Fundamental Elements of Probability and Asymptotic Theory

Class notes for Econ 7818

by

Carlos Brunet Martins-Filho
Department of Economics
256 UCB
University of Colorado at Boulder
Boulder, CO 80309-0256 USA
email: carlos.martins@colorado.edu

Chapter 1

Probability spaces

It is universally accepted, and intuitively understood, that the probability associated with the occurrence of an arbitrary event can be expressed by a number between 0 and 1. For example, we may be informed by a meteorological service that the probability that it will snow tomorrow is 70%. In fact, in many settings we can easily assess the probabilities associated with certain events. Thus, stating that the probability of observing heads after tossing a fair coin is 50% is normally taken to be self-evident. In this chapter we develop a mathematical framework that will allow a formal treatment of the notions of event and probability. The development of this framework, which relies on concepts and results from measure theory, leads us to the notion of probability spaces, foundational to all subsequent topics in this monograph.

1.1 σ-algebras

A set formed by subsets of a fixed set $\mathbb{X}$ is called a system of sets. Systems are commonly described by certain properties that involve taking unions, intersections and differences of their elements. In what follows, we will introduce several systems that will be useful in constructing probability spaces. We start with the definition of the most important of these systems in the study of probability, they are called σ-algebras.

Definition 1.1. *Let $\mathbb{X}$ be an arbitrary set. A σ-algebra $\mathcal{F}$ is a system of subsets of $\mathbb{X}$ having the following properties:*

1. $\mathbb{X} \in \mathcal{F}$
2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$
3. $A_i \in \mathcal{F}$ *for* $i \in \mathbb{N} \implies \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

In this context we say that $\mathcal{F}$ is a σ-algebra *associated* with $\mathbb{X}$. It is evident from this definition that many σ-algebras may be associated with a set $\mathbb{X}$. As a matter of terminology, if $A \in \mathcal{F}$ it is said to be an $\mathcal{F}$-measurable set and the pair $(\mathbb{X}, \mathcal{F})$ is called a measurable space.

Remark 1.1.

1. *Since $\mathbb{X} \in \mathcal{F}$, by property 2, $\mathbb{X}^c = \mathbb{X} - \mathbb{X} = \emptyset \in \mathcal{F}$. Hence, every σ-algebra contains the empty set. Note that complementation is taken with respect to the set $\mathbb{X}$.*
2. *By de Morgan's Laws $\left(\bigcup_{i \in \mathbb{N}} A_i\right)^c = \bigcap_{i \in \mathbb{N}} A_i^c$ and by properties 2 and 3, if $A_i \in \mathcal{F}$ for $i \in \mathbb{N}$, then $A_i^c \in \mathcal{F}$ and $\bigcap_{i \in \mathbb{N}} A_i^c \in \mathcal{F}$.*
3. *Given Definition 1.1 and Remark 1.1.2 we say that $\mathcal{F}$ is "closed" under complementation, countable unions and countable intersections.*
4. *For A_1, $A_2 \in \mathcal{F}$, and given that $A_2 - A_1 = A_2 \cap A_1^c$ we have that $A_2 - A_1 \in \mathcal{F}$. Also, denoting the symmetric difference between sets A_1 and A_2 by $A_1 \Delta A_2 := (A_1 - A_2) \cup (A_2 - A_1)$, we have that $A_1 \Delta A_2 \in \mathcal{F}$.*
5. *A system of subsets of $\mathbb{X}$ is said to be an algebra if properties 1 and 2 in Definition 1.1 hold and if $A_i \in \mathcal{F}$ for $i = 1, \cdots, m$ then $\bigcup_{i=1}^m A_i \in \mathcal{F}$ with $m \in \mathbb{N}$. Clearly, every σ-algebra is also an algebra.*

We now provide examples of σ-algebras.

Example 1.1. *1. For any $\mathbb{X}$, $\mathcal{F} := \{\mathbb{X}, \emptyset\}$ is a σ-algebra. It is called the minimal σ-algebra.*

2. *For any $\mathbb{X}$, the collection $2^{\mathbb{X}}$ of all subsets of $\mathbb{X}$ is a σ-algebra. It is called the maximal σ-algebra.*

3. *Let $A \subset \mathbb{X}$. Then, $\mathcal{F} := \{\mathbb{X}, A, A^c, \emptyset\}$ is a σ-algebra.*

4. *Let $S \subset \mathbb{X}$ and $\mathcal{F}$ a σ-algebra associated with $\mathbb{X}$. Then, $\mathcal{F}_S := S \cap \mathcal{F} := \{S \cap F : F \in \mathcal{F}\}$ is a σ-algebra associated with S. It is called the trace σ-algebra. We verify that $\mathcal{F}_S$ is a σ-algebra by establishing the properties of Definition 1.1:*

   1. $S \in \mathcal{F}_S$.

   Note that since $\mathbb{X} \in \mathcal{F}$, then $S \cap \mathbb{X} = S \in \mathcal{F}_S$.

   2. $A \in \mathcal{F}_S \implies A^c \in \mathcal{F}_S$ (note that $A^c = S - A$, complementation relative to S).

   $A \in \mathcal{F}_S \implies \exists F \in \mathcal{F} \ni A = S \cap F \in \mathcal{F}_S$. Since $F \in \mathcal{F}$ then $F^c \in \mathcal{F}$ and $S \cap F^c \in \mathcal{F}_S$. Furthermore, $S = (S \cap F) \cup (S \cap F^c) = A \cup (S \cap F^c)$. But by definition, $A \cup A^c = S$, hence $A^c = S \cap F^c \in \mathcal{F}_S$.

   3. $A_i \in \mathcal{F}_S$ for $i \in \mathbb{N} \implies \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}_S$.

   $A_i \in \mathcal{F}_S \implies \exists F_i \in \mathcal{F} \ni A_i = S \cap F_i$. Hence, $\bigcup_{i \in \mathbb{N}} A_i = \bigcup_{i \in \mathbb{N}} (S \cap F_i) = S \cap \left(\bigcup_{i \in \mathbb{N}} F_i \right)$. But since $F_i \in \mathcal{F}$, we have $\bigcup_{i \in \mathbb{N}} F_i \in \mathcal{F}$, hence $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}_S$.

5. *Let $f : \mathbb{X} \to \mathbb{Y}$ be a function, $\mathcal{Y}$ be a σ-algebra associated with $\mathbb{Y}$ and $f^{-1}(S) := \{x \in \mathbb{X} : f(x) \in S\}$ denote the inverse image of the set S under f. Then, $\mathcal{F} := f^{-1}(\mathcal{Y}) = \{f^{-1}(S) : S \in \mathcal{Y}\}$ is a σ-algebra associated with $\mathbb{X}$. $\mathcal{F}$ is called the inverse image σ-algebra. Again, we verify that $\mathcal{F}$ is a σ-algebra by establishing the properties of Definition 1.1:*

1. $\mathbb{X} \in \mathcal{F}$.

Since $\mathcal{Y}$ *is a* σ*-algebra associated with* $\mathbb{Y}$, $\mathbb{Y} \in \mathcal{Y}$. $f^{-1}(\mathbb{Y}) = \{x \in \mathbb{X}\colon f(x) \in \mathbb{Y}\} = \mathbb{X}$. *Thus,* $\mathbb{X} \in \mathcal{F}$.

2. $A \in \mathcal{F} \implies A^c \in \mathcal{F}$.

$A \in \mathcal{F} \implies \exists S_A \in \mathcal{Y} \ni A = f^{-1}(S_A)$. *Now,* $S_A \in \mathcal{Y} \implies S_A^c := \mathbb{Y} - S_A \in \mathcal{Y}$ *and* $f^{-1}(\mathbb{Y} - S_A) = \mathbb{X} - f^{-1}(S_A)$. *Thus,* $f^{-1}(\mathbb{Y} - S_A) = \mathbb{X} - A = A^c \in \mathcal{F}$.

3. $A_i \in \mathcal{F}$ *for* $i \in \mathbb{N} \implies \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.

$A_i \in \mathcal{F} \implies \exists\ S_{A_i} \in \mathcal{Y} \ni A_i = f^{-1}(S_{A_i})$. *Now,* $S_{A_i} \in \mathcal{Y}$, $\forall i \in \mathbb{N} \implies \bigcup_{i\in\mathbb{N}} S_{A_i} \in \mathcal{Y}$ *and* $f^{-1}\left(\bigcup_{i\in\mathbb{N}} S_{A_i}\right) = \bigcup_{i\in\mathbb{N}} f^{-1}(S_{A_i}) = \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.

The following theorem shows that the intersection of an arbitrary collection of σ-algebras associated with $\mathbb{X}$ is itself a σ-algebra.

Theorem 1.1. *Let* $F := \{\mathcal{F} : \mathcal{F}$ *is a* σ*-algebra associated with the set* $\mathbb{X}\}$. *Then,* $\mathcal{I} := \bigcap_{\mathcal{F}\in F} \mathcal{F}$ *is a* σ*-algebra associated with* $\mathbb{X}$, *i.e.,* $\mathcal{I} \in F$.

Proof. We verify that $\mathcal{I}$ satisfies Definition 1.1.

1. Since $\mathbb{X} \in \mathcal{F}\ \forall\ \mathcal{F} \in F$ then $\mathbb{X} \in \mathcal{I}$.

2. $A \in \mathcal{I} \implies A \in \mathcal{F}\ \forall\ \mathcal{F} \in F$. Then, $A^c \in \mathcal{F}\ \forall\ \mathcal{F} \in F$. Consequently, $A^c \in \mathcal{I}$.

3. Let $A_i \in \mathcal{I}$ for $i \in \mathbb{N}$. Then, $A_i \in \mathcal{F}\ \forall\ \mathcal{F} \in F$. Hence, $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{F}\ \forall\ \mathcal{F} \in F$, which implies $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{I}$. ■

Since $\mathcal{I} \subset \mathcal{F}\ \forall\ \mathcal{F} \in F$, we can say that $\mathcal{I}$ is the smallest σ-algebra in F.

It is often the case that σ-algebras are obtained from arbitrary systems of sets associated with $\mathbb{X}$ by expanding the systems in such a way that the defining properties in Definition 1.1 are met. In this context it is possible to consider the smallest σ-algebra generated from a system. This motivates the following definition.

Definition 1.2. *Let $\mathcal{C}$ be any collection of subsets of $\mathbb{X}$. The σ-algebra generated by $\mathcal{C}$, denoted by $\sigma(\mathcal{C})$, is a σ-algebra satisfying:*

1. $\mathcal{C} \subset \sigma(\mathcal{C})$
2. *If $\mathcal{F}$ is a σ-algebra such that $\mathcal{C} \subset \mathcal{F}$, then $\sigma(\mathcal{C}) \subset \mathcal{F}$.*

Property 2 characterizes $\sigma(\mathcal{C})$ as the smallest σ-algebra containing $\mathcal{C}$. The existence of this σ-algebra is showed in the next theorem.

Theorem 1.2. *For an arbitrary collection of subsets $\mathcal{C}$ of $\mathbb{X}$, there exists a unique smallest σ-algebra containing $\mathcal{C}$.*

Proof. Let $F = \{\mathcal{F} : \mathcal{F} \text{ is a } \sigma\text{-algebra associated with } \mathbb{X} \text{ and } \mathcal{C} \subset \mathcal{F}\}$ be the set of all σ-algebras containing $\mathcal{C}$. $F \neq \emptyset$ since $2^{\mathbb{X}}$ is a σ-algebra. By Theorem 1.1, $\bigcap_{\mathcal{F}\in F} \mathcal{F}$ is a σ-algebra. Since $\mathcal{C}$ is in all $\mathcal{F}$, $\mathcal{C} \in \bigcap_{\mathcal{F}\in F} \mathcal{F}$. Thus, $\bigcap_{\mathcal{F}\in F} \mathcal{F} \in F$. But by construction it is the smallest σ-algebra in F. $\blacksquare$

Evidently, if $\mathcal{C}$ is a σ-algebra then $\sigma(\mathcal{C}) = \mathcal{C}$. The generation of the smallest σ-algebra associated with a collection of subsets $\mathcal{C}$ of $\mathbb{X}$ is "monotonic" in a sense demonstrated in the following theorem.

Theorem 1.3. *Let $\mathcal{C}$ and $\mathcal{D}$ be two nonempty collections of subsets of $\mathbb{X}$. If $\mathcal{C} \subset \mathcal{D}$ then $\sigma(\mathcal{C}) \subset \sigma(\mathcal{D})$.*

Proof. Let $\mathcal{F}_{\mathcal{C}} := \{\mathcal{H} : \mathcal{H} \text{ is a } \sigma\text{-algebra associated with } \mathbb{X} \text{ and } \mathcal{C} \subset \mathcal{H}\}$ be the collection of all σ-algebras that contain $\mathcal{C}$ and $\mathcal{F}_{\mathcal{D}} := \{\mathcal{G} : \mathcal{G} \text{ is a } \sigma\text{-algebra associated with } \mathbb{X} \text{ and } \mathcal{D} \subset \mathcal{G}\}$ be the collection of all σ-algebras that contain $\mathcal{D}$. Since, $\mathcal{C} \subset \mathcal{D} \subset \mathcal{G}$, $\mathcal{G}$ is a σ-algebra that contains $\mathcal{C}$, therefore $\mathcal{G} \in \mathcal{F}_{\mathcal{C}}$. Hence, $\mathcal{F}_{\mathcal{D}} \subset \mathcal{F}_{\mathcal{C}}$ and $\bigcap_{\mathcal{H}\in\mathcal{F}_{\mathcal{C}}} \mathcal{H} \subset \bigcap_{\mathcal{G}\in\mathcal{F}_{\mathcal{D}}} \mathcal{G}$. By definition, $\sigma(\mathcal{C}) = \bigcap_{\mathcal{H}\in\mathcal{F}_{\mathcal{C}}} \mathcal{H} \subset \bigcap_{\mathcal{G}\in\mathcal{F}_{\mathcal{D}}} \mathcal{G} = \sigma(\mathcal{D})$. $\blacksquare$

Example 1.1.4 shows that if $\mathcal{F}$ is a σ-algebra associated with $\mathbb{X}$ and $S \subset \mathbb{X}$, we can easily obtain a σ-algebra associated with S by taking $S \cap \mathcal{F}$. The next theorem shows that if $\mathcal{F} := \sigma(\mathcal{C})$, then $\mathcal{F} \cap S = \sigma(\mathcal{C} \cap S)$.

Theorem 1.4. *Let $S \subset \mathbb{X}$, $\mathcal{C}$ be a collection of subsets of $\mathbb{X}$ and $\mathcal{C} \cap S = \{A \cap S : A \in \mathcal{C}\}$. Then, $\sigma(\mathcal{C} \cap S) = \sigma(\mathcal{C}) \cap S$ is a σ-algebra associated with S.*

Proof. First, note that since $\mathcal{C} \subset \sigma(\mathcal{C})$ we have $\mathcal{C} \cap S \subset \sigma(\mathcal{C}) \cap S$. From Example 1.1.4, $\sigma(\mathcal{C}) \cap S$ is a σ-algebra associated with S. Then, it follows from Theorem 1.3 that $\sigma(\mathcal{C} \cap S) \subset \sigma(\mathcal{C}) \cap S$. We need only show that $\sigma(\mathcal{C} \cap S) \supset \sigma(\mathcal{C}) \cap S$ to conclude that $\sigma(\mathcal{C} \cap S) = \sigma(\mathcal{C}) \cap S$. To this end, consider the collection of subsets of $\mathbb{X}$ (not necessarily in $\mathcal{C}$) such that their intersection with S is in $\sigma(\mathcal{C} \cap S)$, i.e. $\mathcal{G} := \{B \subset \mathbb{X} : B \cap S \in \sigma(\mathcal{C} \cap S)\}$.

By construction, $\mathcal{C} \subset \mathcal{G}$ since $A \in \mathcal{C} \implies A \cap S \in \mathcal{C} \cap S \subset \sigma(\mathcal{C} \cap S)$. Thus, $A \in \mathcal{G}$ by definition. We will show that $\mathcal{G}$ is a σ-algebra associated with $\mathbb{X}$. If this is the case, $\sigma(\mathcal{C}) \subset \mathcal{G}$. But from the definition of $\mathcal{G}$, if $A \in \sigma(\mathcal{C})$ then $A \cap S \in \sigma(\mathcal{C} \cap S)$. This means that $\sigma(\mathcal{C}) \cap S \subset \sigma(\mathcal{C} \cap S)$.

1. $\mathbb{X} \in \mathcal{G}$ since $\mathbb{X} \cap S = S \in \sigma(\mathcal{C} \cap S)$.

2. $A \in \mathcal{G}$, $A^c = \mathbb{X} - A$ and $A^c \cap S = (\mathbb{X} - A) \cap S = S - (A \cap S)$. But since $A \in \mathcal{G}$, $A \cap S \in \sigma(\mathcal{C} \cap S)$ which implies that $S - (A \cap S) \in \sigma(\mathcal{C} \cap S)$, so $A^c \in \mathcal{G}$.

3. Let $A_i \in \mathcal{G}, i \in \mathbb{N}$ and note that
$$\left(\bigcup_{i \in \mathbb{N}} A_i\right) \cap S = \bigcup_{i \in \mathbb{N}} (A_i \cap S).$$
Since, $A_i \cap S \in \sigma(\mathcal{C} \cap S)$, $\bigcup_{i \in \mathbb{N}} (A_i \cap S) \in \sigma(\mathcal{C} \cap S)$ and $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{G}$.

Thus, $\mathcal{G}$ is a σ-algebra associated with $\mathbb{X}$. ■

In what follows, it will often be the case that $\mathbb{X} := \mathbb{R}^n$ for $n \in \mathbb{N}$. In this case, an important σ-algebra is the one generated by the collection $\mathcal{O}_{\mathbb{R}^n}$ of open sets of $\mathbb{R}^n$, denoted by $\sigma\left(\mathcal{O}_{\mathbb{R}^n}\right)$. The elements of this σ-algebra are called *Borel* sets of $\mathbb{R}^n$ and $\sigma\left(\mathcal{O}_{\mathbb{R}^n}\right)$ is called the Borel σ-algebra, which is denoted by $\mathcal{B}(\mathbb{R}^n)$. If $d_{\mathbb{X}}$ is a metric on $\mathbb{X}$ we say that

$$O \subset \mathbb{X} \text{ is open} \iff \forall x \in O\ \exists\ \epsilon > 0\ \ni B(x, \epsilon) \subset O,$$

where $B(x, \epsilon) := \{y \in \mathbb{X} : d_X(x, y) < \epsilon\}$. In this more general setting, we denote by $\mathcal{O}_{\mathbb{X}}$ the collection of open sets of $\mathbb{X}$. When $\mathbb{X} := \mathbb{R}^n$ a usual choice of metric is $d_{\mathbb{R}^n}(x, y) := \|x - y\| = \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2}$, called the Euclidean metric. The next theorem shows that $\mathcal{B}(\mathbb{R}^n)$ can be generated by systems of rectangles in $\mathbb{R}^n$. Before we prove the theorem we define these rectangles.

Definition 1.3. *Let $a_i, b_i \in \mathbb{R}$ for $i = 1, \cdots, n$, $n \in \mathbb{N}$. Then,*

1. *$R^{n,o} := \times_{i=1}^n (a_i, b_i)$ is called an open rectangle in $\mathbb{R}^n$,*

2. *$R^{n,h} := \times_{i=1}^n [a_i, b_i)$ is called a half-open rectangle in $\mathbb{R}^n$.*

If $b_i \leq a_i$ for some i, $R^{n,o} = R^{n,h} = \emptyset$. When a_i and b_i are restricted to be rational numbers, i.e., a_i, $b_i \in \mathbb{Q}$ we write $R^{n,o}_{\mathbb{Q}}$ and $R^{n,h}_{\mathbb{Q}}$. The collections of all open and half-open rectangles in $\mathbb{R}^n$ are denoted by $\mathcal{I}^{n,o}$ and $\mathcal{I}^{n,h}$. Similarly, $\mathcal{I}^{n,o}_{\mathbb{Q}}$ and $\mathcal{I}^{n,h}_{\mathbb{Q}}$ denote the collections of all open and half-open rectangles in $\mathbb{R}^n$ having rational endpoints.

Theorem 1.5. $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{I}^{n,o}) = \sigma(\mathcal{I}^{n,h}) = \sigma(\mathcal{I}^{n,o}_{\mathbb{Q}}) = \sigma(\mathcal{I}^{n,h}_{\mathbb{Q}})$.

Proof. We start by noting that $R^{n,o}$ is an open set. To verify this, choose any $x \in R^{n,o}$. Since (a_i, b_i) is open for all i, there exists $\delta > 0$ such that $(x_i - \delta, x_i + \delta) \subset (a_i, b_i)$. Let $B(x, \delta) = \{y : \|y - x\| < \delta\}$ and note that $\|y - x\| < \delta \iff \sum_{i=1}^n (y_i - x_i)^2 < \delta^2 \implies (y_i - x_i)^2 < \delta^2 - \sum_{j \neq i}^n (y_j - x_j)^2 < \delta^2 \implies |y_i - x_i| < \delta \iff y_i \in (x_i - \delta, x_i + \delta) \subset (a_i, b_i)$ for all i. Hence, $B(x, \delta) \subset R^{n,o}$. Since, $\mathcal{I}^{n,o}_{\mathbb{Q}} \subset \mathcal{I}^{n,o} \subset \mathcal{O}_{\mathbb{R}^n}$, we have $\sigma(\mathcal{I}^{n,o}_{\mathbb{Q}}) \subset \sigma(\mathcal{I}^{n,o}) \subset \sigma(\mathcal{O}_{\mathbb{R}^n}) := \mathcal{B}(\mathbb{R}^n)$.

Let $O \in \mathcal{O}_{\mathbb{R}^n}$ and consider the set $\bigcup\limits_{R_{\mathbb{Q}}^{n,o} \subset O} R_{\mathbb{Q}}^{n,o}$. If $x \in \bigcup\limits_{R_{\mathbb{Q}}^{n,o} \subset O} R_{\mathbb{Q}}^{n,o}$ then $x \in R_{\mathbb{Q}}^{n,o} \subset O$. Hence, $\bigcup\limits_{R_{\mathbb{Q}}^{n,o} \subset O} R_{\mathbb{Q}}^{n,o} \subset O$.

Now, choose $x \in O$. Since O is open, there exists $B(x, \epsilon) \subset O$. Let $R^{n,o} = \{y \in \mathbb{R}^n : a_i < y_i < b_i \text{ for } i = 1, \cdots, n\}$ be an open rectangle that contains x. Then, $|y_i - x_i| < b_i - a_i$ and $\sum_{i=1}^n (y_i - x_i)^2 < \sum_{i=1}^n (b_i - a_i)^2 < n m_n^2$ where $m_n = \max\limits_{1 \le i \le n} (b_i - a_i)$. If $m_n < \frac{\epsilon}{\sqrt{n}}$, then $\sum_{i=1}^n (y_i - x_i)^2 < \epsilon^2$ and we conclude that $R^{n,o} \subset B(x, \epsilon)$. Since the set of all points in $\mathbb{R}^n$ with rational coordinates is a dense subset of $\mathbb{R}^n$, we can find $R_{\mathbb{Q}}^{n,o} \subset R^{n,o} \subset B(x, \epsilon)$. Hence, every $x \in O$ belongs to a rectangle $R_{\mathbb{Q}}^{n,o} \subset O$ and, consequently, $x \in \bigcup\limits_{R_{\mathbb{Q}}^{n,o} \subset O} R_{\mathbb{Q}}^{n,o}$. Hence, $O \subset \bigcup\limits_{R_{\mathbb{Q}}^{n,o} \subset O} R_{\mathbb{Q}}^{n,o}$. Combining this set containment with the one in the previous paragraph we $O = \bigcup\limits_{R_{\mathbb{Q}}^{n,o} \subset O} R_{\mathbb{Q}}^{n,o}$.

Since the open rectangles in $\bigcup\limits_{R_{\mathbb{Q}}^{n,o} \subset O} R_{\mathbb{Q}}^{n,o}$ have rational endpoints, the union has countably many elements. Furthermore, since σ-algebras are closed under countable unions, we have that $O \in \sigma(\mathcal{I}_{\mathbb{Q}}^{n,o})$. Hence, $\sigma(\mathcal{O}_{\mathbb{R}^n}) \subset \sigma(\mathcal{I}_{\mathbb{Q}}^{n,o})$. Combining this set containment with $\sigma(\mathcal{I}_{\mathbb{Q}}^{n,o}) \subset \sigma(\mathcal{I}^{n,o}) \subset \sigma(\mathcal{O}_{\mathbb{R}^n}) := \mathcal{B}(\mathbb{R}^n)$, we conclude that $\sigma(\mathcal{I}_{\mathbb{Q}}^{n,o}) = \sigma(\mathcal{I}^{n,o}) = \sigma(\mathcal{O}_{\mathbb{R}^n}) := \mathcal{B}(\mathbb{R}^n)$.

Lastly, note that if $a_i, b_i \in \mathbb{Q}$ for all i, $R_{\mathbb{Q}}^{n,h} = \bigcap\limits_{i \in \mathbb{N}} (a_1 - 1/i, b_1) \times \cdots \times (a_n - 1/i, b_n)$ and $R_{\mathbb{Q}}^{n,o} = \bigcup\limits_{i \in \mathbb{N}} [a_1 + 1/i, b_1) \times \cdots \times [a_n + 1/i, b_n)$. Similarly, if $a_i, b_i \in \mathbb{R}$, $R^{n,h} = \bigcap\limits_{i \in \mathbb{N}} (a_1 - 1/i, b_1) \times \cdots \times (a_n - 1/i, b_n)$ and $R^{n,o} = \bigcup\limits_{i \in \mathbb{N}} [a_1 + 1/i, b_1) \times \cdots \times [a_n + 1/i, b_n)$ we have $\sigma(\mathcal{I}^{n,o}) = \sigma(\mathcal{I}^{n,h})$ and $\sigma(\mathcal{I}_{\mathbb{Q}}^{n,o}) = \sigma(\mathcal{I}_{\mathbb{Q}}^{n,h})$, which completes the proof. ■

The collections of rectangles in Definition 1.3 are not the only systems of $\mathbb{R}^n$ that generate the Borel sets. The next theorem shows that the collection of closed sets of $\mathbb{R}^n$, denoted by $\mathcal{C}_{\mathbb{R}^n}$, and the collection of compact sets of $\mathbb{R}^n$, denoted by $\mathcal{K}_{\mathbb{R}^n}$, also generate the Borel sets.

Theorem 1.6. *Let $\mathcal{C}_{\mathbb{R}^n}$, $\mathcal{K}_{\mathbb{R}^n}$ be the collections of closed and compact subsets of $\mathbb{R}^n$. Then, $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{C}_{\mathbb{R}^n}) = \sigma(\mathcal{K}_{\mathbb{R}^n})$.*

Proof. Let $A \subset \mathbb{R}^n$. Then, A compact $\iff$ A closed and bounded. Thus, $\mathcal{K}_{\mathbb{R}^n} \subset \mathcal{C}_{\mathbb{R}^n}$. Hence, by Theorem 1.3, $\sigma(\mathcal{K}_{\mathbb{R}^n}) \subset \sigma(\mathcal{C}_{\mathbb{R}^n})$.

Now, if $C \in \mathcal{C}_{\mathbb{R}^n}$ and $\bar{B}(\theta, k) = \{x \in \mathbb{R}^n : \|x\| \ \leq k, \, k \in \mathbb{N}\}$ is a closed ball with radius k centered at $\theta = (0, \cdots, 0)^T \in \mathbb{R}^n$, then $C_k := C \cap \bar{B}(\theta, k)$ is closed and bounded. Boundedness follows by construction and closeness follows from the fact that complements of open sets are closed, De Morgan's Laws and the fact that arbitrary unions of open sets are open. Hence, $C_k \in \mathcal{K}_{\mathbb{R}^n}$ for all $k \in \mathbb{N}$. By construction, $C = \bigcup_{k \in \mathbb{N}} C_k$, thus $C \in \sigma(\mathcal{K}_{\mathbb{R}^n})$ and $\sigma(\mathcal{C}_{\mathbb{R}^n}) \subset \sigma(\mathcal{K}_{\mathbb{R}^n})$. Hence, combining this set containment with $\sigma(\mathcal{K}_{\mathbb{R}^n}) \subset \sigma(\mathcal{C}_{\mathbb{R}^n})$ we obtain $\sigma(\mathcal{C}_{\mathbb{R}^n}) = \sigma(\mathcal{K}_{\mathbb{R}^n})$.

Since $\mathcal{C}_{\mathbb{R}^n} = (\mathcal{O}_{\mathbb{R}^n})^c$, we have that $\mathcal{C}_{\mathbb{R}^n} \subset \sigma(\mathcal{O}_{\mathbb{R}^n})$ and consequently $\sigma(\mathcal{C}_{\mathbb{R}^n}) \subset \sigma(\mathcal{O}_{\mathbb{R}^n})$. The converse $\sigma(\mathcal{O}_{\mathbb{R}^n}) \subset \sigma(\mathcal{C}_{\mathbb{R}^n})$ follows similarly to give $\sigma(\mathcal{C}_{\mathbb{R}^n}) = \sigma(\mathcal{O}_{\mathbb{R}^n})$. ■

1.2 The structure of $\mathbb{R}$ and its Borel sets

Recall that an open interval on $\mathbb{R}$ is a set $(a, b) := \{x \in \mathbb{R} : a < x < b\}$ and a closed interval is a set $[a, b] := \{x \in \mathbb{R} : a \leq x \leq b\}$. They are said to be finite if $a, b \in \mathbb{R}$ and infinite if $a = -\infty$ or $b = \infty$.

Definition 1.4. *Let S be an open subset of $\mathbb{R}$. An open finite or infinite interval I is called a component interval of S if $I \subset S$ and if $\nexists$ an open interval J such that $I \subset J \subset S$.*

Theorem 1.7. *Let I denote a component interval of the open set S. If $x \in S$, then $\exists I \ni x \in I$. If $x \in I$, then $x \notin J$ where J is any other component interval of S.*

Proof. Since S is open, for any $x \in S$ there exists an open interval I such that $x \in I$ and $I \subset S$. There may be many such intervals, but the largest is $I_x = (a(x), b(x))$, where $a(x) = \inf\{a : (a, x) \subset S\}$, $b(x) = \sup\{b : (x, b) \subset S\}$. Note, a may be $-\infty$ and b may be $+\infty$. There is no open interval $J \ni I_x \subset J \subset S$ and by definition I_x is a component interval of S. If J_x is another component interval containing x, $I_x \cup J_x$ is an open interval with $I_x \subset I_x \cup J_x \subset S$ and $J_x \subset I_x \cup J_x \subset S$. By definition of a component interval $I_x \cup J_x = I_x$ and $I_x \cup J_x = J_x$, so $I_x = J_x$. ■

Theorem 1.8. *Let $S \subset \mathbb{R}$ be open and nonempty. Then, $S = \bigcup_{n \in \mathbb{N}} I_n$ where $\{I_n\}_{n \in \mathbb{N}}$ is a collection of component intervals of S.*

Proof. By Theorem 1.7 if $x \in S$, then x belongs to one, and only one, component interval I_x. Note that $\bigcup_{x \in S} I_x = S$ and by the definition of component intervals and the proof of the previous theorem, the collection of component intervals is disjoint (if x belongs to I_x and J_x, both component intervals, $I_x = J_x$). Let $\{q_1, q_2, \cdots\}$ be the collection of rational numbers (countable). In each component interval, there may be infinitely many of these, but among these there is exactly one with smallest index n. Define a function F, $F(I_x) = n$ if I_x contains the rational number q_n. If $F(I_x) = F(I_y) = n$ then I_x and I_y contain q_n, and $I_x = I_y$. Thus, the collection of component intervals is countable, since F is a bijection between a subset of $\mathbb{N}$ and the intervals I_x. ■

Remark 1.2. *Several collections of subsets of $\mathbb{R}$ generate $\mathcal{B}(\mathbb{R})$. In particular, we have:*

1. *Let $\mathcal{A}_1 = \{I : I = (a, b) \text{ with } -\infty \leq a < b \leq \infty\}$. Since (a, b) is open in $\mathbb{R}$, $\mathcal{A}_1 \subset \mathcal{O}_{\mathbb{R}}$ and $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{O}_{\mathbb{R}}) := \mathcal{B}(\mathbb{R})$. Every nonempty open set $O \subset \mathbb{R}$ can be written as $O = \bigcup_{n \in \mathbb{N}} I_n$, where I_n is a component interval of O. $I_n \in \mathcal{A}_1$ $\forall n$ and $I_n \in \sigma(\mathcal{A}_1)$, hence $O \in \sigma(\mathcal{A}_1)$. Thus, $\mathcal{O}_{\mathbb{R}} \subset \sigma(\mathcal{A}_1)$ and $\sigma(\mathcal{O}_{\mathbb{R}}) \subset \sigma(\mathcal{A}_1)$. Together with $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{O}_{\mathbb{R}})$ gives $\sigma(\mathcal{O}_{\mathbb{R}}) = \sigma(\mathcal{A}_1)$.*

2. *Since $[a, b] = \bigcap_{n \in \mathbb{N}} (a - 1/n, b + 1/n)$, we have $[a, b] \in \sigma(\mathcal{A}_1)$. Hence, the collection of closed intervals $\mathcal{A}_2 = \{I : I = [a, b], a, b \in \mathbb{R}\}$ is such that $\mathcal{A}_2 \subset \sigma(\mathcal{A}_1)$. Hence $\sigma(\mathcal{A}_2) \subset \sigma(\mathcal{A}_1)$. Also, since $(a, b) = \bigcup_{n \in \mathbb{N}} [a + 1/n, b - 1/n]$, we have that $(a, b) \in \sigma(\mathcal{A}_2)$. Hence, the collection of open intervals $\mathcal{A}_1$ is such that $\mathcal{A}_1 \subset \sigma(\mathcal{A}_2)$ and $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{A}_2)$. Hence, $\sigma(\mathcal{A}_1) = \sigma(\mathcal{A}_2)$. But since, $\sigma(\mathcal{A}_1) = \sigma(\mathcal{O}_{\mathbb{R}})$, $\sigma(\mathcal{A}_2) = \sigma(\mathcal{O}_{\mathbb{R}})$.*

3. *Let $\mathcal{A}_3 = \{I : I = (a, b] : -\infty \leq a < b < \infty\}$. Note that since $(a, b) = \bigcup_{n \in \mathbb{N}} (a, b - \frac{1}{n}]$ we have that $(a, b) \in \sigma(\mathcal{A}_3)$. Consequently, $\mathcal{A}_1 \subset \sigma(\mathcal{A}_3)$ and $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{A}_3)$. Also,*

since $(a,b] = \bigcup_{n\in\mathbb{N}} (a, b+\frac{1}{n})$ *we have that* $(a,b] \in \sigma(\mathcal{A}_1)$*. Consequently,* $\mathcal{A}_3 \subset \sigma(\mathcal{A}_1)$ *and* $\sigma(\mathcal{A}_3) \subset \sigma(\mathcal{A}_1)$*. Thus,* $\sigma(\mathcal{A}_3) = \sigma(\mathcal{A}_1)$*.*

4. *Let* $\mathcal{A}_4 = \{I : I = (-\infty, a] : a \in \mathbb{R}\}$*. Note that* $(-\infty, a] = \bigcap_{n\in\mathbb{N}} (-\infty, a+\frac{1}{n}) \in \sigma(\mathcal{A}_1)$*. Hence,* $\mathcal{A}_4 \subset \sigma(\mathcal{A}_1)$ *and* $\sigma(\mathcal{A}_4) \subset \sigma(\mathcal{A}_1)$*. Now, for* $a < b$

$$\begin{aligned}(a,b) &= (-\infty,b) \cap (a,\infty) = (-\infty,b) \cap (-\infty,a]^c \\ &= \left(\bigcup_{n\in\mathbb{N}} (-\infty, b-\frac{1}{n}]\right) \cap (-\infty,a]^c \in \sigma(\mathcal{A}_4).\end{aligned}$$

Hence, $\mathcal{A}_1 \subset \sigma(\mathcal{A}_4)$ *and* $\sigma(\mathcal{A}_1) \subset \sigma(\mathcal{A}_4)$*. Together with the reverse set containment and item 1. in this remark, we have* $\sigma(\mathcal{O}_{\mathbb{R}}) = \sigma(\mathcal{A}_1) = \sigma(\mathcal{A}_4)$*.*

1.3 Measures

Given a measurable space $(\mathbb{X}, \mathcal{F})$, we are ready to define what is meant by a measure. The goal is to associate with a measurable set a non-negative number that conveys an idea of its "size." This general idea of size must inherit the properties we intuitively associate to measures of length, area or volume.

Definition 1.5. *Let* $(\mathbb{X}, \mathcal{F})$ *be a measurable space. A measure* μ *is a function* $\mu : \mathcal{F} \to [0,\infty]$ *having the following properties:*

1. $\mu(\emptyset) = 0$
2. *if* $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{F}$ *is a disjoint collection, i.e.,* $A_i \cap A_j = \emptyset\, \forall\, i \neq j$*,* $\mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \mu(A_i)$*.*

The triple $(\mathbb{X}, \mathcal{F}, \mu)$ is called a measure space. We note that the definition of μ requires the specification of $\mathcal{F}$, and that knowledge of $\mathcal{F}$ implies knowledge of $\mathbb{X}$, its largest element. Hence, knowledge of μ is equivalent to knowledge of the measure space.

A pre-measure is a set function that satisfies the properties of a measure but is defined on a system that is not a σ-algebra. In this case, it must be that $\emptyset$ and $\bigcup_{i\in\mathbb{N}} A_i$ are elements of the system whenever A_i is in the system for $i \in \mathbb{N}$.

Remark 1.3. *1. Property 2 in Definition 1.5 is called σ-additivity or countable additivity of μ.*

2. If $\mu(\mathbb{X}) < \infty$, the measure μ is called a finite measure. In this case, $(\mathbb{X}, \mathcal{F}, \mu)$ is called a finite measure space.

3. A sequence $\{A_1, A_2, \cdots\} \in \mathcal{F}$ such that $A_1 \subset A_2 \subset \cdots$ is said to be exhausting if $\bigcup_{i \in \mathbb{N}} A_i = \mathbb{X}$. A measure μ is called σ-finite if there is an exhausting sequence $\{A_1, A_2, \cdots\} \in \mathcal{F}$ such that $\mu(A_i) < \infty$ for all i.

4. If we assume that for at least one set $A \in \mathcal{F}$ we have $\mu(A) < \infty$, then property 1 follows from property 2 by letting $A_1 = A$ and $A_2 = A_3 = \cdots = \emptyset$.

We are now ready to provide the definition of a probability space and, introduce notation and terminology that will be used henceforth.

Definition 1.6. *Let $(\Omega, \mathcal{F}, P)$ be a measure space such that $P(\Omega) = 1$. We call $(\Omega, \mathcal{F}, P)$ a probability space and P is called a probability measure.*

In the context of probability spaces, Ω is called the outcome space and the elements of $\mathcal{F}$ are called events. The construction of useful measure, or probability, spaces requires some effort as we will soon discover. What follows are simple examples of measure or probability spaces.

Example 1.2. *1. Let $(\mathbb{X}, \mathcal{F})$ be a measurable space and $F \in \mathcal{F}$. Define $\mu_\#(F) = \infty$ if F has infinitely many elements and $\mu_\#(F) =$ number of elements (cardinality) of F (denoted by $\#F$) if F has finitely many elements. $\mu_\#$ is called the counting measure and $(\mathbb{X}, \mathcal{F}, \mu_\#)$ is a measure space.*

We verify that $\mu_\#$ satisfies the defining properties in Definition 1.5. It is evident that for any $F \in \mathcal{F}$, $\mu_\#(F) \in [0, \infty]$, and since the empty set has no elements $\mu_\#(\emptyset) = 0$.

For property 2 in Definition 1.5, consider $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{F}$, a disjoint collection. There are three cases to consider: a) for at least one i, A_i has infinitely many elements. In this case, $\mu_\#(A_i) = \infty$ and since $\bigcup_{i\in\mathbb{N}} A_i$ has infinitely many elements $\mu_\#\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \infty$. Also, $\sum_{i\in\mathbb{N}} \mu(A_i) = \#A_1 + \cdots + \infty + \cdots = \infty$; b) $\forall i$, A_i has finitely many elements and there are only N of these sets that are non-empty. Relabel the sets such that the first N are non-empty. Then, $\mu_\#\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \mu_\#(A_1 \cup \cdots \cup A_N) = \sum_{i=1}^{N} \mu_\#(A_i) = \sum_{i=1}^{\infty} \mu_\#(A_i)$; c) $\forall i$, A_i has finitely many elements and there are only N of these sets that are empty. Then, as in case a) $\mu_\#\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \infty$ and $\sum_{i\in\mathbb{N}} \mu(A_i) = \#A_1 + \#A_2 + \cdots = \infty$.

2. *Let $(\mathbb{X}, \mathcal{F})$ be a measurable space and for $x \in \mathbb{X}$ and $F \in \mathcal{F}$ let $\mu_x(F) = 1$ if $x \in F$ and $\mu_x(F) = 0$ if $x \notin F$. This is called the unit mass at x or Dirac's delta measure. $(\mathbb{X}, \mathcal{F}, \mu_x)$ is a probability space.*

   Clearly, for fixed $x \in \mathbb{X}$ and any $F \in \mathcal{F}$, $\mu_x(F) \in \{0,1\} \subset [0,\infty]$. Also, since the empty set has no elements, $x \notin \emptyset$, hence $\mu_x(\emptyset) = 0$. For property 2 in Definition 1.5, consider $\{A_i\}_{i\in\mathbb{N}} \in \mathcal{F}$, a disjoint collection. If $x \in \bigcup_{i\in\mathbb{N}} A_i$, then it must be that it belongs to one, and only one, A_i. Then, $\mu_x\left(\bigcup_{i\in\mathbb{N}} A_i\right) = 1$ and $\sum_{i=1}^{\infty} \mu_x(A_i) = 1+0+0+\cdots = 1$. If $x \notin \bigcup_{i\in\mathbb{N}} A_i$, then it does not belong to any A_i. Thus, $\mu_x\left(\bigcup_{i\in\mathbb{N}} A_i\right) = 0$ and $\sum_{i=1}^{\infty} \mu_x(A_i) = 0$.

3. *Let $\Omega = \{\omega_i\}_{i\in\mathbb{N}}$ and $p_i \in [0,1]$ for $i \in \mathbb{N}$ with $\sum_{i\in\mathbb{N}} p_i = 1$. Let $(\Omega, 2^\Omega)$ be a measurable space, then the set function*

$$P(A) = \sum_{i:\omega_i\in A} p_i = \sum_{i\in\mathbb{N}} p_i \mu_{\omega_i}(A), \;\; A \subset \Omega$$

   is a probability measure.

   Since every $A \in \mathcal{F}$ is a finite or infinite collection of ω_i's and $\sum_{i\in\mathbb{N}} p_i = 1$,

$$0 \le P(A) = \sum_{i:\omega_i\in A} p_i = \sum_{i\in\mathbb{N}} p_i \mu_{\omega_i}(A) \le 1,$$

where μ_{ω_i} is Dirac's delta measure. Hence, we immediately have that

$$P(\emptyset) = \sum_{i \in \mathbb{N}} p_i \mu_{\omega_i}(\emptyset) = 0.$$

For property 2 in Definition 1.5, consider $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{F}$, a disjoint collection. Then,

$$\begin{aligned} P\left(\bigcup_{i \in \mathbb{N}} A_i\right) &= \sum_{j \in \mathbb{N}} p_j \mu_{\omega_j}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{j \in \mathbb{N}} p_j \sum_{i \in \mathbb{N}} \mu_{\omega_j}(A_i) = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} p_j \mu_{\omega_j}(A_i) \\ &= \sum_{i \in \mathbb{N}} P(A_i) \end{aligned}$$

The second equality follows from the properties of the Dirac measure, and the third follows from the possibility of interchanging infinite sums in this context.

4. *Consider tossing a coin, and define the possible outcomes as heads H or tails T. Hence, the outcome space is $\Omega = \{H, T\}$ and associate with it the following σ-algebra, $\mathcal{F} = \{\emptyset, \Omega, \{H\}, \{T\}\}$. Now, define $P : \mathcal{F} \to [0, 1]$ as follows*

$$P(\emptyset) = 0,\ P(\{H\}) = 0.5,\ P(\{T\}) = 0.5,$$

implying by that $P(\Omega) = 1$ by σ-additivity. $(\Omega, \mathcal{F}, P)$ is a probability space.

1.3.1 Properties and characterization of measures

The following theorem gives properties of measures that follow directly from Definition 1.5 and basic operations with sets.

Theorem 1.9. *Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $\{A_i\}_{i \in \mathbb{N}} \in \mathcal{F}$. Then,*

1. *$A_2 \subset A_1 \implies \mu(A_2) \leq \mu(A_1)$ (monotonicity) and if $\mu(A_2) < \infty$, $\mu(A_1 - A_2) = \mu(A_1) - \mu(A_2)$.*

2. *$\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2) - \mu(A_1 \cap A_2)$*

3. *$\mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) \leq \sum_{i \in \mathbb{N}} \mu(A_i)$ (sub-additivity)*

Proof. 1. Note that $A_1 = A_2 \cup (A_1 - A_2)$ and that A_2 and $A_1 - A_2$ are disjoint sets. Hence, $\mu(A_1) = \mu(A_2 \cup (A_1 - A_2)) = \mu(A_2) + \mu(A_1 - A_2)$, which implies $\mu(A_2) \leq \mu(A_1)$. Now, if $\mu(A_2) < \infty$, $\mu(A_1) - \mu(A_2) = \mu(A_2) - \mu(A_2) + \mu(A_1 - A_2) = \mu(A_1 - A_2)$.

2. $A_2 \cup A_1 = A_2 \cup (A_1 - A_2)$ and $A_1 = (A_2 \cap A_1) \cup (A_1 - A_2)$. By the second equality, given that $(A_2 \cap A_1)$ and $(A_1 - A_2)$ are disjoint, $\mu(A_1) = \mu(A_2 \cap A_1) + \mu(A_1 - A_2)$. By the first, $\mu(A_2 \cup A_1) = \mu(A_2) + \mu(A_1 - A_2)$. Hence, $\mu(A_1) = \mu(A_2 \cap A_1) + \mu(A_2 \cup A_1) - \mu(A_2)$, which gives 2.

3. Let $B_1 = A_1$, $B_2 = A_2 - A_1$, $B_3 = A_3 - \cup_{j=1}^{2} A_j, \cdots$ $\{B_i\}_{i \in \mathbb{N}}$ is a disjoint collection and $B_i \subset A_i$ for all i. Since, $\bigcup_{i \in \mathbb{N}} A_i = \bigcup_{i \in \mathbb{N}} B_i$, $\mu\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \mu\left(\bigcup_{i \in \mathbb{N}} B_i\right) = \sum_{i \in \mathbb{N}} \mu(B_i) \leq \sum_{i \in \mathbb{N}} \mu(A_i)$. ■

Theorem 1.9 establishes for measurable sets and arbitrary measures what seems intuitive for intervals of $\mathbb{R}$ and their lengths. Hence, if we "measure" open or half-open intervals of the type (a, b) or $(a, b]$ by their length, $l = (b - a)$, then it is easily verified l satisfies all properties in Theorem 1.9.

Measures have continuity properties that will play an important role in our study of probability spaces. For this purpose we define what is meant by the limit of a sequence of sets.

Definition 1.7. *Let $\{A_n\}_{n \in \mathbb{N}}$ be a sequence of sets.*

1. *If $A_1 \subset A_2 \subset A_3 \subset \cdots$ then $\lim_{n \to \infty} A_n := \bigcup_{n \in \mathbb{N}} A_n$,*

2. *if $A_1 \supset A_2 \supset A_3 \supset \cdots$ then $\lim_{n \to \infty} A_n := \bigcap_{n \in \mathbb{N}} A_n$,*

3. *if $\{A_i\}_{i \in \mathbb{N}}$ is an arbitrary sequence of sets and $n \in \mathbb{N}$, let $B_n = \bigcap_{i \geq n} A_i$ (note that $B_1 \subset B_2 \subset \cdots$) and $C_n = \bigcup_{i \geq n} A_i$ (note that $C_1 \supset C_2 \supset \cdots$). Then, let $B = \lim_{n \to \infty} B_n = \bigcup_{n \in \mathbb{N}} \bigcap_{i \geq n} A_i$ and $C = \lim_{n \to \infty} C_n = \bigcap_{n \in \mathbb{N}} \bigcup_{i \geq n} A_i$. We say that $A = \lim_{n \to \infty} A_n$ exists if $B = C$,*

and we write $A = B = C$. B is called the limit inferior of $\{A_n\}_{n\in\mathbb{N}}$ and denoted by $\liminf_{n\to\infty} A_n$ and C is called the limit superior of $\{A_n\}_{n\in\mathbb{N}}$ and denoted by $\limsup_{n\to\infty} A_i$.

Theorem 1.10. *Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space. Then,*

1. *if $A_1 \subset A_2 \subset \cdots$, $\mu(A) = \lim_{n\to\infty} \mu(A_n)$, where $A = \lim_{n\to\infty} A_n$, and*
2. *if $A_1 \supset A_2 \supset \cdots$ and $\mu(A_1) < \infty$, $\mu(A) = \lim_{n\to\infty} \mu(A_n)$, where $A = \lim_{n\to\infty} A_n$.*

Proof. 1. Let $B_1 = A_1$, $B_2 = A_2 - A_1$, $B_3 = A_3 - A_2 \cdots$ and note that $A_n = \bigcup_{i=1}^n B_i$. Hence, $\mu(A_n) = \mu\left(\bigcup_{i=1}^n B_i\right)$. Since $B_i \cap B_j = \emptyset$ for all $i \neq j$, $\mu(A_n) = \sum_{i=1}^n \mu(B_i)$. Taking limits on both sides of the last equality gives,

$$\lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \sum_{i=1}^n \mu(B_i) = \sum_{i\in\mathbb{N}} \mu(B_i) = \mu\left(\bigcup_{i\in\mathbb{N}} B_i\right),$$

where the last equality follows from σ-additivity of μ. Since, $\bigcup_{i\in\mathbb{N}} B_i = \bigcup_{i\in\mathbb{N}} A_i = A$, we have $\lim_{n\to\infty} \mu(A_n) = \mu(A)$.
2. Since A_1 is the largest set in the sequence $\{A_i\}_{i\in\mathbb{N}}$, we put $A_i^c := A_1 - A_i$ and note that $A_1^c \subset A_2^c \subset A_3^c \subset \cdots$. Since $A = \bigcap_{i\in\mathbb{N}} A_i$, by de Morgan's Laws $A^c = \bigcup_{i\in\mathbb{N}} A_i^c$ and, consequently, $\mu(A_1 - A) = \mu\left(\bigcup_{i\in\mathbb{N}} A_i^c\right) = \lim_{n\to\infty} \mu(A_1 - A_n)$, where the last equality follows from part 1. By monotonicity of measures, $\mu(A_1) < \infty \implies \mu(A_n), \mu(A) < \infty \, \forall \, n$, and by part 1 of Theorem 1.9 we have

$$\mu(A_1 - A) = \mu(A_1) - \mu(A) = \lim_{n\to\infty} \mu(A_1 - A_n) = \lim_{n\to\infty} (\mu(A_1) - \mu(A_n)) = \mu(A_1) - \lim_{n\to\infty} \mu(A_n),$$

giving $\mu(A) = \lim_{n\to\infty} \mu(A_n)$. ■

As a matter of terminology, we say that part 1 of Theorem 1.10 establishes continuity of measures *from below*, whereas part 2 establishes continuity of measures *from above*.

The next theorem gives necessary and sufficient conditions for a set function $m : \mathcal{F} \to [0, \infty]$ to be a measure.

Theorem 1.11. *Let $(\mathbb{X}, \mathcal{F})$ be a measurable space. A function $m : \mathcal{F} \to [0, \infty]$ is a measure if, and only if,*

1. *$m(\emptyset) = 0$,*

2. *for $A_1, A_2 \in \mathcal{F}$ disjoint $m(A_1 \cup A_2) = m(A_1) + m(A_2)$,*

3. *for $A_1, A_2, \dots \in \mathcal{F}$ and $A_1 \subset A_2 \subset \cdots$ with $A = \lim_{n\to\infty} A_n$ we have*

$$m(A) = \lim_{n\to\infty} m(A_n).$$

Proof. If m is a measure then conditions 1 and 2 in this theorem follow directly from properties 1 and 2 from the definition of measure. Condition 3 follows from part 1 of Theorem 1.10.

Now, assume that m satisfies conditions 1-3 in this theorem. Since condition 1 in this theorem is the same as property 1, we need only show that m satisfies property 2 from the definition of measure. Let $\{B_j\}_{j\in\mathbb{N}}$ be any pairwise disjoint sequence in $\mathcal{F}$ and define $A_n := \bigcup_{j=1}^n B_j$. Then, $A_1 \subset A_2 \subset \cdots$ and $A := \lim\limits_{n\to\infty} A_n = \bigcup\limits_{n\in\mathbb{N}} A_n = \bigcup\limits_{j\in\mathbb{N}} B_j$. By condition 2, we have $m(A_n) = \sum_{j=1}^n m(B_j)$ and from condition 3 we conclude that

$$m\left(\bigcup_{j\in\mathbb{N}} B_j\right) = m(A) = \lim_{n\to\infty} m(A_n) = \lim_{n\to\infty}\left(\sum_{j=1}^n m(B_j)\right) = \sum_{j=1}^\infty m(B_j),$$

establishing that m is σ-additive.■

Remark 1.4. *Condition 3 in Theorem 1.11 can be replaced by the assumption that m is continuous from above if $m(\mathbb{X}) < \infty$. To see this, note that if m is a measure, it is continuous from above by part 2 of Theorem 1.10. Now, assume that m is continuous from above and consider a sequence $\{B_j\}_{j\in\mathbb{N}}$ of disjoint sets in $\mathcal{F}$. Put $A_n = \bigcup_{j=1}^n B_j$ and note that*

$A_1^c \supset A_2^c \supset \cdots$ *and* $m(A_n^c) = m(\mathbb{X} - A_n) = m\left(\mathbb{X} - \bigcup_{j=1}^n B_j\right)$.

$$\lim_{n\to\infty} m(A_n^c) = m\left(\mathbb{X} - \bigcup_{j=1}^{n} B_j\right) = m(\mathbb{X}) - \lim_{n\to\infty} m\left(\bigcup_{j=1}^{n} B_j\right) \text{ since } m(\mathbb{X}) < \infty$$

$$= m(\mathbb{X}) - \lim_{n\to\infty} \sum_{j=1}^{n} m(B_j) = m(\mathbb{X}) - \sum_{j=1}^{\infty} m(B_j) \text{ by additivity of } m. \tag{1.1}$$

Now,

$$\lim_{n\to\infty} m(A_n^c) = m\left(\bigcap_{j\in\mathbb{N}} A_j^c\right) \text{ by continuity of } m \text{ from above}$$

$$= m\left(\left(\bigcup_{j\in\mathbb{N}} A_j\right)^c\right) \text{ by de Morgan's Laws}$$

$$= m(\mathbb{X}) - m\left(\bigcup_{j\in\mathbb{N}} A_j\right) = m(\mathbb{X}) - m\left(\bigcup_{j\in\mathbb{N}} B_j\right). \tag{1.2}$$

Combining (1.1) *and* (1.2) *gives* $m\left(\bigcup_{j\in\mathbb{N}} B_j\right) = \sum_{j\in\mathbb{N}} m(B_j)$.

Similarly, condition 3 in Theorem 1.11 can be replaced by the assumption that m *is continuous at* $\emptyset$ *if* $m(\mathbb{X}) < \infty$. *Continuity at the* $\emptyset$ *means that if* $A_1 \supset A_2 \supset \cdots$ *and* $\lim_{n\to\infty} A_n = \emptyset$ *with* $\mu(A_1) < \infty$ *and* $\lim_{n\to\infty} \mu(A_n) = 0$.

Since probability measures are finite, Theorem 1.11 and Remark 1.4 provide characterizations for probabilities. Consequently, we state the following theorem without proof.

Theorem 1.12. *Let* $(\Omega, \mathcal{F})$ *be a measurable space. A function* $P : \mathcal{F} \to [0, 1]$ *is a probability measure if, and only if,*

1. $P(\emptyset) = 0$,

2. *for* $A_1, A_2 \in \mathcal{F}$ *disjoint* $P(A_1 \cup A_2) = P(A_1) + P(A_2)$,

3. *for* $A_1, A_2, \dots \in \mathcal{F}$ *and* $A_1 \subset A_2 \subset \cdots$ *with* $A = \lim_{n\to\infty} A_n$ *we have*

$$P(A) = \lim_{n\to\infty} P(A_n).$$

Condition 3 can be substituted by either

3'. $A_1, A_2, \dots \in \mathcal{F}$ *and* $A_1 \supset A_2 \supset \cdots$ *with* $A = \lim_{n\to\infty} A_n$ *we have*

$$P(A) = \lim_{n\to\infty} P(A_n)$$

or

3". $A_1, A_2, \dots \in \mathcal{F}$ *and* $A_1 \supset A_2 \supset \cdots$ *with* $\lim_{n\to\infty} A_n = \emptyset$ *we have*

$$\lim_{n\to\infty} P(A_n) = P(\emptyset) = 0.$$

In addition, since in probability spaces $P(\Omega) = 1$, P has properties that general measures do not have. In the next theorem we establish some of these properties.

Theorem 1.13. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space. Then,*

1. $P(A^c) = 1 - P(A)\, \forall\, A \in \mathcal{F}$,

2. $A \subset B \implies P(A) \leq P(B)\, \forall A,\, B \in \mathcal{F}$,

3. *if* $\{A_i\}_{i=1}^n \in \mathcal{F}$ *for* $n \in \mathbb{N}$ *then*

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) + \sum_{1\leq i_1<i_2<i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) \\ + \cdots + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i\right) \end{aligned} \tag{1.3}$$

Proof. 1. $\Omega = A \cup A^c$. Hence, $1 = P(\Omega) = P(A) + P(A^c) \implies P(A^c) = 1 - P(A)$.

2. follows from Theorem 1.9.1.

3. Let $n = 2$. Then, from Theorem 1.9.2 we have

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2). \tag{1.4}$$

Now, let $B_1 = A_1,\ B_2 = B_1 \cup A_2 = A_1 \cup A_2,\ B_3 = B_2 \cup A_3 = A_1 \cup A_2 \cup A_3,\ \cdots,\ B_{n-1} = B_{n-2} \cup A_{n-1} = A_1 \cup \cdots \cup A_{n-1}$. Now, suppose

$$P(B_{n-1}) = P\left(\bigcup_{i=1}^{n-1} A_i\right) = \sum_{i=1}^{n-1} P(A_i) - \sum_{1 \le i_1 < i_2 \le n-1} P(A_{i_1} \cap A_{i_2}) + \sum_{1 \le i_1 < i_2 < i_3 \le n-1} P(A_{i_1} \cap A_{i_2} \cap A_{i_1}) + \cdots + (-1)^n P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \quad (1.5)$$

We will show that (1.4) and (1.5) imply (1.3), establishing 3. by induction. From (1.4) we have that

$$\begin{aligned}
P(B_n) = P\left(\cup_{i=1}^{n} A_i\right) = P(B_{n-1} \cup A_n) &= P(B_{n-1}) + P(A_n) - P(B_{n-1} \cap A_n) \\
&= P(B_{n-1}) + P(A_n) - P((\cup_{i=1}^{n-1} A_i) \cap A_n) \\
&= P(B_{n-1}) + P(A_n) - P(\cup_{i=1}^{n-1}(A_i \cap A_n)) \\
&= P(B_{n-1}) + P(A_n) - P(\cup_{i=1}^{n-1} C_i),\ \text{where } C_i = (A_i \cap A_n).
\end{aligned}$$

But,

$$P\left(\cup_{i=1}^{n-1} C_i\right) = \sum_{i=1}^{n-1} P(C_i) - \sum_{1 \le i_1 < i_2 \le n-1} P(C_{i_1} \cap C_{i_2}) + \sum_{1 \le i_1 < i_2 < i_3 \le n-1} P(C_{i_1} \cap C_{i_2} \cap C_{i_3}) + \cdots + (-1)^n P(C_1 \cap C_2 \cap \cdots \cap C_{n-1}),$$

with

$$\begin{aligned}
\sum_{n=1}^{n-1} P(C_i) &= \sum_{i=1}^{n-1} P(A_i \cap A_n) \\
\sum_{1 \le i_1 < i_2 \le n-1} P(C_{i_1} \cap C_{i_2}) &= \sum_{1 \le i_1 < i_2 \le n-1} P(A_{i_1} \cap A_n \cap A_{i_2} \cap A_n) \\
&= \sum_{1 \le i_1 < i_2 \le n-1} P(A_{i_1} \cap A_{i_2} \cap A_n) \\
\sum_{1 \le i_1 < i_2 < i_3 \le n-1} P(C_{i_1} \cap C_{i_2} \cap C_{i_3}) &= \sum_{1 \le i_1 < i_3 < i_3 \le n-1} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_n) \\
&\vdots \\
P(C_1 \cap C_2 \cap \cdots \cap C_{n-1}) &= P(A_1 \cap \cdots \cap A_n).
\end{aligned}$$

Then, we have

$$\begin{aligned}
P(B_n) = &\sum_{i=1}^{n-1} P(A_i) - \sum_{1\le i_1<i_2\le n-1} P(A_{i_1}\cap A_{i_2}) + \sum_{1\le i_1<i_2<i_3\le n-1} P(A_i\cap A_j\cap A_k) + \\
&\cdots + (-1)^n P(A_1\cap A_2\cap\cdots\cap A_{n-1}) + P(A_n) \\
&- \sum_{i=1}^{n-1} P(A_i\cap A_n) + \sum_{1\le i_1<i_2\le n-1} P(A_{i_1}\cap A_{i_2}\cap A_n) \\
&- \sum_{1\le i_1<i_2<i_3\le n-1} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}\cap A_n) + \cdots + (-1)^{n+1} P(A_{i_1}\cap\cdots\cap A_n) \\
&= \sum_{i=1}^{n} P(A_i) - \sum_{i_1<i_2} P(A_{i_1}\cap A_{i_2}) + \sum_{i_1<i_2<i_3} P(A_{i_1}\cap A_{i_2}\cap A_{i_3}) + \cdots \\
&+ (-1)^{n+1} P(\cap_{i=1}^{n} A_i).
\end{aligned}$$

■

Remark 1.5. *Note that the terms on the right side of* (1.3) *alternate in sign.*

The next theorem shows that probability measures are continuous set functions.

Theorem 1.14. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{A_n\}_{n\in\mathbb{N}} \in \mathcal{F}$. Suppose $A = \lim\limits_{n\to\infty} A_n$ exists. Then, $A \in \mathcal{F}$ and $P(A_n) \to P(A)$ as $n \to \infty$.*

Proof. Since $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ has a limit, there exist $C_1 \supset C_2 \supset C_3 \supset \cdots$ and $B_1 \subset B_2 \subset B_3 \subset \cdots$ as in Definition 1.7. Furthermore, since $\mathcal{F}$ is closed under countable unions and intersections, $B_n, C_n \in \mathcal{F}\ \forall\ n \in \mathbb{N}$. Since A exists, $B = \bigcup\limits_{n\in\mathbb{N}} B_n = \bigcap\limits_{n\in\mathbb{N}} C_n = C = A$ and $A \in \mathcal{F}$. By construction, $B = B_1 \cup (B_2 - B_1) \cup (B_3 - B_2) \cup \cdots = \chi_1 \cup \chi_2 \cup \cdots$. The collection $\{\chi_1, \chi_2, \cdots\}$ is pairwise disjoint. By σ-additivity of measures we have $P(B) = \sum\limits_{i\in\mathbb{N}} P(\chi_i) = \lim\limits_{n\to\infty} \sum_{i=1}^{n} P(\chi_i)$. But, $\sum_{i=1}^{n} P(\chi_i) = P(B_n)$, where $B_n = B_1 \cup (B_2 - B_1) \cup \cdots \cup (B_n - B_{n-1})$. Hence, $P(B) = \lim\limits_{n\to\infty} P(B_n)$.

By De Morgan's Laws $C = \bigcap\limits_{i\in\mathbb{N}} C_i = \left(\bigcup\limits_{i\in\mathbb{N}} C_i^c\right)^c$. Therefore, $P(C) = 1 - P\left(\bigcup\limits_{i\in\mathbb{N}} C_i^c\right)$. Now, $\bigcup\limits_{i\in\mathbb{N}} C_i^c = C_1^c \cup (C_2^c - C_1^c) \cup (C_3^c - C_2^c) \cdots = \theta_1 \cup \theta_2 \cup \theta_3 \cdots$, where the collection $\{\theta_1, \theta_2, \cdots\}$

is pairwise disjoint. Hence, $P\left(\bigcup_{i\in\mathbb{N}} C_i^c\right) = \sum_{i\in\mathbb{N}} P(\theta_i) = \lim_{n\to\infty} \sum_{i=1}^n P(\theta_i)$. But $\sum_{i=1}^n P(\theta_i) = P(C_n^c)$ and $P(C_n^c) = 1 - P(C_n)$. Hence, $P\left(\bigcup_{i\in\mathbb{N}} C_i^c\right) = \lim_{n\to\infty}(1 - P(C_n)) = 1 - \lim_{n\to\infty} P(C_n)$. Consequently, $P(C) = 1 - \left(1 - \lim_{n\to\infty} P(C_n)\right) = \lim_{n\to\infty} P(C_n)$.

Finally, by construction, $B_n \subset A_n \subset C_n$, for all n. Therefore, $P(B_n) \leq P(A_n) \leq P(C_n)$ and $\lim_{n\to\infty} P(B_n) \leq \lim_{n\to\infty} P(A_n) \leq \lim_{n\to\infty} P(C_n)$ or $P(B) \leq \lim_{n\to\infty} P(A_n) \leq P(C)$ and consequently since $A = B = C$, $\lim_{n\to\infty} P(A_n) = P(A)$. ■

Definition 1.8. *Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space. $N \in \mathcal{F}$ is called a μ-null set or, simply, a null set if $\mu(N) = 0$. The collection containing all μ-null sets in $\mathcal{F}$ is denoted by $\mathcal{N}_\mu$.*

Since $\emptyset \in \mathcal{F}$ and $\mu(\emptyset) = 0$ we have that $\emptyset \in \mathcal{N}_\mu$. Also, if $N \in \mathcal{N}_\mu$, $M \subset N$ and $M \in \mathcal{F}$, by monotonicity of measures $0 \leq \mu(M) \leq \mu(N) = 0$. Hence, $M \in \mathcal{N}_\mu$. In addition, if $\{N_j\}_{j\in\mathbb{N}} \in \mathcal{N}_\mu$, by sub-additivity of measures $0 \leq \mu\left(\bigcup_{j\in\mathbb{N}} N_j\right) \leq \sum_{j\in\mathbb{N}} \mu(N_j) = 0$. Hence, $\bigcup_{j\in\mathbb{N}} N_j \in \mathcal{N}_\mu$.

Note that there might be subsets M of μ-null sets that are not in $\mathcal{F}$. This motivates the following definition.

Definition 1.9. *A measure space $(\mathbb{X}, \mathcal{F}, \mu)$ is said to be complete if every subset of μ-null sets are elements of $\mathcal{F}$.*

The next theorem shows that any measure space can be "completed" in such a way that the resulting measure space is complete.

Theorem 1.15. *Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and define:*

1. *$\bar{\mathcal{F}} := \{F \cup M : F \in \mathcal{F}$ and $M \in \mathcal{S}\}$ where $\mathcal{S}$ is the collection of all subsets of μ-null sets,*

2. *$\bar{\mu} : \bar{\mathcal{F}} \to [0, 1]$ such that $\bar{\mu}(F \cup M) = \mu(F)$.*

$(\mathbb{X}, \bar{\mathcal{F}}, \bar{\mu})$ *is a complete measure space and* $\mathcal{F} \subset \bar{\mathcal{F}}$.

Proof. We start by showing that $\bar{\mathcal{F}}$ is a σ-algebra. Note that since $\emptyset \in \mathcal{S}$, we have $\forall\ F \in \mathcal{F}$ that $F \cup \emptyset = F \in \bar{\mathcal{F}}$. Hence, $\mathcal{F} \subset \bar{\mathcal{F}}$. Now, we verify the that $\bar{\mathcal{F}}$ satisfies the defining characteristics for σ-algebras.

1. $\mathbb{X} \in \bar{\mathcal{F}}$. This follows from the fact that $\mathbb{X} \in \mathcal{F} \subset \bar{\mathcal{F}}$.

2. $A \in \bar{\mathcal{F}} \implies A^c \in \bar{\mathcal{F}}$. $A \in \bar{\mathcal{F}} \implies A = F \cup M$ where $F \in \mathcal{F}$ and $M \in \mathcal{S}$ and $M \subset N \in \mathcal{N}_\mu$. $A^c = F^c \cap M^c = F^c \cap M^c \cap \mathbb{X} = F^c \cap M^c \cap (N^c \cup N) = (F^c \cap M^c \cap N^c) \cup (F^c \cap M^c \cap N)$. Since $M \subset N$, $M^c \supset N^c$ and therefore $A^c = (F^c \cap N^c) \cup (F^c \cap M^c \cap N)$. But since $(F^c \cap N^c) \in \mathcal{F}$ and $F^c \cap M^c \cap N \subset N$, by definition $A^c \in \bar{\mathcal{F}}$.

3. $\{A_j\}_{j \in \mathbb{N}} \in \bar{\mathcal{F}} \implies \bigcup\limits_{j \in \mathbb{N}} A_j \in \bar{\mathcal{F}}$. Since $A_j \in \bar{\mathcal{F}}$, $A_j = F_j \cup M_j$ where $F_j \in \mathcal{F}$ and $M_j \in \mathcal{S}$. Now,
$$\bigcup_{j \in \mathbb{N}} A_j = \bigcup_{j \in \mathbb{N}} (F_j \cup M_j) = \left(\bigcup_{j \in \mathbb{N}} F_j \right) \cup \left(\bigcup_{j \in \mathbb{N}} M_j \right).$$
Now, $\bigcup\limits_{j \in \mathbb{N}} F_j \in \mathcal{F}$ and $\bigcup\limits_{j \in \mathbb{N}} M_j \subset \bigcup\limits_{j \in \mathbb{N}} N_j$ where $N_j \in \mathcal{N}_\mu$. Hence, $\bigcup\limits_{j \in \mathbb{N}} N_j \in \mathcal{N}_\mu$ and $\bigcup\limits_{j \in \mathbb{N}} M_j \in \mathcal{S}$. Then, by definition $\bigcup\limits_{j \in \mathbb{N}} A_j \in \bar{\mathcal{F}}$.

We now show that $\bar{\mu}$ is a measure on $\bar{\mathcal{F}}$. Note that $A \in \bar{\mathcal{F}}$ is not uniquely represented as we may have $G \cup O = A = F \cup M$. Note that for $\bar{\mu}$ to be well-defined we need $\mu(G) = \bar{\mu}(G \cup O) = \bar{\mu}(A) = \bar{\mu}(F \cup M) = \mu(F)$, i.e., $\mu(G) = \mu(F)$. Now,

$F \subset F \cup M = G \cup O \subset G \cup N$ where $N \in \mathcal{N}_\mu$ and $G \subset G \cup O = F \cup M \subset F \cup N'$ where $N' \in \mathcal{N}_\mu$.

Consequently, $\mu(F) \leq \mu(G) + \mu(N)$ and $\mu(G) \leq \mu(F) + \mu(N')$. Since $\mu(N) = \mu(N') = 0$ we have $\mu(F) = \mu(G)$.

Now, we verify that $\bar{\mu}$ satisfies the defining properties of measures.

1. Since $\emptyset = \emptyset \cup \emptyset \in \bar{\mathcal{F}}$, we have $\bar{\mu}(\emptyset) = \mu(\emptyset) = 0$.

2. Let $\{A_j\}_{j\in\mathbb{N}} \in \bar{\mathcal{F}}$ be a pairwise disjoint collection. Since $A_j = F_j \cup M_j$, it must be that $\{F_j\}_{j\in\mathbb{N}}$ is a pairwise disjoint collection.

$$\bar{\mu}\left(\bigcup_{j\in\mathbb{N}} A_j\right) = \bar{\mu}\left(\bigcup_{j\in\mathbb{N}}(F_j \cup M_j)\right) = \bar{\mu}\left(\left(\bigcup_{j\in\mathbb{N}} F_j\right) \cup \left(\bigcup_{j\in\mathbb{N}} M_j\right)\right)$$
$$= \mu\left(\bigcup_{j\in\mathbb{N}} F_j\right) = \sum_{j\in\mathbb{N}} \mu(F_j) = \sum_{j\in\mathbb{N}} \bar{\mu}(F_j \cup M_j) = \sum_{\mathbb{N}} \bar{\mu}(A_j).$$

Hence, $(\mathbb{X}, \bar{\mathcal{F}}, \bar{\mu})$ is a measure space. We now verify that it is complete. Take $N \in \mathcal{N}_{\bar{\mu}}$ and $A \subset N$. We need to shaw that $A \in \bar{\mathcal{F}}$. Note that $A \subset N = F \cup M$ where $F \in \mathcal{F}$ and $M \in \mathcal{S}$. Since $0 = \bar{\mu}(N) = \mu(F)$ and M is a subset of a μ-null set (N'), then

$$A \subset N = F \cup M \subset F \cup N' \in \mathcal{F} \text{ and } \mu(F \cup N') \leq \mu(F) + \mu(N') = 0.$$

Hence, A is a subset of a μ-null set and therefore $A \in \mathcal{S}$. In particular, $A = A \cup \emptyset$ and $A \in \bar{\mathcal{F}}$.■

1.4 Independence of events and conditional probability

We start by defining probabilistic independence of events.

Definition 1.10. *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $2 \leq n \in \mathbb{N}$ and $\{E_i\}_{1\leq i\leq n} \subset \mathcal{F}$. The events $E_1, \cdots, E_n \in \mathcal{F}$ are said to be independent if*

$$P\left(\bigcap_{m\in I} E_m\right) = \prod_{m\in I} P(E_m) \text{ for all } I \subset \{1, \cdots, n\} \text{ with } \#I \geq 2\ . \tag{1.6}$$

Remark 1.6. *Note that (1.6) contains $\sum_{i=2}^{n} \binom{n}{i} = 2^n - n - 1$ equations. All of them must hold to characterize independence of the events $E_1, \cdots, E_n \in \mathcal{F}$.*

If two events are independent, their complements are independent and so are any of the events with complement of the other.

Theorem 1.16. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $E_1, E_2 \in \mathcal{F}$ are independent, then:*

1. *E_1 and E_2^c are independent (or E_1^c and E_2 are independent).*

2. *E_1^c and E_2^c are independent.*

Proof. 1. Recall that $E_1 \cup E_2 = E_2 \cup (E_1 \cap E_2^c)$ and $P(E_1 \cup E_2) = P(E_2) + P(E_1 \cap E_2^c)$. The last equality together with Theorem 1.9.2 gives $P(E_1) - P(E_1 \cap E_2) = P(E_1 \cap E_2^c)$. Now, by independence of E_1 and E_2 we have $P(E_1 \cap E_2^c) = P(E_1) - P(E_1)P(E_2)$. Hence, $P(E_1 \cap E_2^c) = P(E_1)(1 - P(E_2)) = P(E_1)P(E_2^c)$.

2. Note that

$$\begin{aligned}
E_1^c \cap E_2^c &= (E_1 \cup E_2)^c \text{ by DeMorgan's Laws. Hence,} \\
P(E_1^c \cap E_2^c) &= P((E_1 \cup E_2)^c) \\
P(E_1^c \cap E_2^c) &= 1 - P(E_1 \cup E_2) \text{ by Theorem 1.13} \\
&= 1 - (P(E_1) + P(E_2) - P(E_1)P(E_2)) \text{ by independence of } E_1 \text{ and } E_2 \\
&= (1 - P(E_1))(1 - P(E_2)) = P(E_1^c)P(E_2^c),
\end{aligned}$$

as desired. ■

There is a useful probability measure that can easily be defined from knowledge of $(\Omega, \mathcal{F}, P)$. It is called conditional probability. What follows is a definition.

Definition 1.11. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Given any $E \in \mathcal{F}$ such that $P(E) > 0$, we define $P(\cdot|E) : \mathcal{F} \to [0, 1]$ as*

$$P(A|E) = \frac{P(A \cap E)}{P(E)} \ \forall A \in \mathcal{F}.$$

Note that $P(\emptyset|E) = P(\emptyset \cap E)/P(E) = P(\emptyset)/P(E) = 0$ and $P(\Omega|E) = P(\Omega \cap E)/P(E) = P(E)/P(E) = 1$. In addition, if $\{E_j\}_{j \in \mathbb{N}}$ forms a pairwise disjoint collection of events

$$P\left(\bigcup_{j \in \mathbb{N}} E_j | E\right) = \frac{P\left(\left(\bigcup_{j \in \mathbb{N}} E_j\right) \cap E\right)}{P(E)} = \frac{P\left(\bigcup_{j \in \mathbb{N}} (E_j \cap E)\right)}{P(E)} = \sum_{j \in \mathbb{N}} \frac{P(E_j \cap E)}{P(E)} = \sum_{j \in \mathbb{N}} P(E_j|E).$$

Hence, $P(\cdot|E)$ is a probability measure on $(\Omega, \mathcal{F})$ and $P(A|E)$ is called the probability of A conditional on E.

The notion of independence between two events is related to the notion of conditional probability. In fact, as the next theorem demonstrates, if knowledge of event E does not change the probability of event A, i.e., if $P(A|E) = P(A)$, A and E are independent.

Theorem 1.17. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $E_1, E_2 \in \mathcal{F}$ such that $P(E_2) > 0$. E_1 and E_2 are independent $\iff$ $P(E_1|E_2) = P(E_1)$.*

Proof. ($\Longrightarrow$) Since E_1 and E_2 are independent $P(E_1 \cap E_2) = P(E_1)P(E_2)$ and since $P(E_1|E_2) = \frac{P(E_1 \cap E_2)}{P(E_2)}$ we have $P(E_1|E_2) = \frac{P(E_1)P(E_2)}{P(E_2)} = P(E_1)$.
($\Longleftarrow$) $P(E_1|E_2) = P(E_1) \implies P(E_1 \cap E_2)/P(E_2) = P(E_1)$. Hence, $P(E_1 \cap E_2) = P(E_1)P(E_2) \implies E_1$ and E_2 are independent. ■

Theorem 1.18. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{E_j\}_{1\leq j\leq n} \subset \mathcal{F}$. If $P\left(\bigcap_{1\leq j\leq n-1} E_j\right) > 0$ then*

$$P\left(\bigcap_{1\leq j\leq n} E_j\right) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2)\cdots P(E_n|E_1 \cap E_2 \cap \cdots \cap E_{n-1}). \tag{1.7}$$

Proof. Note that if $P\left(\bigcap_{1\leq j\leq n-1} E_j\right) > 0$ then $P\left(\bigcap_{1\leq j\leq m} E_j\right) > 0$ for all $m < n-1$. Hence, all conditional probabilities on the right-hand side of (1.7) are well defined.

For $n = 2$, we have that if $P(E_1) > 0$, $P(E_2|E_1) = P(E_1 \cap E_2)/P(E_1)$ which implies

$$P(E_1 \cap E_2) = P(E_1)P(E_2|E_1). \tag{1.8}$$

Now, assume that

$$P\left(\bigcap_{1\leq j\leq n-1} E_j\right) = P(E_1)P(E_2|E_1)P(E_3|E_1 \cap E_2)\cdots P(E_{n-1}|E_1 \cap E_2 \cap \cdots \cap E_{n-2}) \tag{1.9}$$

and define $B_n = (E_1 \cap E_2 \cdots E_{n-1}) \cap E_n$. Then,

$$\begin{aligned}P(B_n) &= P(E_1 \cap \cdots \cap E_{n-1}) P(E_n | E_1 \cap \cdots \cap E_{n-1}) \text{ by } (1.8)\\ &= P(E_1)P(E_2|E_1)\cdots P(E_{n-1}|E_1 \cap E_2 \cap \cdots \cap E_{n-2})P(E_n|E_1 \cap \cdots \cap E_{n-1}) \text{ by } (1.9).\end{aligned}$$

The result follows by induction. ■

The next theorem provides the *total probability* formula for an event. It is the foundation for Bayes' Theorem, which plays an important role in statistics. First, we define a partition of a set Ω.

Definition 1.12. *$\{E_1, E_2, \cdots\}$ is a partition of Ω if $\bigcup\limits_{i \in \mathbb{N}} E_i = \Omega$ and $E_i \cap E_j = \emptyset$, for all $i \neq j$.*

Theorem 1.19. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{E_1, E_2, \cdots\} \in \mathcal{F}$ be a partition of Ω with $P(E_i) > 0$ for all $i \in \mathbb{N}$. If $A \in \mathcal{F}$,*

$$P(A) = \sum_{i \in \mathbb{N}} P(A|E_i)P(E_i).$$

Proof. $A = A \cap \Omega = A \cap \left(\bigcup\limits_{i \in \mathbb{N}} E_i \right) = \bigcup\limits_{i \in \mathbb{N}} (A \cap E_i)$. The collection $\{(A \cap E_1), (A \cap E_2), \cdots\}$ is pairwise disjoint. Therefore, $P(A) = \sum_{i \in \mathbb{N}} P(A \cap E_i) = \sum_{i \in \mathbb{N}} P(A|E_i)P(E_i)$. ■

Theorem 1.20. *(Bayes' Theorem) Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{E_j\}_{j \in \mathbb{N}} \subset \mathcal{F}$ be a partition of Ω with $P(E_i) > 0$ for all $i \in \mathbb{N}$. Let $A \in \mathcal{F}$ such that $P(A) > 0$. Then,*

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{\sum\limits_{j \in \mathbb{N}} P(A|E_j)P(E_j)},$$

Proof. By Theorem (1.19) $P(A) = \sum\limits_{j \in \mathbb{N}} P(A|E_j)P(E_j) \neq 0$. Hence,

$$P(E_i|A) = \frac{P(E_i \cap A)}{P(A)} = \frac{P(A|E_i)P(E_i)}{\sum\limits_{j \in \mathbb{N}} P(A|E_j)P(E_j)}$$

which establishes the desired result. ■

In the context of Bayes' Theorem, $P(E_i)$ is called the *prior* probability of E_i and $P(E_i|A)$ is called the *posterior* probability of E_i given the event A. The following example illustrates how posterior probabilities can be obtained from priors.

Example 1.3. *Suppose that each student in a class can be classified as good G or bad B. The probability of selecting a good student from a class is $P(G) = 0.7$ and, consequently, the probability of selecting a bad student is $P(B) = 0.3$. A student may pass A or fail F a class. The probability that a good student will pass is $P(A|G) = 0.9$ and the probability that a bad student will pass is $P(A|B) = 0.4$. We are interested in the probability that a student that fails is a good student, i.e., $P(G|F)$. From Bayes' Theorem,*

$$P(G|F) = \frac{P(F|G)P(G)}{P(F|G)P(G) + P(F|B)P(B)} = \frac{0.1 \times 0.7}{0.1 \times 0.7 + 0.6 \times 0.3} = 0.28.$$

Taking the prior probabilities as given, minimization $P(G|F)$ involves maximizing $P(F|B)$ and minimizing $P(F|G)$.

Chapter 2

Construction of probability measures

We have revealed a number of properties of measures, but we have not discussed their existence (in general) or how to construct them.

Definition 2.1. *1. A system $\mathcal{P}$ associated with $\mathbb{X}$ is called a π-system if $A, B \in \mathcal{P} \implies A \cap B \in \mathcal{P}$.*

2. A system $\mathcal{D}$ associated with $\mathbb{X}$ is called a Dynkin[1] *system if:*

a) $\mathbb{X} \in \mathcal{D}$

b) $A \in \mathcal{D} \implies A^c \in \mathcal{D}$

c) $\{A_j\}_{j\in\mathbb{N}} \subset \mathcal{D}$ and $A_i \cup A_j = \emptyset \; \forall \; i \neq j, \; i, j \in \mathbb{N} \implies \bigcup_{j\in\mathbb{N}} A_j \in \mathcal{D}$.

It is evident from this definition that a σ-algebra associated with $\mathbb{X}$ is also a Dynkin system associated with $\mathbb{X}$.

Theorem 2.1. *Let $\mathcal{C} \subset 2^{\mathbb{X}}$. There exists a smallest Dynkin system $\delta(\mathcal{C})$ such that $\mathcal{C} \subset \delta(\mathcal{C})$. It is called the Dynkin system generated by $\mathcal{C}$. In addition, $\delta(\mathcal{C}) \subset \sigma(\mathcal{C})$.*

Proof. Existence and characterization of $\delta(\mathcal{C})$ is proved as in Theorem 1.2. Since $\sigma(\mathcal{C})$ is a Dynkin system $\delta(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$. Since $\mathcal{C} \subset \sigma(\mathcal{C})$, $\delta(\mathcal{C}) \subset \delta(\sigma(\mathcal{C})) = \sigma(\mathcal{C})$ as in Theorem 1.3.

■

[1] Eugene Borisovich Dynkin was a Russian mathematician that made important contributions to algebra and probability. He was a student of Andrei Kolmogorov.

The next theorem shows that a Dynkin system is a σ-algebra if, and only if, it is a π-system.

Theorem 2.2. *A Dynkin system $\mathcal{D}$ is a σ-algebra $\iff A, B \in \mathcal{D} \implies A \cap B \in \mathcal{D}$.*

Proof. ($\implies$) If $\mathcal{D}$ is a σ-algebra, then $A, B \in \mathcal{D} \implies A \cap B = (A^c \cup B^c)^c \in \mathcal{D}$.
($\Longleftarrow$) If $\mathcal{D}$ is a Dynkin system it satisfies requirements 1 and 2 for σ-algebras in Definition 1.1. Let $A_i \in \mathcal{D}$ for $i \in \mathbb{N}$, we must show that $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{D}$. Define $B_1 := A_1$, $B_2 := A_2 - B_1 = A_2 \cap B_1^c$, $B_3 := A_3 - \cup_{i=1}^2 B_i = A_3 \cap (\cup_{i=1}^2 B_i)^c \cdots$ $B_n := A_n - \cup_{i=1}^{n-1} B_i = A_n \cap (\cup_{i=1}^{n-1} B_i)^c$. The collection $\{B_i\}_{i\in\mathbb{N}}$ is pairwise disjoint, and since each B_i is the intersection of two sets in $\mathcal{D}$, using closeness under finite intersections, $\bigcup_{i\in\mathbb{N}} B_i = \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{D}$. ■

Theorem 2.3. *If $\mathcal{P} \subset 2^{\mathbb{X}}$ is a π-system, then $\delta(\mathcal{P}) = \sigma(\mathcal{P})$.*

Proof. From Theorem 2.1, $\delta(\mathcal{P}) \subset \sigma(\mathcal{P})$ and from Theorem 2.2 if $\delta(\mathcal{P})$ is a π-system it is a σ-algebra. Since $\sigma(\mathcal{P})$ is the smallest σ-algebra it must be that $\delta(\mathcal{P}) = \sigma(\mathcal{P})$, so it suffices to show that $\delta(\mathcal{P})$ is a π-system. For any $D \in \delta(\mathcal{P})$, let $\mathcal{D}_D = \{A \subset \mathbb{X} : A \cap D \in \delta(\mathcal{P})\}$. First, we show that $\mathcal{D}_D$ is a Dynkin system. We verify conditions a), b) and c) in Definition 2.1.

a) Note that $\mathbb{X} \cap D = D \in \delta(\mathcal{P})$, hence $\mathbb{X} \in \mathcal{D}_D$.

b) If $A \in \mathcal{D}_D$, then $A \cap D \in \delta(\mathcal{P})$. Now, $A^c \cap D = (A^c \cup D^c) \cap D = (A \cap D)^c \cap D = ((A \cap D) \cup D^c)^c$ where $A \cap D$ and D^c are disjoint. Also, since $D \in \delta(\mathcal{P})$ so is D^c, and $A \cap D \in \delta(\mathcal{P})$ by assumption, so $((A \cap D) \cup D^c)^c \in \delta(\mathcal{P})$. Thus $A^c \in \mathcal{D}_D$.

c) Let A_i for $i \in \mathbb{N}$ be pairwise disjoint with $A_i \cap D \in \delta(\mathcal{P})$ and note that $\{(A_i \cap D)\}_{i\in\mathbb{N}}$ forms a disjoint collection. Thus, $\bigcup_{i\in\mathbb{N}} (A_i \cap D) = D \cap \bigcup_{i\in\mathbb{N}} A_i$ and $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{D}_D$. Thus, $\mathcal{D}_D$ is a Dynkin system.

Fix $G \in \mathcal{P}$. Then, $G \in \delta(\mathcal{P})$ and we can define $\mathcal{D}_G = \{A \subset \mathbb{X} : A \cap G \in \delta(\mathcal{P})\}$. Now, consider $G' \in \mathcal{P}$. Since, $\mathcal{P}$ is a π-system, $G' \cap G \in \mathcal{P} \subset \delta(\mathcal{P})$. Hence, $G' \in \mathcal{D}_G$, showing that $\mathcal{P} \subset \mathcal{D}_G$ for all $G \in \mathcal{P}$. But $\mathcal{D}_G$ is a Dynkin system and consequently, by definition $\delta(\mathcal{P}) \subset \mathcal{D}_G$, $\forall\, G \in \mathcal{P}$.

Thus, we have that if $D \in \delta(\mathcal{P})$ and $G \in \mathcal{P}$, then $G \cap D \in \delta(\mathcal{P})$ and $\mathcal{P} \subset \mathcal{D}_D$ (by definition of $\mathcal{D}_D$). Then, $\delta(\mathcal{P}) \subset \mathcal{D}_D$ for all $D \in \delta(\mathcal{P})$ implying that $\delta(\mathcal{P})$ is a π-system by definition of $\mathcal{D}_D$. ■

The following theorem shows that under some conditions, measures that coincide on some generating class $\mathcal{G}$ coincide on $\sigma(\mathcal{G})$.

Theorem 2.4. *Let $(\mathbb{X}, \sigma(\mathcal{P}))$ be a measurable space and $\mathcal{P}$ a collection of subsets of $\mathbb{X}$, such that:*

1. *$\mathcal{P}$ is a π-system,*
2. *there exists $\{P_j\}_{j\in\mathbb{N}} \subset \mathcal{P}$ with $P_1 \subset P_2 \subset \cdots$ such that $\bigcup_{j\in\mathbb{N}} P_j := \lim_{j\to\infty} P_j = \mathbb{X}$ (the sequence $\{P_j\}_{j\in\mathbb{N}}$ is exhausting).*

Then, if μ and v are measures that coincide on $\mathcal{P}$ and are finite for all P_j, $\mu(A) = v(A)$, for all $A \in \sigma(\mathcal{P})$.

Proof. For $j \in \mathbb{N}$ let $\mathcal{D}_j = \{A \in \sigma(\mathcal{P}) : \mu(A \cap P_j) = v(A \cap P_j)\}$. First, we show that $\mathcal{D}_j$ is a Dynkin system.

1. $\mathbb{X} \in \mathcal{D}_j$ since $\mu(\mathbb{X} \cap P_j) = \mu(P_j) = v(P_j) = v(\mathbb{X} \cap P_j)$.
2. Let $A \in \mathcal{D}_j$. Note that $P_j = (A \cap P_j) \cup (A^c \cap P_j)$ and note that the two sets in the union are disjoint. Since μ is a measure $\mu(P_j) = \mu(A \cap P_j) + \mu(A^c \cap P_j)$. Hence, $\mu(A^c \cap P_j) = \mu(P_j) - \mu(A \cap P_j)$. Since μ and v coincide in $\mathcal{P}$ we have that $v(P_j) = \mu(P_j)$ and since $A \in \mathcal{D}_j$ we have that $\mu(A \cap P_j) = v(A \cap P_j)$. Hence,

   $$\mu(A^c \cap P_j) = \mu(P_j) - \mu(A \cap P_j) = v(P_j) - v(A \cap P_j) = v(A^c \cap P_j).$$

   Thus, $A^c \in \mathcal{D}_j$.

3. Let $A_1, A_2, \cdots$ be a pairwise disjoint collection in $\mathcal{D}_j$.

$$\begin{aligned}
\mu\left(\left(\bigcup_{i\in\mathbb{N}} A_i\right)\cap P_j\right) &= \mu\left(\bigcup_{i\in\mathbb{N}}(A_i\cap P_j)\right) = \sum_{i=1}^{\infty}\mu(A_i\cap P_j)\\
&= \sum_{i=1}^{\infty} v(A_i\cap P_j) \text{ since } A_i\in\mathcal{D}_j\\
&= v\left(\bigcup_{i\in\mathbb{N}}(P_j\cap A_i)\right) = v\left(P_j\cap\left(\bigcup_{i\in\mathbb{N}} A_i\right)\right)
\end{aligned}$$

and consequently, $\cup_{i\in\mathbb{N}} A_i \in \mathcal{D}_j$.

Since $\mathcal{P}$ is a π-system, by Theorem 2.3 $\delta(\mathcal{P}) = \sigma(\mathcal{P})$ and $\mathcal{P}\subset\mathcal{D}_j$ by definition of $\delta(\mathcal{P})$, hence $\sigma(\mathcal{P})\subset\mathcal{D}_j$. But by construction $\mathcal{D}_j\subset\sigma(\mathcal{P})$ and we conclude that $\mathcal{D}_j = \sigma(\mathcal{P})$. So, for all $A\in\sigma(\mathcal{P})$ and $j = 1, 2, \cdots$,

$$\mu(A\cap P_j) = v(A\cap P_j). \tag{2.1}$$

By continuity of measures from below and noting that $(A_1\cap P_1)\subset(A\cap P_2)\subset\cdots$, letting $j\to\infty$ in (2.1) we have for all $A\in\sigma(\mathcal{P})$,

$$\begin{aligned}
\lim_{j\to\infty}\mu(A\cap P_j) &= \mu\left(\lim_{j\to\infty}(A\cap P_j)\right) = \mu\left(\underset{j\in\mathbb{N}}{\cup}(A\cap P_j)\right)\\
&= \mu\left(A\cap\left(\underset{j\in\mathbb{N}}{\cup}P_j\right)\right) = \mu\left(A\cap\mathbb{X}\right)\\
&= \mu(A)
\end{aligned}$$

Similarly, $\lim_{j\to\infty} v(A\cap P_j) = v(A)$ and we conclude that $\mu(A) = v(A)$. ■

We take the following path to construct a measure on $\mathcal{F}$. We start with a class of subsets $\mathcal{S}$ of $\mathbb{X}$, such that $\mathcal{F} = \sigma(\mathcal{S})$, and define a pre-measure μ on $\mathcal{S}$. If $\mathcal{S}$ and μ satisfy the requirements of Theorem 2.4, then μ will extend uniquely to $\mathcal{F}$, provided we are able to extend it from $\mathcal{S}$ to $\mathcal{F}$. The result that provides the conditions and possibility for such an extension is known as Carathéodory's Extension Theorem. Before stating this theorem we need the following definition and some remarks.

Definition 2.2. *A semi-ring, denoted by $\mathcal{S}$, is a system associated with $\mathbb{X}$ having the following properties:*

1. $\emptyset \in \mathcal{S}$,

2. $A, B \in \mathcal{S} \implies A \cap B \in \mathcal{S}$,

3. *for all $A, B \in \mathcal{S}$ there exists $m \in \mathbb{N}$ and $\{S_j\}_{j=1}^m \subset \mathcal{S}$ that is pairwise disjoint such that $B - A = \cup_{j=1}^m S_j$.*

Remark 2.1. 1. *A semi-ring is a π-system in view of condition 2.*

2. *Property 3 in Definition 2.2 is equivalent to the following:*

   3'. if $A, B \in \mathcal{S}$ and $A \subset B$, then $B = A \cup \left(\bigcup_{j=1}^m S_j \right)$ where the collection $\{A, S_1, \cdots S_m\} \subset \mathcal{S}$ is pairwise disjoint.

   To verify that 3 $\implies$ 3' note that $A \subset B \implies B = A \cup (B - A) = A \cup \left(\bigcup_{j=1}^m S_j \right)$ by 3, where $\{A, S_1, \cdots S_m\} \subset \mathcal{S}$ is pairwise disjoint. Now, to verify that 3' $\implies$ 3 note that $B = (B \cap A) \cup (B - A)$. Since $(B \cap A) \subset B$, by 3' $B = (B \cap A) \cup \left(\bigcup_{j=1}^m S_j \right)$. Thus, $(B \cap A) \cup (B - A) = (B \cap A) \cup \left(\bigcup_{j=1}^m S_j \right)$ which implies that $B - A = \bigcup_{j=1}^m S_j$ where $\{S_j\}_{j=1}^m \subset \mathcal{S}$ is pairwise disjoint.

3. *A ring $\mathcal{R}$ is a non-empty system of sets associated with $\mathbb{X}$ such that $A, B \in \mathcal{R} \implies A \cup B \in \mathcal{R}$ and $A - B \in \mathcal{R}$. If $A \in \mathcal{R}$ then $A - A = \emptyset \in \mathcal{R}$. Also, if $A, B \in \mathcal{R}$, and noting that $A \cap B = A - (A - B)$, we have that $A \cap B \in \mathcal{R}$. Now let $A \subset B$, $A, B \in \mathcal{R}$. Since $B = A \cup (B - A)$ and $(B - A) \in \mathcal{R}$, we conclude that every ring is a semi-ring using property 3'.*

4. *If $\mathcal{A}$ is an algebra, then for $A, B \in \mathcal{A}$ we have that $A \cup B, A \cap B, B^c \in \mathcal{A}$, and since $A - B = A \cap B^c \in \mathcal{A}$, an algebra is a ring.*

It follows from these remarks that we have the following hierarchy of systems: $\mathcal{A}$ (algebras) are $\mathcal{R}$ (rings) are $\mathcal{S}$ (semi-rings) are π-systems.

Theorem 2.5. *(Carathéodory) Let $\mathcal{S}$ be a semi-ring of subsets of $\mathbb{X}$ and $\mu : \mathcal{S} \to [0, \infty]$ be a pre-measure. Then, μ has an extension to a measure μ on $\sigma(\mathcal{S})$. If there exists $\{E_j\}_{j\in\mathbb{N}} \in \mathcal{S}$ with $E_1 \subset E_2 \cdots$ such that $\lim_{j\to\infty} E_j \to \mathbb{X}$ and $\mu(E_j) < \infty$ for all j, then the extension is unique.*

Proof. **Step 1.** We start by defining the set function $\mu^* : 2^{\mathbb{X}} \to [0, \infty]$. For any $A \subset \mathbb{X}$ define the collection of countable covers for A that are composed of sets in $\mathcal{S}$ by

$$C(A) = \{\{S_j\}_{j\in\mathbb{N}} \subset \mathcal{S} : A \subset \underset{j\in\mathbb{N}}{\cup} S_j\}.$$

If A cannot be covered by some $\underset{j\in\mathbb{N}}{\cup} S_j$, then $C(A) = \emptyset$. Now, define

$$\mu^*(A) := \inf \left\{ \sum_{j\in\mathbb{N}} \mu(S_j) : \{S_j\}_{j\in\mathbb{N}} \in C(A) \right\},$$

where $\inf \emptyset := \infty$. Note that,

a) $\mu^*(\emptyset) = 0$, by taking $S_1 = S_2 = \cdots = \emptyset$

b) $A \subset B$ implies that every cover for B is also a cover for A, i.e., $C(B) \subset C(A)$. Therefore,

$$\mu^*(A) = \inf \left\{ \sum_{j\in\mathbb{N}} \mu(S_j) : \{S_j\}_{j\in\mathbb{N}} \in C(A) \right\} \le \inf \left\{ \sum_{j\in\mathbb{N}} \mu(T_j) : \{T_j\}_{j\in\mathbb{N}} \in C(B) \right\} = \mu^*(B).$$

c) Let $A_n \subset \mathbb{X}$ for $n \in \mathbb{N}$ and, without loss of generality, assume that $\mu^*(A_n) < \infty$ (that is $C(A_n) \neq \emptyset$). Choose $\epsilon > 0$ and let $\{S_{nk}\}_{k\in\mathbb{N}} \in C(A_n)$ be such that

$$\sum_{k\in\mathbb{N}} \mu(S_{nk}) \le \mu^*(A_n) + \epsilon/2^n.$$

Now, $\underset{n\in\mathbb{N}}{\cup} A_n \subset \underset{n\in\mathbb{N}}{\cup}\ \underset{k\in\mathbb{N}}{\cup} S_{nk}$ and by the definition of infimum and sub-additivity of pre-measures

$$\begin{aligned}\mu^*\left(\underset{n\in\mathbb{N}}{\cup} A_n\right) &\leq \sum_{n\in\mathbb{N}}\sum_{k\in\mathbb{N}}\mu(S_{nk}) \\ &\leq \sum_{n\in\mathbb{N}}\left(\mu^*(A_n)+\epsilon/2^n\right) = \sum_{n\in\mathbb{N}}\mu^*(A_n)+\epsilon.\end{aligned}$$

Hence, $\mu^*\left(\underset{n\in\mathbb{N}}{\cup} A_n\right) \leq \sum_{n\in\mathbb{N}}\mu^*(A_n)$. If $\mu^*(A_n)=\infty$ for some n, then the last inequality holds trivially.

Since μ^* satisfies properties a)-c), it is called an outer-measure on $2^{\mathbb{X}}$.

Step 2. We now show that μ^* extends μ (defined on $\mathcal{S}$) to $2^{\mathbb{X}}$. By this we mean that $\mu^*(S)=\mu(S)$ for $S\in\mathcal{S}$.

First, let $\mathcal{S}_U = \{S : S = \cup_{j=1}^m S_j, S_j \in \mathcal{S}, S_i\cap S_j=\emptyset\ \forall i\neq j \text{ and } m\in\mathbb{N}\}$ be the collection of sets that can be written as disjoint finite unions of elements of $\mathcal{S}$ and let $\bar{\mu}(S)=\sum_{j=1}^m \mu(S_j)$ for $S\in\mathcal{S}_U$. Note that $\bar{\mu}(S)$ is invariant to the pairwise disjoint finite union used to represent S. To see this, suppose $S=\cup_{j=1}^m S_j$ and $S=\cup_{k=1}^n T_k$ for $m,n\in\mathbb{N}$. Then, $\cup_{j=1}^m S_j = \cup_{k=1}^n T_k$ and $S_j = S_j\cap(\cup_{k=1}^n T_k) = \cup_{k=1}^n(S_j\cap T_k)$ and $S_j\cap T_k\in\mathcal{S}$, since a semi-ring is a π-system. Since μ is a pre-measure on $\mathcal{S}$, and $\{T_k\}_{k=1}^n$ is a pairwise disjoint collection, $\mu(S_j)=\sum_{k=1}^n\mu(T_k\cap S_j)$. Then,

$$\bar{\mu}(S)=\sum_{j=1}^m\mu(S_j)=\sum_{k=1}^n\sum_{j=1}^m\mu(T_k\cap S_j)=\sum_{k=1}^n\mu(T_k).$$

We now show that $\mathcal{S}_U$ is closed under (arbitrary) finite intersections and unions. If $A,B\in\mathcal{S}_U$ then $A\cap B=(\cup_{j=1}^m S_j)\cap(\cup_{k=1}^n T_k)$ where the two unions are over pairwise disjoint sets. Then, $A\cap B=\cup_{j=1}^m\cup_{k=1}^n(S_j\cap T_k)\in\mathcal{S}_U$ since $S_j\cap T_k\in\mathcal{S}$ for all j,k and $\{S_j\cap T_k\}_{j=1,k=1}^{m,n}$ is pairwise disjoint.

Also, since $S_j,T_k\in\mathcal{S}$, their difference can be written as a finite union of pairwise disjoint elements of $\mathcal{S}$. Hence, $S_j-T_k\in\mathcal{S}_U$. Now,

$$A-B=\cup_{j=1}^m S_j-\cup_{k=1}^n T_k=\cup_{j=1}^m\cap_{k=1}^n(S_j\cap T_k^c)=\cup_{j=1}^m\cap_{k=1}^n(S_j-T_k).$$

Since, $S_j - T_k \in \mathcal{S}_U$ and given that we have shown that $\mathcal{S}_U$ is closed under finite intersections, $\cap_{k=1}^{n}(S_j - T_k) \in \mathcal{S}_U$. Hence, $A - B$ is the finite union of pairwise disjoint elements in $\mathcal{S}_U$ and we conclude that $A - B \in \mathcal{S}_U$, since $\mathcal{S}_U$ is closed under pairwise disjoint unions. Lastly, since $A \cup B = (A - B) \cup (A \cap B) \cup (B - A)$ and all sets in the union are disjoint and in $\mathcal{S}_U$, we conclude that $A \cup B \in \mathcal{S}_U$.

We now show that $\bar{\mu}$ is σ-additive on $\mathcal{S}_U$, i.e., a pre-measure. Let $\{T_k\}_{k\in\mathbb{N}} \subset \mathcal{S}_U$ such that $\{T_k\}_{k\in\mathbb{N}}$ is pairwise disjoint and such that $T := \underset{k\in\mathbb{N}}{\cup} T_k \in \mathcal{S}_U$. Since $T_k \in \mathcal{S}_U$, by definition there exist $\{S_j\}_{j\in\mathbb{N}} \in \mathcal{S}$ and a sequence of $0 = n_0 \leq n_1 \leq \cdots$ of integers such that

$$T_k = S_{n_{(k-1)}+1} \cup S_{n_{(k-1)}+2} \cup \cdots \cup S_{n_k} \text{ for } k \in \mathbb{N},$$

where the collection $\{S_{n_{(k-1)}+1}, S_{n_{(k-1)}+2}, \cdots, S_{n_k}\}$ is pairwise disjoint and

$$T = \bigcup_{k\in\mathbb{N}} \bigcup_{j=n_{(k-1)}+1}^{n_k} S_j.$$

Also, since $T \in \mathcal{S}_U$, it can be written as $T = \bigcup_{l=1}^{N} U_l$ where $N \in \mathbb{N}$ with $U_l \in \mathcal{S}$ and $\{U_l\}_{l=1}^{N}$ a pairwise disjoint collection. Hence,

$$\bigcup_{l=1}^{N} U_l = \bigcup_{k\in\mathbb{N}} \bigcup_{j=n_{(k-1)}+1}^{n_k} S_j.$$

Defining disjoint subsets $J_1, \cdots, J_N$ of $\mathbb{N}$ such that $\cup_{l=1}^{N} J_l = \mathbb{N}$ we write $U_l = \bigcup_{j\in J_l} S_j$ and note that $U_l \in \mathcal{S}$. Now, $T = \bigcup_{k\in\mathbb{N}} T_k = \cup_{l=1}^{N} U_l$ and

$$\begin{aligned}
\bar{\mu}(T) &= \sum_{l=1}^{N} \mu(U_l) \text{ by definition of } \bar{\mu} \\
&= \sum_{l=1}^{N} \sum_{j\in J_l} \mu(S_j) \text{ by } \mu \text{ being a pre-measure on } \mathcal{S} \\
&= \sum_{k\in\mathbb{N}} \sum_{j=n_{(k-1)}+1}^{n_k} \mu(S_j) = \sum_{k\in\mathbb{N}} \bar{\mu}(T_k).
\end{aligned}$$

Now, for any $S \in \mathcal{S}$ and any $\mathcal{S}$-covering of S, i.e., $\{S_j\}_{j\in\mathbb{N}} \in C(S)$

$$\begin{aligned}
\mu(S) = \bar{\mu}(S) = \bar{\mu}\left(\bigcup_{j\in\mathbb{N}} S_j \cap S\right) &\text{ since } S \in \mathcal{S} \implies S \in \mathcal{S}_U \\
&\leq \sum_{j\in\mathbb{N}} \bar{\mu}(S_j \cap S) \text{ since } \bar{\mu} \text{ is a pre-measure and sub-additive} \\
&= \sum_{j\in\mathbb{N}} \mu(S_j \cap S) \leq \sum_{j\in\mathbb{N}} \mu(S_j).
\end{aligned}$$

Taking the infimum over $C(S)$, we have $\mu(S) \leq \mu^*(S)$. Now, taking $(S, \emptyset, \cdots) \in C(S)$ gives $\mu^*(S) \leq \mu(S)$. Combining the two inequalities, we have

$$\mu^*(S) = \mu(S) \text{ for all } S \in \mathcal{S}.$$

Step 3. We will show that $\mathcal{S} \subset \mathcal{A}^*$ where

$$\mathcal{A}^* = \{A \subset \mathbb{X} : \mu^*(Q) = \mu^*(Q \cap A) + \mu^*(Q \cap A^c), \forall\, Q \subset \mathbb{X}\}. \tag{2.2}$$

Let $S, T \in \mathcal{S}$ and note that $T = (T \cap S) \cup (T \cap S^c) = (T \cap S) \cup (T - S) = (T \cap S) \cup (\cup_{j=1}^m S_j)$ with $\{S_j\}_{j=1}^m$ disjoint, $m \in \mathbb{N}$ and where the last equality follows from the third defining property of semi-rings. Since μ is a pre-measure on $\mathcal{S}$ we have

$$\mu(T) = \mu(T \cap S) + \sum_{j=1}^m \mu(S_j).$$

Since μ^* and μ coincide on $\mathcal{S}$ and $T \cap S \in \mathcal{S}$, and since μ^* is sub-additive, from c) in Step 1, we have $\mu^*(T - S) = \mu^*(\cup_{j=1}^m S_j) \leq \sum_{j=1}^m \mu^*(S_j) = \sum_{j=1}^m \mu(S_j)$. Consequently,

$$\mu(T) = \mu(T \cap S) + \sum_{j=1}^m \mu(S_j) \geq \mu^*(T \cap S) + \mu^*(T - S). \tag{2.3}$$

Take $Q \subset \mathbb{X}$ and $\{T_j\}_{j\in\mathbb{N}} \in C(Q)$. Using $\mu^*(T_j) = \mu(T_j)$ and summing (2.3) over j taking $T = T_j$

$$\sum_{j\in\mathbb{N}} \mu^*(S \cap T_j) + \sum_{j\in\mathbb{N}} \mu^*(T_j - S) \leq \sum_{j\in\mathbb{N}} \mu^*(T_j).$$

Sub-additivity and monotonicity of μ^* together with $Q \subset \bigcup_{j\in\mathbb{N}} T_j$ give

$$\begin{aligned}\mu^*(Q\cap S)+\mu^*(Q-S) &\le \mu^*(\cup_{j\in\mathbb{N}}(T_j\cap S))+\mu^*(\cup_{j\in\mathbb{N}}(T_j-S))\\ &\le \sum_{j\in\mathbb{N}}\mu^*(T_j)=\sum_{j\in\mathbb{N}}\mu(T_j).\end{aligned}$$

Taking the infimum over $C(Q)$, $\mu^*(Q\cap S)+\mu^*(Q-S)\le\mu^*(Q)$. The reverse inequality follows easily from sub-additivity of μ^*. Consequently, if $S\in\mathcal{S}$ we have that $S\in\mathcal{A}^*$.

Step 4. We show that $\mathcal{A}^*$ is a σ-algebra and μ^* is a measure on $(\mathbb{X},\mathcal{A}^*)$.

1. For all $Q\subset\mathbb{X}$, $Q\cap\mathbb{X}=Q$ and $Q\cap\mathbb{X}^c=\emptyset$. Since $\mu^*(\emptyset)=0$ we have that $\mathbb{X}\in\mathcal{A}^*$.

2. For all $Q\subset\mathbb{X}$ suppose $A\in\mathcal{A}^*$, i.e.

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

But by symmetry of the right hand side of the equality due to $(A^c)^c=A$, we have $A^c\in\mathcal{A}^*$.

3. If $A, A'\in\mathcal{A}^*$, for all $Q\subset\mathbb{X}$

$$\begin{aligned}&\mu^*(Q\cap(A\cup A'))+\mu^*(Q-(A\cup A'))\\ &\qquad=\mu^*(Q\cap(A\cup(A'-A)))+\mu^*(Q-(A\cup A'))\\ &\qquad=\mu^*((Q\cap A)\cup[Q\cap(A'-A)])+\mu^*(Q-(A\cup A'))\\ &\qquad\le\mu^*(Q\cap A)+\mu^*(Q\cap(A'-A))+\mu^*(Q-(A\cup A'))\\ &\qquad\text{using subadditivity of }\mu^*\\ &\qquad=\mu^*(Q\cap A)+\mu^*((Q-A)\cap A')+\mu^*((Q-A)-A')\\ &\qquad=\mu^*(Q\cap A)+\mu^*(Q-A)=\mu^*(Q)\\ &\qquad\text{using the defining expression for }\mathcal{A}^*\text{ twice, once for }Q-A\text{ and once for }Q.\end{aligned}$$

Thus,

$$\mu^*(Q\cap(A\cup A'))+\mu^*(Q-(A\cup A'))\le\mu^*(Q). \tag{2.4}$$

Now, $Q=\{Q\cap(A\cup A')\}\cup\{Q\cap(A\cup A')^c\}$. By sub-additivity of μ^*

$$\mu^*(Q)\le\mu^*(Q\cap(A\cup A'))+\mu^*(Q-(A\cup A')). \tag{2.5}$$

Combining inequalities (2.4) and (2.5) we conclude that $\mu^*(Q) = \mu^*(Q \cap (A \cup A')) + \mu^*(Q - (A \cup A'))$ and consequently $\mathcal{A}^*$ is closed under finite unions.

If A, $A' \in \mathcal{A}^*$ such that $A \cap A' = \emptyset$, then for $Q = (A \cup A') \cap P$ with $P \subset \mathbb{X}$ the equality $\mu^*(Q \cap A) + \mu^*(Q - A) = \mu^*(Q)$ becomes

$$\mu^*((A \cup A') \cap P) = \mu^*(P \cap A) + \mu^*(P \cap A'), \forall\, P \subset \mathbb{X}.$$

For a disjoint collection $\{A_j\}_{j=1}^m \in \mathcal{A}^*$,

$$\mu^*((\cup_{j=1}^m A_j) \cap P) = \sum_{j=1}^m \mu^*(P \cap A_j).$$

If $A = \cup_{j\in\mathbb{N}} A_j$, where $\{A_j\}$ is a disjoint collection,

$$\mu^*(P \cap A) \geq \mu^*(P \cap (\cup_{j=1}^m A_j)) = \sum_{j=1}^m \mu^*(P \cap A_j).$$

Since $\cup_{j=1}^m A_j \in \mathcal{A}^*$ we have that

$$\begin{aligned}\mu^*(P) &= \mu^*(P \cap (\cup_{j=1}^m A_j)) + \mu^*(P - \cup_{j=1}^m A_j) \\ &\geq \mu^*(P \cap (\cup_{j=1}^m A_j)) + \mu^*(P - A) \\ &= \sum_{j=1}^m \mu^*(P \cap A_j) + \mu^*(P - A).\end{aligned}$$

Let $m \to \infty$, to conclude

$$\mu^*(P) \geq \sum_{j=1}^\infty \mu^*(P \cap A_j) + \mu^*(P - A) \geq \mu^*(P \cap A) + \mu^*(P - A)$$

The reverse inequality follows directly from sub-additivity of μ^*. Thus,

$$\mu^*(P) = \mu^*(P \cap A) + \mu^*(P - A),\ \forall P \subset \mathbb{X}.$$

Consequently, $A = \cup_{j\in\mathbb{N}} A_j$ where the collection $\{A_j\}_{j\in\mathbb{N}}$ is pairwise disjoint is in $\mathcal{A}^*$. Consequently, $\mathcal{A}^*$ is a Dynkin system that is closed under finite unions. By DeMorgan Laws, $\mathcal{A}^*$ is closed under finite intersections, and by Theorem 2.2, $\mathcal{A}^*$ is a σ-algebra.

Now, we show that μ^* is a measure on $\sigma(\mathcal{S})$. From above, $\mathcal{S} \subset \mathcal{A}^*$, so $\sigma(\mathcal{S}) \subset \mathcal{A}^*$. Also, μ^* is a measure on $\mathcal{A}^*$ and on $\sigma(\mathcal{S})$, which extends μ on $\mathcal{S}$. By Theorem 2.4, and under the conditions in the enunciation of this theorem, any two extensions μ^* and v^* of μ coincide on $\sigma(\mathcal{S})$. ■

Remark 2.2. $(\mathbb{X}, \mathcal{A}^*, \mu^*)$ *is a complete measure space. To verify completeness, let* $E \in \mathcal{A}^*$ *such that* $\mu^*(E) = 0$*, and consider* $B \subset E$*. We must verify that* $B \in \mathcal{A}^*$*, i.e., for any* $Q \subset \mathbb{X}$*, it must be that*

$$\mu^*(Q) = \mu^*(Q \cap B) + \mu^*(Q \cap B^c).$$

Now, $Q \cap B \subset Q \cap E \subset E \implies \mu^*(Q \cap B) \leq \mu^*(E) = 0$ *and, consequently* $\mu^*(Q \cap B) = 0$*. Also,* $Q \cap B^c \subset Q \implies \mu^*(Q \cap B^c) \leq \mu^*(Q)$*. Hence,*

$$\mu^*(Q) \geq \mu^*(Q \cap B^c) + \mu^*(Q \cap B). \tag{2.6}$$

By sub-additivity

$$\mu^*(Q) \leq \mu^*(Q \cap B^c) + \mu(Q \cap B) \tag{2.7}$$

Given (2.6) and (2.7) we have $\mu^*(Q) = \mu^*(Q \cap B^c) + \mu^*(Q \cap B)$*. In addition,* $\mu^*(B) = 0$ *follows from monotonicity of measures.*

Theorem 2.6. *Let* $R^{n,h} = \times_{i=1}^{n}[a_i, b_i)$ *for* $n \in \mathbb{N}$ *be a half-open rectangle in* $\mathbb{R}^n$ *and* $\mathcal{I}^{n,h}$ *be the collection formed by all such rectangles with real endpoints.* $\mathcal{I}^{n,h}$ *is a semi-ring.*

Proof. Let $\mathcal{I}^{1,h} = \{[a_i, b_i) : a_i \leq b_i \text{ where } a_i,\, b_i \in \mathbb{R}\}$ and note that:

1. if $b_i = a_i$, $[a_i, b_i) = \emptyset$,

2. if $[a_i, b_i)$, $[a_j, b_j) \in \mathcal{I}^{n,h}$ then $[a_i, b_i) \cap [a_j, b_j) = \begin{cases} \emptyset & \in \mathcal{I}^{1,h} \\ [a_j, b_i) & \in \mathcal{I}^{1,h} \\ [a_i, b_j) & \in \mathcal{I}^{1,h} \\ [a_i, b_i) & \in \mathcal{I}^{1,h} \end{cases}$,

3. if $[a_1, b_1) \subset [a_2, b_2)$ then $[a_2, b_2) = [a_2, a_1) \cup [a_1, b_1) \cup [b_1, b_2)$, where the members in the union are all disjoint.

Hence, $\mathcal{I}^{1,h}$ is a semi-ring.

Now, suppose $\mathcal{I}^{n,h}$ is a semi-ring. We will verify that $\mathcal{I}^{n+1,h}$ is a semi-ring. First, note that $\mathcal{I}^{n+1,h} = \mathcal{I}^{n,h} \times \mathcal{I}^{1,h}$ and since $\emptyset \in \mathcal{I}^{n,h}$ we immediately conclude that $\emptyset \in \mathcal{I}^{n+1,h}$. The intersection of two rectangles in $\mathcal{I}^{n+1,h}$ is given by

$$(R^{n,h} \times R^{1,h}) \cap (I^{n,h} \times I^{1,h}) = (R^{n,h} \cap I^{n,h}) \times (R^{1,h} \cap I^{1,h})$$

where $I^{n,h}$ is a half-open rectangle in $\mathbb{R}^n$ and the righthand side of the equality is an element of $\mathcal{I}^{n+1,h}$. Also, $(R^{n,h} \times R^{1,h}) - (I^{n,h} \times I^{1,h}) = (R^{n,h} \times R^{1,h}) \cap (I^{n,h} \times I^{1,h})^c$ and note that

$$\begin{aligned}(I^{n,h} \times I^{1,h})^c &= \{(x, y) : x \notin I^{n,h}, y \notin I^{1,h}, \text{ or } x \in I^{n,h} \text{ and } y \notin I^{1,h}, \text{ or } x \notin I^{n,h} \text{ and } y \in I^{1,h}\} \\ &= ((I^{n,h})^c \times (I^{1,h})^c) \cup (I^{n,h} \times (I^{1,h})^c) \cup ((I^{n,h})^c \times I^{1,h})\end{aligned}$$

where the components of the union are disjoint. Thus,

$$\begin{aligned}(R^{n,h} \times R^{1,h}) - (I^{n,h} \times I^{1,h}) &= [(R^{n,h} \times R^{1,h}) \cap ((I^{n,h})^c \times (I^{1,h})^c)] \cup [(R^{n,h} \times R^{1,h}) \cap (I^{n,h} \times (I^{1,h})^c)] \\ &\cup [(R^{n,h} \times R^{1,h}) \cap ((I^{n,h})^c \times I^{1,h})] \\ &= [(R^{n,h} - I^{n,h}) \times (R^{1,h} - I^{1,h})] \cup [(R^{n,h} \cap I^{n,h}) \times (R^{1,h} - I^{1,h})] \\ &\cup [(R^{n,h} - I^{n,h}) \times (R^{1,h} \cap I^{1,h})].\end{aligned}$$

By the induction assumption, $R^{n,h} - I^{n,h}$ and $R^{1,h} - I^{1,h}$ can be expressed as finite unions of disjoint rectangles, which completes the proof. ■

Definition 2.3. *Let* $\lambda^n : \mathcal{I}^{n,h} \to [0, \infty)$ *be defined as* $\lambda^n(R^{n,h}) = \prod_{j=1}^{n}(b_j - a_j)$ *whenever* $b_j > a_j$ *for* $j = 1, \cdots, n$ *and* $\lambda^n(R^{n,h}) = 0$ *if* $b_j \leq a_j$ *for some* j.

Theorem 2.7. λ^n *is a pre-measure on* $\mathcal{I}^{n,h}$.

Proof. We start by showing that λ^1 is a pre-measure on $\mathcal{I}^{1,h}$. Let $[a,b) \in \mathcal{I}^{1,h}$ and $[a,b) = \cup_{i=1}^{n}[a_i,b_i)$ with $a_1 = a$, $a_2 = b_1$, $a_3 = b_2, \cdots, a_n = b_{n-1}, b_n = b$. Then,

$$\begin{aligned}\sum_{i=1}^{n} \lambda^1([a_i,b_i)) &= (b_1 - a_1) + (b_2 - a_2) + \cdots + (b_{n-1} - a_{n-1}) + (b_n - a_n) \\ &= (a_2 - a) + (a_3 - a_2) + \cdots + (a_n - a_{n-1}) + (b - a_n) = b - a \\ &= \lambda^1([a,b)) = \lambda^1\left(\cup_{i=1}^{n}[a_i,b_i)\right).\end{aligned}$$

Therefore, λ^1 is finitely additive. For σ-additivity, we need to show that for $[a,b) = \bigcup_{i\in\mathbb{N}} [a_i,b_i)$, where $\{[a_i,b_i)\}_{i\in\mathbb{N}}$ is a pairwise disjoint collection we have $b - a = \sum_{i=1}^{\infty}(b_i - a_i)$.

For any $n \in \mathbb{N}$, let $\{[a_i,b_i)\}_{i=1}^{n}$ be a pairwise disjoint collection. Then, since $\mathcal{I}^{1,h}$ is a semi-ring, we can write

$$[a,b) - \cup_{i=1}^{n}[a_i,b_i) = \cup_{j=1}^{m} I_j,$$

where the last set is the finite union of pairwise disjoint half-open rectangles. Thus, since λ^1 is finitely additive on $\mathcal{I}^{1,h}$

$$\lambda^1([a,b)) = \sum_{i=1}^{n} \lambda^1([a_i,b_i)) + \sum_{j=1}^{m} \lambda^1(I_j) \geq \sum_{i=1}^{n} \lambda^1([a_i,b_i)).$$

Thus, $\lambda^1([a,b)) = b - a \geq \lim_{n\to\infty} \sum_{i=1}^{n} \lambda^1([a_i,b_i)) = \sum_{i=1}^{\infty} \lambda^1([a_i,b_i))$.

We need only show that $b - a \leq \sum_{i=1}^{\infty} \lambda^1([a_i,b_i))$ to complete the proof. Let $0 < \epsilon < b - a$ and note that

$$\begin{aligned}[a, b-\epsilon) \subset [a, b-\epsilon] &\subset \cup_{i=1}^{\infty}(a_i - 2^{-i}\epsilon, b_i) \\ &\subset \cup_{i=1}^{n}(a_i - 2^{-i}\epsilon, b_i) \text{ for some } n \in \mathbb{N}, \text{ by the Heine-Borel Theorem} \\ &\subset \cup_{i=1}^{n}[a_i - 2^{-i}\epsilon, b_i).\end{aligned}$$

But $\lambda^1([a_i, b_i)) = \lambda^1([a_i - 2^{-i}\epsilon, b_i)) - \frac{1}{2^i}\epsilon$. Hence,

$$\begin{aligned}
\lambda^1([a+\epsilon, b)) &\leq \sum_{i=1}^{n} \lambda^1\left([a_i - \frac{1}{2^i}\epsilon, b_i)\right) \text{ by subadditivity} \\
&= \sum_{i=1}^{n} (b_i - a_i + \frac{1}{2^i}\epsilon) \\
b - a - \epsilon &\leq \sum_{i=1}^{n} (b_i - a_i) + \epsilon \sum_{i=1}^{n} \frac{1}{2^i} \text{ or} \\
b - a &\leq \sum_{i=1}^{n} (b_i - a_i) + \epsilon \left(1 + \sum_{i=1}^{n} \frac{1}{2^i}\right).
\end{aligned}$$

Taking limits as $n \to \infty$ on both sides of the last inequality gives $b - a \leq \sum_{i=1}^{\infty}(b_i - a_i)$, which combined with the previously obtained reverse inequality gives $b - a = \sum_{i=1}^{\infty}(b_i - a_i)$. Hence, λ^1 is a pre-measure on $\mathcal{I}^{1,h}$.

Clearly, $\lambda^n(\emptyset) = 0$. The proof is completed by using induction on n, the dimension of the space. Hence, we assume that λ^n is σ-additive on $\mathcal{I}^{n,h}$ for some n and show that λ^{n+1} is σ-additive on $\mathcal{I}^{n+1,h}$. This final step is left as an exercise. ■

Theorem 2.8. *There exists a unique extension of λ^n from $\mathcal{I}^{n,h}$ to a measure on the Borel sets $\mathcal{B}(\mathbb{R}^n)$. This extension is denoted by λ^n and is called Lebesgue measure.*

Proof. We know that $\mathcal{B}(\mathbb{R}^n) = \sigma(\mathcal{I}^{n,h})$ from Theorem 1.5. Since, $[-k,k)^n = [-k,k) \times [-k,k) \cdots \times [-k,k) \uparrow \mathbb{R}^n$ as $k \to \infty$ is an exhausting sequence of n-rectangles, and since $\lambda^n([-k,k)^n) = (2k)^n < \infty$, all conditions of Carathéodory's Theorem are fulfilled. ■

Remark 2.3. *Let $(\mathbb{R}, \sigma(\mathcal{I}^{1,h}) = \mathcal{B}(\mathbb{R}))$ be a measurable space. From Theorem 1.4 if we set $S = [0,1)$ and consider $\mathcal{I} = \mathcal{I}^{1,h} \cap S = \left\{[0,1) \cap A : A \in \mathcal{I}^{1,h}\right\}$ then $\sigma(\mathcal{I}^{1,h} \cap [0,1)) = \mathcal{B}(\mathbb{R}) \cap [0,1)$ is a σ-algebra associated with $[0,1)$. Thus, we define $\mathcal{B}_{[0,1)} := \sigma(\mathcal{I}^{1,h} \cap [0,1))$ and note that*

$$([0,1), \mathcal{B}_{[0,1)} := \sigma(\mathcal{I}))$$

is a measurable space where $\mathcal{I} = \{[a,b) : 0 \leq a \leq b \leq 1\}$. Define the set function $\lambda : \mathcal{I} \to [0,1]$ such that $\lambda(\emptyset) = 0$ and $\lambda([a,b)) = b - a$. Since λ is σ-additive (pre-measure) on $\mathcal{I}$ (a semi-ring), using Carathéodory's Theorem, we can state that

$$([0,1), \mathcal{B}_{[0,1)} := \sigma(\mathcal{I}), \lambda^*)$$

is a measure space, where λ^* is the unique extension of λ from $\mathcal{I}$ to $\sigma(\mathcal{I})$. In addition, $\lambda^*([0,1)) = 1$. Thus, we have constructed a specific probability space.

We will now construct probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. This will be done using distribution functions.

Definition 2.4. *Let $F : \mathbb{R} \to [0,1]$ be a function with the following properties:*

1. $\lim_{h \downarrow 0} F(x+h) := F(x+) = F(x)$ *for all* $x \in \mathbb{R}$ *and* $h > 0$,
2. $F(x) \leq F(y)$ *if* $x < y$,
3. $\lim_{x \to \infty} F(x) = 1$, $\lim_{x \to -\infty} F(x) = 0$.

F is called distribution function (df). If only conditions 1 and 2 are met, F is called a defective df.

Remark 2.4. *1. Let $F(x-) := \lim_{h \downarrow 0} F(x-h)$ for $h > 0$. The left jump of F at x is defined as $LJ_F(x) = F(x) - F(x-)$ and the right jump of F at x is defined as $RJ_F(x) = F(x+) - F(x)$. The jump of F at x is defined as $J_F(x) = LJ_F(x) + RJ_F(x) = F(x+) - F(x-)$. If F is a df, $RJ_F(x) = 0$ for all $x \in \mathbb{R}$ and $J_F(x) = F(x) - F(x-)$. In addition, since F is nondecreasing $J_F(x) \geq 0$. If $J_F(x) = 0$ then F is continuous at x.*
2. For any two $x \leq y \in \mathbb{R}$ we have that $0 \leq F(y) - F(x) \leq 1$

Definition 2.5. *The left (generalized) inverse of a df F, denoted by F^-, is defined as*

$$F^-(p) := \inf\{x : F(x) \geq p \text{ for } p \in (0,1]\}.$$

Note that $\{x : F(x) \geq 0\} = \mathbb{R}$ and the infimum of $\mathbb{R}$ does not exist. Hence, F^- is not defined at 0. Also, $\{x : F(x) \geq 1\} = \{x : F(x) = 1\}$ is either the empty set or $[a, \infty)$, where $a \in \mathbb{R}$. In the first case, $F^-(1) = \inf(\emptyset) = \infty$ and in the second case $F^-(1) = \inf([a, \infty)) = a$.

Theorem 2.9. *Let $S(p) = \{x : F(x) \geq p\}$ for $p \in (0, 1]$. Then,*

1. *$S(p)$ is a closed set.*

2. *$t < F^-(p) \iff F(t) < p$ or $F^-(p) \leq t \iff p \leq F(t)$.*

Proof. 1. If $s_n \in S(p)$ and $s_n \downarrow s$, bp right continuity of F we have $p \leq F(s_n) \downarrow F(s)$. Thus, $p \leq F(s)$ and $s \in S(p)$. If $s_n \in S(p)$ and $s_n \uparrow s$, we have $p \leq F(s_n) \uparrow F(s-) \leq F(s)$. Thus, $p \leq F(s)$ which implies that $s \in S(p)$. Consequently, by a characterization of closed sets, $S(p)$ is closed.

2. Since $S(p)$ is closed, its infimum $F^-(p) \in S(p)$ and therefore $F(F^-(p)) \geq p$. $t < F^-(p) \implies t \notin S(p) \implies F(t) < p$. The reverse implications all apply. ■

Theorem 2.10. *Let $A \subset \mathbb{R}$, $\mathcal{S}_F(A) = \{p \in (0, 1] : F^-(p) \in A\}$ and $\mathcal{I}^1 = \{(a, b] : -\infty \leq a < b < \infty\}$. If $A \in \mathcal{B}(\mathbb{R})$, then $\mathcal{S}_F(A) \in \mathcal{B}_{(0,1]} = \sigma(\mathcal{I}^1) \cap (0, 1]$.*

Proof. Let $\mathcal{G} = \{A \subset \mathbb{R} : \mathcal{S}_F(A) \in \mathcal{B}_{(0,1]}\}$. Note that

$$\begin{aligned}\mathcal{S}_F((a, b]) &= \{p \in (0, 1] : F^-(p) \in (a, b]\} = \{p \in (0, 1] : a < F^-(p) \leq b\} \\ &= \{p \in (0, 1] : F(a) < p \leq F(b)\} \text{ by Theorem 2.9} \\ &= (F(a), F(b)] \in \mathcal{B}_{(0,1]}.\end{aligned}$$

Hence, $(a, b] \in \mathcal{G}$ and $\mathcal{I}^1 \subset \mathcal{G}$. If $\mathcal{G}$ is a σ-algebra, $\sigma(\mathcal{I}^1) = \mathcal{B}(\mathbb{R}) \subset \mathcal{G}$. Hence, $A \in \mathcal{B}(\mathbb{R})$ implies $S_F(A) \in \mathcal{B}_{(0,1]}$. Consequently, we need only show that $\mathcal{G}$ is a σ-algebra associated with $\mathbb{R}$.

1. $\mathcal{S}_F(\mathbb{R}) = \{p \in (0, 1] : F^-(p) \in \mathbb{R}\} = (0, 1) = \bigcup_{n \in \mathbb{N}} (0, 1 - n^{-1}] \in \mathcal{B}_{(0,1]}$, thus $\mathbb{R} \in \mathcal{G}$.

2. By definition of $\mathcal{S}_F$

$$\begin{aligned}\mathcal{S}_F(A^c) &= \{p \in (0,1] : F^-(p) \in A^c\} = \{p \in (0,1] : F^-(p) \notin A\} \\ &= (\mathcal{S}_F(A))^c \in \mathcal{B}_{(0,1]}\end{aligned}$$

where the last inclusion statement follows if $A \in \mathcal{G}$ and the fact that $\mathcal{B}_{(0,1]}$ is a σ-algebra.

3. If $\{A_n\}_{n\in\mathbb{N}} \in \mathcal{G}$ we have by definition of $\mathcal{S}_F$

$$\begin{aligned}\mathcal{S}_F\left(\bigcup_{n\in\mathbb{N}} A_n\right) &= \left\{p \in (0,1] : F^-(p) \in \bigcup_{n\in\mathbb{N}} A_n\right\} = \{p \in (0,1] : F^-(p) \in A_n \text{ for some } n\} \\ &= \bigcup_{n\in\mathbb{N}} \{p \in (0,1] : F^-(p) \in A_n\} = \bigcup_{n\in\mathbb{N}} \mathcal{S}_F(A_n) \in \mathcal{B}_{(0,1]} \qquad (2.8)\end{aligned}$$

where the last inclusion statement follows since $A_n \in \mathcal{G}$ and the fact that $\mathcal{B}_{(0,1]}$ is a σ-algebra.

■

Definition 2.6. *Let $A \in \mathcal{B}(\mathbb{R})$ and define $P_F(A) = \lambda^1(\mathcal{S}_F(A))$ where λ^1 is the Lebesgue measure on $\mathcal{B}_{(0,1]}$.*

Theorem 2.11. *Let P_F be given in Definition 2.6. Then, $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_F)$ is a probability space.*

Proof. First, note that

$$P_F(\emptyset) = \lambda^1(\mathcal{S}_F(\emptyset)) = \lambda^1(\{p \in (0,1] : F^-(p) \in \emptyset\}) = \lambda^1(\emptyset) = 0.$$

Second, if $\{A_n\}_{n\in\mathbb{N}}$ is a pairwise disjoint collection of sets in $\mathcal{B}(\mathbb{R})$ then

$$\begin{aligned}P_F\left(\cup_{n\in\mathbb{N}} A_n\right) &= \lambda^1\left(\mathcal{S}_F\left(\cup_{n\in\mathbb{N}} A_n\right)\right) = \lambda^1\left(\cup_{n\in\mathbb{N}} \mathcal{S}_F(A_n)\right) \text{ by (2.8)} \\ &= \sum_{n=1}^{\infty} \lambda^1(\mathcal{S}_F(A_n)) = \sum_{n=1}^{\infty} P_F(A_n).\end{aligned}$$

where the next to last equality follows from the fact that λ^1 is a measure and $\{\mathcal{S}_F(A_n)\}_{n\in\mathbb{N}}$ is a pairwise disjoint collection.

Lastly,

$$\begin{aligned}
P_F(\mathbb{R}) &= \lambda^1(\mathcal{S}_F(\mathbb{R})) = \lambda^1(\{p \in (0,1] : F^-(p) \in \mathbb{R}\}) = \lambda^1((0,1)) \\
&= \lambda^1\left(\bigcup_{n\in\mathbb{N}}(0, 1-n^{-1}]\right) = \lambda^1\left((0,1/2] \cup (1/2, 2/3] \cup (2/3, 3/4] \cup \cdots\right) \\
&= 1/2 + (2/3 - 1/2) + (3/4 - 2/3) + \cdots = 1.
\end{aligned}$$

■

Remark 2.5. *Note that*

$$\begin{aligned}
P_F\left((-\infty, x]\right) &= \lambda^1\left(\mathcal{S}_F((-\infty, x])\right) = \lambda^1(\{p \in (0,1] : F^-(p) \in (-\infty, x]\}) \\
&= \lambda^1(\{p \in (0,1] : p \leq F(x)\}) = \lambda^1((0, F(x)]) = F(x).
\end{aligned}$$

Chapter 3

Measurable functions

In this chapter we will define measurable functions and study some of their properties. We start with the following definition.

Definition 3.1. *Let* $(\mathbb{X}, \mathcal{F})$ *and* $(\mathbb{E}, \mathcal{E})$ *be two measurable spaces. A function* $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{E}, \mathcal{E})$ *is said to be* $\mathcal{F} - \mathcal{E}$ *measurable if for all* $A \in \mathcal{E}$, $f^{-1}(A) \in \mathcal{F}$.

Remark 3.1.

1. *Since* $f^{-1}(\mathcal{E})$ *is a* σ*-algebra, measurability of* f *is equivalent to stating that* $f^{-1}(\mathcal{E}) \subset \mathcal{F}$. *It is standard notation to write* $\sigma(f) := f^{-1}(\mathcal{E})$ *and call this* σ*-algebra the* σ*-algebra generated by* f.

2. *If* $\mathbb{X} := \Omega$, $(\Omega, \mathcal{F}, P)$ *is a probability space and* f *is* $\mathcal{F} - \mathcal{E}$ *measurable, we say that* f *is a random element. If, in addition,* $(\mathbb{E}, \mathcal{E}) := (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *we will refer to* $f : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *as a random variable. We will normally represent random elements or random variables by uppercase roman letters, e.g.,* X *or* Y.

The next theorem shows that measurability of a function f can be established by examining inverse images of sets in a collection that generates the measurable sets associated with the co-domain of f.

Theorem 3.1. *Let* $\mathcal{C}$ *be a collection of subsets of* $\mathbb{E}$ *such that* $\sigma(\mathcal{C}) = \mathcal{E}$. *Then,* $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{E}, \mathcal{E})$ *is* $\mathcal{F} - \mathcal{E}$ *measurable* $\iff f^{-1}(\mathcal{C}) \subset \mathcal{F}$.

Proof. ($\Longrightarrow$) Assume f is $\mathcal{F}-\mathcal{E}$ measurable. f measurable $\iff$ for all $A \in \mathcal{E}$, $f^{-1}(A) \in \mathcal{F}$. In particular, let A be an element of $\mathcal{C}$, then $f^{-1}(A) \in \mathcal{F}$, hence $f^{-1}(\mathcal{C}) \subset \mathcal{F}$.

($\Longleftarrow$) Assume that $f^{-1}(\mathcal{C}) \subset \mathcal{F}$, i.e., $f^{-1}(C) \in \mathcal{F}$, for all $C \in \mathcal{C}$. We must prove that $\forall A \in \mathcal{E}$, $f^{-1}(A) \in \mathcal{F}$ (or $f^{-1}(\mathcal{E}) \subset \mathcal{F}$). Let $\mathcal{G} = \{A \in \mathcal{E} : f^{-1}(A) \in \mathcal{F}\}$ and by construction $\mathcal{C} \subset \mathcal{G}$. If $\mathcal{G}$ is a σ-algebra, then $\sigma(\mathcal{C}) = \mathcal{E} \subset \mathcal{G}$. Also, by construction $\mathcal{G} \subset \mathcal{E}$, hence $\mathcal{E} = \mathcal{G}$, which is what must be proven.

We need only show that $\mathcal{G}$ is a σ-algebra. Consider a sequence $A_1, A_2, \dots \in \mathcal{E}$ such that $f^{-1}(A_i) \in \mathcal{F}$, i.e., $A_1, A_2 \dots \in \mathcal{G}$. Then, since $\mathcal{E}$ is a σ-algebra, $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{E}$. And since $f^{-1}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \bigcup_{i\in\mathbb{N}} f^{-1}(A_i)$, which is the union of elements in $\mathcal{F}$, $f^{-1}\left(\bigcup_{i\in\mathbb{N}} A_i\right) \in \mathcal{F}$.

Now, if $A \in \mathcal{E}$ is such that $f^{-1}(A) \in \mathcal{F}$, i.e., $A \in \mathcal{G}$, then $A^c \in \mathcal{E}$ and $f^{-1}(A^c) = f^{-1}(\mathbb{E}) - f^{-1}(A) = \mathbb{X} - f^{-1}(A)$ which is in $\mathcal{F}$. Hence $\mathcal{G}$ is a σ-algebra. ■

Example 3.1. *Let $\mathcal{A}_4 = \{(-\infty, a] : a \in \mathbb{R}\}$ be the collection $\mathcal{A}_4$ in Remark 1.2.4. Since $\sigma(\mathcal{A}_4) = \mathcal{B}(\mathbb{R})$,*

$$X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \sigma(\mathcal{A}_4) = \mathcal{B}(\mathbb{R}))$$

is a random variable if, and only if, $X^{-1}(\mathcal{A}_4) \subset \mathcal{F}$. Equivalently we can state X is a random variable if, and only if, $X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F} \ \forall a \in \mathbb{R}$.

The next theorem shows that continuous functions are measurable.

Theorem 3.2. *Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be collections of open sets associated with $\mathbb{X}_1$ and $\mathbb{X}_2$. If $f : (\mathbb{X}_1, \sigma(\mathcal{O}_1)) \to (\mathbb{X}_2, \sigma(\mathcal{O}_2))$ is continuous, then f is measurable.*

Proof. $f^{-1}(\mathcal{O}_2) \subset \mathcal{O}_1$ by continuity. But $\mathcal{O}_1 \subset \sigma(\mathcal{O}_1)$. Thus, by Theorem 3.1, f is measurable. ■

The composition of measurable functions is measurable.

Theorem 3.3. *Let $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{X}_1, \mathcal{F}_1)$ and $g : (\mathbb{X}_1, \mathcal{F}_1) \to (\mathbb{X}_2, \mathcal{F}_2)$ be measurable functions. Let $(g \circ f) : (\mathbb{X}, \mathcal{F}) \to (\mathbb{X}_2, \mathcal{F}_2)$. Then, $(g \circ f)$ is $\mathcal{F} - \mathcal{F}_2$ measurable.*

Proof. Let $F_2 \in 2^{\mathbb{X}_2}$.

$$(g \circ f)^{-1}(F_2) = \{x \in \mathbb{X} : g(f(x)) \in F_2\} = \{x \in \mathbb{X} : f(x) \in g^{-1}(F_2)\}$$
$$= \{x \in \mathbb{X} : x \in f^{-1}(g^{-1}(F_2))\}.$$

If $F_2 \in \mathcal{F}_2$, and given that g is measurable, $g^{-1}(F_2) \in \mathcal{F}_1$. Since f is measurable, $f^{-1}(g^{-1}(F_2)) \in \mathcal{F}$. Hence, $(g \circ f)$ is $\mathcal{F} - \mathcal{F}_2$ measurable. ■

The next theorem shows that measurable functions can be used to transfer measures between spaces.

Theorem 3.4. *Let* $(\mathbb{X}, \mathcal{F}, \mu)$ *be a measure space,* $(\mathbb{E}, \mathcal{E})$ *be a measurable space and* $f : \mathbb{X} \to \mathbb{E}$ *be a* $\mathcal{F} - \mathcal{E}$ *measurable function. Then,*

$$m(E) := \mu(f^{-1}(E)) \text{ for all } E \in \mathcal{E}$$

is a measure on $(\mathbb{E}, \mathcal{E})$*.*

Proof. We verify the two defining properties of measures. First, note that if $E = \emptyset$, $m(\emptyset) = \mu(f^{-1}(\emptyset)) = \mu(\emptyset) = 0$ since μ is a measure. Second, if $\{E_n\}_{n \in \mathbb{N}}$ is a pairwise disjoint collection of sets in $\mathcal{E}$ then

$$m\left(\bigcup_{n \in \mathbb{N}} E_n\right) = \mu\left(f^{-1}\left(\bigcup_{n \in \mathbb{N}} E_n\right)\right) = \mu\left(\bigcup_{n \in \mathbb{N}} f^{-1}(E_n)\right) = \sum_{n \in \mathbb{N}} \mu(f^{-1}(E_n)) = \sum_{n \in \mathbb{N}} m(E_n),$$

where the next to last equality follows from the fact that μ is a measure and the last equality follows from the definition of m. ■

Example 3.2. *Let* $(\mathbb{X}, \mathcal{F}, \mu) := (\Omega, \mathcal{F}, P)$*,* $(\mathbb{E}, \mathcal{E}) := (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*, and* $f := X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *then* $P_X(B) := P(X^{-1}(B))$ *is a measure on* $\mathcal{B}(\mathbb{R})$*.*

The measurability of real valued functions can be characterized differently. In Example 3.1 it is shown that a function $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is said to be $\mathcal{F} - \mathcal{B}(\mathbb{R})$ measurable if for all $a \in \mathbb{R}$, the set $S_a = \{x \in \mathbb{X} : f(x) \leq a\} \in \mathcal{F}$. But since $S_a \in \mathcal{F}$ and $\mathcal{F}$ is a σ-algebra, $S_a^c \in \mathcal{F}$. Hence, f is measurable if $S_a^c = \{x \in X : f(x) > a\} \in \mathcal{F}$. Also, consider $S_{a-1/n}^c = \{x \in \mathbb{X} : f(x) > a - 1/n\}$ and let $S_a' = \bigcap_{n \in \mathbb{N}} \{x \in \mathbb{X} : f(x) > a - 1/n\} = \{x \in \mathbb{X} : f(x) \geq a\}$. Clearly, by the properties of σ-algebras $S_a' \in \mathcal{F}$. Hence, f is measurable if $\{x \in X : f(x) \geq a\} \in \mathcal{F}$. Since, $\{x \in \mathbb{X} : f(x) < a\} = \{x \in \mathbb{X} : f(x) \geq a\}^c$, measurability could also be defined in terms of $\{x \in \mathbb{X} : f(x) < a\}$.

Example 3.3. 1. *Let $f : \mathbb{X} \to \mathbb{R}$, such that for all $x \in \mathbb{X}$, $f(x) = c$, $c \in \mathbb{R}$. Let $a \in \mathbb{R}$ and consider $S_a^c = \{x \in \mathbb{X} : f(x) > a\} = \{x \in \mathbb{X} : c > a\}$. If $a \geq c$, $S_a^c = \emptyset$, and if $c > a$, $S_a^c = \mathbb{X}$. Since σ-algebras always contain $\emptyset$ and $\mathbb{X}$, $f(x) = c$ is measurable.*

2. *Let $E \in \mathcal{F}$ ($\mathcal{F}$ a σ-algebra). Recall that the indicator function of E is*

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

If $a \geq 1$, $S_a^c = \emptyset$; if $0 \leq a < 1$, $S_a^c = E$; if $a < 0$ $S_a^c = \mathbb{X}$. Since $\mathbb{X}, \emptyset \in \mathcal{F}$ (always) and $E \in \mathcal{F}$ by construction, I_E is measurable.

3. *Let $\mathbb{X} = \mathbb{R}$ and $\mathcal{F} = \mathcal{B}(\mathbb{R})$. If f is monotone increasing, i.e., $\forall\, x < x'$, $f(x) \leq f(x')$, f is measurable. Note that in this case, $S_a^c = \{x : x > y \text{ for some } y \in \mathbb{R}\} = (y, \infty)$ or $S_a^c = \{x : x \geq y\} = [y, \infty)$, which are Borel sets.*

Theorem 3.5. *Let f and g be measurable real valued functions and let $c \in \mathbb{R}$. Then, $cf, f^2, f+g, fg, |f|$ are measurable.*

Proof. If $c = 0$, $cf = 0$ is a constant and consequently, measurable. If $c > 0$, then $\{x \in \mathbb{X} : cf(x) > a\} = \{x \in \mathbb{X} : f(x) > a/c\} \in \mathcal{F}$. Similarly for $c < 0$. If $a < 0$, $\{x \in \mathbb{X} : (f(x))^2 > a\} = \mathbb{X}$ and $\mathbb{X} \in \mathcal{F}$. If $a \geq 0$, $\{x \in \mathbb{X} : f^2(x) > a\} = \{x \in \mathbb{X} : f(x) > a^{1/2} \text{ or } f(x) < -a^{1/2}\} = \{x \in \mathbb{X} : f(x) > a^{1/2}\} \cup \{x \in \mathbb{X} : f(x) < -a^{1/2}\}$. The first set in the union is in $\mathcal{F}$ by assumption (f is measurable) and the second is in $\mathcal{F}$ by the comments following Example 3.2.

Now, $g(x) + f(x) > a \implies f(x) > a - g(x)$ which implies that there exists a rational number r such that $f(x) > r > a - g(x)$. Hence, $\{x \in \mathbb{X} : g(x) + f(x) > a\} = \underset{r\in\mathbb{Q}}{\cup}\{x \in \mathbb{X} : f(x) > r\} \cap \{x \in \mathbb{X} : g(x) > a - r\}$. Since the rational numbers are countable $\underset{r\in\mathbb{Q}}{\cup}$ is countable. Since f and g are measurable, and unions of countable measurable sets are measurable $\{x \in \mathbb{X} : g(x) + f(x) > a\} \in \mathcal{F}$. Note that $-f = (-1)f$. Hence if f is measurable, $-f$ is also measurable and so is $f + (-g) = f - g$.

Now, $fg = 1/2[(f+g)^2 - (f^2 + g^2)]$. Since $f^2, g^2, f+g, f-g$ and cf are measurable, if f, g are measurable, so is fg.

Lastly, $\{x \in \mathbb{X} : |f(x)| > a\} = \{x \in \mathbb{X} : f(x) > a \text{ or } f(x) < -a\} = \{x \in \mathbb{X} : f(x) > a\} \cup \{x \in \mathbb{X} : f(x) < -a\} = \{x \in \mathbb{X} : f(x) > a\} \cup \{x \in \mathbb{X} : -f(x) > a\}$. Since f and $-f$ are measurable, $\{x \in \mathbb{X} : |f(x)| > a\} \in \mathcal{F}$. ■

Recall that if $\{x_n\}_{n\in\mathbb{N}}$ is a sequence of real numbers

$$\liminf_{n\to\infty} x_n := \sup_{k\in\mathbb{N}} \inf_{j\geq k}\{x_j\} \text{ and } \limsup_{n\to\infty} x_n := \inf_{k\in\mathbb{N}} \sup_{j\geq k}\{x_j\}.$$

Theorem 3.6. *Let $f_i(x) : \mathbb{X} \to \mathbb{R}$ for $i = 1, 2, \cdots$ be measurable. Then $\sup\{f_1, \cdots, f_n\}$, $\inf\{f_1, \cdots, f_n\}$, $\sup_n f_n$, $\inf_n f_n$, $\limsup_n f_n$ and $\liminf_n f_n$ are all measurable functions.*

Proof. Let $h(x) = \sup\{f_1(x), \cdots, f_n(x)\}$. Then, $S_a = \{x \in \mathbb{X} : h(x) > a\} = \cup_{i=1}^n \{x : f_i(x) > a\}$. Consequently, since f_i is measurable, $S_a \in \mathcal{F}$. Similarly if $g(x) = \sup_{n\in\mathbb{N}} f_n(x)$, $S_a = \{x \in \mathbb{X} : g(x) > a\} = \bigcup_{n\in\mathbb{N}} \{x : f_n(x) > a\} \in \mathcal{F}$. The same argument can be made for inf. Since $\limsup_{n\to\infty} f_n = \inf_{n\geq 1} \sup_{k\geq n} f_k$, $\limsup f_n$ is measurable. The same for $\liminf_{n\to\infty} f_n$. ■

Definition 3.2. *Let $i \in I$ an arbitrary index set and $f_i : (\mathbb{X}, \mathcal{F}) \to (\mathbb{X}_i, \mathcal{F}_i)$ be $\mathcal{F} - \mathcal{F}_i$ measurable. If $\mathcal{G} \subset \mathcal{F}$ is a σ-algebra, we say that f_i is measurable with respect to $\mathcal{G}$ if $\sigma(f_i) \subset \mathcal{G}$. The smallest σ-algebra $\mathcal{G}$ that makes all f_i measurable with respect to $\mathcal{G}$ is $\sigma\left(\bigcup_{i\in I} f_i^{-1}(\mathcal{F}_i)\right)$ and is denoted by $\sigma(f_i : i \in I)$.*

Chapter 4

Integration

4.1 Simple functions

Often, it is necessary to use the symbols $-\infty$ or ∞ in calculations. In these cases we work with the extended real line, i.e., $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\} = [-\infty, \infty]$. Functions that take values in $\bar{\mathbb{R}}$ are called *numerical* functions. The Borel sets associated with the extended real line are denoted by $\bar{\mathcal{B}} := \mathcal{B}(\bar{\mathbb{R}})$ and are defined as the collection of sets $\bar{B}$ such that $\bar{B} = B \cup S$ where $B \in \mathcal{B}(\mathbb{R})$ and $S \in \{\emptyset, \{-\infty\}, \{\infty\}, \{-\infty, \infty\}\}$. It can be verified that $\bar{\mathcal{B}}$ is a σ-algebra and that $\mathcal{B}(\mathbb{R}) = \mathbb{R} \cap \mathcal{B}(\bar{\mathbb{R}}) := \{\mathbb{R} \cup B : B \in \mathcal{B}(\mathbb{R})\}$. In addition, $\bar{\mathcal{B}}$ is generated by a collection of sets of the form $[a, \infty]$ (or $(a, \infty]$, $[-\infty, a]$, $[-\infty, a)$) where $a \in \mathbb{R}$.

Theorem 4.1. *$\bar{\mathcal{B}} = \sigma(\mathcal{C})$, where $\mathcal{C} := \{[a, \infty] : a \in \mathbb{R}\}$.*

Proof. Let $\mathcal{C} := \{[a, \infty] : a \in \mathbb{R}\}$ and $\mathcal{G} := \sigma(\mathcal{C})$. Note that since $[a, \infty] = [a, \infty) \cup \{\infty\}$, $[a, \infty] \in \bar{\mathcal{B}}$ and $\mathcal{C} \subset \bar{\mathcal{B}}$. Then, since $\bar{\mathcal{B}}$ is a σ-algebra $\sigma(\mathcal{C}) := \mathcal{G} \subset \bar{\mathcal{B}}$. Now, let $\mathcal{C}_1 = \{[a, b) : -\infty < a \leq b < \infty\}$ and note that $[a, b) = [a, \infty] - [b, \infty] \in \mathcal{G}$. Hence, $\mathcal{C}_1 \subset \mathcal{G}$ and $\sigma(\mathcal{C}_1) = \mathcal{B}(\mathbb{R}) \subset \mathcal{G}$ since $\mathcal{G}$ is a σ-algebra.

Note that $\{\infty\} = \bigcap_{n \in \mathbb{N}} [n, \infty]$, $\{-\infty\} = \bigcap_{n \in \mathbb{N}} [-\infty, -n) = \bigcap_{n \in \mathbb{N}} [-n, \infty]^c$ and, consequently, $\{\infty\}, \{-\infty\} \in \mathcal{G}$. Then, for all $B \in \mathcal{B}(\mathbb{R})$ and $S \in \{\emptyset, \{-\infty\}, \{\infty\}, \{-\infty, \infty\}\}$ we have $B \cup S \in \mathcal{G}$, showing that $\bar{\mathcal{B}} \subset \mathcal{G}$. ■

Let $(\mathbb{X}, \mathcal{F})$ and $(\mathbb{R}, \mathcal{B})$ be measurable spaces. Since the indicator function of a measurable set is a measurable function, it follows from Theorem 3.5 that if $\{A_j\}_{j=1}^n$ with $n \in \mathbb{N}$ is a pairwise disjoint collection in $\mathcal{F}$ and $a_j \in \mathbb{R}$ for $j = 1, \cdots, n$, the linear combination

$$f(x) = \sum_{i=1}^{n} a_j I_{A_j}(x) \tag{4.1}$$

is a $\mathcal{F} - \mathcal{B}$-measurable function.

Definition 4.1. *A real-valued function on a measurable space $(\mathbb{X}, \mathcal{F})$ is said to be simple if it has the representation (4.1). A standard representation of a simple function is given by*

$$f(x) = \sum_{j=0}^{n} a_j I_{A_j}(x) \text{ with } a_0 = 0 \text{ and } A_0 = (\cup_{j=1}^n A_j)^c. \tag{4.2}$$

Remark 4.1.

1. *If $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ is measurable and takes on finitely many values, say $\{a_j\}_{j=1}^n$ then it is a simple function. To see this, note that $B_j = \{x : f(x) = a_j\}$ is measurable, since $B_j = \{x : f(x) \leq a_j\} - \{x : f(x) < a_j\}$ and f is measurable. Also, note that the collection $\{B_j\}_{j=1}^n$ is pairwise disjoint. Hence,*

$$f(x) = \sum_{j=1}^{n} a_j I_{B_j}(x) = \sum_{j=1}^{n} a_j I_{\{x: f(x) = a_j\}}(x). \tag{4.3}$$

   Conversely, if f is simple it takes on finitely many values.

2. *Representation (4.2) is not unique, but a simple function has at least one representation such as (4.2) .*

The next theorem shows that certain functions of simple functions are simple functions.

Theorem 4.2. *Let $f : (\mathbb{X}, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ and $g : (\mathbb{X}, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ be simple functions. Then, $f \pm g$, cf for $c > 0$, fg, $f^+ = \max\{f, 0\}$, $f^- = -\min\{f, 0\}$ and $|f|$ are simple functions.*

4.2 Integral of simple functions

Definition 4.2. *Let* $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ *be a non-negative simple function with standard representation* (4.2)*. The integral of* f *with respect to* μ*, denoted by* $\int_{\mathbb{X}} f d\mu$*, is given by*

$$\int_{\mathbb{X}} f d\mu := \sum_{j=0}^{n} a_j \mu(A_j) \in [0, \infty]. \tag{4.4}$$

By definition $a_j \in \mathbb{R}$ for $j = 0, 1, \cdots, n$, but since μ takes values in $[0, \infty]$ we can have $\int_{\mathbb{X}} f d\mu = \infty$. If μ is a finite measure, e.g., a probability measure P, then it must be that $\int_{\mathbb{X}} f d\mu \in \mathbb{R}$. When $\mathbb{X} := \Omega$ an outcome space, $f := X$ is a random variable and $\mu := P$ is a probability measure we write $E_P(X) := \int_{\Omega} X dP$ and call it the expectation of X given probability P.

It will be convenient, in the case of simple functions, to write $I_\mu(f) := \int_{\mathbb{X}} f d\mu$.

Remark 4.2. *Since the representation* (4.2) *is not unique, for uniqueness, the definition of integral requires that it be invariant to the representation used. To see this, suppose that* $f(x) = \sum_{j=0}^{n} a_j I_{A_j}(x) = \sum_{k=0}^{m} b_k I_{B_k}(x)$*. Then,* $\mathbb{X} = \cup_{j=0}^{n} A_j = \cup_{k=0}^{m} B_k$ *and*

$$A_j = \cup_{k=0}^{m} (A_j \cap B_k),\ B_k = \cup_{j=0}^{n} (A_j \cap B_k).$$

Since μ *is finitely additive and the sets in the above unions are disjoint we have that*

$$\sum_{j=0}^{n} a_j \mu(A_j) = \sum_{j=0}^{n} a_j \sum_{k=0}^{m} \mu(A_j \cap B_k) = \sum_{j=0}^{n} \sum_{k=0}^{m} a_j \mu(A_j \cap B_k).$$

Similarly,

$$\sum_{k=0}^{m} b_k \mu(B_k) = \sum_{k=0}^{m} b_k \sum_{j=0}^{n} \mu(A_j \cap B_k) = \sum_{j=0}^{n} \sum_{k=0}^{m} b_k \mu(A_j \cap B_k).$$

But $a_j = b_k$ *whenever* $A_j \cap B_k \neq \emptyset$*, and when* $A_j \cap B_k = \emptyset$*,* $\mu(A_j \cap B_k) = 0$*. Thus,* $a_j \mu(A_j \cap B_k) = b_k \mu(A_j \cap B_k)$ *for all pairs* (j, k)*, and* $I_\mu(f)$ *is invariant to the representation of the simple function.*

Theorem 4.3. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ and $g : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ be simple non-negative functions. Then,*

1. $\int\limits_{\mathbb{X}} cf d\mu = c \int\limits_{\mathbb{X}} f d\mu$ *for* $c \geq 0$ *and* $\int\limits_{\mathbb{X}} I_E d\mu = \mu(E)$ *for* $E \in \mathcal{F}$.
2. $\int\limits_{\mathbb{X}} (f+g) d\mu = \int\limits_{\mathbb{X}} f d\mu + \int\limits_{\mathbb{X}} g d\mu$,
3. *If for* $E \in \mathcal{F}$, *we define* $m(E) = \int\limits_{\mathbb{X}} f I_E d\mu$, *then* m *is a measure on* $\mathcal{F}$.
4. $f \leq g \implies \int\limits_{\mathbb{X}} f d\mu \leq \int\limits_{\mathbb{X}} g d\mu$.

Proof. For 1., note that $c \geq 0 \implies cf \geq 0$ with representation $cf(x) = \sum_{j=0}^n c a_j I_{A_j}(x)$. Therefore, $\int_{\mathbb{X}} cf d\mu = \sum_{j=0}^n c a_j \mu(A_j) = c \sum_{j=0}^n a_j \mu(A_j) = c \int_{\mathbb{X}} f d\mu$. For the second part, note that $I_E(x) = I_E(x) + 0 I_{E^c}(x)$. Hence, $\int\limits_{\mathbb{X}} I_E d\mu = \mu(E)$.

For 2., let $f(x) = \sum_{j=0}^n a_j I_{A_j}(x)$ and $g(x) = \sum_{k=0}^m b_k I_{B_k}(x)$. Then, $f(x) + g(x) = \sum_{j=0}^n \sum_{k=0}^m (a_j + b_k) I_{A_j \cap B_k}(x)$ with $(A_j \cap B_k) \cap (A_{j'} \cap B_{k'}) = \emptyset$ whenever $(j,k) \neq (j',k')$. Then,

$$\begin{aligned}
\int_{\mathbb{X}} (f+g)\, d\mu &= \sum_{j=0}^n \sum_{k=0}^m (a_j + b_k) \mu(A_j \cap B_k) \\
&= \sum_{j=0}^n a_j \sum_{k=0}^m \mu(A_j \cap B_k) + \sum_{k=0}^m b_k \sum_{j=0}^n \mu(A_j \cap B_k) \\
&= \sum_{j=0}^n a_j \mu(A_j) + \sum_{k=0}^m b_k \mu(B_k),
\end{aligned}$$

since $\mathbb{X}$ is the union of both $\{A_j\}$ and $\{B_k\}$. Then, by definition $\int_{\mathbb{X}} (f+g) d\mu = \int_{\mathbb{X}} f d\mu + \int_{\mathbb{X}} g d\mu$.

For 3., note that $f(x) I_E(x) = \sum_{j=0}^n a_j I_{A_j \cap E}(x)$. From b) and a),

$$\lambda(E) = \int_{\mathbb{X}} f I_E d\mu = \sum_{j=0}^n a_j \int_{\Omega} I_{A_j \cap E}(x) d\mu = \sum_{j=0}^n a_j \mu(A_j \cap E).$$

But $\mu(A_j \cap E)$ is a measure, and we have expressed $\lambda(E)$ as a linear combination of measures on $\mathcal{F}$, hence λ is a measure on $\mathcal{F}$.

For 4., write $g = f + (g - f)$. Note that $g - f$ is simple and non-negative since $g \geq f$. Hence, $I_\mu(g) = I_\mu(f) + I_\mu(g - f) \geq I_\mu(f)$. ■

4.3 Integral of non-negative functions

We start with the following fundamental theorem.

Theorem 4.4. *Let $f : (\Omega, \mathcal{F}) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a non-negative measurable function. Then, there exists a sequence $\varphi_n : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B})$ of simple non-negative functions such that:*

1. *$\varphi_n(\omega) \leq \varphi_{n+1}(\omega)$, for all $\omega \in \Omega$ and $n \in \mathbb{N}$*

2. *$\lim_{n\to\infty} \varphi_n(\omega) = f(\omega)$, for all $\omega \in \Omega$.*

Proof. 1. For each $n = 1, 2, \cdots$ define the sets

$$E_{k,n} = \begin{cases} \{\omega \in \Omega : \frac{k}{2^n} \leq f(\omega) < \frac{k}{2^n} + \frac{1}{2^n}\} = f^{-1}([\frac{k}{2^n}, \frac{k}{2^n} + \frac{1}{2^n})) \text{ for } k = 0, 1, \cdots, n2^n - 1 \\ \{\omega \in \Omega : f(\omega) \geq n\} = f^{-1}([n, \infty]) \text{ for } k = n2^n. \end{cases}$$

For each n, the sets $\{E_{k,n} : k = 0, 1, \cdots, n2^n\}$ are disjoint by construction, belong to $\mathcal{F}$ since f is measurable and $\cup_{k=0}^{n2^n} E_{k,n} = \Omega$. Now, let

$$\varphi_n(\omega) = \sum_{k=0}^{n2^n} \frac{k}{2^n} I_{E_{k,n}}(\omega).$$

Fix $\omega \in \Omega$ and for any $n \in \mathbb{N}$ we note that $\omega \in E_{k_0,n}$ for some k_0. By definition

$$\varphi_n(\omega) = \begin{cases} \frac{k_0}{2^n} \text{ if } k_0 = 0, 1, \cdots, n2^n - 1 \\ n \text{ if } k_0 = n2^n. \end{cases}$$

First, let $k_0 \in \{0, 1, \cdots, n2^n - 1\}$ and consider $n+1$. The lower bound on $[\frac{k_0}{2^n}, \frac{k_0}{2^n} + \frac{1}{2^n})$ must coincide with $\frac{k}{2^{n+1}}$, which gives $k = 2k_0$. Thus, $E_{k,n+1} = E_{2k_0,n+1} = f^{-1}\left([\frac{2k_0}{2^{n+1}}, \frac{2k_0}{2^{n+1}} + \frac{1}{2^{n+1}})\right) = f^{-1}\left([\frac{k_0}{2^n}, \frac{k_0}{2^n} + \frac{1}{2^{n+1}})\right)$ and

$$E_{k+1,n+1} = E_{2k_0+1,n+1} = f^{-1}\left([\frac{k_0}{2^n} + \frac{1}{2^{n+1}}, \frac{k_0}{2^n} + \frac{2}{2^{n+1}})\right) = f^{-1}\left([\frac{k_0}{2^n} + \frac{1}{2^{n+1}}, \frac{k_0}{2^n} + \frac{1}{2^n})\right).$$

Consequently, $E_{k_0,n} = E_{k,n+1} \cup E_{k+1,n+1} = E_{2k_0,n+1} \cup E_{2k_0+1,n+1}$. If $\omega \in E_{2k_0,n+1} \subset E_{k_0,n}$ then $\varphi_{n+1}(\omega) = \frac{2k_0}{2^{n+1}}$ and $\varphi_{n+1}(\omega) - \varphi_n(\omega) = \frac{2k_0}{2^{n+1}} - \frac{k_0}{2^n} = 0$. Alternatively, if $\omega \in E_{2k_0+1,n+1}$ then $\varphi_{n+1}(\omega) = \frac{2k_0+1}{2^{n+1}}$ and $\varphi_{n+1}(\omega) - \varphi_n(\omega) = \frac{2k_0+1}{2^{n+1}} - \frac{k_0}{2^n} = \frac{1}{2^{n+1}} > 0$. Consequently, if $\omega \in E_{k_0,n}$ then $\varphi_{n+1}(\omega) - \varphi_n(\omega) \geq 0$.

Second, if $k_0 = n2^n$ then $E_{k_0,n} = f^{-1}([n,\infty])$. Now, if $\omega \in f^{-1}([n+1,\infty])$ then $\varphi_{n+1}(\omega) = n+1$ and $\varphi_n(\omega) = n$. Consequently, $\varphi_{n+1}(\omega) - \varphi_n(\omega) = 1 > 0$. If $\omega \in f^{-1}([n, n+1])$ then $\varphi_n(\omega) = n$ and $\varphi_{n+1}(\omega) = \frac{k}{2^{n+1}}$ if $\omega \in f^{-1}([\frac{k}{2^{n+1}}, \frac{k}{2^{n+1}} + \frac{1}{2^{n+1}}))$. Setting the lower bound of the interval equal to n gives $k = n2^{n+1}$ and $\varphi_{n+1}(\omega) = n$ if $\omega \in f^{-1}([n, n + \frac{1}{2^{n+1}}))$, giving $\varphi_{n+1}(\omega) - \varphi_n(\omega) = 0$. If $\omega \in f^{-1}([n + \frac{1}{2^{n+1}}, n + \frac{2}{2^{n+1}}))$ then $\varphi_{n+1}(\omega) = \frac{n2^{n+1}+1}{2^{n+1}}$ and consequently $\varphi_{n+1}(\omega) - \varphi_n(\omega) = \frac{1}{2^{n+1}} > 0$. Continuing in this fashion for subsequent sub-intervals of $[n, n+1]$ gives $\varphi_{n+1}(\omega) - \varphi_n(\omega) \geq 0$.

2. From item 1, we have that $\varphi_1(\omega) \leq \varphi_2(\omega) \leq \cdots \leq f(\omega)$ for all $\omega \in \Omega$. Hence, $\lim_{n\to\infty} \varphi_n(\omega) = \sup_{n\in\mathbb{N}} \varphi_n(\omega)$. But $0 \leq f(\omega) - \varphi_n(\omega) \leq \frac{1}{2^n}$ and taking limits as $n \to \infty$ we have $f(\omega) = \lim_{n\to\infty} \varphi_n(\omega) = \sup_{n\in\mathbb{N}} \varphi_n(\omega)$. ■

Definition 4.3. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a non-negative measurable function. The integral of f with respect to μ is given by*

$$\int_{\mathbb{X}} f d\mu := \sup_{\varphi} \int_{\mathbb{X}} \varphi(x) d\mu := \sup_{\varphi} I_\mu(\varphi) \in [0,\infty], \tag{4.5}$$

where the sup *is taken over all simple functions φ which are non-negative satisfying $\varphi(x) \leq f(x)$ for all $x \in \mathbb{X}$.*

Remark 4.3. *If f is a non-negative simple function $\int_{\mathbb{X}} f d\mu = I_\mu(f)$.*

Theorem 4.5. *(Beppo-Levi Theorem) Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and $\{f_j\}_{j\in\mathbb{N}}$ be an increasing sequence of non-negative measurable functions $f_j : (\mathbb{X}, \mathcal{F}) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$. Then $f = \sup_{j\in\mathbb{N}} f_j$ is a non-negative measurable function and*

$$\int_{\mathbb{X}} f d\mu := \int_{\mathbb{X}} \sup_{j\in\mathbb{N}} f_j d\mu = \sup_{j\in\mathbb{N}} \int_{\mathbb{X}} f_j d\mu.$$

Proof. That f is a non-negative measurable function follows from Theorem 3.6. Note that if g and h are non-negative measurable functions, we have by definition that

$$\int_{\mathbb{X}} g d\mu := \sup_{\varphi} \int_{\mathbb{X}} \varphi d\mu \text{ where } \varphi \leq g,\ \varphi \text{ a simple function.}$$

But if $g \leq h$,

$$\int_{\mathbb{X}} g d\mu \leq \sup_{\varphi} \int_{\mathbb{X}} \varphi d\mu = \int_{\mathbb{X}} h d\mu \text{ where } \varphi \leq h.$$

Now, $f_j \leq f := \sup_{j \in \mathbb{N}} f_j$. By the monotonicity of integrals, which we just established,

$$\int_{\mathbb{X}} f_j d\mu \leq \int_{\mathbb{X}} f d\mu.$$

Taking $\sup_{j \in \mathbb{N}}$ on both sides gives $\sup_{j \in \mathbb{N}} \int_{\mathbb{X}} f_j d\mu \leq \int_{\mathbb{X}} f d\mu$.

Now, we establish the reverse inequality, i.e., $\sup_{j \in \mathbb{N}} \int_{\mathbb{X}} f_j d\mu \geq \int_{\mathbb{X}} f d\mu$. Let $\varphi(x)$ be a simple non-negative function such that $\varphi \leq f$. If we can show that

$$I_{\mu}(\varphi) = \int_{\mathbb{X}} \varphi d\mu \leq \sup_{j \in \mathbb{N}} \int_{\mathbb{X}} f_j d\mu \tag{4.6}$$

we will have the desired inequality since we can take sup over all simple functions on both sides of (4.6) to give

$$\sup_{\varphi} \int_{\mathbb{X}} \varphi d\mu := \int_{\mathbb{X}} f d\mu \ \leq \sup_{j \in \mathbb{N}} \int_{\mathbb{X}} f_j d\mu.$$

Let φ be a simple non-negative function such that $\varphi \leq f$. Since $f(x) := \sup_{j \in \mathbb{N}} f_j(x)$, for every $x \in \mathbb{X}$ and $\epsilon \in (0,1)$, there exists $N_{(x,\epsilon)}$ such that

$$f_j(x) \geq \epsilon \varphi(x) \text{ whenever } j \geq N_{(x,\epsilon)}.$$

Now, if $A_j = \{x : f_j(x) \geq \epsilon \varphi(x)\}$ we note that the sets A_j increase as $j \to \infty$ since $f_1 \leq f_2 \cdots$. Furthermore, these sets are measurable by measurability of f_j and φ. By definition of A_j

$$\epsilon I_{A_j}(x) \varphi(x) \leq I_{A_j}(x) f_j(x) \leq f_j(x). \tag{4.7}$$

Since φ is a simple function it has a standard representation $\varphi(x) = \sum_{i=0}^{m} y_i I_{B_i}(x)$ and

$$\epsilon I_{A_j}(x) \sum_{i=0}^{m} y_i I_{B_i}(x) = \epsilon \sum_{i=0}^{m} y_i I_{B_i \cap A_j}(x).$$

Thus, the integral of the simple function in this expression is given by $\epsilon \sum_{i=0}^{m} y_i \mu(B_i \cap A_j)$. By monotonicity of integrals and using (4.7) we have

$$\epsilon \sum_{i=0}^{m} y_i \mu(B_i \cap A_j) \leq \int_{\mathbb{X}} f_j d\mu \leq \sup_{j \in \mathbb{N}} \int_{\mathbb{X}} f_j d\mu.$$

Since $\varphi \leq f$, the collection $\{A_j\}$ grows to $\mathbb{X}$ as $j \to \infty$. Thus, by the fact that μ is continuous from below

$$\mu(B_i \cap A_j) \uparrow \mu(B_i \cap \mathbb{X}) = \mu(B_i) \text{ as } j \to \infty$$

and

$$\epsilon \sum_{i=0}^{m} y_i \mu(B_i) = \epsilon \int_{\mathbb{X}} \varphi d\mu \leq \sup_{j \in \mathbb{N}} \int_{\mathbb{X}} f_j d\mu.$$

Now, just let ϵ be arbitrarily close to 1 to finish the proof. ■

Remark 4.4. 1. *If we take $f_j = \varphi_j$ where φ_j are non-negative simple functions and $f = \sup_{j \in \mathbb{N}} \varphi_j$, then*

$$\int_{\mathbb{X}} f d\mu = \sup_{j \in \mathbb{N}} \int_{\mathbb{X}} \varphi_j d\mu.$$

Note that sup *can be replaced with* $\lim_{j \to \infty}$.

2. *If $E \in \mathcal{F}$, then $I_E(x) f(x)$ is a non-negative measurable function if $f \geq 0$. We define*

$$\int_{E} f d\mu = \int_{\mathbb{X}} I_E f d\mu. \tag{4.8}$$

Theorem 4.6. *Let $(\mathbb{X}, \mathcal{F}, \mu)$ be a measure space and f, $g : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be numerical non-negative measurable functions. Then*

1. $\int_{\mathbb{X}} a f d\mu = a \int_{\mathbb{X}} f d\mu$ *for $a \geq 0$,*

2. $\int\limits_{\mathbb{X}}(f+g)d\mu = \int\limits_{\mathbb{X}} f d\mu + \int\limits_{\mathbb{X}} g d\mu$,

3. *If* $E, F \in \mathcal{F}$ *and* $E \subset F$, *then* $\int_E f d\mu \leq \int_F f d\mu$.

Proof. 1. If $a > 0$, let φ_n be an increasing sequence of measurable non-negative simple functions converging to f (such sequence exists by Theorem 4.4). Then, $a\varphi_n$ is an increasing sequence converging point wise to af. By Theorem 4.5 and the fact that $I_\mu(a\varphi_n) = aI_\mu(\varphi_n)$

$$\int\limits_{\mathbb{X}} af d\mu = \lim_{n\to\infty} \int\limits_{\mathbb{X}} a\varphi_n d\mu = a \lim_{n\to\infty} \int\limits_{\mathbb{X}} \varphi_n(\omega) d\mu = a \int\limits_{\mathbb{X}} f d\mu$$

2. Let φ_n, ψ_n be non-negative increasing simple functions converging to f and g. Then $\varphi_n + \psi_n$ is an increasing sequence converging to $f + g$. Again, by Theorem 4.5

$$\begin{aligned}
\int\limits_{\mathbb{X}} (f+g)d\mu &= \lim_{n\to\infty} \int\limits_{\mathbb{X}} (\varphi_n + \psi_n) d\mu \text{ by Beppo-Levi's Theorem} \\
&= \lim_{n\to\infty} \int\limits_{\mathbb{X}} \varphi_n d\mu + \lim_{n\to\infty} \int\limits_{\mathbb{X}} \psi_n d\mu \text{ by Theorem 4.3} \\
&= \int\limits_{\mathbb{X}} f d\mu + \int\limits_{\mathbb{X}} g d\mu. \text{ by Beppo-Levi's Theorem}
\end{aligned}$$

3. Since f is non-negative $fI_E \leq fI_F$ therefore

$$\int_E f d\mu = \int\limits_{\mathbb{X}} f I_E d\mu \leq \int\limits_{\mathbb{X}} f I_F d\mu = \int_F f d\mu.$$

■

Corollary 4.1. *Let* $\{f_j\}_{j\in\mathbb{N}}$ *be a sequence of measurable non-negative numerical functions, i.e.,* $f_j : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$. *Then,* $\sum_{j=1}^{\infty} f_j$ *is measurable and*

$$\int_{\mathbb{X}} \left(\sum_{j=1}^{\infty} f_j \right) d\mu = \sum_{j=1}^{\infty} \int_{\mathbb{X}} f_j d\mu.$$

Proof. Let $S_m = \sum_{j=1}^{m} f_j$, $S = \lim\limits_{m\to\infty} \sum_{j=1}^{m} f_j = \sum_{j=1}^{\infty} f_j$ and note that $0 \leq S_1 \leq S_2 \leq \cdots$. Then, by Theorem 4.6.3 we have that

$$\int_{\mathbb{X}} S_m d\mu = \sum_{j=1}^{m} \int_{\mathbb{X}} f_j d\mu.$$

Taking limits as $m \to \infty$ and using Theorem 4.5, we have

$$\lim_{m\to\infty}\int_{\mathbb{X}} S_m d\mu = \lim_{m\to\infty}\sum_{j=1}^{m}\int_{\mathbb{X}} f_j d\mu = \sum_{j=1}^{\infty}\int_{\mathbb{X}} f_j d\mu = \int_{\mathbb{X}} S d\mu = \int_{\mathbb{X}}\left(\sum_{j=1}^{\infty} f_j\right) d\mu.$$

■

Theorem 4.7. *(Fatou's Lemma): Let $\{f_j\}_{j\in\mathbb{N}}$ be a sequence of measurable non-negative numerical functions $f_j : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$. Then, $f := \liminf_{j\to\infty} f_j$ is measurable and*

$$\int_{\mathbb{X}} f d\mu \leq \liminf_{j\to\infty}\int_{\mathbb{X}} f_j d\mu.$$

Proof. First, f is measurable by Theorem 3.6. Let $g_n = \inf\{f_n, f_{n+1}, \cdots\}$ for $n = 1, 2, \cdots$, and note that $g_1 \leq f_1$, $g_1 \leq f_2, \cdots$. Also, $g_2 \leq f_2$, $g_2 \leq f_3 \cdots$. Thus, $g_n \leq f_j$ for all $n \leq j$. Furthermore, $g_1 \leq g_2 \leq \cdots$. Now, recall that $f := \liminf_{j\to\infty} f_j := \sup_{n\in\mathbb{N}} \inf_{j\geq n} f_j$ and

$$\lim_{n\to\infty} g_n = \liminf_{j\to\infty} f_j := f.$$

Also, $\int_{\mathbb{X}} g_n d\mu \leq \int_{\mathbb{X}} f_j d\mu$ for all $n \leq j$ and

$$\int_{\mathbb{X}} g_n d\mu \leq \liminf_{j\to\infty}\int_{\mathbb{X}} f_j d\mu.$$

Since the sequence $g_n \uparrow \liminf_{j\to\infty} f_j$, by Theorem 4.5

$$\lim_{n\to\infty}\int_{\mathbb{X}} g_n d\mu = \int_{\mathbb{X}} f d\mu \leq \liminf_{j\to\infty}\int_{\mathbb{X}} f_j(\omega) d\mu.$$

■

4.4 Integral of functions

Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a measurable numerical function and $f^+ = \max\{f, 0\}$ and $f^- = -\min\{f, 0\}$.

Definition 4.4. *Let* $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ *be a measurable numerical function such that* $\int_{\mathbb{X}} f^+ d\mu < \infty$ *and* $\int_{\mathbb{X}} f^- d\mu < \infty$. *In this case, we say that* f *is* μ*-integrable and we write*

$$\int_{\mathbb{X}} f d\mu := \int_{\mathbb{X}} f^+ d\mu - \int_{\mathbb{X}} f^- d\mu.$$

We note that $\int_{\mathbb{X}} f d\mu \in \mathbb{R}$ and denote by $\mathcal{L}_{\mathbb{R}}$ the set of integrable real functions and $\mathcal{L}_{\bar{\mathbb{R}}}$ the set of integrable numerical functions. A non-negative function f is said to be integrable if, and only if, $\int_{\mathbb{X}} f d\mu < \infty$. If $(\mathbb{X}, \mathcal{F}, \mu) := (\mathbb{R}^n, \mathcal{B}^n, \lambda^n)$ we call $\int_{\mathbb{R}^n} f d\lambda^n$ the Lebesgue integral.

Theorem 4.8. *Let* $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ *be a measurable function. Then, the following statements are equivalent:*

1. $f \in \mathcal{L}_{\bar{\mathbb{R}}}$,

2. $|f| \in \mathcal{L}_{\bar{\mathbb{R}}}$,

3. *there exists* $0 \leq g \in \mathcal{L}_{\bar{\mathbb{R}}}$ *such that* $|f| \leq g$.

Proof. $(1 \implies 2)$ Since, $|f| = f^+ + f^-$ and since integrability of f implies $\int_{\mathbb{X}} f^+ d\mu < \infty$ and $\int_{\mathbb{X}} f^- d\mu < \infty$ we have $\int_{\mathbb{X}} |f| d\mu = \int_{\mathbb{X}} f^+ d\mu + \int_{\mathbb{X}} f^- d\mu < \infty$.
$(2 \implies 3)$ Just take $g = |f|$.
$(3 \implies 1)$ Since $f^+ \leq |f| \leq g$ and $f^- \leq |f| \leq g$, we have by the monotonicity of the integral of non-negative functions and the integrability of g that $f^+, f^- \in \mathcal{L}_{\bar{\mathbb{R}}}$. Hence, $f \in \mathcal{L}_{\bar{\mathbb{R}}}$. ■

Theorem 4.9. *Let* $f, g : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ *be measurable functions such that* $f, g \in \mathcal{L}_{\bar{\mathbb{R}}}$ *and* $a \in \mathbb{R}$. *Then,*

1. $af \in \mathcal{L}_{\bar{\mathbb{R}}}$ *and* $\int_{\mathbb{X}} af d\mu = a \int_{\mathbb{X}} f d\mu$,

2. $(f + g) \in \mathcal{L}_{\bar{\mathbb{R}}}$ *and* $\int_{\mathbb{X}} (f + g) d\mu = \int_{\mathbb{X}} f d\mu + \int_{\mathbb{X}} g d\mu$,

3. $\max\{f, g\}, \min\{f, g\} \in \mathcal{L}_{\bar{\mathbb{R}}}$,

4. *if $f \leq g$ then $\int_{\mathbb{X}} f d\mu \leq \int_{\mathbb{X}} g d\mu$.*

Proof. Homework. Use Theorems 4.8 and 4.6. ■

Remark 4.5. *Note that*

$$\left|\int_{\mathbb{X}} f d\mu\right| \leq \left|\int_{\mathbb{X}} f^{+} d\mu\right| + \left|\int_{\mathbb{X}} f^{-} d\mu\right| = \int_{\mathbb{X}} f^{+} d\mu + \int_{\mathbb{X}} f^{-} d\mu = \int_{\mathbb{X}} (f^{+} + f^{-}) d\mu = \int_{\mathbb{X}} |f| d\mu.$$

Theorem 4.10. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be a non-negative measurable function such that $f \in \mathcal{L}_{\bar{\mathbb{R}}}$ and*

$$m(E) = \int_{E} f d\mu \text{ for all } E \in \mathcal{F}.$$

Then, m is a measure on $\mathcal{F}$.

Proof. Since $f \geq 0$, $m(E) \geq 0$. If $E = \emptyset$, then $f I_E = 0$ and

$$m(\emptyset) = \int_{\emptyset} f d\mu = \int_{\mathbb{X}} f I_{\emptyset} d\mu = \int_{\mathbb{X}} 0 d\mu = 0.$$

Now, let $\{E_j\}_{j \in \mathbb{N}}$ be a disjoint collection of sets in $\mathcal{F}$ such that $\cup_{j=1}^{\infty} E_j = E$ and let $f_n(x) = \sum_{j=1}^{n} f(x) I_{E_j}(x)$. By Theorem 4.6 $\int_{\mathbb{X}} f_n d\mu = \sum_{j=1}^{n} \int_{\mathbb{X}} f I_{E_j} d\mu$. Thus, $\int_{\mathbb{X}} f_n d\mu = \sum_{j=1}^{n} m(E_j)$. Note that $f_1 \leq f_2 \leq \cdots$ and converges to $f I_E$. Hence, by Theorem 4.5

$$m(E) = \int_{\mathbb{X}} f I_E d\mu = \lim_{n \to \infty} \int_{\mathbb{X}} f_n d\mu = \lim_{n \to \infty} \sum_{j=1}^{n} m(E_j) = \sum_{j=1}^{\infty} m(E_j).$$

■

Remark 4.6. 1. *Suppose $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a random variable and P_X is the probability measure induced by X on $\mathcal{B}(\mathbb{R})$ as in Example 3.2. Then, in Theorem 4.10 letting $(\mathbb{X}, \mathcal{F}, \mu) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), P_X)$, we conclude that*

$$m_X(B) = \int_{B} f dP_X \text{ for all } B \in \mathcal{B}(\mathbb{R})$$

is a measure on $\mathcal{B}(\mathbb{R})$. In particular, if $B = (-\infty, x]$ for $x \in \mathbb{R}$, $m_X((-\infty, x]) = \int_{(-\infty, x]} f dP_X$.

2. *m is called the measure with density function f with respect to μ and is denoted by $m = f\mu$. If m has a density with respect to μ it is traditional in mathematics to write $dm/d\mu$ for the the density function. We note that with a little more work we can recognize f as the Radon-Nikodým derivative of m with respect to the measure μ.*

Chapter 5

Lebesgue's convergence theorems and $\mathcal{L}^p$ spaces

In this chapter we study two important convergence theorems and some of their uses and applications.

5.1 Convergence theorems

Theorem 5.1. *(Lebesgue's Monotone Convergence Theorem) Let* $f_n : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ *for* $n \in \mathbb{N}$ *be integrable functions such that* $f_1 \leq f_2 \leq \cdots$ *and* $f := \lim_{n\to\infty} f_n = \sup_{n\in\mathbb{N}} f_n$*. Then,*

$$f \in \mathcal{L}_{\bar{\mathbb{R}}}(\mu) \iff \sup_{n\in\mathbb{N}} \int_{\mathbb{X}} f_n d\mu < \infty.$$

In this case,

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{X}} f_n d\mu = \int_{\mathbb{X}} \sup_{n\in\mathbb{N}} f_n d\mu = \int_{\mathbb{X}} f d\mu.$$

Proof. Since $f_n \in \mathcal{L}_{\bar{\mathbb{R}}}$ and $f_1 \leq f_2 \leq \cdots$ we have that $0 \leq f_n - f_1 \in \mathcal{L}_{\bar{\mathbb{R}}}$ forms an increasing sequence of nonnegative measurable functions. Hence, by Theorem 4.5

$$0 \leq \sup_{n\in\mathbb{N}} \int_{\mathbb{X}} (f_n - f_1) d\mu = \int_{\mathbb{X}} \sup_{n\in\mathbb{N}} (f_n - f_1) d\mu. \tag{5.1}$$

Now, suppose $f \in \mathcal{L}_{\bar{\mathbb{R}}}$ and note that from the left side of equation (5.1)

$$\sup_{n\in\mathbb{N}} \int_{\mathbb{X}} f_n d\mu - \int_{\mathbb{X}} f_1 d\mu = \int_{\mathbb{X}} (f - f_1) d\mu, \text{ or}$$
$$\sup_{n\in\mathbb{N}} \int_{\mathbb{X}} f_n d\mu = \int_{\mathbb{X}} f_1 d\mu + \int_{\mathbb{X}} (f - f_1) d\mu$$
$$= \int_{\mathbb{X}} f_1 d\mu + \int_{\mathbb{X}} f d\mu - \int_{\mathbb{X}} f_1 d\mu = \int_{\mathbb{X}} f d\mu < \infty.$$

If $\sup_{n\in\mathbb{N}} \int_{\mathbb{X}} f_n d\mu < \infty$, then from equation (5.1) we have $\int_{\mathbb{X}}(f - f_1)d\mu < \infty$ and since f_1 is integrable $f = (f - f_1) + f_1$ is integrable. Therefore,

$$\int f d\mu = \int (f - f_1) d\mu + \int f_1 du = \sup_{n\in\mathbb{N}} \int_{\mathbb{X}} f_n d\mu < \infty.$$

■

We now prove a useful inequality.

Theorem 5.2. *(Markov's Inequality) Let* $(\mathbb{X}, \mathcal{F}, \mu)$ *be a measure space and* $f \in \mathcal{L}_{\bar{\mathbb{R}}}$*. Then, for all* $E \in \mathcal{F}$ *and* $a > 0$

$$\mu\left(\{|f| \geq a\} \cap E\right) \leq \frac{1}{a} \int_E |f| d\mu.$$

Proof. Note that, $aI_{\{|f|\geq a\}\cap E} = aI_{\{|f|\geq a\}} I_E \leq |f| I_E$ and consequently, integrating both sides, $a\mu(\{|f| \geq a\} \cap E) \leq \int_E |f| d\mu$. Therefore,

$$\mu(\{|f| \geq a\} \cap E) \leq \frac{1}{a} \int_E |f| d\mu.$$

■

Remark 5.1. *Note that if* $E = \mathbb{X}$ *we have* $\mu(\{|f| \geq a\}) \leq \frac{1}{a} \int_{\mathbb{X}} |f| d\mu$*. When* $(\mathbb{X}, \mathcal{F}, \mu) = (\Omega, \mathcal{F}, P)$ *a probability space and* $f := X$ *a random variable, the last result is commonly stated as*

$$P(\{|X| \geq a\}) \leq \frac{1}{a} E_P(|X|).$$

Also, note that if $f := (X - E_P(X))^2$ *we have*

$$P(\{(X - E_P(X))^2 \geq a\}) = P(\{|X - E_P(X)| \geq a^{1/2}\}) \leq \frac{1}{a} E_P((X - E_P(X))^2),$$

and letting $b = a^{1/2}$ *we have*

$$P(\{|X - E_P(X)| \geq b\}) \leq \frac{1}{b^2} E_P((X - E_P(X))^2),$$

which is known as Chebyshev's Inequality.

Recall that for a measure space $(\mathbb{X}, \mathcal{F}, \mu)$, N is a null set if $N \in \mathcal{F}$ and $\mu(N) = 0$. If a certain property $\mathcal{P}(x)$ that depends on $x \in \mathbb{X}$ holds for all $x \in \mathbb{X}$ except $x \in N_{\mathcal{P}} \subset N$, where N is a null set, we say that the property is true almost everywhere (ae) or almost surely (as). Note the set $N_{\mathcal{P}}$ where the property does not hold need not be a measurable set.

Theorem 5.3. *Let* $(\mathbb{X}, \mathcal{F}, \mu)$ *be a measure space and* $f \in \mathcal{L}_{\bar{\mathbb{R}}}$. *Then,*

1. *if* N *is a null set* $\int_N f d\mu = 0$,

2. $\int_{\mathbb{X}} |f| d\mu = 0 \iff |f| = 0$ *ae.*

Proof. 1. For $j \in \mathbb{N}$, let $f_j = \min\{|f|, j\}$ and note $0 \leq f_1 \leq f_2 \leq \cdots$ with $\lim_{j \to \infty} f_j = |f|$. Hence, by Theorem 4.5

$$\begin{aligned} 0 \leq \left| \int_N f d\mu \right| &= \left| \int_{\mathbb{X}} I_N f d\mu \right| \leq \int_{\mathbb{X}} I_N |f| d\mu \\ &= \lim_{j \to \infty} \int_{\mathbb{X}} I_N f_j d\mu = \lim_{j \to \infty} \int_{\mathbb{X}} I_N \min\{|f|, j\} d\mu \leq \lim_{j \to \infty} \int_{\mathbb{X}} j I_N d\mu \\ &= \lim_{j \to \infty} j \int_{\mathbb{X}} I_N d\mu = \lim_{j \to \infty} j \mu(N) = 0. \end{aligned}$$

2. ($\Leftarrow$) $\int_{\mathbb{X}} |f| d\mu = \int_{\{|f|=0\}} |f| d\mu + \int_{\{|f| \neq 0\}} |f| d\mu = \int_{\{|f| \neq 0\}} |f| d\mu = 0$ by item 1.

($\Rightarrow$) Note that by the fact that μ is a measure

$$\begin{aligned} \mu(\{|f| > 0\}) &= \mu\left(\bigcup_{j \in \mathbb{N}} \{|f| \geq 1/j\} \right) \leq \sum_{j \in \mathbb{N}} \mu(\{|f| \geq 1/i\}) \\ &\leq \sum_{j \in \mathbb{N}} j \int_{\mathbb{X}} |f| d\mu = 0 \end{aligned}$$

by Markov's Inequality and the assumption that $\int_{\mathbb{X}} |f| d\mu = 0$. ■

Remark 5.2. *1. If f, $g \geq 0$ are measurable, integrable and $f = g$ μ-ae then $\int_{\mathbb{X}} f d\mu = \int_{\{x:f(x)\neq g(x)\}} f d\mu + \int_{\{x:f(x)=g(x)\}} f d\mu$. But by Theorem 5.3.1, the first integral in this sum is equal to zero. Consequently, $\int_{\mathbb{X}} f d\mu = \int_{\{x:f(x)=g(x)\}} f d\mu = \int_{\{x:f(x)=g(x)\}} g d\mu = \int_{\{x:f(x)\neq g(x)\}} g d\mu + \int_{\{x:f(x)=g(x)\}} g d\mu = \int_{\mathbb{X}} g d\mu$.*

2. If $f \in \mathcal{L}_{\bar{\mathbb{R}}}$ and $f = g$ μ-ae then $g \in \mathcal{L}_{\bar{\mathbb{R}}}$. To see this, note that $f = g$ μ-ae implies $f^+ = g^+$ and $f^- = g^-$ μ-ae. Using the previous remark on f^+ and f^- we have $\int_{\mathbb{X}} f^+ d\mu = \int_{\mathbb{X}} g^+ d\mu$ and $\int_{\mathbb{X}} f^- d\mu = \int_{\mathbb{X}} g^- d\mu$. Hence, $g \in \mathcal{L}_{\bar{\mathbb{R}}}$ and $\int_{\mathbb{X}} f d\mu = \int_{\mathbb{X}} g d\mu$.

3. If f is measurable and $0 \leq g \in \mathcal{L}_{\bar{\mathbb{R}}}$ with $|f| \leq g$ ae, then

$$f^+ \leq |f| \leq g \text{ ae and } f^- \leq |f| \leq g \text{ ae}.$$

Hence, $\int_{\mathbb{X}} f^+ d\mu \leq \int_{\mathbb{X}} g d\mu$, $\int_{\mathbb{X}} f^- d\mu \leq \int_{\mathbb{X}} g d\mu$ and f is integrable.

Theorem 5.4. *Let $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\bar{\mathbb{R}}, \bar{\mathcal{B}})$ be integrable. Then f is real valued almost everywhere.*

Proof. Let $\{x : |f(x)| = \infty\} = \{x : f(x) = \infty\} \cup \{x : f(x) = -\infty\} \in \mathcal{B}$. Note that $\bigcap_{n\in\mathbb{N}} \{x : |f(x)| \geq n\} := \bigcap_{n\in\mathbb{N}} B_n$ with $B_1 \supset B_2 \supset \cdots$. Hence, $\lim_{n\to\infty} B_n = \bigcap_{n\in\mathbb{N}} B_n = N$. Also, note that by Markov's Inequality and integrability of f

$$\mu(B_1) = \mu(\{x : |f(x)| \geq 1\}) \leq \int_{\mathbb{X}} |f| d\mu < \infty.$$

Hence, by continuity of measures from above, and Markov's Inequality

$$\mu(N) = \lim_{n\to\infty} \mu(B_n) = \lim_{n\to\infty} \mu(\{x : |f(x)| \geq n\}) \leq \lim_{n\to\infty} \frac{1}{n} \int_{\mathbb{X}} |f| d\mu = 0.$$

■

Theorem 5.5. *(Lebesgue's Dominated Convergence Theorem) Let* $(\mathbb{X}, \mathcal{F}, \mu)$ *be a measure space and* $\{f_n\}_{n\in\mathbb{N}}$ *be a sequence of integrable functions such that* $|f_n| \leq g$ *for all* n*, almost everywhere, where* g *is some integrable nonnegative function. If* $\lim_{n\to\infty} f_n(x) = f(x)$ *exists almost everywhere in* $\bar{\mathbb{R}}$*, then* f *is integrable and*

$$\lim_{n\to\infty} \int_{\mathbb{X}} f_n d\mu = \int_{\mathbb{X}} \lim_{n\to\infty} f_n d\mu := \int_{\mathbb{X}} f d\mu.$$

Proof. We start by observing that since the f_n and g are measurable, the set

$$N = \{x : \lim_{n\to\infty} f_n(x) \text{ does not exist}\} \cup \left(\bigcup_{n\in\mathbb{N}} \{x : |f(x)| > g(x)\} \right)$$

is measurable and $\mu(N) = 0$. Thus, we proceed by taking $N = \emptyset$ as it does not contribute to any of the integrals in the proof of the Theorem. By the point wise limit of the sequence f_n, for any $\epsilon > 0$ there exists $N_{(\epsilon,x)}$ such that for all $n > N_{(\epsilon,x)}$

$$\begin{aligned} |f| &= |f - f_n + f_n| \leq |f_n| + |f - f_n| \\ &\leq g + |f - f_n| \text{ by } |f_n| < g \\ &\leq g + \epsilon. \end{aligned}$$

Therefore, $\int_{\mathbb{X}} f d\mu < \infty$ provided $g \in \mathcal{L}_{\bar{\mathbb{R}}}(\mu)$. Also, $|f_n| \leq g \iff -g \leq f_n \leq g$. Hence, $f_n + g \geq 0$. By Fatou's Lemma,

$$\begin{aligned} \int \liminf_{n\to\infty} (f_n + g) d\mu = \int (f + g) d\mu &\leq \liminf_{n\to\infty} \int (f_n + g) d\mu \\ &= \liminf_{n\to\infty} \int f_n d\mu + \int g d\mu. \end{aligned}$$

Therefore,

$$\int f d\mu \leq \liminf_{n\to\infty} \int f_n d\mu. \tag{5.2}$$

Also, $g - f_n \geq 0$ and again by Fatou's Lemma,

$$
\begin{aligned}
0 \leq \int \liminf_{n\to\infty}(g - f_n)d\mu = \int g d\mu - \int f d\mu \\
&\leq \liminf_{n\to\infty} \int (g - f_n)d\mu \\
&= \int g d\mu + \liminf_{n\to\infty} - \int f_n d\mu \\
&= \int g d\mu - \limsup_{n\to\infty} \int f_n d\mu.
\end{aligned}
$$

The second inequality together with the last equality imply that

$$\int f d\mu \geq \limsup_{n\to\infty} \int f_n d\mu. \tag{5.3}$$

Combining (5.2) and (5.3) completes the proof. ■

We now consider a measurable function that is indexed by a parameter $\theta \in (a, b)$ for $a < b$. As such, we define $f(x, \theta) : (\mathbb{X}, \mathcal{F}, \mu) \times (a, b) \to (\mathbb{R}, \mathcal{B})$ where f is measurable for all $\theta \in (a, b)$.

Theorem 5.6. *Let $f(x, \theta) : (\mathbb{X}, \mathcal{F}, \mu) \times (a, b) \to (\mathbb{R}, \mathcal{B})$ where f is measurable and $f \in \mathcal{L}_{\mathbb{R}}$ for all $\theta \in (a, b)$. Also, assume that $f(x, \theta)$ is continuous for every $x \in \mathbb{X}$ and $|f(x, \theta)| \leq g(x)$ for all $(x, \theta) \in \mathbb{X} \times (a, b)$ and some nonnegative integrable function g. Then, the function $h : (a, b) \to \mathbb{R}$ given by*

$$h(\theta) := \int_{\mathbb{X}} f(x, \theta)d\mu$$

is continuous.

Proof. The function h is well defined because of integrability of $f(x, \theta)$. It suffices to show that for any sequence $\{\theta_n\}_{n\in\mathbb{N}} \subset (a, b)$ such that $\theta_n \to \theta$ we have $h(\theta_n) \to h(\theta)$. By continuity of $f(x, \theta)$, for every x, we have $f(x, \theta_n) \to f(x, \theta)$ and $|f(x, \theta_n)| \leq g(x)$. By Lebesgue's Dominated Convergence Theorem,

$$\lim_{n\to\infty} h(\theta_n) = \int_{\mathbb{X}} \lim_{n\to\infty} f(x, \theta_n)d\mu = \int_{\mathbb{X}} f(x, \theta)d\mu = h(\theta).$$

■

Theorem 5.7. *Let $f(x,\theta) : (\mathbb{X}, \mathcal{F}, \mu) \times (a,b) \to (\mathbb{R}, \mathcal{B})$ where f is measurable and $f \in \mathcal{L}_{\mathbb{R}}$ for all $\theta \in (a,b)$. Also, assume that $f(x,\theta)$ is differentiable on (a,b) for every $x \in \mathbb{X}$ and $|\frac{d}{d\theta} f(x,\theta)| \leq g(x)$ for all $(\theta, x) \in (a,b) \times \mathbb{X}$ and some nonnegative integrable function g. Then, the function $h : (a,b) \to \mathbb{R}$ given by*

$$h(\theta) := \int_{\mathbb{X}} f(x,\theta) d\mu$$

is differentiable and its derivative is given by

$$\frac{d}{d\theta} h(\theta) = \int_{\mathbb{X}} \frac{d}{d\theta} f(x,\theta) d\mu.$$

Proof. Recall that $\theta, \theta_n \in (a,b)$ with $\theta_n \to \theta$ and $\theta_n \neq \theta$.

$$\frac{d}{d\theta} f(x,\theta) = \lim_{n\to\infty} \frac{f(x,\theta_n) - f(x,\theta)}{\theta_n - \theta}$$

for all $x \in \mathbb{X}$ and consequently $\frac{d}{d\theta} f(x,\theta)$ is measurable. By the Mean Value Theorem, $f(x,\theta_n) - f(x,\theta) = \frac{d}{d\theta} f(x, \theta_{z,n})(\theta_n - \theta)$ with $\theta_{z,n} = \lambda\theta_n + (1-\lambda)\theta$, $\lambda \in (0,1)$, $\theta_{z,n} \in (a,b)$. Consequently,

$$\left| \frac{f(x,\theta_n) - f(x,\theta)}{\theta_n - \theta} \right| = \left| \frac{d}{d\theta} f(x, \theta_{z,n}) \right| \leq g(x)$$

so that $\left| \frac{f(x,\theta_n) - f(x,\theta)}{\theta_n - \theta} \right|$ is integrable. Thus,

$$\frac{h(\theta_n) - h(\theta)}{\theta_n - \theta_0} = \int_{\mathbb{X}} \frac{f(x,\theta_n) - f(x,\theta)}{\theta_n - \theta} d\mu.$$

Hence, by the Lebesgue's Dominated Convergence Theorem

$$\lim_{n\to\infty} \frac{h(\theta_n) - h(\theta)}{\theta_n - \theta} = \frac{d}{d\theta} h(\theta) = \int_{\mathbb{X}} \lim_{n\to\infty} \frac{f(x,\theta_n) - f(x,\theta)}{\theta_n - \theta} d\mu = \int_{\mathbb{X}} \frac{d}{d\theta} f(x,\theta) d\mu.$$

■

5.2 $\mathcal{L}^p$ spaces

Definition 5.1. *The collection of measurable functions* $f : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ *such that* $\int_{\mathbb{X}} |f|^p d\mu < \infty$ *for* $p \in [1, \infty)$ *is denoted by* $\mathcal{L}^p_{\mathbb{R}}(\mu)$ *or* $\mathcal{L}^p_{\mathbb{R}}(\mathbb{X}, \mathcal{F}, \mu)$.

Let $f, g \in \mathcal{L}^p_{\mathbb{R}}(\mathbb{X}, \mathcal{F}, \mu)$ and define $s : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ as $s(x) = f(x) + g(x)$ for all $x \in \mathbb{X}$. Then, $|s(x)| \leq |f(x)| + |g(x)| \leq 2 \max\{|f(x)|, |g(x)|\}$ and

$$|s(x)|^p \leq 2^p \max\{|f(x)|, |g(x)|\}^p = 2^p \max\{|f(x)|^p, |g(x)|^p\} \leq 2^p(|f(x)|^p + |g(x)|^p).$$

Consequently, $\int_{\mathbb{X}} |s|^p d\mu \leq 2^p(\int_{\mathbb{X}} |f|^p d\mu + \int_{\mathbb{X}} |g|^p d\mu) < \infty$. Also, if $a \in \mathbb{R}$ and $m : (\mathbb{X}, \mathcal{F}, \mu) \to (\mathbb{R}, \mathcal{B})$ is defined as $m(x) = af(x)$ for all $x \in \mathbb{X}$, m is measurable and we have $|m(x)|^p = |a|^p |f(x)|^p$ and $\int_{\mathbb{X}} |m|^p d\mu = |a|^p \int_{\mathbb{X}} |f|^p d\mu < \infty$. Lastly, if we take $\theta(x) = 0$ for all $x \in \mathbb{X}$ to be the null vector in $\mathcal{L}^p_{\mathbb{R}}(\mathbb{X}, \mathcal{F}, \mu)$, then $\mathcal{L}^p_{\mathbb{R}}(\mathbb{X}, \mathcal{F}, \mu)$ is a vector space.

If $f \in \mathcal{L}^p_{\mathbb{R}}(\mathbb{X}, \mathcal{F}, \mu)$ we define the function $\| \cdot \|_p : \mathcal{L}^p_{\mathbb{R}}(\mathbb{X}, \mathcal{F}, \mu) \to [0, \infty)$ as $\|f\|_p = \left(\int_{\mathbb{X}} |f|^p d\mu\right)^{1/p}$ and prove the following inequality.

Theorem 5.8. *(Hölder's Inequality) If* $1 < p < \infty$, $p^{-1} + q^{-1} = 1$, $f \in \mathcal{L}^p_{\mathbb{R}}$, $g \in \mathcal{L}^q_{\mathbb{R}}$, *then* $fg \in \mathcal{L}_{\mathbb{R}}$ *and* $\int_{\mathbb{X}} |fg| d\mu \leq \|f\|_p \|g\|_q$.

Proof. If $\|f\|_p = 0$ then, by Theorem 5.3 $|f| = 0$ ae, so $|fg| = 0$ ae. Hence, $\int |fg| d\mu = 0$ and the inequality holds. Likewise for $\|g\|_q = 0$. So, assume $\|f\|_p, \|g\|_q \neq 0$. Let $x = f/\|f\|_p$, $y = g/\|g\|_q$ and note that $\|x\|_p = 1$ and $\|y\|_q = 1$. It suffices to prove $\int |xy| d\mu \leq 1$.

Now, note that for any $a, b > 0$ and $0 < \alpha < 1$,

$$a^\alpha b^{1-\alpha} \leq \alpha a + (1 - \alpha) b.$$

To see this, divide by b to obtain $(\frac{a}{b})^\alpha \leq \alpha \frac{a}{b} + (1 - \alpha)$. It suffices to show $u^\alpha \leq \alpha u + (1 - \alpha)$, for $u > 0$.

The inequality holds for $u = 1$. Now, $\frac{d}{du} u^\alpha = \alpha u^{\alpha - 1} = \alpha \frac{1}{u^{1-\alpha}}$. Since $\alpha \in (0, 1)$ we have that $u^{1-\alpha} < 1$ if $u < 1$. Consequently, in this case, $u^{\alpha-1} > 1$ and $\frac{d}{du} u^\alpha > \alpha$. Also, using the same arguments, if $u > 1$ we have that $\frac{d}{du}u^\alpha < \alpha$. By the Mean Value Theorem, for $\lambda \in (0,1)$

$$u^\alpha - 1 = \alpha(\lambda u + (1-\lambda))^{\alpha-1}(u-1) < \alpha(u-1) \implies u^\alpha < 1 - \alpha + \alpha u \text{ if } u > 1.$$

Also,

$$u^\alpha - 1 = \alpha(\lambda u + (1-\lambda))^{\alpha-1}(u-1) < \alpha(u-1) \implies u^\alpha < 1 + \alpha u - \alpha, \text{ if } u < 1.$$

Thus, $u^\alpha \leq \alpha u + (1-\alpha)$ for $u > 0$.

Now, let $\alpha = 1/p$, $a(\omega) = |x(\omega)|^p$, $b(\omega) = |y(\omega)|^q$ and $1 - \alpha = 1/q$. Then,

$$(|x(\omega)|^p)^{1/p}(|y(\omega)|^q)^{1/q} \leq \alpha|x(\omega)|^p + (1-\alpha)|y(\omega)|^q, \text{ or}$$
$$|x(\omega)y(\omega)| \leq \alpha|x(\omega)|^p + (1-\alpha)|y(\omega)|^q.$$

Thus, integrating both sides of the inequality we obtain $\int |xy|d\mu \leq \alpha\|x\|_p + (1-\alpha)\|y\|_q = 1$.

■

Theorem 5.9. *(Minkowski-Riez Inequality) For $1 \leq p < \infty$, if f and g are in $\mathcal{L}^p$ we have $\|f+g\|_p \leq \|f\|_p + \|g\|_p$.*

Proof. By the triangle inequality

$$\begin{aligned} \|f+g\|_p^p &= \int |f+g||f+g|^{p-1}d\mu \leq \int \left(|f||f+g|^{p-1} + |g||f+g|^{p-1}\right) d\mu \\ &= \int |f||f+g|^{p-1}d\mu + \int |g||f+g|^{p-1}d\mu, \text{ and if } p = 1 \text{ the proof is complete.} \\ &\text{If } p > 1, \text{ by Hölder's Inequality} \\ &\leq \|f\|_p\||f+g|^{p-1}\|_q + \|g\|_p\||f+g|^{p-1}\|_q, \end{aligned}$$

where $1/p + 1/q = 1$ which implies $1/q = 1 - 1/p \implies q = \frac{p}{p-1}$. Thus,

$$\|f+g\|_p^p \leq \|f\|_p\||f+g|^{p/q}\|_q + \|g\|_p\||f+g|^{p/q}\|_q = (\|f\|_p + \|g\|_p)\||f+g|^{p/q}\|_q. \tag{5.4}$$

Now,

$$\||f+g|^{p/q}\|_q = \left(\int (|f+g|^{p/q})^q d\mu\right)^{1/q} = \left(\int |f+g|^p d\mu\right)^{1/q}$$
$$= \left(\int |f+g|^p d\mu\right)^{\frac{p-1}{p}} = \|f+g\|_p^{p-1}$$

Using this in inequality (5.4) we obtain $\|f+g\|_p^{p-(p-1)} = \|f+g\|_p \leq \|f\|_p + \|g\|_p$. ■

Remark 5.3. *1. The Minkowski-Riez Inequality and the fact that for $a \in \mathbb{R}$, $\|af\|_p = |a|\|f\|_p$ and $\|f\|_p \geq 0$ shows that $\|\cdot\|_p$ has almost all of the properties of a norm. The exception is that $\|f\|_p = 0$ does not imply that $f(x) = 0$ for all $x \in \mathbb{X}$. It only implies that $f(x) = 0$ almost everywhere.*

2. $f, g \in \mathcal{L}^p_{\mathbb{R}}(\mathbb{X}, \mathcal{F}, \mu)$ are taken to be equivalent if they differ at most on a set of μ-measure zero (null set), i.e., $f \sim g$ if $\{x : f(x) \neq g(x)\}$ is a null set. Then, for every $f \in \mathcal{L}^p_{\mathbb{R}}(\mathbb{X}, \mathcal{F}, \mu)$ we can define an equivalence class (reflexive, symmetric and transitive) of $\mathcal{L}^p_{\mathbb{R}}$ functions induced by f, which will be denoted by $[f]_p$. The space of all equivalence classes $[f]_p$ of functions $f \in \mathcal{L}^p_{\mathbb{R}}$ is denoted by $L^p_{\mathbb{R}}$ with norm $\|[f]_p\|_p := \inf\{\|g\|_p : g \in \mathcal{L}^p_{\mathbb{R}}$ and $g \sim f\}$. $(L^p, \|f_{[p]}\|_p)$ is a norm vector space and in what follows we will dispense with these technicalities and identify $[f]_p$ with f.

A commonly encountered case, treated in the next theorem, has $p = 2$ and $X, Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ being random variables such that $X, Y \in \mathcal{L}^2_{\mathbb{R}}(\Omega, \mathcal{F}, P)$.

Theorem 5.10. *Let $X, Y : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ be random variables such that $X, Y \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$.*

1. *$XY \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and $|\int_\Omega XY dP| \leq \left(\int_\Omega X^2 dP\right)^{1/2} \left(\int_\Omega Y^2 dP\right)^{1/2}$,*

2. *If $X \in \mathcal{L}^2(\Omega, \mathcal{F}, P)$ then $X \in \mathcal{L}(\Omega, \mathcal{F}, P)$ and $\left(\int_\Omega X dP\right)^2 \leq \int_\Omega X^2 dP$,*

3. *$\mathcal{L}^2(\Omega, \mathcal{F}, P)$ is a vector space.*

Proof. 1. This is just a special case of Hölder's Inequality with $p = q = 2$. 3. follows from the comments after Definition 5.1. 2. Let $X \in \mathcal{L}^2$ and note that $I_\Omega \in \mathcal{L}^2$ with $\int_\Omega I_\Omega dP = \int_\Omega dP$. Then,

$$\left|\int_\Omega X I_\Omega dP\right| \leq \left(\int_\Omega X^2 dP\right)^{1/2} \left(\int_\Omega dP\right)^{1/2}.$$

Since $\int_\Omega dP = 1$, we have

$$\left|\int_\Omega X dP\right| \leq \left(\int_\Omega X^2 dP\right)^{1/2} \text{ or } \left(\int_\Omega X dP\right)^2 \leq \int_\Omega X^2 dP.$$

■

Remark 5.4. *If $X \in \mathcal{L}^2$ we define $V_P(X) = \int_\Omega (X - E_P(X))^2 dP = \int_\Omega X^2 dP - (\int_\Omega X dP)^2$ and call it the variance of X (under P).*

Theorem 5.11. *Let X be a random variable defined on the probability space $(\Omega, \mathcal{F}, P)$ taking values in $(\mathbb{R}, \mathcal{B})$ and $h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ be measurable.*

1. *$f := h \circ X$ is integrable in $(\Omega, \mathcal{F}, P)$ if, and only if, h is integrable in $(\mathbb{R}, \mathcal{B}, P_X)$.*

2. *$E_P(h(X)) := \int_\Omega f dP = \int_\mathbb{R} h dP_X$.*

Proof. First, let h be a non-negative simple function. Then we have that $f(\omega) = \sum_{j=0}^m y_j I_{A_j}(\omega)$ where $A_j \in \mathcal{F}$. Consequently,

$$\begin{aligned} I_P(f) = \int_\Omega f dP &= \sum_{j=0}^m y_j P(A_j) = \sum_{j=0}^m y_j P(X^{-1}(B_j)) \text{ where } B_j = \{x \in \mathbb{R} : h(x) = y_j\} \\ &= \sum_{j=0}^m y_j (P \circ X^{-1})(B_j) = \sum_{j=0}^m y_j P_X(B_j) = \int_\mathbb{R} h dP_X = I_{P_X}(h). \end{aligned}$$

Second, let $h \geq 0$. Then, by Theorem 4.4 there exists a sequence of increasing non-negative simple function ϕ_n such that $\phi_n \to h$ as $n \to \infty$. Hence, if we define $f_n(\omega) = \phi_n(X(\omega)) =$

$(\phi_n \circ X)(\omega)$, it is a sequence of increasing simple function that converges to f.

$$\begin{aligned}\int_\Omega f dP = \int_\Omega (h \circ X) dP &= \int_\Omega \lim_{n\to\infty} (\phi_n \circ X) dP \\ &= \lim_{n\to\infty} \int_\Omega (\phi_n \circ X) dP \text{ by Beppo-Levi's Theorem} \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} \phi_n dP_X \text{ by the first part of the argument for simple functions} \\ &= \int_{\mathbb{R}} h dP_X, \text{ by Beppo-Levi's Theorem.}\end{aligned}$$

This proves 2. for simple and non-negative h. If h takes values in $\mathbb{R}$, consider $|h|$ and let ϕ_n be a sequence of increasing non-negative simple function such that $\phi_n \to |h|$ as $n \to \infty$. Then, we have from above that

$$\int_\Omega |f| dP = \int_{\mathbb{R}} |h| dP_X.$$

But from Remark 4.5, if $|h|$ is integrable in $(\mathbb{R}, \mathcal{B}, P_X)$ then h is integrable in $(\mathbb{R}, \mathcal{B}, P_X)$, establishing 1. Now, for arbitrary h we can prove the rest of part 2 by applying the same arguments to h^+ and h^- and using the fact that $h = h^+ - h^-$. $\blacksquare$

Clearly, taking $h(x) = x$ in the previous theorem gives $E_P(X) := \int_\Omega X dP = \int_{\mathbb{R}} x dP_X(x)$ where in the last integral we emphasize that the "variable" in integration is taking values in $\mathbb{R}$. In this proof, there is no requirement that $P(\Omega) = 1$. Hence, we can take $(\Omega, \mathcal{F}, P)$ to be an arbitrary measure space.

Definition 5.2. *The density of a probability measure P_X associated with a random variable X defined on a probability space $(\Omega, \mathcal{F}, P)$ is a non-negative Borel measurable function f_X that satisfies*

$$P_X((-\infty, a]) = \int_{(-\infty,a]} f_X d\lambda = \int_{\mathbb{R}} I_{(-\infty,a]} f_X d\lambda$$

where λ is Lebesgue measure on $\mathbb{R}$.

Theorem 5.12. *f_X is a density $\iff$ $\int_{\mathbb{R}} f_X d\lambda = 1$, f_X is unique almost everywhere.*

Proof. ($\Longrightarrow$) f_X a density implies $F_X(a) = P_X((-\infty, a]) = \int_{(-\infty,a]} f_X d\lambda$. $\lim_{a\to\infty} P_X((-\infty, a]) = 1 = \lim_{a\to\infty} \int_{(-\infty,a]} f_X d\lambda$, where the first equality follows from Definition 2.4 and continuity of probability measures.

($\Longleftarrow$) Suppose f_X is a non-negative Borel measurable function such that $\int_{\mathbb{R}} f_X d\lambda = 1$. For all $A \in \mathcal{B}$, we put

$$P_X(A) = \int_A f_X d\lambda = \int_{\mathbb{R}} I_A f_X d\lambda.$$

By Theorem 4.10, P_X is a measure on $\mathcal{B}$ with $P_X(\mathbb{R}) = 1$, by assumption. Taking $A = (-\infty, a]$,

$$P_X((-\infty, a]) = \int_{(-\infty,a]} f_X d\lambda$$

and f_X is a density for F_X.

Now, suppose g_X is another density for F_X. Then, $P_X(A) = \int_A g_X d\lambda = \int_{\mathbb{R}} g_X I_A d\lambda$. Let $A_n = \{x : g_X(x) \geq f_X(x) + 1/n\}$. For all $n \in \mathbb{N}$, $\int_{A_n} g_X d\lambda \geq \int_{A_n} (f_X + \frac{1}{n}) d\lambda = \int_{A_n} f_X d\lambda + \frac{1}{n}\lambda(A_n)$. Since $\int_{A_n} f_X d\lambda = \int_{A_n} g_X d\lambda$ it must be that $\lambda(A_n) = 0$.

Note that $A_1 \subset A_2 \subset \cdots$. $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n = A = \{x : g_X(x) > f_X(x)\}$ and $\lambda(A) = \lim_{n\to\infty} \lambda(A_n) = 0$. Similarly, we have $\lambda(B) = 0$ for $B = \{x : g_X(x) < f_X(x)\}$. So, $\lambda(\{x : g_X = f_X\}) = 1$. ∎

Theorem 5.13. *Let* $X : (\Omega, \mathcal{F}, P) \to (\mathbb{R}, \mathcal{B})$ *be a random variable with density* f_X *and* $h : (\mathbb{R}, \mathcal{B}) \to (\mathbb{R}, \mathcal{B})$ *be a measurable function such that* $\int_\Omega |h \circ X| dP < \infty$, *i.e.,* $f = h \circ X$ *is integrable. Then,*

$$\int_\Omega (h \circ X) dP = \int_{\mathbb{R}} h dP_X = \int_{\mathbb{R}} h(x) f_X(x) d\lambda(x)$$

Proof. Homework. ∎