# Measure Theory and the Radon–Nikodym Theorem

Sai Nallani

July 10, 2024

Abstract: This paper plans on motivating and explaining measure theory and the Lebesgue integral. After those initial topics have been covered, the paper will go to deeper theorems such as the Monotone Convergence theorem, the Hahn Decomposition theorem, and the heart of the paper, the Radon–Nikodym theorem.

## Contents



## 1 Introduction

Measure theory had its root in mathematicians yearning to generalize the integral in the second-half of the nineteenth century. When the question "what functions are integrable" was asked, mathematicians saw that functions had to be nice everywhere but sets with zero "measure". Initially, they saw that sets with zero "measure" are finite and countable sets. Mathematicians wanted to generalize the notion of a "measure".

The first definition of a measure was given by Cantor (in 1883) and Stolz (in 1884). Then, Peano (in 1887) and Jordan (in 1892) added more substance to the definitions. See [Tao11] for an excellent historical motivation of measure theory through the Jordan measure.

Émile Borel (in 1898) formulated some postulates of what a measure should do in his book entitled "Lecons sur la Théorie des Fonctions". A few years later in 1902, Henri Lebesgue published his dissertation "Intégrale, Longueur, Aire". Amusingly, he was advised by Borel. Lebesgue added rigor to the ideas of a measure that have been floating in the past half-century, as well as generalized the integral.

In the 1930s, Johann Radon and Otto Nikodym further advanced the field with the Radon-Nikodym theorem. This theorem provides conditions under which a measure can be derived from another measure, introducing the concept of the Radon-Nikodym derivative. Their work was crucial in formalizing the relationship between different measures and has had profound implications in functional analysis, probability, and statistics. See [Pes14] for more on the history of measure theory and the Lebesgue integral.

We will first introduce measure theory, and then we will develop Lebesgue integration theory. Next, we will go over some of the limit theorems. Then, we will explore how different measures are related to each other with the Radon–Nikodym theorem. Finally, we will end with some applications of the Radon–Nikodym theorem in probability theory.

## 2 Preliminaries

There are a few definitions to be made before proceding to measure theory. The **extended real line** refers to the set $\mathbb{R}\cup\{-\infty,+\infty\}$, which is also denoted as the interval $[-\infty,+\infty]$. In this context, we have

$$x+(+\infty)=(+\infty)+x=+\infty$$

and

$$x+(-\infty)=(-\infty)+x=-\infty$$

to hold for all $x\in\mathbb{R}$. We have

$$x\cdot(+\infty)=(+\infty)\cdot x=+\infty$$

and

$$x\cdot(-\infty)=(-\infty)\cdot x=-\infty$$

to hold for all positive $x\in\mathbb{R}$. We have

$$x\cdot(+\infty)=(+\infty)\cdot x=-\infty$$

and

$$x\cdot(\infty)=(\infty)\cdot x=+\infty$$

to hold for all negative $x\in\mathbb{R}$. Finally, we declare

$$(+\infty)+(+\infty)=+\infty,$$

$$(-\infty)+(-\infty)=-\infty,$$

$$(+\infty)\cdot(+\infty)=(-\infty)\cdot(-\infty)=+\infty,$$

$$(+\infty) \cdot (-\infty) = (-\infty) \cdot (+\infty) = -\infty,$$

and

$$0 \cdot (+\infty) = (+\infty) \cdot 0 = 0 \cdot (-\infty) = (-\infty) \cdot 0 = 0.$$

If $X$ is a set, the **power set** of $X$, $\mathcal{P}(X)$, is defined as the set of all subsets of $X$. In other words, $U \in \mathcal{P}(X)$ if and only if $U \subseteq X$.

A **metric space** $(X, d)$ is a pair with a set $X$ and a **metric** $d : X \times X \to \mathbb{R}$ satisfying the following properties for $x, y, z \in X$:

1. We have $d(x, y) = 0$ if and only if $x = y$.
2. If $x \neq y$, then $d(x, y) > 0$.
3. $d(x, y) = d(y, x)$.
4. $d$ satisfies the triangle inequality, meaning $d(x, y) \leq d(x, z) + d(z, y)$.

Intuitively, a metric space describes distances between points.

Given a metric space $(X, d)$, an open ball of radius $\epsilon$ around $x \in X$ is denoted by $B_\epsilon(x) = \{y \in X : d(x, y) < \epsilon\}$.

An $n$**-dimensional Euclidean space** is a pair $(\mathbb{R}^n, d)$ where $d$ is the $n$**-dimensional Euclidean metric or norm** defined as

$$d((x_1, x_2, \ldots, x_n), (y_1, y_2, \ldots, y_n)) = \sum_{i=1}^{n} (x_i - y_i)^2.$$

Euclidean spaces are examples of metric spaces. The one dimensional metric space is the real line and absolute value function.

**Intervals** in $\mathbb{R}$ are subsets in the form $[a, b], (a, b), (a, b]$, and $[a, b)$ for $-\infty \leq a \leq b \leq +\infty$. A $n$**-dimensional interval** $I$ in $\mathbb{R}^n$ is in the form $I = I_1 \times I_2 \times \cdots \times I_n$ where each $I_i$ is an interval in $\mathbb{R}$ for $1 \leq i \leq n$. Define the **length** of an interval $I \subset \mathbb{R}$ as $b - a$, and define the **volume** of a $n$-dimensional interval $I = I_1 \times I_2 \times \cdots \times I_n$ as the product of the lengths of the intervals $I_i$ and denote the volume as $\mathrm{vol}(I)$.

# 3 Measures and $\sigma$-algebras

## 3.1 Introduction

Measure theory is well motivated by considering the problem of how to assign things like lengths, areas, and volumes to sets. For intervals in the standard real line $\mathbb{R}$, we can just consider its length. For example, a closed interval $[a, b] \subseteq \mathbb{R}$ has length $b - a$ for $b \geq a$. This gets trickier when trying to assign a "length" to subsets like $\mathbb{Q}$ or $\mathbb{N}$. A **measure** is a map from subsets of a set to the extended positive real line. An ideal measure $\mu$ on $\mathbb{R}$ should satisfy the following properties:

1. The domain of $\mu$ is $\mathcal{P}(\mathbb{R})$. In other words, $\mu$ can take any subset of $\mathbb{R}$ as an input.
2. If $I \subset \mathbb{R}$ is an interval, then $\mu(I)$ is the length of the interval.
3. If $S_1, S_2, \ldots$ is a countable collection of pairwise disjoint sets, then

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

4. (Translation Invariant) Let $A \subseteq \mathbb{R}$ and define $A + x = \{a + x : a \in A\}$ for all $x \in \mathbb{R}$. Then we want $\mu(A) = \mu(A + x)$. Moving the set on the number line shouldn't change its measure.

Unfortunately, we can only dream of such a measure; see [Fol99, Chapter 1] for a proof of the nonexistence of such measure. Defining $\mu$ on $\mathcal{P}(\mathbb{R})$ is not possible, so we will need to define $\mu$ on a smaller or coarser set of subsets of $\mathbb{R}$. This is where the notion of a **$\sigma$-algebra** comes in, as we restrict the domain of a measure to be a specific $\sigma$-algebra.

## 3.2 $\sigma$-algebras and Measures

**Definition 3.1.** A **$\sigma$-algebra** $\mathcal{A} \subseteq \mathcal{P}(X)$ on $X$ is a set such that

1. $\varnothing \in \mathcal{A}$
2. If $S \in \mathcal{A}$, then $S^c \in \mathcal{A}$, where $S^c = X \setminus S$ is the complement of $S$.
3. If $S_1, S_2, \ldots$ is a countable collection of elements in $\mathcal{A}$, then

$$\bigcup_{n=1}^{\infty} S_n \in \mathcal{A} \text{ and } \bigcap_{n=1}^{\infty} S_n \in \mathcal{A}.$$

The pair $(X, \mathcal{A})$ is called a **measurable space**.

**Definition 3.2.** Given a set $X$ and a $\sigma$-algebra $\mathcal{A}$, a **measure** on $(X, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, +\infty]$ such that

1. $\mu(\varnothing) = 0$
2. If $S_1, S_2, \ldots$ is a countable collection of pairwise disjoint elements in $\mathcal{A}$, then

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

**Definition 3.3.** Let $X$ be a set, $\mathcal{A}$ a $\sigma$-algebra on $X$, and $\mu$ a measure on $\mathcal{A}$. The triple $(X, \mathcal{A}, \mu)$ is a **measure space**.

**Example 3.4.** Let $X$ be non-empty, and take $x \in X$. Let us again take $\mathcal{A} = \mathcal{P}(X)$ Then the **point-mass measure** $\mu_x : \mathcal{A} \to [0, \infty]$ is defined as

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

The measure space is $(X, \mathcal{P}(X), \mu_x)$.

**Example 3.5.** Given a set $X$ and $\sigma$-algebra $\mathcal{A} = \mathcal{P}(X)$ we can define a measure $\mu : \mathcal{A} \to [0, \infty]$ by $\mu(A) = |A|$ for finite sets $A$. If $A$ is infinite, we can let $\mu(A) = \infty$. $\mu$ is known as the **counting measure**. The measure space is $(X, \mathcal{P}(X), \mu)$.

**Example 3.6.** Let $X$ be finite, and let the $\sigma$-algebra be $\mathcal{A} = \mathcal{P}(X)$. Let $\mu : \mathcal{A} \to [0, 1]$ be defined as $\mu(A) = \frac{|A|}{|X|}$. This is an example of something called a **probabilty measure**!

A big application of measure theory is probability theory. Since the probability of *anything* happening is 1, we want the probability measure of the entire space to be 1. In fact, that is qualifying condition for a probability space.

**Definition 3.7.** Let $(X, \mathcal{A}, \mu)$ be a measure space. If $\mu(X) = 1$, then $(X, \mathcal{A}, \mu)$ is a **probability space** and $\mu$ is a **probability measure**.

**Example 3.8.** Say we roll a six-sided die three times. The space we are working with is $(X = \{1, 2, 3, 4, 5, 6\}^3, \mathcal{A} = \mathcal{P}(X), \mu : \mathcal{A} \to [0, 1])$ where $\mu(A) = \frac{|A|}{216}$. Since $\mu(X) = \frac{|X|}{216} = 1$, we are working in a probability space.

**Remark 3.9.** While the framework of a probability space may be overkill in discrete scenarios, it is certainly useful in continuous situations. See [Example after borel sigma algebra] for how this is useful.

**Definition 3.10.** Let $(X, \mathcal{A}, \mu)$ be a measure space. If $A \in \mathcal{A}$ and $\mu(A) = 0$, then $A$ is called a **null set**.

## 3.3 Borel $\sigma$-algebra

We now turn to a useful $\sigma$-algebra called the Borel $\sigma$-algebra, named after French mathematician Émile Borel. First, we define what open and closed sets are.

**Definition 3.11.** Given a set $X$ endowed with a metric $d$, a set $U \subseteq X$ is **open** if for all $x \in U$, there exists an $\epsilon > 0$ such that $B_\epsilon(x) \subseteq U$, where $B_\epsilon(x)$ is the open ball of radius $\epsilon$ around $x$. If $V \subseteq X$, and $V^c$ is open, then $V$ is **closed**.

In $\mathbb{R}$, this means a set $A \subseteq \mathbb{R}$ is open if for all $x \in U$, we have $(x - \epsilon, x + \epsilon) \subseteq A$ for some $\epsilon > 0$. For higher dimensional Euclidean spaces such as $\mathbb{R}^2$ and $\mathbb{R}^3$, we use the metric space definition of open sets under the Euclidean norm.

**Definition 3.12.** The **Borel $\sigma$-algebra** $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ is the smallest $\sigma$-algebra generated by all open sets of $\mathbb{R}$. In general, the Borel $\sigma$-algebra $\mathcal{B}(\mathbb{R}^n)$ on the $n$-dimensional Euclidean space $(\mathbb{R}^n, d)$, where $d$ is the $n$-dimensional Euclidean norm is the smallest $\sigma$-algebra generated by all open sets of $\mathbb{R}^n$. A set is a **Borel** set if it is in the Borel $\sigma$-algebra.

**Remark 3.13.** For those of us with a little bit of a background in topology, we need not define a Borel $\sigma$-algebra just on $\mathbb{R}^n$. We can define it on all topologies by saying that the Borel $\sigma$-algebra on a set $X$ is the smallest $\sigma$-algebra containing all open sets in $X$. The open sets are defined by the topology over $X$. Note that the Borel $\sigma$-algebra over $\mathbb{R}^n$ is the most useful under the standard topology.

What does "smallest $\sigma$-algebra generated by all open sets" mean? First, we start by saying $S$ is the set of all open sets in $\mathbb{R}^n$. We generate the $\sigma$-algebra $\mathcal{B}(\mathbb{R}^n)$ by taking complements, countable unions, and countable intersections of all sets in $S$. The Borel $\sigma$-algebra over $\mathbb{R}$ contains all the typical intervals that we encounter.

**Proposition 3.14.** If $I \subseteq \mathbb{R}$ is an interval, then $I \in \mathcal{B}(\mathbb{R})$.

*Proof.* We split this into three cases. Since $I$ is an interval, it is in the form $[a, b]$, $(a, b)$, $(a, b]$, or $[a, b)$ for $-\infty \leq a \leq b \leq +\infty$. Since $(a, b)$ is open, it is already in $\mathcal{B}(\mathbb{R})$. Therefore, we can just concern ourselves with the other cases. Note that $[a, b] = ((-\infty, a) \cup (b, +\infty))^c$. We also notice $(-\infty, b] \in \mathcal{B}(\mathbb{R})$ as $(b, +\infty)$ is open and $(-\infty, b]$ is the complement of $(b, +\infty)$. A similar argument applies to $[a, +\infty)$. With this, we note that $(a, b] = (-\infty, b] \cap (-\infty, a]^c$ and $[a, b) = [a, +\infty) \cap (-\infty, b)$. By closure, we find all the intervals to be in $\mathcal{B}(\mathbb{R})$. □

Intuitively, the Borel $\sigma$-algebra seems like a good domain for a measure as we start with intervals, and intervals are easy to measure. We will construct a measure on $\mathcal{B}(\mathbb{R}^n)$ in the next section. While $\mathcal{B}(\mathbb{R}^n)$ is a good starting point, it is not the most useful for us as it is rather small.

### 3.4 Outer measure and Lebesgue Measure

The motivation for this section is the search for a measure on the Borel $\sigma$-algebra such that the measure of any interval on $\mathbb{R}$ is its length, the measure of any 2-dimensional interval on $\mathbb{R}$ is its area, and the measure of any 3-dimensional interval on $\mathbb{R}^3$ is its volume. First, we will work just on $\mathbb{R}$, and then extend it to $\mathbb{R}^n$. Before we construct that measure, we need to construct a function on $\mathcal{P}(\mathbb{R})$ that is similar to a measure. While this function won't be a measure because it will not fulfill all the properties of a measure, if we restrict the domain of the function to the Borel $\sigma$-algebra, then it will be a measure.

**Definition 3.15** (Outer measure)**.** Let $X$ be a set. A function $\mu^* : \mathcal{P}(X) \to [0, \infty]$ is an **outer measure** if it has the following properties:

1. $\mu^*(\varnothing) = 0$
2. $\mu^*$ is **monotonic**, meaning if $A \subseteq B \subseteq X$, then $\mu^*(A) \leq \mu^*(B)$.
3. If $A_1, A_2, A_3, \dots$ is a countable collection of sets, then we have
$$\mu^*\left(\bigcup_{n=1}^{\infty} A_n\right) \leq \sum_{n=1}^{\infty} \mu^*(A_n).$$
   This property is called countable subadditivity.

To show that an outer measure $\mu^*$ is a measure, we restrict the domain of $\mu^*$ to a $\sigma$-algebra $\mathcal{A}$. In other words, we define $\mu(A) = \mu^*(A)$ if and only if $A \in \mathcal{A}$. Otherwise, it is undefined. Then to show that the measure of a disjoint union of sets is the sum of the measures of each of the disjoint sets, or $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$, we just need to show that $\mu(\bigcup_n A_n) \geq \sum_n \mu(A_n)$ because we already have the other direction due to countable subadditivity of $\mu^*$. The most important outer measure we are concerned with is the **Lebesgue outer measure**.

**Definition 3.16.** (Lebesgue outer measure on $\mathbb{R}$) The **Lebesgue outer measure** on $\mathbb{R}$, denoted by $\lambda^*$, is defined as follows. For each subset $A$ of $\mathbb{R}$, let $\mathcal{C}_A$ be the set of all infinite sequences $\{(a_i, b_i)\}$ of bounded open intervals such that $A \subseteq \cup_i (a_i, b_i)$. In other words, $\mathcal{C}_A$ contains sequences of all open intervals such that their union contains $A$. Then, $\lambda^* : \mathcal{P}(\mathbb{R}) \to [0, \infty]$ is defined by
$$\lambda^*(A) = \inf\left\{\sum_i (b_i - a_i) : \{(a_i, b_i)\} \in \mathcal{C}_A\right\}.$$

We will now prove that the Lebesgue outer measure is really an outer measure.

**Proposition 3.17.** The Lebesgue outer measure on $\mathbb{R}$ is an outer measure, and it assigns to each interval of $\mathbb{R}$ its length.

*Proof.* Let us first verify that $\lambda^*$ is an outer measure. Since for all $\epsilon > 0$, we can choose a sequence of open intervals $\{(a_i, b_i)\}$ (whose union includes $\varnothing$) such that $\sum_i (b_i - a_i) < \epsilon$ (for example, we can let $b_i - a_i = \frac{\epsilon}{n^i}$ for an integer $n$, so we have a convergent geometric series), we have that $\lambda^*(\varnothing) = 0$.

To prove that $\lambda^*$ is monotonic, let $A \subseteq B \subseteq \mathbb{R}$. Then if a sequence of open intervals covers $B$, it covers $A$. Thus, we have $\lambda^*(A) \leq \lambda^*(B)$.

Finally, we prove countable subadditivity. Let $\{A_n\}$ be an arbitrary sequence of sets in $\mathbb{R}$. If $\sum_n \lambda^*(A_n) = +\infty$, then we have $\lambda^*(\cup_n A_n) \leq \sum_n \lambda^*(A_n) = +\infty$. So suppose that $m = \sum_n \lambda^*(A_n)$ is finite, and choose arbitrary $\epsilon > 0$. For each $n$, choose a sequence $\{(a_{n,i}, b_{n,i})\}_{i=1}^{\infty}$ that covers $A_n$ and satisfies

$$\sum_{i=1}^{\infty} (b_{n,i} - a_{n,i}) < \lambda^*(A_n) + \frac{\epsilon}{2^n}. \tag{1}$$

Such a sequence must exist because $\lambda^* A_n \leq \sum_{i=1}^{\infty} (b_{n,i} - a_{n,i})$, as $\lambda^* A_n$ is the infinum of the sum of lengths of the intervals, and if such a sequence $\{(a_{n,i}, b_{n,i})\}$ didn't exist, we would have $\lambda^*(A_n) + \frac{\epsilon}{2^n} > \lambda^*(A_n)$ to be the infinum instead, leading to a contradiction. Enumerate the pairs $\{(n,i) : \{(a_{n,i}, b_{n,i}\}$ covers $A_n\}$ by $\{j_1, j_2, \dots\}$. Then we see that each $A_n$ is covered by some $\{(a_{j_k}, b_{j_k})\}$, so we have

$$\bigcup_{n=1}^{\infty} A_n \subseteq \bigcup_{k=1}^{\infty} (a_{j_k}, b_{j_k}).$$

By monotonicity and (1), we have

$$\lambda^* \left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{k=1}^{\infty} (b_{j_k} - a_{j_k}) < \sum_{n=1}^{\infty} \left( \lambda^*(A_n) + \frac{\epsilon}{2^n} \right) = \sum_{n=1}^{\infty} \lambda^*(A_n) + \epsilon.$$

Since $\epsilon$ is arbitrary, we conclude that $\lambda^* \left( \bigcup_{n=1}^{\infty} A_n \right) \leq \sum_{n=1}^{\infty} \lambda^*(A_n)$. Therefore $\lambda^*$ satisfies countable subadditivity and it is an outer measure.

Now we make sure the Lebesgue outer measure of intervals is their length. Let $I$ be an interval with endpoints $a$ and $b$. For any $\epsilon > 0$, $I$ is contained in $(a - \epsilon, b + \epsilon)$ with length $b - a + 2\epsilon$. Since this is one of the lengths in the infinum in the definitions of $\lambda^*(I)$, we have $\lambda^*(I) \leq b - a + 2\epsilon$. Since $\epsilon$ is arbitrary, we have $\lambda^*(I) \leq b - a$. Now, cover $I$ by a sequence of open intervals $\{(x_i, y_i)\}$ such that $\sum_i (y_i - x_i)$ is finite. Let $\epsilon > 0$. and find an $n$ such that the sum of the lengths of the subintervals of $I$ not covered by the first $n$ intervals $(x_1, y_1), \dots, (x_n, y_n)$ is less than $\epsilon$, i.e. the sum of the lengths of the subintervals in $I \setminus \bigcup_{i=1}^{n} (x_i, y_i)$ is less than $\epsilon$. Then $\sum_{i=1}^{n} (y_i - x_i)$ must be at least $(b - a) - \epsilon$. Since this holds for any $\epsilon > 0$, we must have $\sum_{i=1}^{\infty} (y_i - x_i) \geq b - a$. Thus we have $\lambda^*(I) \geq b - a$, as desired. Therefore the Lebesgue outer measure of an interval in $\mathbb{R}$ is simply its length! $\square$

**Definition 3.18.** (Lebesgue outer measure on $\mathbb{R}^n$) Let $A \subseteq \mathbb{R}^n$ and let $\mathscr{S}_A$ be the set of all sequences $\{R_i\}$ of bounded and open $n$-dimensional intervals for which $A \subset \bigcup_{i=1}^{\infty} R_i$. Then the Lebesgue outer measure of $A$ is

$$\lambda^*(A) = \inf \left\{ \sum_{i=1}^{\infty} \operatorname{vol}(R_i) : \{R_i\} \in \mathscr{S}_A \right\}.$$

Let us note the following analogue of the previous proposition.

**Proposition 3.19.** The Lebesgue outer measure on $\mathbb{R}^n$ is an outer measure, and it assigns to each $n$-dimensional interval its respective volume.

*Proof.* Most of the details are omitted because they are very similar to those in the proof of Proposition 3.17. In the arguments, instead of using lengths of the intervals, we instead use the volumes of the $n$-dimensional intervals. $\square$

We have successfully constructed a natural outer measure on $\mathbb{R}^n$. However, we are not done yet, as we are trying to get to a measure. Greek mathematician Constantin Carathéodory came up with an elegant way to describe measurable sets under any outer measure $\mu^*$!

**Definition 3.20.** (Carathéodory's Condition or $\mu^*$-measurability) Let $X$ be a set, and let $\mu^* : \mathcal{P}(X) \to [0, \infty]$ be an outer measure on $X$. A set $A \subseteq X$ is **$\mu^*$-measurable** (or just measurable if the measure we are using in the context is clear), if for each arbitrary set $S \subseteq X$, we have

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

This condition may seem like it came out of nowhere, but the underlying intuition is that a set is measurable if and only if we can divide the set in such a way that the sizes (as measured by $\mu^*$) of the pieces add properly. We will now prove that all sets in the Borel $\sigma$-algebra satisfy Carathéodory's condition. Furthermore, if we are trying to prove that $\mu^*(S) = \mu^*(S \cap B) + \mu^*(S \cap B^c)$, since by countable subadditivity we already have $\mu^*(S) \leq \mu^*(S \cap B) + \mu^*(S \cap B^c)$, we will only need to prove that $\mu^*(S) \geq \mu^*(S \cap B) + \mu^*(S \cap B^c)$

**Proposition 3.21.** Let $B \subseteq X$, and let $\mu^*$ be an outer measure on $X$. If $\mu^*(B) = 0$ or $\mu^*(B^c) = 0$, then $B$ is $\mu^*$-measurable.

*Proof.* Suppose $\mu^*(B) = 0$. Let $S \subseteq X$. Then by monotonicity, we have $\mu^*(S \cap B) = 0$. By monotonicity again, we have $\mu^*(S) \geq \mu^*(S^c \cap B)$. Thus, we have $\mu^*(S) \geq \mu^*(S \cap B) + \mu^*(S \cap B^c)$, as desired. The case where $\mu^*(B^c) = 0$ is very similar. □

**Theorem 3.22.** Let $X$ be a set and $\mu^*$ be an outer measure on $X$. Let us denote $\mathscr{M}_{\mu^*}$ as the set of $\mu^*$-measurable sets. Then $\mathscr{M}_{\mu^*}$ is a $\sigma$-algebra. Furthermore, $\mu^*$ restricted to $\mathscr{M}_{\mu^*}$ is a measure.

*Proof.* The rigorous proof will be omitted as it is technical, but an outline will be provided. We have $\varnothing \in \mathscr{M}_{\mu^*}$ because $\mu^*(\varnothing) = 0$, and $\varnothing$ is $\mu^*$-measurable by Proposition 3.21. If $S \in \mathscr{M}_{\mu^*}$, then $S^c$ is also $\mu^*$-measurable. We can see this by choosing any set $A \subseteq X$ and seeing $\mu^*(A) = \mu^*(S^c \cap A) + \mu^*((S^c)^c \cap A) = \mu^*(S^c \cap A) + \mu^*(S \cap A)$ because $S$ is $\mu^*$-measurable.

Finally, we need to prove that $\mathscr{M}_{\mu^*}$ is closed under countable and finite unions. We must first prove that $\mathscr{M}_{\mu^*}$ is closed under the union of two sets (therefore finitely many sets). Then, we take a sequence of $\mu^*$-measurable sets $B_1, B_2, \ldots$ that are pairwise disjoint. We show by induction that for any set $A \subseteq X$, we have that $\mu^*(A) \geq \sum_{i=1}^n \mu^*(A \cap B_i) + \mu^*\left(A \cap \bigcup_{i=1}^n B_i^c\right)$ holds for any $n$, so it holds when the sum goes up to $+\infty$ as well. We showed that a countable union of pairwise disjoint $\mu^*$-measurable, but for $\sigma$-algebras, we need to deal with countable unions of arbitrary sets. This is fine, because a sequence of arbitrary sets can be written as a sequence of pairwise disjoint sets; if $B_1, B_2, \ldots$ is an arbitrary sequence of sets, then if

$$C_i = B_1^c \cap B_2^c \cap \cdots \cap B_{i-1}^c \cap B_i,$$

we have all $C_i$ to be $\mu^*$-measurable and we have $C_1, C_2, \ldots$ to be a sequence of pairwise disjoint sets with the same union as that of the $B_i$'s. By proving countable additivity, we can also show that $\mu^*$ restricted to $\mathscr{M}_{\mu^*}$ is a measure. □

**Proposition 3.23.** All sets in the Borel $\sigma$-algebra over $\mathbb{R}$ are $\lambda^*$-measurable, where $\lambda^*$ is the Lebesgue outer measure.

*Proof.* The Borel $\sigma$-algebra is the smallest $\sigma$-algebra containing intervals of the form $(-\infty, b]$. This is because all open intervals in the form $(a, b)$ can be represented as

$$(a, b) = (-\infty, a]^c \cap \bigcup_{n=1}^{\infty} (-\infty, b - 1/n].$$

Let $b$ be an arbitrary real number, and let $B = (-\infty, b]$. We need to check for all sets $S \subseteq \mathbb{R}$, we have

$$\lambda^*(B) \geq \lambda^*(B \cap S) + \lambda^*(B \cap S^c).$$

Let us safely assume $\lambda^*(S) < +\infty$, since otherwise both sides are $+\infty$. Take an arbitrary $\epsilon > 0$. Let $\{(a_n, b_n)\}$ be a sequence of sets covering $S$ such that the total length $(\sum_n (b_n - a_n))$ is less than $\lambda^*(S) + \epsilon$ (such a sequence always exists because $\lambda^*(S)$ is the infinum of all the lengths). For each $n$, the sets $(a_n, b_n) \cap B$ and $(a_n, b_n) \cap B^c$ are disjoint intervals whose union is $(a_n, b_n)$. Thus we have

$$b_n - a_n = \lambda^*((a_n, b_n)) = \lambda^*((a_n, b_n) \cap B) + \lambda^*((a_n, b_n) \cap B^c).$$

Since the sequence $\{(a_n, b_n) \cap B\}$ covers $S \cap B$ and $\{(a_n, b_n) \cap B^c\}$ covers $S \cap B^c$, countable subadditivity of $\lambda^*$ implies that

$$\lambda^*(S \cap B) + \lambda^*(S \cap B^c) \leq \sum_n ((a_n, b_n) \cap B) + \sum_n ((a_n, b_n) \cap B^c)$$

$$= \sum_n b_n - a_n < \lambda^*(S) + \epsilon.$$

Since $\epsilon$ is arbitrary, we have that $\lambda^*(S \cap B) + \lambda^*(S \cap B^c) = \lambda^*(S)$. □

**Proposition 3.24.** All sets in the Borel $\sigma$-algebra over $\mathbb{R}^n$ are $\lambda^*$-measurable.

*Proof.* An analogous argument to one presented in Proposition 3.22 can be modified to prove this proposition, so the proof will be omitted. □

We will be mainly working in the Lebesgue $\sigma$-algebra, denoted by $\mathscr{M}_{\lambda^*}$, which contains all sets that are $\lambda^*$-measurable over $\mathbb{R}^n$. The Lebesgue $\sigma$-algebra is **finer** than the Borel $\sigma$-algebra over $\mathbb{R}^n$, meaning $\mathscr{M}_{\lambda^*} \supset \mathcal{B}(\mathbb{R}^n)$, where $\lambda^*$ is the Lebesgue outer measure on $\mathbb{R}^n$. We call $\lambda : \mathscr{M}_{\lambda^*} \to [0, +\infty]$ defined by $\lambda(A) = \lambda^*(A)$ the **Lebesgue measure**. We call sets $S \in \mathscr{M}_{\lambda^*}$ **Lebesgue sets**.

**Remark 3.25.** There are sets $S \subseteq \mathbb{R}$ that are Lebesgue measurable but are not Borel sets. See [Coh13, Pg. 55] regarding the construction of such set.

# 4 Measurable Functions and the Lebesgue Integral

The Lebesgue measure brings life to the heart of measure theory: integration! It kind of makes sense to why measure theory is going to be related to integration. After all, we are assigning lengths, areas, and volumes to sets, so it is natural for us to try to extend this to functions.

## 4.1 The Riemann Integral

In a Calculus course, we may have learnt about the Riemann integral. If we have a bounded, non-negative valued function $f$ defined on an interval $[a, b]$, the Riemann integral is $\int_a^b f(x)dx$ measuring the area bounded by the lines $y = 0$, $x = a$, and $x = b$. If $f$ is negative, that portion of the area under $y = 0$ will have negative area. Let us just assume that $f(x) \geq 0$ for simplicity.

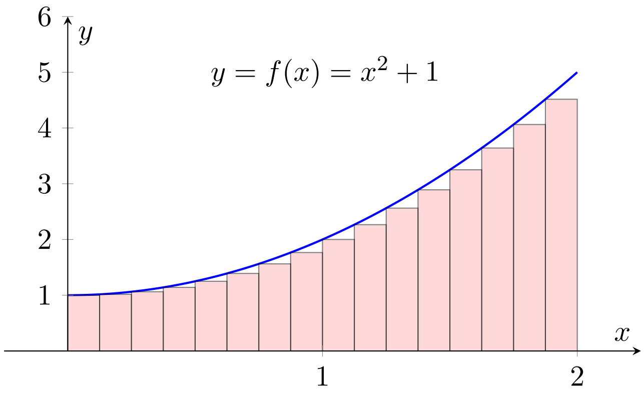



Figure 1: As the width of the rectangles gets smaller, the sum of the areas approaches the area under the curve.

**Definition 4.1** (Riemann Sum)**.** Let $[a, b]$ be an interval on $\mathbb{R}$, and let $f : [a, b] \to \mathbb{R}_{\geq 0}$. Consider any partition $P$ of $[a, b]$, i.e. a finite collection of points $t_0, t_1, \ldots, t_n$ with $a = t_0 < t_1 < t_2 < \cdots < t_n$, and define

$$\mathcal{R}_P(f)_a^b = \sum_{i=1}^{n} f(t_i)(t_i - t_{i-1}).$$

We say that $f$ is **Riemann-integrable** on $[a, b]$ if

$$\lim_{\|P\| \to 0} \mathcal{R}_P(f)_a^b$$

exists, where $\|P\| = \max(t_i - t_{i-1})$. If $f$ is Riemann-integrable on $[a, b]$, we define its **Riemann integral** to be

$$\int_a^b f(x)dx = \lim_{\|P\| \to 0} \mathcal{R}_P(f)_a^b.$$

We essentially approximate the area under $f(x)$ with vertical rectangles with very small widths. Any continuous function is Riemann-integrable. See [Pug03, Chapter 3] for a proof. More generally, a function is Riemann-integrable if it is continuous **almost everywhere** (see [Coh13, Chapter 2] for a proof).

**Definition 4.2** (Almost everywhere and almost nowhere)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $f : X \to Y$ be a function, where $Y$ is any set. Define a **property on** $f$ $P_f : X \to \{\text{TRUE}, \text{FALSE}\}$ to be a true or false function. An example of a property on $f$ is continuity; $P_f(x)$ is TRUE if $f$ is continuous at $x$ and FALSE otherwise. A property on $f$, $P_f$, holds **almost everywhere** if $\mu(\{x \in X : P_f(x) = \text{FALSE}\}) = 0$. In other words, a property holds almost everywhere if the set of points it doesn't hold has measure zero. Likewise, a property on $f$ holds **almost nowhere** if the set of points it is TRUE has measure zero.

**Example 4.3** (Single point of discontinuity)**.** A function is continuous almost everywhere if the set of discontinuous points has measure zero. On the measure space $(\mathbb{R}, \mathscr{M}_{\mu^*}, \lambda)$, which is the real line under the Lebesgue measure, an example of a function that is continuous almost everywhere is the function $f$ defined as $f(x) = \begin{cases} \sin(x) & x \neq 0 \\ 1729 & x = 0 \end{cases}$. It is discontinuous at $x = 0$, and $\lambda(0) = 0$.

An example of a function that isn't Riemann integrable is the **indicator function** on $\mathbb{Q}$.

**Definition 4.4.** Let $X$ be a set, and let $A \subseteq X$. An **indicator function** $\mathbb{1}_A$ is defined as follows:

$$\mathbb{1}_A = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$$

**Example 4.5** (Function that is not Riemann-integrable)**.** The indicator function on $\mathbb{Q}$ denoted by $\mathbb{1}_{\mathbb{Q}} : \mathbb{R} \to \mathbb{R}$ is only continuous at $x = 0$. Therefore, it is continuous almost nowhere! This function is not Riemann-integrable, but we do expect $\int_0^1 \mathbb{1}_{\mathbb{Q}}(x)dx = 0$, as there are only a few points $x \in [0,1]$ with $\mathbb{1}_{\mathbb{Q}}(x) = 1$. On most points from $[0,1]$, we seem to just have the function $f(x) = 0$.

Riemann's theory of integration seems to have some limitations, so we may need a better notion of an integral to be able to integrate functions like $\mathbb{1}_{\mathbb{Q}}$.

## 4.2 The Lebesgue Integral

Measure theory was developed to integrate functions like the ones above. Some functions should be integrable, but aren't under the Riemann definition of integration. In this section, we will build up to the Lebesgue integral in steps. We will start by defining integrals on a nice set of functions known as **simple functions**. Afterwards, we can extend the definition to a more general class of functions.

**Definition 4.6** (Integral of an indicator function)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $A \in \mathcal{A}$. The **(Lebesgue) integral** of $\mathbb{1}_A$ is defined as

$$\int_X \mathbb{1}_A d\mu = \mu(A).$$

To prove that this definition makes sense to us, we can consider the integral of $\mathbb{1}_{[0,1]}$. It is zero everywhere but the interval $[0,1]$, so the area bounded by $\mathbb{1}_{[0,1]}$ and the horizontal axis is simply $\mu([0,1]) = 1$, as the height there is also 1. This resolves our motivating problem of integrating $\mathbb{1}_{\mathbb{Q}}$, as

$$\int_{\mathbb{R}} \mathbb{1}_{\mathbb{Q}} d\lambda = \lambda(\mathbb{Q}) = 0.$$

How can we extend this from indicator functions? Since we want this new theory of integration to be better than Riemann's, we still want to preserve the Riemann's integrals nice properties of linearity; namely, we want

$$\int_X a\mathbb{1}_A d\mu = a \int_X \mathbb{1}_A d\mu = a\mu(A)$$

for real $a \in \mathbb{R}$ and we also want

$$\int_X (\mathbb{1}_A + \mathbb{1}_B) d\mu = \int_X \mathbb{1}_A d\mu + \int_X \mathbb{1}_B d\mu = \mu(A) + \mu(B).$$

Therefore, we can extend our (and Lebesgue's) integral by defining **simple functions**, which are finite linear combinations of indicator functions!

**Definition 4.7** (Simple functions)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space. A **simple function** is a function $f : X \to \mathbb{R}$ such that there exist finitely many measurable subsets $A_1, A_2, \ldots, A_n \in \mathcal{A}$ and real numbers $c_1, \ldots, c_n$ such that

$$f(x) = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}(x).$$

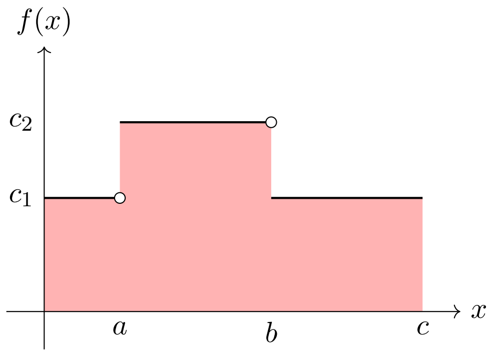
The integral of a simple function $f$ agrees with out intuition laid out previously.

**Definition 4.8** (Lebesgue Integral of a simple function)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $f(x) = \sum_{i=1}^{n} c_i \mathbb{1}_{A_i}(x)$ be a simple function. Then we define the integral of $f$ to be

$$\int_X f d\mu = \sum_{i=1}^{n} c_i \mu(A_i).$$

Simple functions look like piecewise horizontal functions. For example, below is the graph of

$$f = c_1 \mathbb{1}_{[0,a)} + c_2 \mathbb{1}_{[a,b)} + c_1 \mathbb{1}_{[b,c]}.$$




A typical theme in mathematics when dealing with functions is "if a function $f : X \to Y$ is a function we care about, for 'nice' subsets $V \subseteq Y$, we want $f^{-1}(V) \subseteq X$ to also be 'nice', where $f^{-1}(V)$ is the preimage of $V$." We define a measurable function this way as well.

**Definition 4.9** (Measurable function)**.** Let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be measurable spaces. A function $f : X \to Y$ is **measurable** if for all sets $V \in \mathcal{A}_Y$, we have $f^{-1}(V) \in \mathcal{A}_X$.

This definition seems very random, but it is a little more intuitive when we look through the lens of probability. Again, let $(X, \mathcal{A}_X)$ and $(Y, \mathcal{A}_Y)$ be measurable spaces, and let $f : X \to Y$. Think of the $(X, \mathcal{A}_X)$ as the space of "actions", and the space $(Y, \mathcal{A}_Y)$ as the space of possible outcomes given the actions. The function $f : X \to Y$ takes a possible action and outputs a possible outcome. Let us say we want to calculate the probability of outcomes in $Y$. The possible outcomes are sets $V \in \mathcal{A}_Y$, and the probability of that outcome happening is the measure of the set $f^{-1}(V)$, which is the set of all actions that lead to outcomes in $V$. Therefore, we want $f^{-1}(V)$ to be in $\mathcal{A}_X$. For real valued functions however, it is sufficient to consider a special case of measurability

**Definition 4.10** (Measurable extended real-valued functions)**.** Let $(X, \mathcal{A})$ be a measurable space, and let $f : X \to [-\infty, +\infty]$. We say $f$ is **$\mathcal{A}$-measurable** (or just **measurable**) if for each real number $t \in \mathbb{R}$ the set $\{x \in A : f(x) \leq t\}$ belongs to $\mathcal{A}$.

Now we may define the Lebesgue integral of arbitrary non-negative, extended real-valued measurable functions.

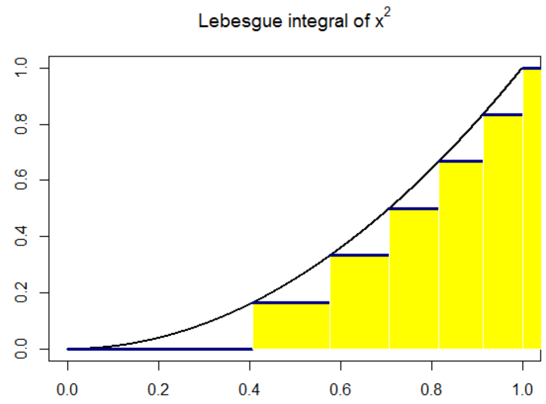



Figure 2: Bad approximation of integral using a simple function

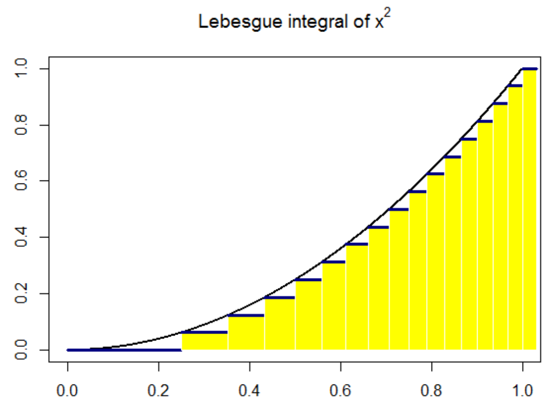



Figure 3: Decent approximation of integral using a simple function

**Definition 4.11** (Lebesgue integral of measurable functions $f : X \to [0, +\infty)$. $]Let($X, A, $\mu$) be a measure space, and let $f : X \to [0, +\infty]$ to be a measurable function. We define its **Lebesgue integral** to be

$$\int_X f d\mu = \sup \left\{ \int_X s d\mu : s \text{ is a simple function and } s(x) \leq f(x) \text{ for all } x \right\},$$

if this suprenum is finite. When the suprenum is finite, we say that $f$ is **integrable**.

See figures 2, 3, and 4 for a visualization of how the Lebesgue integral works. Through those figures, we can see that the simple functions are all "under" the function we are integrating, and as the simple function is a better approximation, the area is getting closer to the full area under the function. We have defined integrals only over the entire space, but let us say we want to integrate over a measurable subset $A$ of $X$. Then we define the Lebesgue integral of $f$ over $A$ as

$$\int_A f d\mu = \int_X f \cdot \mathbb{1}_A d\mu.$$

This works because $f \cdot \mathbb{1}_A$ vanishes when $x \notin A$, so when we integrate over it, it does not change the value of the integral.

What if we want to integrate over functions that take up negative values? Suppose $f : X \to [-\infty, +\infty]$ is measurable. Define two functions, splitting $f$ into a positive part

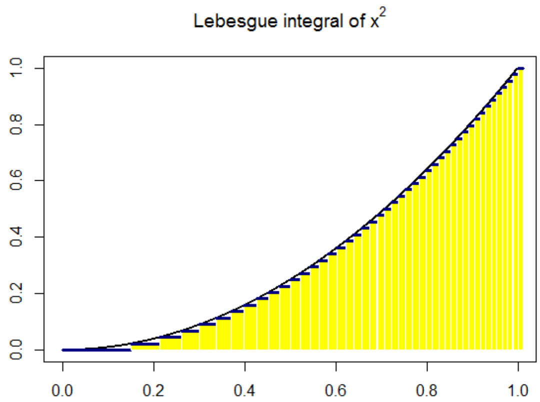



Figure 4: Good approximation of integral using a simple function

and negative part, as follows:

$$f^+(x) = \max(f(x), 0), \qquad f^-(x) = \max(0, -f(x)).$$

Now we define the integral of $f$ to be

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

## 5 Limit Theorems

We want to be able to swap limits and integrals. In the early 1800s, mathematicians switched around summations, integrals, and limits without much thought. After the foundations of real analysis was thoroughly investigated, mathematicians realized we need to prove when we can swap integrals and limits without any harm.

**Definition 5.1.** Let $(X, \mathcal{A}, \mu)$ be a measure space. Denote $\mathscr{S}$ to be the set of measurable real-valued simple functions, and denote $\mathscr{S}_+ \subseteq \mathscr{S}$ to be the set of measurable non-negative simple functions.

The following lemma shows that simple functions are a good framework for defining Lebesgue integrals the way we did. It shows that measurable functions can be approximated by simple functions.

**Lemma 5.2** (Measurable functions are limits of simple functions)**.** Let $(X, \mathcal{A})$ be a measurable space, let $A$ be a subset of $X$ that belongs to $\mathcal{A}$ , and let $f$ be a $[0, +\infty]$-valued measurable function on $A$. Then there is a sequence $\{f_n\}$ of simple $[0, +\infty)$-valued measurable functions on $A$ that satisfy

$$f_1(x) \leq f_2(x) \leq \ldots$$

and

$$f(x) = \lim_n f_n(x)$$

at each $x$ in $A$.

*Proof.* For each positive integer $n$ and for $k = 1, 2, \ldots, n2^n$ let $A_{n,k} = \{x \in A : (k-1)/2^n \leq f(x) < k/2^n\}$. The measurability of $f$ implies that each $A_{n,k}$ belongs to $\mathcal{A}$. Define a sequence $\{f_n\}$ of functions from $A$ to $\mathbb{R}$ by requiring $f_n$ to have value $(k-1)/2^n$ at each point in $A_{n,k}$ (for $k = 1, 2, \ldots, n2^n$) and to have value $n$ at each point in $A - \bigcup_k A_{n,k}$. The functions so defined are simple and measurable, and it is easy to check that they satisfy (1) and (2) at each $x$ in $A$. $\square$

**Proposition 5.3** (Swapping integrals and limits)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f$ belong to $\mathscr{S}_+$, and let $\{f_n\}$ be a non-decreasing sequence of functions belonging to $\mathscr{S}_+$ for which $f(x) = \lim_n f_n(x)$ holds at each $x$ in $X$. Then

$$\int_X f d\mu = \lim_n \int f_n d\mu.$$

*Proof.* By monotonicity and linearity of integrals of simple functions, we have

$$\int_X f_1\, d\mu \leq \int_X f_2\, d\mu \leq \cdots \leq \int_X f\, d\mu;$$

hence $\lim_n \int_X f_n\, d\mu$ exists and satisfies $\lim_n \int_X f_n\, d\mu \leq \int_X f\, d\mu$. We turn to the reverse inequality. Let $\epsilon$ be a number such that $0 < \epsilon < 1$. We will construct a nondecreasing sequence $\{g_n\}$ of functions in $\mathscr{S}_+$ such that $g_n \leq f_n$ holds for each $n$ and such that $\lim_n \int_X g_n\, d\mu = (1-\epsilon)\int_X f\, d\mu$. Since $\int_X g_n\, d\mu \leq \int_X f_n\, d\mu$, this will imply that $(1-\epsilon)\int_X f\, d\mu \leq \lim_n \int_X f_n\, d\mu$ and, since $\epsilon$ is arbitrary, that $\int_X f\, d\mu \leq \lim_n \int_X f_n\, d\mu$. Consequently $\int_X f\, d\mu = \lim_n \int_X f_n\, d\mu$.

We turn to the construction of the sequence $\{g_n\}$. Suppose that $a_1, \ldots, a_k$ are the nonzero values of $f$ and that $A_1, \ldots, A_k$ are the sets on which these values occur. Thus $f = \sum_{i=1}^k a_i \mathbb{1}_{A_i}$. For each $n$ and $i$ let

$$A_{(n,i)} = \{x \in A_i : f_n(x) \geq (1-\epsilon)a_i\}.$$

Then each $A_{(n,i)}$ belongs to $\mathcal{A}$, and for each $i$ the sequence $\{A_{(n,i)}\}_{n=1}^\infty$ is nondecreasing and satisfies $A_i = \bigcup_n A_{(n,i)}$. If we let $g_n = \sum_{i=1}^k (1-\epsilon)a_i \mathbb{1}_{A_{(n,i)}}$, then $g_n$ belongs to $\mathscr{S}_+$ and satisfies $g_n \leq f_n$, and we can use the fact that $\lim_n \mu(A_{(n,i)}) = \mu(A_i)$ to conclude that

$$\lim_n \int_X g_n\, d\mu = \lim_n \sum_{i=1}^k (1-\epsilon)a_i\mu(A_{(n,i)}) = \sum_{i=1}^k (1-\epsilon)a_i\mu(A_i) = (1-\epsilon)\int_X f\, d\mu.$$

$\square$

**Proposition 5.4.** Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f$ be a $[0, +\infty]$-valued $\mathcal{A}$-measurable function on $X$, and let $\{f_n\}$ be a non-decreasing sequence of functions in $\mathscr{S}_+$ such that $f(x) = \lim_n f_n(x)$ holds at each $x \in X$. Then

$$\int_X f\, d\mu = \lim_n \int_X f_n\, d\mu.$$

*Proof.* It is clear that

$$\int_X f_1\, d\mu \leq \int_X f_2\, d\mu \leq \cdots \leq \int_X f\, d\mu;$$

hence $\lim_n \int_X f_n\, d\mu$ exists and satisfies $\lim_n \int_X f_n\, d\mu \leq \int_X f\, d\mu$.

We turn to the reverse inequality. Recall that $\int_X f\, d\mu$ is the supremum of those elements of $[0, +\infty]$ of the form $\int_X g\, d\mu$, where $g$ ranges over the set of functions that belong to $\mathscr{S}_+$

and satisfy $g \leq f$. Thus to prove that $\int_X f\, d\mu \leq \lim_n \int_X f_n\, d\mu$, it is enough to check that if $g$ is a function in $\mathscr{S}_+$ that satisfies $g \leq f$, then $\int_X g\, d\mu \leq \lim_n \int_X f_n\, d\mu$. Let $g$ be such a function. Define $(a \wedge b)(x) = \min(a(x), b(x))$ for any $[-\infty, +\infty]$-valued functions $a$ and $b$. Then $\{g \wedge f_n\}$ is a nondecreasing sequence of functions in $\mathscr{S}_+$ for which $g = \lim_n (g \wedge f_n)$, and so we have that $\int_X g\, d\mu = \lim_n \int_X (g \wedge f_n)\, d\mu$. Since $\int_X (g \wedge f_n)\, d\mu \leq \int_X f_n\, d\mu$, it follows that $\int_X g\, d\mu \leq \lim_n \int_X f_n\, d\mu$, and the proof is complete. $\square$

For a sanity check, we will prove monotonicity and linearity of the Lebesgue integral.

**Proposition 5.5.** (Properties of Lebesgue Integral) Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f$ and $g$ be $[0, +\infty]$-valued $\mathcal{A}$-measurable functions on $X$, and let $\alpha$ be a non-negative number. Then

1. $\int_X \alpha f d\mu = \alpha \int_X f d\mu$,
2. $\int_X (f + g) d\mu = \int_X f d\mu + \int_X g d\mu$, and
3. if $f(x) \leq g(x)$ for all $x \in X$, then $\int_X f d\mu \leq \int_X g d\mu$.

*Proof.* Choose nondecreasing sequences $\{f_n\}$ and $\{g_n\}$ of functions in $\mathscr{S}_+$ such that $f = \lim_n f_n$ and $g = \lim_n g_n$ (see Lemma 5.2). Then $\{\alpha f_n\}$ and $\{f_n + g_n\}$ are nondecreasing sequences of functions in $\mathscr{S}_+$ that satisfy $\alpha f = \lim_n \alpha f_n$ and $f + g = \lim_n (f_n + g_n)$. Hence, we can use Proposition 5.4, together with the linearity of the integral on $\mathscr{S}_+$, to conclude that

$$\int_X \alpha f\, d\mu = \lim_n \int_X \alpha f_n\, d\mu = \lim_n \alpha \int_X f_n\, d\mu = \alpha \int_X f\, d\mu$$

and

$$\begin{aligned} \int_X (f+g)\, d\mu &= \lim_n \int_X (f_n + g_n)\, d\mu \\ &= \lim_n \left( \int_X f_n\, d\mu + \int_X g_n\, d\mu \right) = \int_X f\, d\mu + \int_X g\, d\mu. \end{aligned}$$

Thus, parts (a) and (b) are proved. For part (c), note that if $f \leq g$, then the class of functions $h$ in $\mathscr{S}_+$ that satisfy $h \leq f$ is included in the class of functions $h$ in $\mathscr{S}_+$ that satisfy $h \leq g$; it follows that

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

$\square$

**Proposition 5.6.** (Properties of Lebesgue Integral) Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f$ and $g$ be real-valued $\mathcal{A}$-measurable functions on $X$, and let $\alpha$ be a real number. Then

1. $\alpha f$ and $f + g$ are integrable,
2. $\int_X \alpha f\, d\mu = \alpha \int_X f\, d\mu$,
3. $\int_X (f + g)\, d\mu = \int_X f\, d\mu + \int_X g\, d\mu$, and
4. if $f(x) \leq g(x)$ for all $x \in X$, then $\int_X f\, d\mu \leq \int_X g\, d\mu$.

*Proof.* The proof is omitted as we just break up $f$ and $g$ into positive and negative parts and use Proposition 5.5. $\square$

**Theorem 5.7.** (Monotone Convergence Theorem) Let $(X, \mathcal{A}, \mu)$ be a measure space, let $f$ and $f_1, f_2, \ldots$ be $[0, +\infty]$-valued $\mathcal{A}$-measurable functions on $X$, Suppose that the relations

$$f_1(x) \leq f_2(x) \leq \ldots \tag{1}$$

and

$$f(x) = \lim_n f_n(x) \tag{2}$$

hold at almost every $x$ in $X$. Then $\int_X f\,d\mu = \lim_n \int f_n\,d\mu$.

*Proof.* First, suppose that relations (1) and (2) hold at each $x$ in $X$. The monotonicity of the integral implies that

$$f_1\,d\mu \leq f_2\,d\mu \leq \cdots \leq f\,d\mu;$$

hence the sequence $\{\int_X f_n\,d\mu\}$ converges (perhaps to $+\infty$), and its limit satisfies $\lim_n \int_X f_n\,d\mu \leq \int_X f\,d\mu$. We turn to the reverse inequality. For each $n$, choose a nondecreasing sequence $\{g_{n,k}\}_{k=1}^{\infty}$ of simple $[0, +\infty)$-valued measurable functions such that $f_n = \lim_k g_{n,k}$ (Lemma 5.2). For each $n$, define a function $h_n$ by

$$h_n = \max(g_{1,n}, g_{2,n}, \ldots, g_{n,n}).$$

Then $\{h_n\}$ is a nondecreasing sequence of simple $[0, +\infty)$-valued measurable functions that satisfy $h_n \leq f_n$ and $f = \lim_n h_n$. It follows from these remarks, Proposition 5.4, and the monotonicity of the integral that

$$\int_X f\,d\mu = \lim_n \int_X h_n\,d\mu \leq \lim_n \int_X f_n\,d\mu.$$

Hence $\int_X f\,d\mu = \lim_n \int_X f_n\,d\mu$.

Now suppose that we only require that relations (1) and (2) hold for almost every $x$ in $X$. Let $N$ be a set that belongs to $\mathcal{A}$, has measure zero under $\mu$, and contains all points at which one or more of these relations fails. The function $f\mathbb{1}_{N^c}$ and the sequence $\{f_n\mathbb{1}_{N^c}\}$ satisfy the hypotheses made in the first part of the proof, and so

$$\int_X f\mathbb{1}_{N^c}\,d\mu = \lim_n \int_X f_n\mathbb{1}_{N^c}\,d\mu. \tag{3}$$

Since $f_n\mathbb{1}_{N^c}$ agrees with $f_n$ almost everywhere and $f\mathbb{1}_{N^c}$ agrees with $f$ almost everywhere, Eq. (3) and the fact that if two functions agree almost everywhere they have the same integral imply that

$$\int_X f\,d\mu = \lim_n \int_X f_n\,d\mu.$$

$\square$

For completeness, we will provide the following corollary. It is unnecessary for the Radon–Nikodym theorem, but it is nice to know.

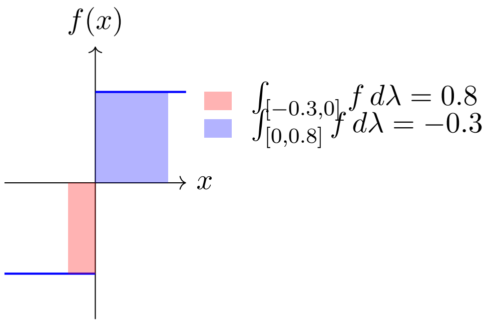



Figure 5: A visualization of a signed measure.

**Corollary 5.8** (Beppo Levi's theorem)**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $\sum_{k=1}^{\infty} f_k$ be an infinite series whose terms are $[0, +\infty]$-valued $\mathcal{A}$-measurable functions on $X$. Then

$$\int_X \sum_{k=1}^{\infty} f_k d\mu = \sum_{k=1}^{\infty} \int_X f_k d\mu.$$

*Proof.* Use the linearity of the integral, and apply the Monotone Convergence theorem to the sequence $\{\sum_{k=1}^{n} f_k\}_{n=1}^{\infty}$ of partial sums of the series $\sum_{k=1}^{\infty} f_k$. □

## 6 Signed and Complex Measures

**Definition 6.1** (Signed and Complex Measures)**.** Let $(X, \mathcal{A})$ be a measurable space, and let $\mu : \mathcal{A} \to [-\infty, +\infty]$. If the function $\mu$ is **countably additive**, meaning the identity

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

holds for each sequence of pairwise disjoint sets $A_1, A_2, \ldots$ and $\mu(\varnothing) = 0$, then it is a **signed measure**. Let $\nu : \mathcal{A} \to \mathbb{C}$. If $\nu$ is countably additive and $\nu(\varnothing) = 0$, then $\nu$ is a **complex measure**.

**Example 6.2.** An example of a signed measure is integrals of functions. For example, let $f(x) = x^{1/3}$, and take the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then we can define a measure $\nu(A) = \int_A f\, d\mu$.

**Example 6.3.** Let $(\mathbb{R}, \mathcal{A}, \lambda)$ be a measure space under the Lebesgue measure. Define $f(x) = 1$ if $x \geq 0$ and $f(x) = -1$ if $x < 0$. Then, we can see that $\nu(A) = \int_A f\, d\lambda$ is a measure. Furthermore, we can interpret $\nu$ as being a measure that gives negative weight to sets with negative numbers. For example, we can see that $\nu([-0.3, 0.8]) = 0.5$. We might ask ourselves if every signed measure can be split up into a positive and negative measure. See Figure 5 for a visualization.

**Definition 6.4** (Positive and negative sets)**.** Let $\mu$ be a signed measure on the measurable space $(X, \mathcal{A})$. A subset $A$ of $X$ is a **positive set** for $\mu$ if $A \in \mathcal{A}$ and each $\mathcal{A}$-measurable subset $E$ of $A$ satisfies $\mu(E) \geq 0$. Likewise $A$ is a **negative set** for $\mu$ if $A \in \mathcal{A}$ and for each $\mathcal{A}$-measurable subsets $E$ of $A$ we have $\mu(E) \leq 0$.

**Lemma 6.5** (Set of negative measure contains negative set)**.** Let $\mu$ be a signed measure on $(X, \mathcal{A})$, and let $A$ be a subset of $X$ that belongs to $\mathcal{A}$ and satisfies $-\infty < \mu(A) < 0$. Then there is a negative set $B$ that is included in $A$ and satisfies $\mu(B) \leq \mu(A)$.

*Proof.* We will remove a suitable sequence of subsets from $A$ and then let $B$ consist of the points of $A$ that remain. To begin, let

$$\delta_1 = \sup\{\mu(E) : E \in \mathcal{A} \text{ and } E \subseteq A\},$$

and choose an $\mathcal{A}$-measurable subset $A_1$ of $A$ that satisfies

$$\mu(A_1) \geq \min\left(\frac{1}{2}\delta_1, 1\right).$$

Then $\delta_1$ and $\mu(A_1)$ are nonnegative (note that $\mu(A_1) \geq \min\left(\frac{1}{2}\delta_1, 1\right)$ implies that $\delta_1 \geq \mu(\emptyset) = 0$). We proceed by induction, constructing sequences $\{\delta_n\}$ and $\{A_n\}$ by letting

$$\delta_n = \sup\left\{\mu(E) : E \in \mathcal{A} \text{ and } E \subseteq \left(A - \bigcup_{i=1}^{n-1} A_i\right)\right\},$$

and then choosing an $\mathcal{A}$-measurable subset $A_n$ of $A - \bigcup_{i=1}^{n-1} A_i$ that satisfies

$$\mu(A_n) \geq \min\left(\frac{1}{2}\delta_n, 1\right).$$

Now define $A_\infty$ and $B$ by $A_\infty = \bigcup_{n=1}^{\infty} A_n$ and $B = A - A_\infty$.

We require that $\mu(A_1)$ be at least $\min(\delta_1/2, 1)$, rather than at least $\delta_1/2$, because we have not yet proved that $\delta_1$ is finite (see Exercise 4).

Let us check that $B$ has the required properties. Since the sets $A_n$ are disjoint and satisfy $\mu(A_n) \geq 0$, it follows that $\mu(A_\infty) \geq 0$ and hence that

$$\mu(A) = \mu(A_\infty) + \mu(B) \geq \mu(B).$$

Thus we have that $\mu(B) \leq \mu(A)$.

We turn to the negativity of $B$. The finiteness of $\mu(A)$ implies the finiteness of $\mu(A_\infty)$ and so, since $\mu(A_\infty) = \sum_n \mu(A_n)$, implies that $\lim_n \mu(A_n) = 0$. Consequently, $\lim_n \delta_n = 0$. Since an arbitrary $\mathcal{A}$-measurable subset $E$ of $B$ satisfies $\mu(E) \leq \delta_n$ for each $n$ and so satisfies $\mu(E) \leq 0$, $B$ must be a negative set for $\mu$. □

**Theorem 6.6** (Hahn Decomposition Theorem)**.** Let $(X, \mathcal{A})$ be a measurable space, and let $\mu$ be a signed measure on $(X, \mathcal{A})$. Then there are disjoint subsets $P$ and $N$ of $X$ such that $P$ is a positive set for $\mu$, $N$ is a negative set for $\mu$, and $X = P \cup N$.

*Proof.* A signed measure $\mu$ cannot include both $+\infty$ and $-\infty$ among its values, we can for definiteness assume that $-\infty$ is not included. Let

$$L = \inf\{\mu(A) : A \text{ is a negative set for } \mu\}.$$

Choose a sequence $\{A_n\}$ of negative sets for $\mu$ for which $L = \lim_n \mu(A_n)$, and let $N = \bigcup_{n=1}^{\infty} A_n$. It is easy to check that $N$ is a negative set for $\mu$ (each $\mathcal{A}$-measurable subset of $N$ is the union of a sequence of disjoint $\mathcal{A}$-measurable sets, each of which is included in some $A_n$). Hence $L \leq \mu(N) \leq \mu(A_n)$ holds for each $n$, and so $L = \mu(N)$. Furthermore, since $\mu$ does not attain the value $-\infty$, $\mu(N)$ must be finite. Let $P = N^c$. Our only remaining task is to check that $P$ is a positive set for $\mu$. If $P$ included an $\mathcal{A}$-measurable set $A$ such that $\mu(A) < 0$, then it would include a negative set $B$ such that $\mu(B) < 0$ (Lemma 6.5), and $N \cup B$ would be a negative set such that $\mu(N \cup B) = \mu(N) + \mu(B) < \mu(N) = L$ (recall that $\mu(N)$ is finite). However this contradicts how $L$ was initially defined, and so $P$ must be a positive set for $\mu$.

□

**Theorem 6.7** (Jordan Decomposition Theorem)**.** Every signed measure is the difference of two positive measures, at least one of which is finite.

*Proof.* Let $\mu$ be a signed measure on $(X, \mathcal{A})$. Choose a Hahn decomposition $(P, N)$ for $\mu$, and then define functions $\mu^+$ and $\mu^-$ on $\mathcal{A}$ by

$$\mu^+(A) = \mu(A \cap P) \quad \text{and} \quad \mu^-(A) = -\mu(A \cap N).$$

It is clear that $\mu^+$ and $\mu^-$ are positive measures such that $\mu = \mu^+ - \mu^-$. Since $+\infty$ and $-\infty$ cannot both occur among the values of $\mu$, at least one of the values $\mu(P)$ and $\mu(N)$, and hence at least one of the measures $\mu^+$ and $\mu^-$, must be finite. □

**Corollary 6.8.** Let $\nu$ be a complex measure on a measurable space $(X, \mathcal{A})$. Then $\nu = \mu_1 + i\mu_2$ where $\mu_1$ and $\mu_2$ are finite signed measures on $(X, \mathcal{A})$. Hence by the Jordan decomposition theorem we have that $\nu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)$.

# 7 Radon–Nikodym theorem

The Jordan decomposition theorem gives us an insight on classifying measures. Since we know that every signed or complex measure can be broken down into positive measures, we may ask ourselves how else we can "break down" measures into another measure. The Radon–Nikodym theorem does exactly that. It shows us that we can represent a measure $\nu$ in terms of a measure $\mu$ through an integral of a function as long as $\nu$ is **absolutely continuous** with respect to $\mu$.

**Definition 7.1** (Absolute continuity)**.** Let $(X, \mathcal{A})$ be a measurable space and let $\mu$ and $\nu$ be measures on $(X, \mathcal{A})$. Then for all $A \in \mathcal{A}$, if $\mu(A) = 0$ implies $\nu(A) = 0$, then $\nu$ is absolutely continuous with respect to $\nu$.

**Example 7.2.** Define $\nu(A) = \int_A f \, d\mu$, then $\nu$ is absolutely continuous with respect to $\mu$. This is because if $\mu(A) = 0$, the integral vanishes; $\nu(A) = \int_X f \mathbb{1}_A \, d\mu$ and $f \mathbb{1}_A = 0$ almost everywhere, so $\nu(A) = 0$ too.

**Definition 7.3** ($\sigma$-finite measures)**.** Let $(X, \mathcal{A})$ be a measurable space, and let $\mu$ be a positive measure on $(X, \mathcal{A})$. The measure $\mu$ is $\sigma$-finite if $X$ is a countable union of sets of finite measure. That is, $X = \bigcup_{i=1}^{\infty} A_i$ where $\mu(A_i) < +\infty$ for all $i$.

**Example 7.4.** The Lebesgue measure is $\sigma$-finite over $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. This is because $\mathbb{R} = \bigcup_{n=1} [-n, n]$.

Before we prove the Radon–Nikodym theorem, we will need a small lemma showing that integrable functions are finite-valued almost everywhere.

**Lemma 7.5.** Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $f : X \to [-\infty, +\infty]$ be integrable. Then we must have $|f(x)| < +\infty$ holds for almost every $x \in X$.

*Proof.* First we observe that if $g : X \to [0, +\infty]$ is an $\mathcal{A}$-measurable function, $t$ is a positive real number, and $A_t = \{x \in X : f(x) \geq t\}$, then

$$\mu(A_t) \leq \frac{1}{t} \int_{A_t} f d\mu \leq \frac{1}{t} \int_X f d\mu.$$

This is because $0 \leq t \cdot \mathbb{1}_{A_t} \leq f \cdot \mathbb{1}_{A_t} \leq f$, and the monotonicity and linearity of the integral implies that

$$t\mu(A_t) = \int_X t \mathbb{1}_{A_t} d\mu \leq \int_{A_t} f d\mu \leq \int_X f d\mu.$$

Applying that observation to the function $|f|$, we have

$$\mu(\{x \in X : |f(x)|n\}) \leq \frac{1}{n}\int_X |f| d\mu$$

holds for each positive integer $n$. Thus

$$\mu(\{x \in X : |f(x)| = +\infty\}) \leq \mu(\{x \in X : |f(x)| \geq n\}) \leq \frac{1}{n}\int_X |f| d\mu$$

holds for each $n$, and so $\mu(\{x \in X : |f(x)| = +\infty\}) = 0$. (Note that $\int_X |f| d\mu$ must be finite as $f$ is integrable.) □

From Example 7.2, we know that if we make a new measure from an integral, that the new measure is absolutely continuous with respect to our old measure. How about the converse? Does absolute continuity imply that we can put a measure as an integral? Turns out the answer is yes, given a few more conditions.

**Theorem 7.6** (Radon–Nikodym theorem)**.** Let $(X, \mathcal{A})$ be a measurable space, and let $\mu$ and $\nu$ be $\sigma$-finite positive measures on $(X, \mathcal{A})$. If $\nu$ is absolutely continuous with respect to $\mu$, then there is an $\mathcal{A}$-measurable function $g : X \to [0, +\infty)$ such that $\nu(A) = \int_A g d\mu$ holds for each $A$ in $\mathcal{A}$. The function $g$ is unique up to $\mu$-almost everywhere equality.

*Proof.* First consider the case where $\mu$ and $\nu$ are both finite. Let $\mathscr{F}$ be the set consisting of those $\mathcal{A}$-measurable functions $f : X \to [0, +\infty]$ that satisfy $\int_A f\, d\mu \leq \nu(A)$ for each $A \in \mathcal{A}$. We will show first that $\mathscr{F}$ contains a function $g$ such that

$$\int g\, d\mu = \sup\left\{\int f\, d\mu : f \in F\right\} \tag{1}$$

and then that this function $g$ satisfies

$$\nu(A) = \int_A g\, d\mu \quad \text{for each } A \in \mathcal{A}.$$

Finally, we will show that $g$ can be modified so as to have only finite values.

We begin by checking that if $f_1$ and $f_2$ belong to $\mathscr{F}$, then $f_1 \vee f_2$ defined as $(f_1 \vee f_2)(x) = \max f_1(x), f_2(x)$ belongs to $\mathscr{F}$; to see this note that if $A$ is an arbitrary set in $\mathcal{A}$, if $A_1 = \{x \in A : f_1(x) > f_2(x)\}$, and if $A_2 = \{x \in A : f_2(x) \geq f_1(x)\}$, then

$$\int_A (f_1 \vee f_2)\, d\mu = \int_{A_1} f_1\, d\mu + \int_{A_2} f_2\, d\mu \leq \nu(A_1) + \nu(A_2) = \nu(A).$$

Furthermore, $\mathscr{F}$ is not empty (the constant 0 belongs to it). Now choose a sequence $\{f_n\}$ of functions in $\mathscr{F}$ for which

$$\lim_n \int f_n\, d\mu = \sup\left\{\int f\, d\mu : f \in F\right\}.$$

By replacing $f_n$ with $f_1 \vee \cdots \vee f_n$, we can assume that the sequence $\{f_n\}$ is increasing. Let $g = \lim_n f_n$. The monotone convergence theorem implies that the relation

$$\int_A g\, d\mu = \lim_n \int_A f_n\, d\mu \leq \nu(A)$$

holds for each $A$ and hence that $g$ belongs to $\mathscr{F}$. It also implies that

$$\int g\, d\mu = \sup\left\{\int f\, d\mu : f \in F\right\}.$$

Thus $g$ has the first of the properties claimed for it.

We turn to the proof that

$$\nu(A) = \int_A g \, d\mu \quad \text{holds for each } A \in \mathcal{A}.$$

Since $g$ belongs to $\mathscr{F}$, the formula

$$\nu_0(A) = \nu(A) - \int_A g \, d\mu$$

defines a positive measure on $\mathcal{A}$. We need only show that $\nu_0 = 0$. Assume the contrary. Then, since $\mu$ is finite, there is a positive number $\epsilon$ such that

$$\nu_0(X) > \epsilon\mu(X).$$

Let $(P, N)$ be a Hahn decomposition for the signed measure $\nu_0 - \epsilon\mu$. Note that for each $A \in \mathcal{A}$ we have

$$\nu_0(A \cap P) \geq \epsilon\mu(A \cap P),$$

and hence we have

$$\nu(A) = \int_A g \, d\mu + \nu_0(A) \geq \int_A g \, d\mu + \nu_0(A \cap P)$$

$$\geq \int_A g \, d\mu + \epsilon\mu(A \cap P) = \int_A (g + \epsilon\chi_P) \, d\mu.$$

Note also that $\mu(P) > 0$, since if $\mu(P) = 0$, then $\nu_0(P) = 0$, and so

$$\nu_0(X) - \epsilon\mu(X) = (\nu_0 - \epsilon\mu)(N) \leq 0,$$

contradicting the assumption. It follows from this, the relation $\int g \, d\mu \leq \nu(X) < +\infty$, and the above inequality that $g + \epsilon\chi_P$ belongs to $\mathscr{F}$ and satisfies

$$\int (g + \epsilon\chi_P) \, d\mu > \int g \, d\mu.$$

This, however, contradicts the supremum property of $g$ and so implies that $\nu_0 = 0$. Hence

$$\nu(A) = \int_A g \, d\mu \quad \text{holds for each } A \in \mathcal{A}.$$

Since $g$ can have an infinite value only on a $\mu$-null set (Lemma 7.5), it can be redefined so as to have only finite values. With this, we have constructed the required function in the case where $\mu$ and $\nu$ are finite.

Now suppose that $\mu$ and $\nu$ are $\sigma$-finite. Then $X$ is the union of a sequence $\{B_n\}$ of disjoint $\mathcal{A}$-measurable sets, each of which has finite measure under $\mu$ and under $\nu$. For each $n$ the first part of this proof provides an $\mathcal{A}$-measurable function $g_n : B_n \to [0, +\infty)$ such that

$$\nu(A) = \int_A g_n \, d\mu \quad \text{holds for each } \mathcal{A}\text{-measurable subset } A \text{ of } B_n.$$

The function $g : X \to [0, +\infty)$ that agrees on each $B_n$ with $g_n$ is then the required function.

We turn to the uniqueness of $g$. Let $g, h : X \to [0, +\infty)$ be $\mathcal{A}$-measurable functions that satisfy

$$\nu(A) = \int_A g\, d\mu = \int_A h\, d\mu \quad \text{for each } A \in \mathcal{A}.$$

First consider the case where $\nu$ is finite. Then $g - h$ is integrable and

$$\int_A (g-h)\, d\mu = 0$$

holds for each $A \in \mathcal{A}$; since in this equation $A$ can be the set where $g > h$ or the set where $g < h$, it follows that

$$\int (g-h)^+\, d\mu = 0 \quad \text{and} \quad \int (g-h)^-\, d\mu = 0$$

and hence that $(g-h)^+$ and $(g-h)^-$ vanish $\mu$-almost everywhere. Thus $g$ and $h$ agree $\mu$-almost everywhere. If $\nu$ is $\sigma$-finite and if $\{B_n\}$ is a sequence of $\mathcal{A}$-measurable sets that have finite measure under $\nu$ and satisfy $X = \cup_n B_n$, then the preceding argument shows that $g$ and $h$ agree $\mu$-almost everywhere on each $B_n$ and hence $\mu$-almost everywhere on $X$. □

**Theorem 7.7** (Radon–Nikodym theorem for finite signed or complex measure)**.** Let $(X, \mathcal{A})$ be a measurable space, let $\mu$ be $\sigma$-finite positive measure on $(X, \mathcal{A})$, and let $\nu$ be a finite signed or complex measure on $(X, \mathcal{A})$. If $\nu$ is absolutely continuous with respect to $\mu$, then there is a function $g$ that belongs to $\mathscr{L}^1(X, \mathcal{A}, \mu, \mathbb{R})$ (or to $\mathscr{L}^1(X, \mathcal{A}, \mu, \mathbb{C})$) and satisfies $\nu(A) = \int_A g d\mu$ for each $A$ in $\mathcal{A}$. The function $g$ is unique up to $\mu$-almost everywhere equality.

*Proof.* If $\nu$ is a complex measure that is absolutely continuous with respect to $\mu$, then it can be written in the form

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

where $\nu_1$, $\nu_2$, $\nu_3$, and $\nu_4$ are finite positive measures that are absolutely continuous with respect to $\mu$. Then by the Radon–Nikodym theorem for two $\sigma$-finite positive measures, we have functions $g_j$, $j = 1, \ldots, 4$, that satisfy

$$\nu_j(A) = \int_A g_j\, d\mu$$

for each $A \in \mathcal{A}$. The required function $g$ is now given by

$$g = g_1 - g_2 + ig_3 - ig_4.$$

The case of a finite signed measure is similar. The uniqueness of $g$ can be proved with the method used in the proof of the previous theorem; in case $\nu$ is a complex measure, the real and imaginary parts of $g$ should be considered separately. □

Let $(X, \mathcal{A})$ be a measurable space, let $\mu$ be a $\sigma$-finite positive measure on $(X, \mathcal{A})$, and let $\nu$ be a finite signed, complex, or $\sigma$-finite positive measure on $(X, \mathcal{A})$. Suppose that $\nu$ is absolutely continuous with respect to $\mu$. An $\mathcal{A}$-measurable function $g$ on $X$ that satisfies

$$\nu(A) = \int_A g\, d\mu \quad \text{for each } A \in \mathcal{A}$$

is called a Radon–Nikodym derivative of $\nu$ with respect to $\mu$ or, in view of its uniqueness up to $\mu$-null sets, the Radon–Nikodym derivative of $\nu$ with respect to $\mu$. A Radon–Nikodym derivative of $\nu$ with respect to $\mu$ is sometimes denoted by $\frac{d\nu}{d\mu}$.

# 8 Applications of Radon–Nikodym Theorem

The Radon–Nikodym theorem gives a quick way to switch between different measures. It is used a ton in finance, probability, and physics, but the easiest way to see its significance is through probability theory.

## 8.1 Probability Density Functions

Remember that a probability space is a measure space $(\Omega, \mathcal{A}, \mathbb{P})$. The elements of $\Omega$ are called the **elementary outcomes** or the **sample points** of our experiment, and the members of $\mathcal{A}$ are called **events**. If $A \in \mathcal{A}$, then $\mathbb{P}(A)$ is the probability of the event $A$.

**Definition 8.1.** A **real-valued random variable** on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ is an $\mathcal{A}$-measurable function from $\Omega$ to $\mathbb{R}$.

Let $X$ be a real-valued random variable on a probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Such a variable represents a numerical observation or measurement whose value depends on on the outcome of the random event represented by $(\Omega, \mathcal{A}, \mathbb{P})$. The distribution of $X$ is represented by the measure $PX^{-1}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined by $(PX^{-1})(A) = P(X^{-1}(A))$ where $X^{-1}$ is the pre-image of $A$ under $X$. We usually denote $PX^{-1}$ as $P_X$.

Now let us look at the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P})$ with $\mathbb{P}(A) = \int_A f\, d\lambda$ where $f$ is a Borel measurable function such that $\int_{\mathbb{R}} f\, d\lambda = 1$. We call $f$ the **density** of $\mathbb{P}$, but we can also see that $f$ is actually a Radon–Nikodym derivative of $\mathbb{P}$ with respect to $\mu$! This means that some of the distributions we go through in a introductory statistics course, such as the Gaussian or Normal distribution, are actually Radon–Nikodym derivatives!

## 8.2 Conditional Expectation

When learning probability in grade school, we may have learnt about **conditional probability**. It is something like "the probabilty of event A given event B". It even has a nice formula:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

In terms of probability spaces, we can use the same relation. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $A$ and $B$ be events in $\mathcal{A}$, and assume that $P(B) \neq 0$. We still use the same formula as above for the conditional probability. We will generalize this idea a bit further to something called **conditional expectation**.

**Definition 8.2.** Let $X$ be a real-valued random variable on $(\Omega, \mathcal{A}, \mathbb{P})$, and let $X$ be integrable with respect to $\mathbb{P}$. The **expected value** or **expectation** of $X$ is written $E(X)$ and is defined by $E(X) = \int_{\mathbb{R}} X\, d\mathbb{P}$. If $X$ is integrable, one also says that $X$ has **finite expected value**.

**Definition 8.3.** Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, and let $\mathcal{B}$ be a sub-algebra of $\mathcal{A}$ (meaning that $\mathcal{B} \subseteq \mathcal{A}$ and $\mathcal{B}$ is a $\sigma$-algebra). Suppose that $X$ is a real-valued random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ that has finite expected value. A **conditional expectation of** $X$ **given** $\mathcal{B}$ is a random variable $Y$ that is $\mathcal{B}$-measurable, is integrable (that is, has finite expected value), and satisfies

$$\int_B Y\, d\mathbb{P} = \int_B X\, d\mathbb{P}.$$

One generally writes $E(X \mid \mathcal{B})$ for a conditional expectation of $X$ given $\mathcal{B}$. When one needs to be more precise, one sometimes calls an integrable $\mathcal{B}$-measurable function $Y$

that satisfies

$$\int_B Y\,dP = \int_B X\,dP \quad \text{for all } B \in \mathcal{B}$$

a version of the conditional expectation of $X$ given $\mathcal{B}$ or a version of $\mathbb{E}(X|\mathcal{B})$. Now for the main result of this section.

**Theorem 8.4.** Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, let $X$ be a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ with finite expected value, and let $\mathcal{B}$ be a $\sigma$-algebra such that $\mathcal{B} \subseteq \mathcal{A}$. Then

1. $X$ has a conditional expectation given $\mathcal{B}$, and
2. the conditional expectation of $X$ given $\mathcal{B}$ is unique, in the sense that if $Y_1$ and $Y_2$ are versions of $E(X|\mathcal{B})$

*Proof.* The formula

$$\mu(B) = \int_B X\,dP$$

defines a finite signed measure on $(\Omega, \mathcal{B})$; it is absolutely continuous with respect to the restriction of $P$ to $\mathcal{B}$. Thus, the Radon–Nikodym theorem, applied to $\mu$ and the restriction of $P$ to $\mathcal{B}$, gives a $\mathcal{B}$-measurable random variable $Y$ such that

$$\int_B Y\,dP = \mu(B) = \int_B X\,dP$$

holds for each $B \in \mathcal{B}$. Thus, $Y$ is a conditional expectation of $X$ given $\mathcal{B}$. The uniqueness assertion in the Radon–Nikodym theorem gives the uniqueness of the conditional expectation.

□

See [Bro02] for more information on conditional expectation.

## Acknowledgements

The author would like to thank his teaching assistant Ari Krishna, and his teacher Simon Rubinstein-Salzedo for their assistance with this paper. This paper would not have been possible without the Euler Circle Independent Research and Paper Writing class.

## References

[Bro02] Radon–nikodym and conditional expectation. `https://www.dam.brown.edu/people/huiwang/classes/Am264/Archive/cond_expe.pdf`, 2002.

[Coh13] Donald L Cohn. *Measure theory*, volume 2. Springer, 2013.

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

[Pes14] Ivan N Pesin. *Classical and modern integration theories*. Academic Press, 2014.

[Pug03] C.C. Pugh. *Real Mathematical Analysis*. Undergraduate Texts in Mathematics. Springer New York, 2003.

[Tao11] Terence Tao. *An introduction to measure theory*, volume 126. American Mathematical Soc., 2011.