mu = \lim \int_E gw \sum_{j=0}^n g^j \,\mathrm{d}\mu = \lim \int_E (1 - g^{n+1}) \,\mathrm{d}\lambda = \int_{E\cap A} \mathrm{d}\lambda = \lambda_{\mathrm{ac}}(E)$$

Thus for all E we have

$$\varphi_{\mu,h}(E) = \int_E h \,\mathrm{d}\mu = \lambda_{\mathrm{ac}}(E)$$

To see that $h \in L^1(\mu)$, since λ is finite we have

$$\int_X h \,\mathrm{d}\mu = \lambda_{\mathrm{a}}(X) < \infty$$

This also shows that $\lambda_{\mathrm{a}} \ll \mu$.

To demonstrate uniqueness for the Lebesgue decomposition theorem, we first need to show that the decomposition of λ into mutually singular components is unique. Indeed if we have some other $\tilde{\lambda}_{\mathrm{ac}}, \tilde{\lambda}_{\mathrm{s}}$ with

$$\tilde{\lambda}_{\mathrm{ac}} + \tilde{\lambda}_{\mathrm{s}} = \lambda = \lambda_{\mathrm{ac}} + \lambda_{\mathrm{s}}$$

and $\tilde{\lambda}_{\mathrm{ac}} \ll \mu$, $\tilde{\lambda}_{\mathrm{s}} \perp \mu$. It follows that

$$\tilde{\lambda}_{\mathrm{ac}} - \lambda_{\mathrm{ac}} = \lambda_{\mathrm{s}} - \tilde{\lambda}_{\mathrm{s}}$$

Since the LHS is absolutely continuous with respect to μ and the RHS mutually singular, both sides are zero. To show that the choice of h is unique μ-almost everywhere. If it is true that for all $E \in \mathcal{M}$,

$$\int_E h \,\mathrm{d}\mu = \int_E \tilde{h} \,\mathrm{d}\mu \implies \frac{\int_E \tilde{h} \,\mathrm{d}\mu}{\lambda_{\mathrm{ac}}(E)} = 1$$

then by Lemma 3.13 it follows that $h = \tilde{h}$ μ-almost everywhere. $\square$

3.4 Complex Measures

Our construction of the Radon-Nikodym derivative allows us to better understand complex measures. In this section, we will first consider ways to construct nontrivial complex measures. Afterward, we will consider the ways that any two complex measures may be related to each other.

First note that any finite positive measure is also a complex measure. Thus one way to construct a complex measure is to simply consider $\mu \pm i\nu$ for μ, ν finite positive measures. Alternatively, and mroe helpfully, we can consider measures which are constructed by integrated functions against other measures. For instance, let ν be a finite measure on X, and let $f \in L^1(X \to \mathbb{C}, \nu)$. Then we can define a new, complex measure by

$$\mu(S) = \int_S f \,\mathrm{d}\nu$$

Given this construction, we can ask whether any complex measure on X arises through this construction for appropriate f, ν.

Also, we can ask for more control over these functions. If μ, ν are two positive measures, then $\mu, \nu \ll \mu + \nu$. By Radon-Nikodym, we are then given f such that $f = \frac{\mathrm{d}\mu}{\mathrm{d}(\mu+\nu)}$, which we will also write as $\mathrm{d}\mu = f\,\mathrm{d}(\mu + \nu)$. So given two measures, we can write them as integrated measures with respect to a common measure.

The above construction does not directly work for complex measures, and will instead require the development of some additional tools. First, we will define a positive measure produced from any complex measure, called the variation of the measure. The intuitive idea is that if $\mathrm{d}\mu = f\,\mathrm{d}\nu$, with $f \in L^1(\nu)$, then we will define $|\mu|$ to be $\mathrm{d}|\mu| = |f|\,\mathrm{d}\nu$. However, we need to show that such a ν even exists, and also that the definition of $|\mu|$ is independent of the choice of f, ν.

Definition 3.13

Let μ be a complex measure on $(X, \mathcal{M})$. Then the **variation** of μ is the nonnegative measure $|\mu| : \mathcal{M} \to [0, \infty]$ defined by

$$|\mu|(S) \coloneqq \sup \left\{ \sum_{j=1}^{\infty} |\mu(E_j)| : \{E_n\} \text{ partitions } S \right\}$$

Proposition 3.15

The variation of a complex measure is indeed a measure.

Proof. Let $S_1, S_2, \ldots \in \mathcal{M}$ be a countable sequence of disjoint sets. We want to show that

$$|\mu| \left(\bigcup_j S_j \right) = \sum_j |\mu|(S_j)$$

($\geq$) We can assume $|\mu|(S_j) < \infty$ for all j, since otherwise we are done. Fix $\varepsilon > 0$. Then for each S_j, take a partition $\{E_{jk}\}_k \subseteq \mathcal{M}$ such that

$$\sum_k |\mu(E_{jk})| \geq |\mu|(S_j) - \frac{\varepsilon}{2^j}$$

Then the entire collection $\{E_{jk}\}_{jk}$ is a partition of $S = \bigcup_j S_j$, and

$$|\mu(S)| \geq \sum_{j,k} |\mu(E_{jk})| = \sum_j |\mu|(S_j) - \varepsilon$$

Taking $\varepsilon \to 0$, we obtain one direction.

($\leq$) Assume first that $|\mu|(S) \neq \infty$. Consider a partition $E_1, E_2, \ldots$ of S such that

$$\sum_j |\mu(E_j)| \geq |\mu|(S) - \varepsilon$$

Define $E_{jk} \coloneqq S_j \cap E_k$. Then $\{E_{jk}\}_k$ partitions S_j for each j, so by definition

$$\sum_k |\mu(E_{jk})| \leq |\mu|(S_j)$$

Summing over all j, we have

$$\sum_j |\mu|(S_j) \geq \sum_{j,k} |\mu(E_{j,k})|$$

By the triangle inequality, for any fixed k we have

$$|\mu(E_k)| = \left|\mu\left(\sum_j E_{jk}\right)\right| \leq \sum_j |\mu(E_{jk})|$$

Then it follows that

$$\sum_j |\mu|(S_j) \geq \sum_{j,k} |\mu(E_{j,k})| \geq \sum_k |\mu(E_k)| \geq |\mu|(S) - \varepsilon$$

If $|\mu|(S) = \infty$, then instead of writing $\sum_j |\mu(E_j)| \geq |\mu|(S) - \varepsilon$, we will write $\sum_j |\mu(E_j)| \geq M$ and take $M \to \infty$. □

In the proof above, we needed to consider the case $|\mu|(S) = \infty$ separately. We can show that the variation is in fact a finite measure, but it will take a few steps to show.

Lemma 3.16

Let μ be a complex measure. Then the image of μ in $\mathbb{C}$ is bounded.

Proof. Proof in Rudin. □

Lemma 3.17

There exists $C > 0$ such that for any $z_1, z_2, \ldots, z_n \in \mathbb{C}$, there exists a subset $J \subseteq \{1, 2, \ldots, n\}$ with

$$\left|\sum_{j \in J} z_j\right| \geq C \sum_{j=1}^{n} |z_j|$$

(Note that the sharp bound is $C = \frac{1}{\pi}$, but here we will prove the statement for $C = \frac{1}{4\sqrt{2}}$.)

Proof. First note we can assume $z_j \neq 0$ since it doesn't affect any value in the inequality. Each z_i lies in one of the four quadrants, and the sum of $|z_j|$ over each quadrant is at least $\frac{1}{4}\sum_{j=1}^{n} |z_j|$ for at least one of the quadrants. Since rotating the z_j does not change any of the relevant values, we can assume each z_j lies in the quadrant $\arg z \in (-\frac{\pi}{4}, \frac{\pi}{4})$ (note that rotating also allows us to assume no point lies on the boundary of the quadrants). Now if

$z_j = x_j + iy_j$, we have $|y_j| \leq |x_j|$, and also $x_j > 0$. So letting J be the set of indices of these points, we have

$$\sum_{j \in J} |z_j| = \sum_j |x_j + iy_j| \leq \sum_j \sqrt{2} x_j$$

On the other hand, we have

$$\left| \sum_{j \in J} z_j \right| = \left| \sum_{j \in J} x_j + i \sum_{j \in J} y_j \right| \geq \sum_j x_j$$

Since

$$\sum_{j \in J} |z_j| \geq \frac{1}{4} \sum_{j=1}^{n} |z_j|$$

we conclude. ☐

Proposition 3.18

The variation of any complex measure is a finite positive measure.

Proof. Let μ be a complex measure. Then by Lemma 3.16, there exists $M > 0$ such that $|\mu(A)| < M$. Let C be as in Lemma 3.17. We claim that

$$|\mu|(S) \leq MC^{-1}$$

To see this, suppose there is some $S \in \mathcal{M}$ with a partition $E_1, E_2, \ldots$ such that

$$\sum_j |\mu(E_j)| > MC^{-1}$$

Since this is true for the infinite sum we can restrict to a finite sum

$$\sum_{j=1}^{N} |\mu(E_j)| > MC^{-1}$$

Then by Lemma 3.17, there is a subset of the indices such that

$$\left| \sum_{j \in J} \mu(E_j) \right| > M$$

But the E_j are disjoint, so we have

$$\left| \mu \left(\bigsqcup_{j \in J} E_j \right) \right| > M$$

contradicting the definition of M. ☐

We can now lay out some easy properties of the variation of a measure.

Proposition 3.19

Let μ be a complex measure. Then:

1. $|\mu|(S) \geq |\mu(S)|$.
2. If ν is any positive measure satisfying $\nu(S) \geq |\mu(S)|$ for all S, then $\nu(S) \geq |\mu|(S)$ for all S.
3. If μ is positive and finite then $|\mu| = \mu$.
4. If μ is concentrated on A then so is $|\mu|$.
5. If $\nu \perp \mu$ then $|\nu| \perp |\mu|$.
6. If $\mu \ll \nu$ and ν is a positive measure, then $|\mu| \ll \nu$. Moreover if ν is σ-finite, then
$$\frac{\mathrm{d}|\mu|}{\mathrm{d}\lambda} = \left|\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}\right|$$

Note that in general we may separate a complex measure $\nu : \mathcal{M} \to \mathbb{C}$ into its real and imaginary parts, both of which are necessarily also measures:

$$\nu = \mathrm{Re}(\nu) + i\,\mathrm{Im}(\nu)$$

These measures are real-valued measures.

Definition 3.14

A **signed measure** is a measure $\mu : \mathcal{M} \to \mathbb{R}$.

Now, for signed measures we can then consider the total variation measure $|\nu|$. Using this, we can define two new measures by

$$\nu^{\pm} := \frac{1}{2}\left(|\nu| \pm \nu\right)$$

It is immediate that $\nu^{\pm}$ are both signed, but because $|\nu| \geq \nu$, they are in fact both positive measures. Moreover, we then have
$$\nu = \nu^{+} - \nu^{-}$$

which is known as the **Jordan decomposition** of ν. Then a similar decomposition works for complex measures by decomposing the real and imaginary parts separately, so that we know a general complex measure can be written as

$$\nu = \nu_1 - \nu_2 + i\nu_3 - i\nu_4$$

where each ν_j is positive. This also finally allows us to define integration against a complex measure.

Definition 3.15

Let ν be a complex measure and let $\nu = \nu_1 - \nu_2 + i\nu_3 - i\nu_4$ be its Jordan decomposition into positive measures. If $f \in L^1(X \to \mathbb{C}, |\mu|)$, then the integral of f with respect to ν is defined as

$$\int_X f \, \mathrm{d}\nu = \int_X f \, \mathrm{d}\nu_1 - \int_X f \, \mathrm{d}\nu_2 + i \int_X f \, \mathrm{d}\nu_3 - i \int_X f \, \mathrm{d}\nu_4$$

We also have the identity

$$|\nu| = \nu^+ + \nu^-$$

Moreover, while it is not in general the case that this decomposition of ν into the difference between two positive measures is unique, if $\nu = \mu^+ - \mu^-$ for some other positive measures μ^+, μ^-, then it is the case that $\mu^+ \geq \nu^+$ and $\mu^- \geq \nu^-$. In this way we can say that the Jordan decomposition is the simplest such decomposition.

A stronger theorem, called the Jordan decomposition theorem, provides a uniqueness criterion for this decomposition, which is related to the Hahn decomposition of a measure, which we will see after the next theorem.

This finally allows us to conclude the proof of the Lebesgue Decomposition Theorem and Radon-Nikodym Theorem.

Proof of Lebesgue Decomposition and Radon-Nikodym. For $\lambda = \lambda_1 - \lambda_2 + i\lambda_3 - i\lambda_4$, we observe that both Lebesgue decomposition and Radon-Nikodym derivatives are linear for positive measures, so that we can calculate the decompositions and derivatives for the positive measures first and then add them. $\square$

Theorem 3.20

Let $\mu : \mathcal{M} \to [0, \infty]$ be σ-finite, and let $\nu : \mathcal{M} \to \mathbb{C}$. Then $\nu \ll \mu$ if and only if for all $\varepsilon > 0$ there exists $\delta > 0$ such that if $A \in \mathcal{M}$ and $\mu(A) < \delta$ then $|\nu(A)| < \varepsilon$.

Proof. ($\Longleftarrow$) If $\mu(A) = 0$ then for any $\varepsilon > 0$, $|\nu(A)| < \varepsilon$. Thus $\nu(A) = 0$.

($\Longrightarrow$) Suppose That $\nu \ll \mu$ but there exists $\varepsilon > 0$ and a sequence $\{A_i\} \subseteq \mathcal{M}$ such that $\mu(A_i) > 2^{-i}$ and $|\nu(A_i)| \geq \varepsilon$. Note that in general $|\nu|(\cdot) \geq |\nu(\cdot)|$, so we also have $|\nu|(A_i) \geq \varepsilon$. Then for any n we have

$$\mu\left(\bigcup_{i=n}^{\infty} A_i\right) \leq \sum_{i=n}^{\infty} \mu(A_i) \leq \sum_{i=n}^{\infty} 2^{-i} = 2^{-n+1}$$

It follows then that $\{\bigcup_{i=n}^{\infty} A_i\}_n$ is a decreasing sequence of sets, with at least having finite measure. Thus we can see that by approximation from the outside,

$$\mu\left(\bigcap_{n=1}^{\infty} \bigcup_{i=1}^{n} A_i\right) = \lim_{n\to\infty} \mu\left(\bigcup_{i=n}^{\infty} A_i\right) \leq 0$$

Also
$$|\nu|\left(\bigcap_{n=1}^{\infty}\bigcup_{i=n}^{\infty}A_i\right)=\lim_{n\to\infty}|\nu|\left(\bigcup_{i=n}^{\infty}A_i\right)\geq\varepsilon>0$$
Therefore $|\nu|\not\ll\mu$ and thus $\nu\not\ll\mu$, contradicting the assumption. $\square$

Definition 3.16

If $A,B\subseteq X$, then the **symmetric difference** is defined as
$$A\triangle B\coloneqq(A\setminus B)\cup(B\setminus A)$$

Theorem 3.21: Hahn Decomposition Theorem

Let $\mu:\mathcal{M}\to\mathbb{R}$ be a measure. Then there exist two sets $A^\pm\in\mathcal{M}$ such that $A^+\sqcup A^-=X$ and the Jordan decomposition of $\mu=\mu^+-\mu^-$ satisfies $\mu^\pm=\mu(A^\pm\cap\cdot)$. Moreover, if B^+,B^- are two other such sets then it is the case that $\mu(A^+\triangle B^+)=\mu(A^-\triangle B^-)=0$.

Proof. We know that $\mu\lll|\mu|$, and that $\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}\in L^1(X,|\mu|)$, and also $\left|\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}\right|=1$ almost everywhere. Since μ is real, we can choose the derivaitve so that its range is ±1. Then we will set
$$\begin{aligned}A^+&\coloneqq\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}^{-1}(\{1\})\\A^-&\coloneqq\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}^{-1}(\{-1\})\end{aligned}$$
From here, if E is any measurable set then we have
$$\begin{aligned}\mu^+(E)&=\frac{1}{2}\left(|\mu|(E)+\mu(E)\right)\\&=\frac{1}{2}\int_E\left(1+\frac{\mathrm{d}\mu}{\mathrm{d}|\mu|}\right)\mathrm{d}|\mu|\\&=\frac{1}{2}\int_{E\cap A^+}2\,\mathrm{d}|\mu|+\frac{1}{2}\int_{E\cap A^-}0\,\mathrm{d}|\mu|\\&=\mu(E\cap A^+)\end{aligned}$$
and similarly for A^-. The symmetric difference clearly must have measure zero since if not, $\mu(A^-\cap(A^+\triangle B^+))\neq\mu(B^-\cap(A^+\triangle B^+))$. $\square$

Corollary 3.22

If $\mu:\mathcal{M}\to\mathbb{R}$ and $\lambda_1,\lambda_2:\mathcal{M}\to[0,\infty)$ are positive finite measures such that $\mu=\lambda_1-\lambda_2$, then $\mu^+\leq\lambda_1$ and $\mu^-\leq\lambda_2$.

Chapter 4

Lebesgue Differentiation

This chapter considers the specific setting $(X, \mathcal{M}, \mu) = (\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n), \lambda)$. We then consider how we can understand differentiation through parameters of an integral, including results such as the fundamental theorem of calculus. For instance, if f is continuous then the correspondence between the Riemann and Lebesgue integrals allows us to use the fundamental theorem of calculus to write

$$\lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \left[\int_a^{x+\varepsilon} f \, \mathrm{d}\lambda - \int_a^x f \, \mathrm{d}\lambda \right] = \lim_{\varepsilon \to 0^+} \frac{1}{\varepsilon} \int_x^{x+\varepsilon} f \, \mathrm{d}\lambda = f(x)$$

Otherwise, we can more generally write this integral in terms of the pushforward measure:

$$\int_x^{x+\varepsilon} f \, \mathrm{d}\lambda = \varphi_{\lambda,f}([x, x+\varepsilon])$$

with $\lambda([x, x+\varepsilon]) = \varepsilon$. This leads us to ask the question of which Borel measures on $\mathbb{R}^n$ it is true that

$$\lim_{\varepsilon \to 0^+} \frac{\mu(B_\varepsilon(x))}{\lambda(B_\varepsilon(x))}$$

exists.

Definition 4.1

If $\mu : \mathscr{B}(\mathbb{R}^n) \to \mathbb{C}$ is a measure then the **symmetric derivative** of μ at x with respect to λ is defined as

$$(D_\lambda \mu)(x) = \lim_{\varepsilon \to 0^+} \frac{\mu(B_\varepsilon(x))}{\lambda(B_\varepsilon)}$$

if the limit exists. The **Hardy-Littlewood maximal function** is defined as

$$(M_\lambda \mu)(x) \coloneqq \sup_{\varepsilon > 0} \frac{|\mu|(B_\varepsilon(x))}{\lambda(B_\varepsilon)}$$

In the case that $\mu = \varphi_{\lambda,f}$ for some f, we use the shorthand $M_\lambda f = M_\lambda \varphi_{\lambda,f}$.

Definition 4.2

A function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is said to be **lower semicontinuous** if for all a it is the case that
$$f^{-1}((a, \infty])$$
is open.

To see that $M_\lambda \mu$ is measurable, we can show that it is lower semicontinous. Then we observe that the proof we used for showing that continuous functions are measurable in fact only needed lower semicontinuity. Thus $M_\lambda \mu$ is measurable.

Example 4.1

$M_\lambda \lambda = 1$, since the inside of the limit is always 1. In contrast $M_\lambda \delta_{x_0} = \infty$ at x_0, and decays polynomially away from x_0:

$$(M_\lambda \delta_{x_0})(x) = \begin{cases} \sim \|x - x_0\|^{-n}, & x \neq x_0 \\ \infty, & x = x_0 \end{cases}$$
$$(D_\lambda \delta_{x_0})(x) = \begin{cases} 0, & x \neq x_0 \\ \infty, & x = x_0 \end{cases}$$

For continuous functions $f : \mathbb{R}^n \to \mathbb{C}$,

$$(D_\lambda \varphi_{\lambda,f})(x) = f(x)$$

On the other hand, we would expect the maximal function to be equal to $|f|$, (it would be if it was defined as the $\limsup_{\varepsilon \to 0^+}$), but in general it is possible for the average to be greater on some radius away from zero. So all that we can say in general is that
$$(M_\lambda \varphi_{\lambda,f})(x) \geq |f(x)|$$

Example 4.2

For fixed ε, the function
$$f(x) = \begin{cases} \frac{\|x\|}{\varepsilon}, & \|x\| < \varepsilon \\ 1, & \|x\| > \varepsilon \end{cases}$$
satisfies
$$(M_\lambda f)(0) = 1$$
but
$$f(0) = 0$$

Lemma 4.1: Vitali Covering Lemma

Let $\{x_i\}_{i=1}^N \subseteq \mathbb{R}^n, \{r_i\}_{i=1}^n \subseteq (0,\infty)$. Then there exists a collection of indices $S \subseteq \{1,\dots,N\}$ such that:

1. The collection of balls $\{B_i = B_{r_i}(x_i)\}_{i\in S}$ are pairwise disjoint.
2. $\bigcup_{i=1}^N B_i \subseteq \bigcup_{i\in S} 3B_i$ where $3B_i = B_{3r_i}(x_i)$.
3. $\lambda\left(\bigcup_{i=1}^N B_i\right) \le 3^n \sum_{i\in S}\lambda(B_i)$.

Proof. The idea is to choose the balls using a greedy algorithm, and to show that it works. First we assume without loss of generality that the r_i are increasing, so that $r_i \le r_j$ if $i \ge j$. Add 1 to S. Then add the next smallest index j such that B_j is pairwise disjoint from the elements of S, and continue until all indices have been exhausted (this process stops since N is finite).

To see that this works, pick $i \in \{1,\dots,N\}\setminus S$. Then there exists $j \in S$ with $j < i$ and $B_i \cap B_j \neq \varnothing$. Then by assumption $r_j \ge r_i$, so $\|x_j - x_i\| \le 2r_j$, and in particular $3B_j \supseteq B_i$ by a geometric argument. The measure result is clear based on the scaling of λ. ☐

Now the Vitali covering lemma will allow us to provide boudns on the maximal function (which in turn bounds the symmetric derivative).

Definition 4.3

Let $\mu : X \to \mathbb{C}$ be a measure. Then the **total variation** of μ on X is $\|\mu\| \coloneqq |\mu|(X)$.

Theorem 4.2

Let $\mu : \mathscr{B}(\mathbb{R}^n) \to \mathbb{C}$ and $a > 0$. Then

$$\lambda\left(\{x \in \mathbb{R}^n : (M_\lambda\mu)(x) > a\}\right) \le 3^n \frac{\|\mu\|}{a}$$

Proof. Write $E_a = \{(M_\lambda\mu(x) > a)\}$. Since $M_\lambda\mu$ is lower semicontinuous, $E_a \in \mathbb{R}^n$. Then by regularity of λ,

$$\lambda(E_a) = \sup_{\substack{K\in\text{Compact}(\mathbb{R}^n)\\ K\subseteq E_a}} \lambda(K)$$

Now for any $K \subseteq E_a$ compact, for all $x \in K$ we have $(M_\lambda\mu)(x) > a$ by definition of E_a. Thus by the definition of $M_\lambda\mu$ there exists $\varepsilon > 0$ such that

$$\frac{|\mu|(B_{\varepsilon_x}(x))}{\lambda(B_{\varepsilon_x})} > a \iff \lambda(B_{\varepsilon_x}) < \frac{|\mu|(B_{\varepsilon_x})(x)}{a}$$

The collection $\{B_x = B_{\varepsilon_x}(x)\}_{x\in K}$ is an open cover for K so we can select a finite subcover $B_1,\dots,B_n$. We then apply the Vitali covering lemma to pick a subcollection $S \subseteq \{1,\dots,N\}$.

This gives us

$$\lambda(K) \le \lambda\left(\bigcup_{i=1}^{N} B_i\right) \le 3^n \sum_{i\in S} \lambda(B_i) \le \frac{3^n}{a} \sum_{i\in S} |\mu|(B_{\varepsilon_x}(x)) = \frac{3^n}{a} |\mu| \left(\bigcup_{i\in S} B_i\right) \le \frac{3^n}{a} \|\mu\|$$

Taking the supremum over all K proves the result. ☐

Corollary 4.3

If $f \in L^1$ then

$$\lambda\left(\{x \in \mathbb{R}^n : (M_\lambda f)(x) > a\}\right) \le 3^n \frac{\|f\|_{L^1}}{a}$$

Proof. Follows since $\|\varphi_{\lambda,f}\| = \|f\|_{L^1}$. ☐

Definition 4.4

Let $f \in L^1(\mathbb{R}^n \to \mathbb{C}, \lambda)$. Then $x \in \mathbb{R}^n$ is said to be a **Lebesgue point** if

$$\lim_{\varepsilon\to 0^+} \frac{1}{\lambda(B_\varepsilon)} \int_{B_\varepsilon(x)} |f - f(x)| \, \mathrm{d}\lambda = 0$$

This is a generalization of the requirement that

$$D_\lambda f = \lim_{\varepsilon\to 0^+} \frac{1}{\lambda(B_\varepsilon)} \int_{B_\varepsilon(x)} f \, \mathrm{d}\lambda = f(x)$$

which is certainly the case if x is a Lebesgue points. Now, the following important theorem is an extension of Lebesgue's criterion for Riemann integrable functions to L^1 functions.

Theorem 4.4: Lebesgue Differentation Theorem

If $f \in L^1(\lambda)$ then λ-almost-all points in $\mathbb{R}^n$ are Lebesgue points.

The significance of the differentiation theorem is that it provides a method to calculate the Radon-Nikodym derivative with respect to λ:

Corollary 4.5

Let $\mu : \mathscr{B}(\mathbb{R}^n) \to \mathbb{C}$ satisfy $\mu \ll \lambda$. Then

$$\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} = D_\lambda \mu$$

λ-a.e.

Proof. $\mu \ll \lambda$, so $\frac{d\mu}{d\lambda}$ exists and is L^1. By the Lebesgue differentaiton theorem, λ-almost-all points in $\mathbb{R}^n$ are Lebesgue points. Let $x \in \mathbb{R}^n$ be such a point. Then

$$\frac{d\mu}{d\lambda}(x) = \lim_{\varepsilon \to 0^+} \frac{1}{\lambda(B_\varepsilon)} \underbrace{\int_{B_\varepsilon(x)} \frac{d\mu}{d\lambda}\, d\lambda}_{\mu(B_\varepsilon(x))} = (D_\lambda \mu)(x)$$

□

Theorem 4.6: Fundamental Theorem of Calculus I

Let $f \in L^1(\mathbb{R} \to \mathbb{C}, \lambda)$. If $x \in \mathbb{R}$ is a Lebesgue point of f, then

$$\left(\partial \int_{-\infty}^{*} f\, d\lambda\right)(x) = f(x)$$

Theorem 4.7: Fundamental Theorem of Calculus II

Let $f : [a, b] \to \mathbb{C}$ be absolutely continuous. Then f is differentiable λ-a.e. and

$$f = f(a) + \int_a^{*} f'\, d\lambda$$

4.1 Change of Variables in $\mathbb{R}^n$

Now that we have developed tools to calculate derivatives of measures more explicitly, we will return to the question of the change of variables formula so that it can be more practically useful for computation. Here we consider the case where $(X, \mathcal{M}, \mu)$ is a measure space, $(Y, \mathcal{N})$ a measurable space, and $\varphi : X \to Y, f : Y \to \mathbb{C}$. Previously we found that

$$\int_X f \circ \varphi\, d\mu = \int_Y f\, d\mu_\varphi$$

where $\mu_\varphi : \mathcal{N} \to \mathbb{C}$ is defined by $\mu_\varphi(A) = \mu(\varphi^{-1}(A))$. Also we saw that if φ was injective then it was moreover true that

$$\int_A f \circ \varphi\, d\mu = \int_{\varphi(A)} f\, d\mu_\varphi$$

In such a case, if we write $\eta = \varphi^{-1}$, then for $B \in \mathcal{N}$ it is true that

$$\int_{\eta(B)} g\, d\mu = \int_B g \circ \eta\, d\mu_{\eta^{-1}}$$

If we additionally know that $\mu_{\eta^{-1}} \ll \nu$, with $\nu : \mathcal{N} \to [0, \infty]$ σ-finite, then

$$\int_B g \circ \eta\, d\mu_{\eta^{-1}} = \int_B g \circ \eta \frac{d\mu_{\eta^{-1}}}{d\nu}\, d\nu$$

Now, in the case $\mu, \nu = \lambda$ on $X, Y = \mathbb{R}^n$, then we find (in agreement with multivariable analysis) that

$$\frac{\mathrm{d}\lambda_{\eta^{-1}}}{\mathrm{d}\lambda} = \left|\det(D\eta^{-1})\right|$$

To show this, we need to prove some scaling properties of λ under linear transformations. We have had the tools to prove this locally, which we will produce now, but the Radon-Nikodym derivative allows us to connect this with the global change of variables formula.

Proposition 4.8

Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be an invertible linear transformation. Then there exists a (unique) positive constant c_T such that

$$\lambda(T\cdot) = c_T\lambda$$

Proof. By linearity and translation invariance, we observe that

$$\lambda(M(S + x)) = \lambda(MS + Mx) = \lambda(MS)$$

so $\lambda(M\cdot)$ is a translation invariant Borel measure and hence equal to λ, up to a constant. □

Corollary 4.9

If T is orthogonal then $c_T = 1$. Hence λ is invariant under orthogonal transformations.

Proof. The unit ball is fixed under T and has finite measure. □

Proposition 4.10

Let M be an $n \times n$ matrix. Then

$$\lambda(M\cdot) = |\det(M)|\lambda$$

Proof. The proof is clearly true if M is not invertible since both sides are zero.

If M is invertible then $\lambda(M\cdot)$ is translation invariant by linearity, so the measure

$$\mu := \frac{\lambda(M\cdot)}{\lambda(M[0, 1]^n)}$$

is a normalized, translation invariant Borel measure on $\mathbb{R}^n$ and hence equal to λ. Now it just remains to show $|\det(M)| = \lambda(M[0, 1]^n)$. This is clear for diagonalizable M and in the general case it is proved using singular value decomposition. □

Lemma 4.11

Let $E \subseteq \mathbb{R}^n$ have measure zero. Let $\varphi : E \to \mathbb{R}^n$ be such that

$$\lim_{\substack{y \to x \\ y \in E}} \frac{\|\varphi(x) - \varphi(y)\|}{\|x - y\|} < \infty$$

Then $\lambda(\varphi(E)) = 0$.

Theorem 4.12

Let $\varphi : V \to \mathbb{R}^n$ be continuous with $V \subseteq \mathbb{R}^n$ open. If φ is differentiable at $x \in V$ then

$$\lim_{\varepsilon \to 0^+} \frac{\lambda(\varphi(B_\varepsilon(x)))}{\lambda(B_\varepsilon)} = |\det(D\varphi)_x|$$

Theorem 4.13

Let $V \subseteq \mathbb{R}^n$ be open. Let $\varphi : V \to \mathbb{R}^n$ be continuously differentiable and injective such that φ^{-1} is also continuously differentiable. Then $\lambda_\varphi \ll \lambda$ and

$$\frac{\mathrm{d}\lambda_\varphi}{\mathrm{d}\lambda} = |\det(D\varphi)|$$

so that for $f \in L^1 9\lambda$, $A \in \mathscr{B}(\mathbb{R}^n)$,

$$\int_{\varphi(A)} f \, \mathrm{d}\lambda = \int_A f \circ \varphi |\det(D\varphi)| \, \mathrm{d}\lambda$$

Chapter 5

Probability Theory

Now we discuss one of the chief applications of measure theory, which is probability. Measures allow us a way of defining the notation of "probability" which is by definition compatible with the boolean operations AND (disjoint unions) and NOT (complements). Note that in order to ensure compatibility with complements we need to assume that the entire space has finite measure. Thus probability theory emerges as the study specifically of finite measure spaces.

5.1 Preliminaries

Definition 5.1

A **probability space** is a measure space $(\Omega, \mathcal{M}, \mathbb{P})$ such that $\mathbb{P}(\Omega) = 1$.

Note that we can assume from the outset that the measure of the entire space is 1, since otherwise we simply normalize the measure. Note also that the convention is to use Ω to denote the underlying set, and $\mathbb{P}$ as the measure. The measurable sets take the interpretation of observable events, or events where we can calculate a probability.

Definition 5.2

A measurable function $f : \Omega \to \mathbb{C}$ for a probability space $(\Omega, \mathcal{M}, \mathbb{P})$ is called a **random variable**.

Definition 5.3

Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space. Then the **expectation** of a random variable X is

$$\mathbb{E}[X] = \int_{\Omega} X \, \mathrm{d}\mathbb{P}$$

Definition 5.4

Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space, X a random variable, and $A \in \mathscr{B}(\mathbb{C})$. Then the **probability** of $X \in A$, denoted $\mathbb{P}(X \in A)$, is given by the pushforward measure:

$$\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A)) = \mathbb{P}_X(A)$$

In the case that X is a real valued random variable, we have a natural ordering on $\mathbb{R}$:

Definition 5.5

Let $X : \Omega \to \mathbb{R}$ be a random variable. Then the **cumulative distribution function** of X is given by

$$\mathrm{cdf}(X)(t) = \mathbb{P}_X[(-\infty, t]]$$

Moreover if $\mathbb{P}_X \ll \lambda$ then the **probability density function** of X is given by

$$\mathrm{pdf}(X) = \frac{\mathrm{d}\mathbb{P}_X}{\mathrm{d}\lambda}$$

Clearly if the pdf exists then

$$\mathrm{cdf}(X)(t) = \int_{-\infty}^{t} \mathrm{pdf}(X)\,\mathrm{d}\lambda$$

Definition 5.6

Let $X : \Omega \to \mathbb{C}$ be a random variable. Then the **variance** of X is defined as

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$

The **standard deviation** is defined as

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

Definition 5.7

Let $\mathbb{P}$ be a probability measure on $\mathbb{R}$. Then the nth **moment** of $\mathbb{P}$ is defined as

$$M_n(\mathbb{P}) = \int_{\mathbb{R}} x^n\,\mathrm{d}\mathbb{P}$$

The nth moment of a random variable $X\Omega \to \mathbb{C}$ is

$$M_n(X) = \mathbb{E}[X^n]$$

The **moment generating function** of X is defined as

$$M_X(t) \coloneqq \mathbb{E}[\exp(tX)]$$

The **characteristic function** of X is defined as

$$\varphi_X(t) \coloneqq \mathbb{E}[\exp(itX)]$$

The **cumulant generating function** of X is defined as

$$K_X(t) \coloneqq \log(M_x(t))$$

The nth **cumulant** is defined as

$$K_X^{(n)}(t) = \partial_t^n \Big|_{t=0} \log(\mathbb{E}[\exp(tX)])$$

The point of the moment generating function is that

$$M_X^{(n)}(0) = \mathbb{E}[X^n] = M_n(X)$$

recovers the nth moment of X.

One way to construct random variables with prescribed pdfs is to set X to be the identity on the probability space $(\mathbb{R}, \mathscr{B}(\mathbb{R}), f\, \mathrm{d}\lambda)$, where f is the prescribed pdf.

Example 5.1

A random variable X is said to be **standard normal**, written $X \sim \mathcal{N}(0, 1)$, if X is the identity on the probability space $(\Omega, \mathcal{M}, \mathbb{P}) = (\mathbb{R}, \mathscr{B}(\mathbb{R}), f\, \mathrm{d}\lambda)$ with

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$$

Example 5.2

A random variable X is **standard Cauchy**, written $X \sim \mathrm{Cauchy}(0, 1)$, if X is the identity on the probability space $(\mathbb{R}, \mathscr{B}(\mathbb{R}), f\, \mathrm{d}\lambda)$ with

$$f(x) = \frac{1}{\pi} \frac{1}{x^2 + 1}$$

Note that no moments of X are finite.

Example 5.3

X is **uniform** on $[a, b]$, written $X \sim U(a, b)$, if it has pdf

$$f(x) = \chi_{[a,b]}(x) \frac{1}{b - a}$$

5.2 Independence

So far we have only worked with one random variable at a time, but one of the most important aspects of probability theory is its ability to analyze linked random variables. Here we will develop the notion of independent random variables.

Intuitively independence means that the outcome of one variable does not affect the other. In other words, the likelihood of the two variables assuming some given values is just given by finding the likelihood that each assumes its given value individually, and multiplying the results.

Definition 5.8

Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a measure space and $\{E_\alpha\}_{\alpha \in A} \subseteq \mathcal{M}$ be a collection of events. This collection is said to be **independent** if for any subcollection of indices $S \subseteq A$, we have

$$\mathbb{P}\left(\bigcap_{\alpha \in S} E_\alpha\right) = \prod_{\alpha \in S} \mathbb{P}(E_\alpha)$$

A collection $\{X_\alpha : \Omega \to \mathbb{C}\}_{\alpha \in A}$ of random variables is independent if for any choice of Borel sets $\{B_\alpha\}_{\alpha \in A} \subseteq \mathscr{B}(\mathbb{C})$, the collection $\{X_\alpha^{-1}(B_\alpha)\}_{\alpha \in A}$ is independent.

There is a more convenient way to characterize independence in terms of distributions. For instance, consider two random variables $X, Y : \Omega \to \mathbb{C}$. Since they are defined together on the same domain, we can package them into a new function $(X, Y) : \Omega \to \mathbb{C}^2$, which is a vector-valued random variable.

Definition 5.9

Let $X, Y : \Omega \to \mathbb{C}$ be random variables. Then the **joint probability distribution** of X, Y is defined as

$$\mathbb{P}_{(X,Y)} = \mathbb{P} \circ (X, Y)^{-1}$$

The definition is similar for any finite collection of variables.

Proposition 5.1

Let $\{X_\alpha : \Omega \to \mathbb{C}\}_{\alpha \in A}$ be a collection of random variables. Then this collection is independent if and only if for any finite subcollection $X_{\alpha_1}, \ldots, X_{\alpha_n}$,

$$\mathbb{P}_{(X_{\alpha_1}, \ldots, X_{\alpha_n})} = \prod_{i=1}^{n} \mathbb{P}_{X_{\alpha_i}}$$

where the right side is the product measure on $\mathbb{C}^n$.

Proposition 5.2

If $\{X_\alpha\}$ is a collection of independent random variables, then for any subcollection $X_{\alpha_1}, \ldots, X_{\alpha_n}$,

$$\mathbb{E}[X_{\alpha_1} \cdots X_{\alpha_n}] = \mathbb{E}[X_{\alpha_1}] \cdots \mathbb{E}[X_{\alpha_n}]$$

Proposition 5.3

If X, Y are independent then $\operatorname{Cov}(X, Y) = 0$.

One of the useful aspects of phrasing probability theory in terms of measure theory is that probability theoretic notions such as expectation, variance, and measure can be easily connected with ideas such as function spaces.

Proposition 5.4

If X, Y are both L^1, then XY is also L^1 and

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

If X, Y are nonnegative random variables then being L^1 is the same as having finite first moment.

Proposition 5.5

If X, Y are both L^2, then $X + Y$ is also L^2 and

$$\operatorname{Var}(X + Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2\operatorname{Cov}(X, Y)$$

In particular if X, Y are independent then the covariance term disappears, so we see that variances of independent variables add.

5.3 Asymptotic Analysis

Next we investigate how sequences of independent random variables display emergent order properties. This is phrased through the limit theorems.

The first goal we have is the central limit theorem, which characterizes the distribution of running average of independent variables. Essentially what it says is that if $\{X_n : \Omega \to \mathbb{R}\}$ is a collection of independent and identically distributed (IID) random variables, with

$$\mathbb{E}[X_n] = \mu \qquad \operatorname{Var}(X_n) = \sigma^2$$

Then we define

$$A_n \coloneqq \frac{1}{N} \sum_{n=1}^{N} X_n$$

Then by linearity of expectation and independence,

$$\mathbb{E}[A_N] = \mu$$
$$\operatorname{Var}(A_N) = \frac{1}{N^2} \operatorname{Var}\left(\sum_{n=1}^{N} X_n\right) = \frac{1}{N^2} \sum_{n=1}^{N} \sigma^2 = \frac{\sigma^2}{N} \overset{N\to\infty}{\longrightarrow} 0$$

Since the variance disappears in the limit, we would expect that A_N tends to the constant value μ. However, there are stronger ways that we can quantify exactly how A_N tends to the mean. In other wrods, we want to classify $\mathbb{P}_{A_N}$ given the distribution $\nu = \mathbb{P}_{X_n}$. To universalize our analysis it is helpful to standard A_N as

$$Z_N := \frac{A_N - \mu}{\sigma/\sqrt{N}}$$

so that

$$\mathbb{E}[Z_N] = 0 \qquad \operatorname{Var}[Z_N] = 1$$

Then it turns out that the distribution of Z_N converges to the standard normal distribution, regardless of the input distribution (though we have not yet specified what it means for a sequence of distributions to converge).

$$\mathbb{P}_{Z_N} \overset{N\to\infty}{\longrightarrow} \mathcal{N}(0,1)$$

The first mode of convergence that we will study is called convergence in probability.

Definition 5.10

Let $\{Y_N\}_{n\in\mathbb{N}}$ be a sequence of random variables. Then we say that Y_n **converges to Y in probability**, denoted $Y_n \overset{P}{\longrightarrow} Y$, if for any $\varepsilon > 0$,

$$\mathbb{P}(|Y_n - Y| \geq \varepsilon) \overset{N\to\infty}{\longrightarrow} 0$$

Theorem 5.6: Markov's Inequality

Let $X : \Omega \to \mathbb{R}$ be a random variable and $\varphi : \mathbb{R} \to [0,\infty)$ nondecreasing such that $\varphi \circ X \in L^1$. Then for any $a \in \mathbb{R}$,

$$\varphi(a)\mathbb{P}(X \geq a) \leq \mathbb{E}[\varphi(X)]$$

Proof. By definition,

$$\mathbb{E}[\varphi(X)] = \int_{w\in\mathbb{R}} \varphi(X(w))\,\mathrm{d}\mathbb{P}(w) = \int_{x\in\mathbb{R}} \varphi(x)\,\mathrm{d}\mathbb{P}_X(x)$$

For any $a \in \mathbb{R}$ we will split this integral:

$$\int_{x\in\mathbb{R}} \varphi(x)\,\mathrm{d}\mathbb{P}_X(x) = \underbrace{\int_{x<a} \varphi(x)\,\mathrm{d}\mathbb{P}_X(x)}_{\geq 0} + \int_{x\geq a} \varphi(x)\,\mathrm{d}\mathbb{P}_X(x)$$

$$\geq \int_{x\geq a} \varphi(a)\,\mathrm{d}\mathbb{P}_X(x) = \varphi(a) \int_{x\in a} \mathrm{d}\mathbb{P}_X(x) = \varphi(a)\mathbb{P}(X \geq a)$$ □

Theorem 5.7: Chebyshev's Inequality

Let $X : \Omega \to \mathbb{R}$ be an L^1 random variable. Then for any $\varepsilon > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \varepsilon) \leq \frac{\mathrm{Var}(X)}{\varepsilon^2}$$

Proof. We take $Y = (X - \mathbb{E}[X])^2$ and $\varphi = \mathrm{id}$, with $a = \varepsilon^2$. Then by Markov's inequality,

$$\varepsilon^2 \mathbb{P}(|X - \mathbb{E}[X]| > \varepsilon) \leq \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathrm{Var}(X)$$ □

Theorem 5.8: Weak Law of Large Numbers

Let $\{X_n : \Omega \to \mathbb{R}\}$ be a collection of IID random variables with mean μ and

$$A_N := \frac{1}{N} \sum_{n=1}^{N} X_n$$

Then $A_n \xrightarrow{P} \mu$.

Proof. By Chebyshev's inequality,

$$\mathbb{P}(|A_N - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon N} \xrightarrow{N\to\infty} 0$$ □

There are ways to weaken the assumptions of the theorem.

Theorem 5.9

Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space, $\{X_n : \Omega \to \mathbb{R}\}$ a sequence of IID random variables such that

$$\lim_{x\to\infty} x\mathbb{P}(|X_n| > x) = 0$$

and write

$$A_N - \mathbb{E}[X\chi_{[-N,N]}(X)] \xrightarrow{P} 0$$

Definition 5.11

Let $\{Y_n\}$ be a sequence of random variables and Y another random variable. We say that Y_n **converges to** Y **almost surely**, written $Y_n \xrightarrow{a.s.} Y$, if

$$\mathbb{P}\left(\lim_{n\to\infty} Y_n(x) = Y(x)\right) = 1$$

This is the same as a sequence of measurable functions converging pointwise almost everywhere.

Proposition 5.10

If $Y_N \xrightarrow{a.s.} Y$ then $Y_N \xrightarrow{P} Y$.

Proof. Suppose $Y_N \xrightarrow{a.s.} Y$. It suffices to show that for any $\varepsilon > 0$ we have

$$\mathbb{P}(|Y_N - Y| \geq \varepsilon) \xrightarrow{N\to\infty} 0$$

Fix $\varepsilon > 0$ and define

$$A_n \coloneqq \bigcup_{m\geq n} \{|Y_m - Y| \geq \varepsilon\}$$

Then $\{A_n\}$ defines a decreasing sequence of sets, so by approximation from the outside (which is always valid on a probability space),

$$\lim_{n\to\infty} \mathbb{P}(A_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$$

By monotonicity,

$$\mathbb{P}(A_n) = \mathbb{P}\left(\bigcup_{m\geq n} \{|Y_m - Y| \geq \varepsilon\}\right) \geq \mathbb{P}\left(\{Y_n - Y\} \geq \varepsilon\right)$$

So in the limit,

$$\lim_{n\to\infty} \mathbb{P}\left(|Y_n - Y| \geq \varepsilon\right) \leq \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right)$$

We just need to show the right hand side is zero. Indeed, pick $w \in \{\lim_{n\to\infty} Y_n = Y\}$. Then we have N such that for any $n \geq N$,

$$|Y_n(w) - Y(w)| < \varepsilon$$

Then by definition, $w \notin A_n$. So

$$\bigcap_{n=1}^{\infty} A_n \subseteq \left\{\lim_{n\to\infty} Y_n = Y\right\}^c$$

The right hand side has measure zero by assumption, so we are done. $\square$

Lemma 5.11: Borel-Cantelli

Let $\{E_n\}_n \subseteq \mathcal{M}$ be a sequence such that

$$\sum_{n\in\mathbb{N}} \mathbb{P}(E_n) < \infty$$

Then

$$\mathbb{P}\left(\bigcup_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k\right) = 0$$

Informally, the Borel-Cantelli lemma says that if a collection of events has finite total probability then with probability one only finitely many of them occur.

Proof. Set $F_N := \bigcup_{n\geq N} E_n$. Then

$$F_N \searrow \bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_n = \bigcap_{N\in\mathbb{N}} F_N$$

so by approximation from the outside and subadditivity,

$$\mathbb{P}\left(\bigcap_{N\in\mathbb{N}} F_N\right) = \lim_{N\to\infty} \mathbb{P}(F_N) \leq \lim_{N\to\infty} \sum_{n\geq N} \mathbb{P}(E_n)$$

Since the series converges by assumption, the limit of the tails is zero and we are done. $\square$

Lemma 5.12: Second Borel-Cantelli

Let $\{E_n\}_n \subseteq \mathcal{M}$ be an independent sequence such that

$$\sum_{n\in\mathbb{N}} \mathbb{P}(E_n) = \infty$$

Then

$$\mathbb{P}\left(\bigcap_{n\in\mathbb{N}} \bigcup_{k\geq n} E_k\right) = 1$$

Proof. We use the complement:

$$\begin{aligned}
1 - \mathbb{P}\left(\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} E_k\right) &= \mathbb{P}\left(\left(\bigcap_{n\in\mathbb{N}}\bigcup_{k\geq n} E_k\right)^c\right) \\
&= \mathbb{P}\left(\bigcup_{n\in\mathbb{N}}\bigcap_{k\geq n} E_k^c\right) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geq n} E_k^c\right)
\end{aligned}$$

Since the E_k are independent, we can take a finite subset and turn the intersection into a product. Taking the limit to infinity and applying a convergence theorem, we have

$$\begin{aligned}
\lim_{n\to\infty} \mathbb{P}\left(\bigcap_{k\geq n} E_k^c\right) &= \lim_{n\to\infty} \prod_{k\geq n} (1 - \mathbb{P}(E_k)) \\
&= \lim_{n\to\infty} \exp\left[\sum_{k\geq n} \log\left(1 - \mathbb{P}(E_k)\right)\right] \\
&\leq \lim_{n\to\infty} \exp\left[-\sum_{k\geq n} \mathbb{E}_k\right] \\
&= e^{-\infty} \\
&= 0
\end{aligned}$$

□

Lemma 5.13: Kolmogorov's Inequality

Let $\{X_n\}_n$ be a sequence of independent random variables such that $\mathbb{E}[X_n] = 0$. Let $S_n \coloneqq X_1 + \ldots + X_n$. Then for any $\varepsilon > 0, n \in \mathbb{N}$,

$$\mathbb{P}\left(\left(\max_{k\in\{1,\ldots,n\}} |S_k|\right) \geq \varepsilon\right) \leq \frac{\mathrm{Var}(S_n)}{\varepsilon^2}$$

Proof. Fix $\varepsilon > 0$ and define

$$A_k \coloneqq \{|S_k| \geq \varepsilon, |S_j| < \varepsilon \forall j < k\}$$

In other words, A_k is the set of $\omega \in \Omega$ where k is the first time that the random walk $(S_n(\omega))_n$ leaves $(-\varepsilon, \varepsilon)$.

Then

$$\left\{\left(\max_{k\leq n} |S_k|\right) \geq \varepsilon\right\} = \bigsqcup_{k=1}^{n} A_k$$

so

$$\mathbb{P}\left(\left(\max_{k\leq n}|S_k|\right) \geq \varepsilon\right) = \sum_{k=1}^{n}\mathbb{P}(A_k) = \sum_{k=1}^{n}\mathbb{E}[\chi_{A_k}] \leq \varepsilon^{-2}\sum_{k=1}^{n}\mathbb{E}[\chi_{A_k}S_k^2]$$

Now we have

$$\mathbb{E}[S_n^2] \geq \sum_{k=1}^{n}\mathbb{E}[S_n^2\chi_{A_k}] = \sum_{k=1}^{n}\mathbb{E}[\chi_{A_k}(S_k - (S_n - S_k))^2]$$
$$= \sum_{k=1}^{n}\mathbb{E}[\chi_{A_k}S_k^2] + 2\mathbb{E}[\chi_{A_k}S_k(S_n - S_k)] + \mathbb{E}[\chi_{A_k}(S_n - S_k)^2]$$

The third term is nonnegative. The second term is zero, because $\chi_{A_k}S_k$ is independent of $S_n - S_k$, and $\mathbb{E}[S_n - S_k] = 0$. Thus

$$\mathrm{Var}(S_n) = \mathbb{E}[S_n^2] \geq \sum_{k=1}^{n}\mathbb{E}[\chi_{A_k}S_k^2]$$ □

Theorem 5.14: Strong Law of Large Numbers (Kolmogorov)

Let $(\Omega, \mathcal{M}, \mathbb{P})$ be a probability space and $\{X_n : \Omega \to \mathbb{C}\}$ a collection of independent L^2 random variables such that

$$\lim_{N\to\infty}\frac{1}{N^2}\sum_{n=1}^{N}\mathrm{Var}(X_n) < \infty$$

Then defining

$$A_N = \frac{1}{N}\sum_{n=1}^{N}X_n$$

we have

$$A_n - \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[X_n] \xrightarrow{a.s.} 0$$

Proof. We first center A_N by defining

$$B_N = A_N - \frac{1}{N}\sum_{n=1}^{N}\mathbb{E}[X_n]$$

Fix $\varepsilon > 0$ and $k \in \mathbb{N}$. Then by monotonicity and Kolmogorov's inequality,

$$\mathbb{P}\left(\max_{n\in[2^{k-1},2^k]}|B_n| \geq \varepsilon n\right) \leq \mathbb{P}\left(\max_{n\in[1,2^k]}|B_n| \geq \varepsilon 2^{k-1}\right) \leq (\varepsilon 2^{k-1})^{-2}\sum_{n=1}^{2^k}\mathrm{Var}(X_n)$$

Summing over all k, we have

$$\sum_{k=1}^{\infty} \mathbb{P}\left(\max_{n \in 2^{k-1}, 2^k} |B_n| \geq \varepsilon n\right) \leq \sum_{k=1}^{\infty} \frac{1}{(\varepsilon 2^{k-1})^2} \frac{1}{4^k} \sum_{n=1}^{2^k} \operatorname{Var}(X_n)$$
$$\leq \sum_{n=1}^{\infty} \sum_{k \sim \log_2 n}^{\infty} \frac{4}{\varepsilon^2 n^2 2^{-2k}} \operatorname{Var}(X_n) = \frac{8}{\varepsilon^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \operatorname{Var}(X_n) < \infty$$

So by the first Borel-Cantelli lemma,

$$\mathbb{P}\left(\limsup_k \left\{\max_{n \in [2^{k-1}, 2^k]} |B_n| \geq \varepsilon n\right\}\right) = 0$$

Taking $\varepsilon \to 0$, we see that

$$B_N \xrightarrow{a.s.} 0$$

□

The law of large numbers (in both weak and strong forms) describes the zeroth-order asymptotic behavior of the running averages. The first-order behavior is calculated by the central limit theorem and the normal distribution.

We first note that by the law of large numbers, we can write A_N in terms of fluctuations about the mean by some variable which has variance tending to 0 as $N \to \infty$. If we remove a square-root scaling factor in N analogous to standard deviation, we write this as

$$A_N = \mu + \frac{\sigma}{\sqrt{N}} Z_N$$

where we want to argue that Z_N is asymptotically independent of N. In fact, we will see that

$$Z_n \xrightarrow{N \to \infty} \mathcal{N}(0,1)$$
$$\mathbb{P}_{Z_N} \xrightarrow{N \to \infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right) \mathrm{d}\lambda$$

The mode of convergence is the in which this occurs is the weakest which we have studied so far.

Definition 5.12

Let $\{Y_n\}_n$ be a sequence of complex random variables and Y another random variable. We say Y_n **converges in distribution** to Y, written $Y_n \xrightarrow{d} Y$, if for any bounded continuous function $f : \mathbb{C} \to \mathbb{C}$,

$$\mathbb{E}[f(Y_N)] \xrightarrow{N \to \infty} \mathbb{E}[f(Y)]$$

Proposition 5.15

If $Y_n \xrightarrow{P} Y$ then $Y_n \xrightarrow{d} Y$.

It follows then that $Y_n \xrightarrow{a.s.} Y \implies Y_n \xrightarrow{P} Y \implies Y_n \xrightarrow{d} Y$.

The next theorems characterize convergence in distribution in a manner that will make it easier to prove the central limit theorem.

Theorem 5.16: Levy's Continuity Theorem

$Y_n \xrightarrow{d} Y$ if and only if

$$\mathbb{E}[e^{itY_n}] \to \mathbb{E}[e^{itY}]$$

converges pointwise in t.

Theorem 5.17

If $\{Y_n\}_n$ and Y are real random variables then $Y_n \xrightarrow{d} Y$ if and only if the cumulative distribution functions converge pointwise in t:

$$\mathbb{P}(Y_n \leq t) \to \mathbb{P}(Y \leq t)$$

Theorem 5.18: Central Limit Theorem (IID)

Let $\{X_n : \Omega \to \mathbb{R}\}_n$ be an IID sequence of random variables with mean μ and standard deviation σ, and define

$$Z_N = \frac{A_N - \mu}{\sigma/\sqrt{N}}$$

Then

$$Z_N \xrightarrow{d} \mathcal{N}(0,1)$$

Proof. We will use Levy's continuity theorem, so that it suffices to show that

$$\varphi(Z_N)(t) = \mathbb{E}[\exp(itZ_N)] \to \varphi_Z(t) = \exp\left(-\frac{1}{2}t^2\right)$$

pointwise in t, where $Z \sim \mathcal{N}(0,1)$. We will then write

$$Y_n = \frac{X_n - \mu}{\sigma}$$

so that

$$\mathbb{E}[Y_n] = 0$$
$$\mathrm{Var}(Y_n) = 1$$
$$Z_N = \frac{\sum_{n=1}^N Y_n}{\sqrt{N}}$$

This allows us to separate the exponential and apply independence:

$$\mathbb{E}[\exp(itZ_N)] = \mathbb{E}\left[\exp\left(itN^{-\frac{1}{2}}\sum_{n=1}^{N} Y_n\right)\right]$$

$$= \mathbb{E}\left[\prod_{n=1}^{N}\exp\left(itN^{-\frac{1}{2}}Y_n\right)\right] = \prod_{n=1}^{N}\mathbb{E}\left[\exp\left(itN^{-\frac{1}{2}}Y_n\right)\right]$$

We evaluate the expectation by Taylor expanding the exponential:

$$\mathbb{E}\left[\exp\left(itN^{-\frac{1}{2}}Y_n\right)\right] = \mathbb{E}\left[1 - itN^{-\frac{1}{2}}Y_n + \frac{1}{2}\frac{t^2}{N}Y_n^2 + O\left(N^{-\frac{3}{2}}\right)\right]$$

$$= 1 - \frac{t^2}{2N} + O\left(N^{-\frac{3}{2}}\right)$$

Thus we have

$$\mathbb{E}\left[\exp\left(itZ_N\right)\right] = \exp\left(\sum_{n=1}^{N}\log\left(1 - \frac{t^2}{2N} + O\left(N^{-\frac{3}{2}}\right)\right)\right)$$

Applying the taylor expansion of $\log(1-x)$, we obtain

$$\exp\left(\sum_{n=1}^{N} -\frac{t^2}{2N} + O\left(N^{-\frac{3}{2}}\right)\right) = \exp\left(-\frac{t^2}{2} + O\left(N^{-\frac{1}{2}}\right)\right)$$

Note that the $O(\sqrt{N})$ term contains dependence on t and the moments of Y_n; however pointwise in t this term vanishes as $N \to \infty$. $\square$

The importance of this is that because we did not make any assumptions on the distribution of Y_n, we can use analysis of the Gaussian in order to study arbitrary random variables. One way that we will do this is to study tail bounds of the averages:

$$\mathbb{P}\left[|A_N - \mu| > t\right] = \mathbb{P}\left[\left|\frac{\sigma}{\sqrt{N}}Z_N\right| > t\right] = \mathbb{P}\left[|Z_N| > \frac{\sqrt{N}t}{\sigma}\right]$$

However, because there is N dependence on the right side, we cannot easily apply the central limit theorem on the distribution of Z_N to make arguments about this value (since we don't know anything about the asypmtotic convergence of $Z_N \to Z$). One way to address this is to ask questions about small deviations from the mean:

$$\mathbb{P}\left[|A_N - \mu| > \frac{t}{\sqrt{N}}\right] = \mathbb{P}\left[|Z_N| > \frac{t}{\sigma}\right]$$

In this case the central limit theorem tells us that

$$\Phi\left(\frac{s}{\sigma}\right) \sim \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{s^2}{\sigma^2}\right)$$

In order to return to the original question we need to study patterns of large deviation.

Another question that we can investigate is the higher-order asymptotic behavior of A_N. Such an expansion is called an **Edgeworth expansion**.

5.4 Large Deviations

In this section we are motivated by considering asymptotic analysis on probability measures. If $\{\mathbb{P}_N\}_N$ is a sequence of probability measures and X is a random variable which is not dependent on N. If the measures $\mathbb{P}_N$ have a density which is also a Laplace transform, then

$$\mathbb{E}_N[X] = \int_\Omega X(\omega)\,\mathrm{d}\mathbb{P}_N(\omega) = \int_\Omega X(\omega) Z_N^{-1} e^{-N\mathcal{I}(\omega)}\,\mathrm{d}\lambda(\omega)$$

This is a setup which can be analyzed with existing techniques.

Theorem 5.19: Laplace Asymptotic

Let $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{C}$. Ase f has continuous Hessian $\mathbb{H}f : \mathbb{R}^n \to M_{n\times n}(\mathbb{R})$ at some point $x_0 \in \mathbb{R}^n$, where

- g is continuous and nonvanishing at x_0,
- $(\nabla f)(x_0) = 0$,
- $(\mathbb{H}f)(x_0)$ is positive definite,
- There exists $\eta_* > 0$ such that

$$\int_{x\in\mathbb{R}^n} e^{-\eta_* f(x)} g(x)\,\mathrm{d}\lambda(x) < \infty$$

Then it follows that

$$\lim_{\eta\to\infty} \frac{\int_{x\in\mathbb{R}^n} e^{-\eta f(x)} g(x)\,\mathrm{d}\lambda(x)}{\eta^{-\frac{n}{2}} e^{-\eta f(x_0)}} = \frac{g(x_0)}{\sqrt{\det\left(\frac{1}{2\pi}(\mathbb{H}f)(x_0)\right)}}$$

The goal of large deviations principles is to understand how this analysis can be adapted for distributions where the densities are not Laplace transforms, or where a density does not exist.

The setting that we study this is on a fixed measurable space $(\Omega, \mathscr{B}(\Omega))$, where Ω is a complete separable metric space, and a family of probability measures $\{\mathbb{P}_\varepsilon\}_{\varepsilon\in(0,\infty)}$.

Definition 5.13

A **rate function** is a map $\mathcal{I} : \Omega \to [0,\infty]$ which is lower semicontinuous such that for any $\ell \in (0,\infty)$, $\mathcal{I}^{-1}([0,\ell]) \subseteq \mathrm{Compact}(\Omega)$.

Then the large deviation principles tell us under which rate functions we have an exponential decay relation

Definition 5.14

A sequence of probability measures $\{\mathbb{P}_\varepsilon\}_{\varepsilon\in(0,\infty)}$ obeys a **large deviation principle** with rate function $I : \Omega \to [0,\infty]$ if:

- For all $F \subseteq \Omega$ closed,

$$\limsup_{\varepsilon\to 0^+} \varepsilon \log\left(\mathbb{P}_\varepsilon(F)\right) \le -\inf_{\omega\in F} \mathcal{I}(\omega)$$

- For all $U \subseteq \Omega$ open,

$$\liminf_{\varepsilon\to 0^+} \varepsilon \log\left(\mathbb{P}_\varepsilon(U)\right) \ge -\inf_{\omega\in U} \mathcal{I}(\omega)$$

Proposition 5.20

Let $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ be a family of probability measures obeying a large deviation principle with rate function I. If $A \in \mathscr{B}(\Omega)$ is such that

$$\inf_{\omega\in \operatorname{int} A} I(\omega) = \inf_{\omega\in A} I(\omega) = \inf_{\omega\in \overline{A}} I(\omega)$$

then

$$\lim_{\varepsilon\to 0^+} \varepsilon \log(\mathbb{P}_\varepsilon[A]) = -\inf_{\omega\in A} I(\omega)$$

Lemma 5.21: Varadhan's Lemma

Let $\{\mathbb{P}_\varepsilon\}_{\varepsilon>0}$ satisfy an LDP with rate function I. Then for any bounded continuous random variable $X : \Omega \to \mathbb{R}$,

$$\lim_{\varepsilon\to 0^+} \varepsilon \log\left(\mathbb{E}_\varepsilon\left[\exp\left(\frac{1}{\varepsilon}X\right)\right]\right) = \sup_{\omega\in\Omega}(X(\omega) - I(\omega))$$

Schilder's theorem tells us that Brownian motion $\{B_t\}_{T>0}$ obeys a large deviation principle. The Feynman-Kac formula then tells us that

$$\exp(-T(-\Delta + V))(x,y) = \mathbb{E}\left[\delta(B_0 - x)\delta(B_T - y)\exp\left(-\int_0^T V\circ B_t \,\mathrm{d}t\right)\right]$$

where Δ is the Laplacian operator on $L^2(\mathbb{R})$ and the exponential of the operator is defined in the typical way. This gives a nice connection between PDEs and probability.

5.5 Kolmogorov Extension Theorem

Consider some fixed probability space $(\Omega, \mathcal{M}, \mathbb{P})$ and sequence of random variables $\{X_n : \Omega \to \mathbb{R}\}_n$. Then for any tuple of distinct indices $\alpha_1, \ldots, \alpha_n$, the pushforward measure defines a new probability measure on $\mathbb{R}^n$ by

$$\mathbb{P}_{(X_{\alpha_1},\ldots,X_{\alpha_n})} = \mathbb{P} \circ (X_{\alpha_1}, \ldots, X_{\alpha_n})^{-1}$$

In the reverse direction, we consider how we can build a probability space given a collection of prescribed joint distributions, such that the joint distributions indeed are the distributions of random variables on the space.

For instance, if we want to prescribe two random variables X, Y, then we need to designate the marginal distributions $\mathbb{P}_X, \mathbb{P}_Y$, then try to find a joint distribution $\mathbb{P}_{(X,Y)}$ such that the marginals are indeed $\mathbb{P}_X, \mathbb{P}_Y$.

Note that the joint distribution naturally satisfies the properties

$$\mathbb{P}_{(X,Y)}[A \times B] = \mathbb{P}_{(Y,X)}[B \times A]$$
$$\mathbb{P}_{(X,Y)}[A \times \mathbb{R}] = \mathbb{P}_X[A]$$

To formalize this idea, we let A be an arbitrary index set. For any set of distinct tuples (an injective map $\alpha : \{1, \ldots, n\} \to A$), we are prescribed a probability measure $\mu_\alpha : \mathscr{B}(\mathbb{R}^n) \to [0, 1]$. Then we want to produce a probability space $(\Omega, \mathcal{M}, \mathbb{P})$ and a collection of random variables $\{X_\beta : \Omega \to \mathbb{R}\}_{\beta \in A}$ such that for any $\alpha : \{1, \ldots, n\} \to A$ injective,

$$\mu_\alpha = \mathbb{P}_{(X_{\alpha_1},\ldots,X_{\alpha_n})}$$

The Kolmogorov consistency conditions tell us when this assembly is possible. Essentially we need the joint distributions to be compatible with each other under permutations and marginalization.

Definition 5.15

Let A be an index set. Let $\{\mu_\alpha\}_\alpha$ be a collection of Borel measures on $\mathbb{R}^n$, where $\alpha : \{1, \ldots, n\} \to A$ is injective and n ranges over all $n \leq |A|$. We say that $\{\mu_\alpha\}_\alpha$ satisfies the **Kolmogorov consistency conditions** if for any $\pi \in S_n$ and $B_1, \ldots, B_n \in \mathscr{B}(\mathbb{R})$,

$$\mu_\alpha(B_1 \times \cdots \times B_n) = \mu_{\alpha \circ \pi}(B_{\pi(1)} \times \cdots \times B_{\pi(n)})$$

and whenever $k < n$, the restriction $\alpha|_k$ of α to the first k indices satisfies

$$\mu_\alpha(B_1 \times \cdots \times B_k \times \mathbb{R}^{n-k}) = \mu_{\alpha|_k}(B_1 \times \cdots \times B_k)$$

Theorem 5.22: Kolmogorov's Extension Theorem

If $\{\mu_\alpha\}_\alpha$ satisfes the Kolmogorov consistency conditions, then there exists a unique probability space $(\Omega, \mathcal{M}, \mathbb{P})$ such that for any α,

$$\mu_\alpha = \mathbb{P} \circ (X_{\alpha_1}, \ldots, X_{\alpha_n})^{-1}$$

Moreover $\mathbb{P}$ is regular.

Proof. If $|A| = n$ is finite, then we can simply take the product measure space $(\mathbb{R}^n, \mathscr{B}(\mathbb{R}^n))$, with X_i the projection onto the ith coordinate and $\mathbb{P} = \mu_{\alpha_{\max}}$, where $\alpha_{\max}$ is the unique map from $\{1, \dots, n\}$. For arbitrary A we require more caution since we have yet to make sense of the infinite product of a measure.

We first set $\Omega = \mathbb{R}^A = \{f : A \to \mathbb{R}\}$ to be the space of A-indexed real tuples. We the σ-algebra to be the A-indexed product algebra

$$\otimes_{a \in A} \mathscr{B}(\mathbb{R}) = \sigma\left(\left\{\pi_\beta^{-1}(E_\beta) : B_\beta \in \mathscr{B}(\mathbb{R}), a \in A\right\}\right)$$

Note that this equivalent to the σ-algebra

$$\sigma\left(\left\{\prod_{a \in A} E_a : E_a \in \mathscr{B}(\mathbb{R}) \text{ and } E_a \neq \mathbb{R} \text{ for only finitely many } a\right\}\right)$$

For any finite subcollection of indices $S \subseteq A$, we define

$$\mathcal{F}_s := \sigma\left(\left\{\prod_{a \in A} E_a : E_a \in \mathscr{B}(\mathbb{R}) \text{ and } E_a \neq \mathbb{R} \text{ only for } a \in S\right\}\right)$$

We then let

$$\mathscr{A} := \bigcup_{\substack{S \subseteq A \\ |S| < \infty}} \mathcal{F}_S$$

Now, we define a premeasure $p : \mathscr{A} \to [0, 1]$. For any $E \in \mathscr{A}$, by definition there is a finite index set $S \subseteq A$ such that $E \in \mathcal{F}_S$. We suppose that $\alpha : \{1, \dots, |S|\} \to A$ enumerates S. Let $\pi_S : \Omega \to \mathbb{R}^S$ be the projection map. Then we define

$$p(E) := \mu_\alpha (\pi_S(E))$$

We claim without proof that $\mathscr{A}$ is an algebra, and also that this definition of p is well-defined as a result of the Kolmogorov consistency conditions.

Then by the Caratheodory extension theorem, there exists a unique probability measure μ_{φ_ρ} on $\sigma(\mathscr{A}) = \mathscr{B}(\Omega)$ (which is defined by the product topology on $\Omega = \mathbb{R}^A$). We take $\mathbb{P}$ to be μ_{φ_ρ}. By construction the marginal distributions agree with those prescribed. $\square$

Note that the above proof does not make it clear that μ_{φ_ρ} is regular. A proof by Folland using the Kakutani-Markov-Riesz theorem more directly develops this fact by instead considering the compact space $\Omega = (\mathbb{R} \cup \{\infty\})^A$.

We can use the Kolmogorov extension theorem to define stochastic processes such the simple random walk. First, we define $\mu_0 : \mathscr{B}(\mathbb{R}) = [0, 1]$ to be some probability measure, for instance the Bernoulli measure $\mu_0 = \frac{1}{2}(\delta_{-1} + \delta_1)$. Then we take $A = \mathbb{N}$ and for any $\alpha : \{1, \dots, n\} \to \mathbb{N}$ injective, we specify $\mu_\alpha = \mu_0^n$. The result of applying the Kolmogorov extension theorem to this initial data is $\Omega = \mathbb{R}^N$, $X_n = \pi_n$, and $\{X_n\}_n$ are IID RVs, each with distribution μ_0. Then the sum

$$S_N = \sum_{n=1}^{N} X_n$$

is called a **simple random walk** with measure μ_0.

To define Brownian motion, we take $A = [0,\infty)$ and for any $\alpha : \{1,\dots,n\} \to A$ with $(t_1,\dots,t_n) = (\alpha_1,\dots,\alpha_n)$, we specify the density for μ_α

$$\frac{\mathrm{d}\mu_\alpha}{\mathrm{d}\lambda}(x \in \mathbb{R}^n) = \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\det(K)}} \exp\left(-\frac{1}{2}\langle x, K^{-1}x\rangle_{\mathbb{R}^n}\right)$$

where the covariance matrix K is positive definite and defined by $K_{i,j} = \min(\{t_i, t_j\})$. The result of the Kolmogorov extension theorem on this data is a stochastic process $(B_t)_{t\geq 0}$. (B_t) has the property that for any finite sequence of times $0 \leq t_1 < \ldots < t_n$, $(B_{t_1},\dots,B_{t_n})$ is a random vector in $\mathbb{R}^n$ which by construction has the above distribution. From this we conclude that:

1. $B_0 = 0$ almost surely.
2. $\{B_{t_j} - B_{t_{j-1}}\}_{j=2}^n$ an independent sequence with $B_{t_j} - B_{t_{j-1}} \sim \mathcal{N}(0, \sqrt{t_j - t_{j-1}})$.

Definitions

absolutely continuous, 45
algebra, 25

Banach space, 42, 43
Borel
  measure, 31
  σ-algebra, 8
  sets, 8
bounded operator, 44

characteristic function, 66
complete, 24
concentrated, 46
conjugate pairs, 43
convergence almost surely, 71
convergence in distribution, 75
convergence in probability, 69
counting measure, 14
cumulant, 66
cumulant generating function, 66
cumulative distribution function, 65

Dirac delta measure, 14

Edgeworth expansion, 77
elementary family, 27
expectation, 64
extended real line, 9

F_σ, 8

G_δ, 8

Hardy-Littlewood maximal function, 57
Hausdorff space, 31
Hilbert space, 43

independent, 67

joint probability distribution, 67
Jordan decomposition, 54

Kolmogorov consistency conditions, 80

L^p space, 42
L^1 space, 19
large deviation principle, 79
Lebesgue
  integral, 14, 15
  measurable, 29
  measure, 29
  premeasure, 29
Lebesgue point, 60
locally compact, 32
lower semicontinuous, 58

μ-inner regular, 31
μ-outer regular, 31
measurable
  function, 6
  φ-measurable, 22
  set, 6
  space, 6
measure
  complex, 13
  locally finite, 32
  nonnegative, 13
measure space, 13
measure zero, 4
moment, 65
moment generating function, 65
monotone class, 36
  generated, 36
mutually singular, 46

norm, 41

operator norm, 44
orthogonal, 44
outer measure, 21

parallelogram law, 44
polarization identity, 44
premeasure, 25
probability, 65
probability density function, 65
probability space, 64
product σ-algebra, 33
product space, 33
pushforward measure, 39

Radon-Nikodym derivative, 48
random variable, 64
rate function, 78
rectangular subset, 33
regular, 31

σ-algebra, 6
    generated, 7
σ-compact, 31
σ-finite, 26
second countable, 31
section, 35
signed measure, 54
simple function, 11
simple random walk, 82
standard Cauchy, 66
standard deviation, 65
standard normal, 66
symmetric derivative, 57
symmetric difference, 56

topology, 6
total variation, 59
translation, 20
    invariant, 20

uniform, 66

variance, 65
variation, 51