Fundamental Elements of Probability and Asymptotic Theory

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 a certain *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 concept of a probability space, foundational to all subsequent topics in this monograph.

1.1 σ-algebras

A set formed by subsets of a given set $\mathbb{X}$ is called a system of sets associated with $\mathbb{X}$. 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, it is called a σ-algebra.

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. The word "measurable," in this very general setting, suggests that a notion of measure (or size) will be subsequently attached to the sets in $\mathcal{F}$, but for now it is just a label given to the members of $\mathcal{F}$.

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* $\bigcap_{i \in \mathbb{N}} A_i^c \in \mathcal{F}$*. Hence, countable intersections of sets in a* σ*-algebra are measurable sets.*

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* $A_1 \Delta A_2 \in \mathcal{F}$*. Hence,* $\mathcal{F}$ *is closed under set difference and under symmetric difference.*

5. *A system associated with $\mathbb{X}$ is said to be an algebra if properties 1 and 2 in Definition 1.1 hold and if $A_i \in \mathcal{F}$ for $i = 1, \cdots, m$ implies $\bigcup_{i=1}^m A_i \in \mathcal{F}$ with $m \in \mathbb{N}$. Clearly, every σ-algebra is also an algebra.*

We now provide examples of important 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 that it satisfies 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 $A^c = S - A$, i.e., complementation is 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 that it satisfies the properties of Definition 1.1:*

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

   Since $\mathcal{Y}$ *is a* σ*-algebra associated with* $\mathbb{Y}$*,* $\mathbb{Y} \in \mathcal{Y}$*.* $f^{-1}(\mathbb{Y}) = \{x \in \mathbb{X}\text{: } f(x) \in \mathbb{Y}\} = \mathbb{X}$*. 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 $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{I}$. ■

By the fact that $\mathcal{I}$ is an intersection of σ-algebras, $\mathcal{I} \subset \mathcal{F} \;\forall\; \mathcal{F} \in F$, and we can say that $\mathcal{I}$ is the "smallest" σ-algebra in F.

It is often instructive to consider σ-algebras that are obtained from smaller systems associated with $\mathbb{X}$ by expanding these 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 by such a system. This motivates the following definition.

Definition 1.2. *Let $\mathcal{C}$ be a system 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 of Definition 1.2 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 $F = \{\mathcal{F} : \mathcal{F}$ is a σ-algebra associated with $\mathbb{X}$ 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. ■

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 systems 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, $\sigma(\mathcal{C}) = \bigcap_{\mathcal{H} \in \mathcal{F}_\mathcal{C}} \mathcal{H} \subset \bigcap_{\mathcal{G} \in \mathcal{F}_\mathcal{D}} \mathcal{G} = \sigma(\mathcal{D})$. ■

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, we often have $\mathbb{X} = \mathbb{R}^n$ 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 $\sigma(\mathcal{O}_{\mathbb{R}^n})$. The elements of this σ-algebra are called the *Borel* sets of $\mathbb{R}^n$ and $\sigma(\mathcal{O}_{\mathbb{R}^n})$ is called the Borel σ-algebra, which is commonly 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. But first, recall that an open interval on $\mathbb{R}$ is a set $(a, b) := \{x \in \mathbb{R} : a < x < b\}$, a closed interval is a set $[a, b] := \{x \in \mathbb{R} : a \le x \le b\}$ and a half-open interval is a set $[a, b) := \{x \in \mathbb{R} : a \le x < b\}$ or $(a, b] := \{x \in \mathbb{R} : a < x \le b\}$. They are said to be finite if $a, b \in \mathbb{R}$ and infinite if $a = -\infty$ or $b = \infty$.

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 $b_i \le a_i$ 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 $R_{\mathbb{Q}}^{n,o}$ and $R_{\mathbb{Q}}^{n,h}$. 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, $\mathcal{I}_{\mathbb{Q}}^{n,o}$ and $\mathcal{I}_{\mathbb{Q}}^{n,h}$ 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 $\bigcup\limits_{R^{n,o}_{\mathbb{Q}} \subset O} R^{n,o}_{\mathbb{Q}}$. If $x \in \bigcup\limits_{R^{n,o}_{\mathbb{Q}} \subset O} R^{n,o}_{\mathbb{Q}}$ then $x \in R^{n,o}_{\mathbb{Q}} \subset O$. Hence, $\bigcup\limits_{R^{n,o}_{\mathbb{Q}} \subset O} R^{n,o}_{\mathbb{Q}} \subset O$.

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 $m_n = \max\limits_{1 \leq i \leq n} (b_i - a_i)$. 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 $R^{n,o}_{\mathbb{Q}} \subset R^{n,o} \subset B(x, \epsilon)$. Hence, every $x \in O$ belongs to a rectangle $R^{n,o}_{\mathbb{Q}} \subset O$ and, consequently, $x \in \bigcup\limits_{R^{n,o}_{\mathbb{Q}} \subset O} R^{n,o}_{\mathbb{Q}}$. Hence, $O \subset \bigcup\limits_{R^{n,o}_{\mathbb{Q}} \subset O} R^{n,o}_{\mathbb{Q}}$. Combining this set containment with the one in the previous paragraph we $O = \bigcup\limits_{R^{n,o}_{\mathbb{Q}} \subset O} R^{n,o}_{\mathbb{Q}}$.

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

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

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 $\mathcal{A}_1 = \{I : I = (a, b)$ 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 will now 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. For example, if we are interested in measuring the area of two surfaces on a plane that don’t intersect, the area of the two surfaces should be the sum of the areas of each of the surfaces. Similarly, if we are interested on the length of two line segments that don’t intersect, the length of the two segments should be the sum of the length of each segment.

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* $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$ *is a disjoint collection, i.e.,* $A_i \cap A_j = \emptyset \, \forall \, i \neq j$, $\mu\left(\bigcup\limits_{i\in\mathbb{N}} A_i\right) = \sum\limits_{i\in\mathbb{N}} \mu(A_i)$.

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

4. *If we assume that for at least one set* $A \in \mathcal{F}$ *we have* $\mu(A) < \infty$, *then property 1 in Definition 1.5 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 $F \in \mathcal{F}$, $\mu_{\#}(F) \in \{0, \infty, \mathbb{N}\} \subset [0, \infty]$, and since the empty set has no elements $\mu_{\#}(\emptyset) = 0$. For property 2 in Definition 1.5, consider $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$, 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 $\bigcup\limits_{i\in\mathbb{N}} A_i$ has infinitely many elements $\mu_{\#}\left(\bigcup\limits_{i\in\mathbb{N}} A_i\right) = \infty$. Also, $\sum\limits_{i\in\mathbb{N}} \mu(A_i) = \#A_1 + \cdots + \infty + \cdots = \infty$; 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, $\mu_{\#}\left(\bigcup\limits_{i\in\mathbb{N}} A_i\right) = \mu_{\#}(A_1 \cup \cdots \cup A_N) = \sum_{i=1}^{N} \mu_{\#}(A_i) = \sum_{i=1}^{\infty} \mu_{\#}(A_i)$; 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) $\mu_{\#}\left(\bigcup\limits_{i\in\mathbb{N}} A_i\right) = \infty$ and $\sum\limits_{i\in\mathbb{N}} \mu(A_i) = \#A_1 + \#A_2 + \cdots = \infty$.

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

$\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$, *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 $A \in 2^\Omega$ *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* $\{A_i\}_{i\in\mathbb{N}} \subset \mathcal{F}$*. 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. $\mu\left(\bigcup\limits_{i\in\mathbb{N}} A_i\right) \leq \sum\limits_{i\in\mathbb{N}} \mu(A_i)$ *(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, $\bigcup\limits_{i\in\mathbb{N}} A_i = \bigcup\limits_{i\in\mathbb{N}} B_i$, $\mu\left(\bigcup\limits_{i\in\mathbb{N}} A_i\right) = \mu\left(\bigcup\limits_{i\in\mathbb{N}} B_i\right) = \sum_{i\in\mathbb{N}} \mu(B_i) \leq \sum_{i\in\mathbb{N}} \mu(A_i)$. ■

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 $\{A_i\}_{i\in\mathbb{N}}$ be a sequence of sets.*

1. *If $A_1 \subset A_2 \subset A_3 \subset \cdots$ then $\lim_{i\to\infty} A_i := \bigcup_{i\in\mathbb{N}} A_i$,*
2. *if $A_1 \supset A_2 \supset A_3 \supset \cdots$ then $\lim_{i\to\infty} A_i := \bigcap_{i\in\mathbb{N}} A_n$,*
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 $\{A_i\}_{i\in\mathbb{N}}$ and denoted by $\liminf_{i\to\infty} A_i$ and C is called the limit superior of $\{A_i\}_{i\in\mathbb{N}}$ and denoted by $\limsup_{i\to\infty} A_i$.*

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 $\{A_n\}_{n\in\mathbb{N}}$, consider complements relative to A_1 putting $A_n^c := A_1 - A_n$, and note that $A_1^c \subset A_2^c \subset A_3^c \subset \cdots$. Since $A = \bigcap_{n\in\mathbb{N}} A_n$, by

de Morgan's Laws $A_1 - A = A^c = \bigcup_{n\in\mathbb{N}} A_n^c$ and, consequently, $\mu(A_1 - A) = \mu\left(\bigcup_{n\in\mathbb{N}} A_n^c\right) = \mu\left(\lim_{n\to\infty} A_n^c\right) = \lim_{n\to\infty} \mu(A_n^c) = \lim_{n\to\infty} \mu(A_1 - A_n)$, where the next to 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

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

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 $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ 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. ($\implies$) 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.

($\Longleftarrow$) 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 $A := \lim_{n\to\infty} A_n = \bigcup_{n\in\mathbb{N}} A_n = \bigcup_{j\in\mathbb{N}} B_j$. 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 $m(A_n^c) = m(\mathbb{X} - A_n) = m\left(\mathbb{X} - \bigcup_{j=1}^n B_j\right)$.*

$$\begin{aligned}\lim_{n\to\infty} m(A_n^c) &= m\left(\mathbb{X} - \bigcup_{j=1}^n B_j\right) = m(\mathbb{X}) - \lim_{n\to\infty} m\left(\bigcup_{j=1}^n B_j\right) \quad \textit{since } m(\mathbb{X}) < \infty \\ &= m(\mathbb{X}) - \lim_{n\to\infty}\sum_{j=1}^n m(B_j) = m(\mathbb{X}) - \sum_{j=1}^\infty m(B_j) \quad \textit{by additivity of } m. \qquad (1.1)\end{aligned}$$

Now,

$$\begin{aligned}\lim_{n\to\infty} m(A_n^c) &= m\left(\bigcap_{j\in\mathbb{N}} A_j^c\right) \quad \textit{by continuity of } m \textit{ from above} \\ &= m\left(\left(\bigcup_{j\in\mathbb{N}} A_j\right)^c\right) \quad \textit{by de Morgan's Laws} \\ &= m(\mathbb{X}) - m\left(\bigcup_{j\in\mathbb{N}} A_j\right) = m(\mathbb{X}) - m\left(\bigcup_{j\in\mathbb{N}} B_j\right). \qquad (1.2)\end{aligned}$$

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 $\emptyset$ means that if $A_1 \supset A_2 \supset \cdots$ and $\lim_{n\to\infty} A_n = \emptyset$ with $\mu(A_1) < \infty$, then $\lim_{n\to\infty} \mu(A_n) = \mu(\emptyset) = 0$.

Since probability measures are finite, Theorem 1.11 and Remark 1.4 provide characterizations for probability measures. 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* $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ *with* $A_1 \subset A_2 \subset \cdots$ *and* $A = \lim_{n\to\infty} A_n$ *we have*

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

Condition 3 can be substituted by either

3'. *for* $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ *with* $A_1 \supset A_2 \supset \cdots$ *and* $A = \lim_{n\to\infty} A_n$ *we have* $P(A) = \lim_{n\to\infty} P(A_n)$ *or*

3". *for* $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$ *with* $A_1 \supset A_2 \supset \cdots$ *and* $\lim_{n\to\infty} A_n = \emptyset$ *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 $\{A_i\}_{i=1}^n \subset \mathcal{F}$ *for* $n \in \mathbb{N}$ *then*

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

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

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

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

$$\sum_{n=1}^{n-1} P(C_i) = \sum_{i=1}^{n-1} P(A_i \cap A_n)$$

$$\sum_{1 \le i_1 < i_2 \le n-1} P(C_{i_1} \cap C_{i_2}) = \sum_{1 \le i_1 < i_2 \le n-1} P(A_{i_1} \cap A_n \cap A_{i_2} \cap A_n)$$

$$= \sum_{1 \le i_1 < i_2 \le n-1} P(A_{i_1} \cap A_{i_2} \cap A_n)$$

$$\sum_{1 \le i_1 < i_2 < i_3 \le n-1} P(C_{i_1} \cap C_{i_2} \cap C_{i_3}) = \sum_{1 \le i_1 < i_3 < i_3 \le n-1} P(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap A_n)$$

$$\vdots$$

$$P(C_1 \cap C_2 \cap \cdots \cap C_{n-1}) = P(A_1 \cap \cdots \cap A_n).$$

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, $\{A_n\}_{n \in \mathbb{N}} \subset \mathcal{F}$ and suppose $A = \lim_{n \to \infty} A_n$ 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, $P\left(\bigcup\limits_{i\in\mathbb{N}} C_i^c\right) = \sum\limits_{i\in\mathbb{N}} P(\theta_i) = \lim\limits_{n\to\infty} \sum_{i=1}^n P(\theta_i)$. But $\sum_{i=1}^n P(\theta_i) = P(C_n^c)$ and $P(C_n^c) = 1 - P(C_n)$. Hence, $P\left(\bigcup\limits_{i\in\mathbb{N}} C_i^c\right) = \lim\limits_{n\to\infty}(1 - P(C_n)) = 1 - \lim\limits_{n\to\infty} P(C_n)$. Consequently, $P(C) = 1 - \left(1 - \lim\limits_{n\to\infty} P(C_n)\right) = \lim\limits_{n\to\infty} P(C_n)$.

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

1.4 Null sets and complete measure spaces

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 $0 \le \mu(M) \le \mu(N) = 0$. Hence, $M \in \mathcal{N}_\mu$. In addition, if $\{N_j\}_{j\in\mathbb{N}} \subset \mathcal{N}_\mu$, by sub-additivity of measures $0 \le \mu\left(\bigcup\limits_{j\in\mathbb{N}} N_j\right) \le \sum\limits_{j\in\mathbb{N}} \mu(N_j) = 0$. Hence, $\bigcup\limits_{j\in\mathbb{N}} N_j \in \mathcal{N}_\mu$.

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 is an element 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. First, 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. $\{A_j\}_{j \in \mathbb{N}} \subset \bar{\mathcal{F}} \implies \bigcup_{j \in \mathbb{N}} A_j \in \bar{\mathcal{F}}$: 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, $\bigcup_{j \in \mathbb{N}} F_j \in \mathcal{F}$ and $\bigcup_{j \in \mathbb{N}} M_j \subset \bigcup_{j \in \mathbb{N}} N_j$ where $N_j \in \mathcal{N}_\mu$. Hence, $\bigcup_{j \in \mathbb{N}} N_j \in \mathcal{N}_\mu$ and $\bigcup_{j \in \mathbb{N}} M_j \in \mathcal{S}$. Then, by definition $\bigcup_{j \in \mathbb{N}} A_j \in \bar{\mathcal{F}}$.

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,

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

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 $\{A_j\}_{j \in \mathbb{N}} \subset \bar{\mathcal{F}}$ 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.

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

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 show 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 $A \in \bar{\mathcal{F}}$.■

1.5 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 the 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 $(\Omega, \mathcal{F}, P)$, given a certain event E. It is called conditional probability on E. 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 $P(A|E) = P(A)$, then 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* $\{E_j\}_{1\le j\le n} \subset \mathcal{F}$*. If* $P\left(\bigcap_{1\le j\le n-1} E_j\right) >$ 0 *then*

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

Proof. Note that if $P\left(\bigcap_{1\le j\le n-1} E_j\right) > 0$ then $P\left(\bigcap_{1\le j\le m} E_j\right) > 0$ 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

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

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* $\bigcup_{i\in\mathbb{N}} E_i = \Omega$ *and* $E_i \cap E_j = \emptyset$*, for all* $i \ne j$*.*

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. $A = A \cap \Omega = A \cap \left(\bigcup_{i \in \mathbb{N}} E_i \right) = \bigcup_{i \in \mathbb{N}} (A \cap E_i)$. 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 $P(F|G)$.

1.6 Exercises

1. Let $f : \mathbb{N} \times \mathbb{N} \to \mathbb{R}$ be a double sequence with typical value given by $f(m, n)$. Assume that

   (a) for every $n \in \mathbb{N}$, $f(m_1, n) \leq f(m_2, n)$ whenever $m_1 \leq m_2$,

   (b) for every $m \in \mathbb{N}$, $f(m, n_1) \leq f(m, n_2)$ whenever $n_1 \leq n_2$.

   Show that $\lim_{n\to\infty} \left(\lim_{m\to\infty} f(m, n) \right) = \lim_{m\to\infty} \left(\lim_{n\to\infty} f(m, n) \right) = \lim_{n\to\infty} f(n, n)$.

   As a corollary, show that if $f(m, n) \geq 0$ then $\sum_{n\in\mathbb{N}} \sum_{m\in\mathbb{N}} f(m, n) = \sum_{m\in\mathbb{N}} \sum_{n\in\mathbb{N}} f(m, n)$.

2. Let $\mathbb{X}$ be an arbitrary set and consider the collection of all subsets of $\mathbb{X}$ that are countable or have countable complements. Show that this collection is a σ-algebra. Use this fact to obtain the σ-algebra generated by $\mathcal{C} = \{\{x\} : x \in \mathbb{R}\}$.

3. Denote by $B(x, r)$ an open ball in $\mathbb{R}^n$ centered at x and with radius r. Show that the Borel sets are generated by the collection $B = \{B_r(x) : x \in \mathbb{R}^n,\ r > 0\}$.

4. Let $(\Omega, \mathcal{F})$ be a measurable space. Show that: a) if μ_1 and μ_2 are measures on $(\Omega, \mathcal{F})$, then $\mu_c(F) := c_1\mu_1(F) + c_2\mu_2(F)$ for $F \in \mathcal{F}$ and all $c_1, c_2 \geq 0$ is a measure; b) if $\{\mu_i\}_{i\in\mathbb{N}}$ are measures on $(\Omega, \mathcal{F})$ and $\{\alpha_i\}_{i\in\mathbb{N}}$ is a sequence of positive numbers, then $\mu_\infty(F) = \sum_{i\in\mathbb{N}} \alpha_i\mu_i(F)$ for $F \in \mathcal{F}$ is a measure.

5. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space and $\mathcal{G} \subset \mathcal{F}$ be a σ-algebra. In this case, we call $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Let $\nu := \mu|_{\mathcal{G}}$ be the restriction of μ to $\mathcal{G}$. That is, $\nu(G) = \mu(G)$ for all $G \in \mathcal{G}$. Is ν a measure? If μ is finite, is ν finite? If μ is a probability, is ν a probability?

6. Show that a measure space $(\Omega, \mathcal{F}, \mu)$ is σ-finite if, and only if, there exists $\{F_n\}_{n\in\mathbb{N}} \in \mathcal{F}$ such that $\cup_{n\in\mathbb{N}} F_n = \Omega$ and $\mu(F_n) < \infty$ for all n.

7. Let $(\Omega, \mathcal{F}, P)$ be a probability space and $\{E_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$. Show that if $\sum_{n=1}^{\infty} P(E_n) < \infty$ then $P\left(\limsup_{n\to\infty} E_n\right) = 0$.

8. Let $\{E_j\}_{j\in J}$ be a collection of pairwise disjoint events. Show that if $P(E_j) > 0$ for each $j \in J$, then J is countable.

9. Consider the extended real line, i.e., $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{\infty\}$. Let $\bar{\mathcal{B}} := \mathcal{B}(\bar{\mathbb{R}})$ be defined as the collection of sets $\bar{B}$ such that $\bar{B} = B \cup S$ where $B \in \mathcal{B}(\mathbb{R})$ and $S \in \{\emptyset, \{-\infty\}, \{\infty\}, \{-\infty, \infty\}\}$. Show that $\bar{\mathcal{B}}$ is a σ-algebra and that it is generated by a collection of sets of the form $[a, \infty]$ where $a \in \mathbb{R}$.

10. If $E_1, E_2, \cdots, E_n$ are independent events, show that the probability that none of them occur is less than or equal to $\exp\left(-\sum_{i=1}^{n} P(E_i)\right)$.

11. Let $\{A_n\}_{n\in\mathbb{N}}$ and $\{B_n\}_{n\in\mathbb{N}}$ be events (measurable sets) in a probability space with measure P with $\lim A_n = A$, $\lim B_n = B$, $P(B_n), P(B) > 0$ for all n. Show that $P(A_n|B) \to P(A|B)$, $P(A|B_n) \to P(A|B)$, $P(A_n|B_n) \to P(A|B)$ as $n \to \infty$.

12. Let $(\mathbb{X}, \bar{\mathcal{F}}, \bar{\mu})$ be the measure space defined in Theorem 1.15 and $\mathcal{C} = \{G \in \mathbb{X} : \exists\, A,\ B \in \mathcal{F} \ni A \subset G \subset B \text{ and } \mu(B - A) = 0\}$. Show that $\bar{\mathcal{F}} = \mathcal{C}$.