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

August 22, 2022

Chapter 1

Probability spaces

1.1 σ-algebras

We begin by defining σ-algebras and investigating some of their properties.

Definition 1.1. *Let $\mathbb{X}$ be an arbitrary set. A σ-algebra (or σ-field) is a collection of subsets $\mathcal{F}$ 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* $i \in \mathbb{N} \implies \underset{i\in\mathbb{N}}{\cup} A_i \in \mathcal{F}$.

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

Remark 1.1. 1. *Since $\mathbb{X} \in \mathcal{F}$, by property 2 we conclude that the empty set $\emptyset \in \mathcal{F}$.*

2. *By de Morgan's Laws $\left(\underset{i\in\mathbb{N}}{\cup} A_i\right)^c = \underset{i\in\mathbb{N}}{\cap} A_i^c$ and by properties 2 and 3, if $A_i \in \mathcal{F}$ for $i \in \mathbb{N}$, $A_i^c \in \mathcal{F}$ and $\cap_{i=1}^{\infty} A_i^c \in \mathcal{F}$.*

3. *Given Remarks 1.1.1, 1.1.2 and Definition 1.1, we say that $\mathcal{F}$ is "closed" under countable unions, intersections and complementation.*

4. *A collection 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* $i = 1, \cdots m$ *implies* $\cup_{i=1}^{m} A_i \in \mathcal{F}$ *with* $m \in \mathbb{N}$.

5. *If* $A_1, A_2 \in \mathcal{F}$ *then* $A_2 - A_1 \in \mathcal{F}$. *This follows from the fact that* $A_2 - A_1 = A_2 \cap A_1^c$.

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* $S \subset \mathbb{X}$ *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* $\mathbb{X} \in \mathcal{F}, S \cap \mathbb{X} = S \in \mathcal{F}_S$.

   2. $A \in \mathcal{F}_S \implies A^c \in \mathcal{F}_S$ *(note that* $A^c = S - A$, *complementation relative to* S*).* $A \in \mathcal{F}_S \implies \exists\, F \in \mathcal{F}$ *such that* $A = S \cap F \in \mathcal{F}_S$. *Since* $F \in \mathcal{F}$, $F^c \in \mathcal{F}$ *and* $S \cap F^c \in \mathcal{F}_S$. *Furthermore,* $(S \cap F) \cup (S \cap F^c) = S$ *or* $A \cup (S \cap F^c) = S$. *But by definition,* $A \cup A^c = S$, *so* $A^c = S \cap F^c \in \mathcal{F}_S$.

   3. *If* $A_i \in \mathcal{F}_S$ *for* $i \in \mathbb{N}$, $U = \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}_S$. $A_i \in \mathcal{F}_S \implies \exists F_i \in \mathcal{F} \ni A_i = S \cap F_i$. $U = \cup_{i \in \mathbb{N}}(S \cap F_i) = S \cap \cup_{i \in \mathbb{N}} F_i$, *but* $\cup_{i \in \mathbb{N}} F_i \in \mathcal{F}$, *so* $U \in \mathcal{F}_S$.

4. *Let* $f : \mathbb{X} \to \mathbb{Y}$ *and* $\mathcal{Y}$ *be a* σ*-algebra associated with* $\mathbb{Y}$. *Then* $\mathcal{F} := f^{-1}(\mathcal{Y}) = \{f^{-1}(S) : S \in \mathcal{Y}\}$ *is a* σ*-algebra associated with* $\mathbb{X}$. *We need to verify that:*

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

2. *If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$. $A \in \mathcal{F} \implies \exists S_A \in \mathcal{Y} \ni A = f^{-1}(S_A)$. Now, $S_A \in \mathcal{Y} \implies \mathbb{Y} - S_A \in \mathcal{Y}$ 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. *If $A_i \in \mathcal{F}$ for $i \in \mathbb{N}$, $\cup_{i\in\mathbb{N}} A_i \in \mathcal{F}$. $A_i \in \mathcal{F} \implies \exists\, S_{A_i} \in \mathcal{Y} \ni A_i = f^{-1}(S_{A_i})$. Now, $S_{A_i} \in \mathcal{Y}$, $\forall i \in \mathbb{N} \implies \cup_{i\in\mathbb{N}} S_{A_i} \in \mathcal{Y}$ and $f^{-1}(\cup_{i\in\mathbb{N}} S_{A_i}) = \cup_{i\in\mathbb{N}} f^{-1}(S_{A_i}) = \cup_{i\in\mathbb{N}} A_i \in \mathcal{F}$.*

$\mathcal{F}$ is called the inverse image σ-algebra.

These examples demonstrate that multiple σ-algebras can be associated with a set. The following theorem shows that the intersection of an arbitrary collection of σ-algebras 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 $\mathcal{I} := \underset{\mathcal{F}\in F}{\cap} \mathcal{F}$ is a σ-algebra.*

Proof. 1. Since $\mathbb{X} \in \mathcal{F}$ for all $\mathcal{F} \in F$ then $\mathbb{X} \in \underset{\mathcal{F}\in F}{\cap} \mathcal{F}$. 2. $A \in \mathcal{I} \implies A \in \mathcal{F}$ for all $\mathcal{F} \in F$. Then, $A^c \in \mathcal{F}$ for all $\mathcal{F} \in F$. Consequently, $A^c \in \mathcal{I}$. 3. Let $A_i \in \mathcal{I}$ for $i \in \mathbb{N}$. Then, $A_i \in \mathcal{F}$ for all $\mathcal{F} \in F$. Hence, $\underset{i\in\mathbb{N}}{\cup} A_i \in \mathcal{F}$ for all $\mathcal{F} \in F$, which implies $\underset{i\in\mathbb{N}}{\cup} A_i \in \mathcal{I}$. ■

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

1. *$\mathcal{B} \subseteq \sigma(\mathcal{B})$*

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

The defining properties of $\sigma(\mathcal{B})$ allow us to view it as the smallest σ-algebra containing $\mathcal{B}$ as is made explicit in the next theorem.

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

Proof. Let $F = \{\mathcal{F} : \mathcal{F}$ is a σ-algebra associated with $\mathbb{X}$ and $\mathcal{B} \subseteq \mathcal{F}\}$ be the set of all σ-algebras containing $\mathcal{B}$. $F \neq \emptyset$ since $2^{\mathbb{X}}$ is a σ-algebra. By Theorem 1.1, $\underset{\mathcal{F}\in F}{\cap} \mathcal{F}$ is a σ-algebra.

Since $\mathcal{B}$ is in all $\mathcal{F}$, $\mathcal{B} \in \bigcap_{\mathcal{F}\in F} \mathcal{F}$. Thus, $\bigcap_{\mathcal{F}\in F} \mathcal{F} \in F$, but by construction it is the smallest σ-algebra in F, since all others contain or are equal to $\bigcap_{\mathcal{F}\in F} \mathcal{F}$. ■

The generation of the smallest σ-algebra associated with a collection of subsets $\mathcal{B}$ of $\mathbb{X}$ is monotonic as demonstrated in the following theorem.

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

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

Remark 1.2. 1. *A topology of $\mathbb{X}$, denoted by $\mathcal{O}_{\mathbb{X}}$, is a collection of subsets of $\mathbb{X}$ that satisfies the following properties:*

   (a) $\mathbb{X}, \emptyset \in \mathcal{O}_{\mathbb{X}}$

   (b) $O_i \in \mathcal{O}_{\mathbb{X}}$ *for* $i = 1, \cdots, n$ *and* $n \in \mathbb{N} \implies \cap_{i=1}^{n} O_i \in \mathcal{O}_{\mathbb{X}}$

   (c) $O_i \in \mathcal{O}_{\mathbb{X}}$ *for* $i \in I$ *(an arbitrary index set)* $\implies \bigcup_{i\in I} O_i \in \mathcal{O}_{\mathbb{X}}$.

   The elements of $\mathcal{O}_{\mathbb{X}}$ are called the open sets of $\mathbb{X}$ and the pair $(\mathbb{X}, \mathcal{O}_{\mathbb{X}})$ is called a topological space. The σ-algebra generated by the open sets $\sigma(\mathcal{O}_{\mathbb{X}})$ is called the Borel σ-algebra associated with $\mathbb{X}$. The elements of $\sigma(\mathcal{O}_{\mathbb{X}})$ are called the Borel sets of $\mathbb{X}$.

2. *If we define a metric*[1] $d_{\mathbb{X}}$ *on* $\mathbb{X}$ *we say that*

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

[1] A metric on $\mathbb{X}$ is a function $d_{\mathbb{X}} : \mathbb{X} \times \mathbb{X} \to \mathbb{R}$ such that for all $x, y, z \in \mathbb{X}$ it satisfies a) $d_{\mathbb{X}}(x,y) \geq 0$, $d_{\mathbb{X}}(x,y) = 0$ if, and only if, $x = y$; b) $d_{\mathbb{X}}(x,y) = d_{\mathbb{X}}(y,x)$; c) $d_{\mathbb{X}}(x,z) \leq d_{\mathbb{X}}(x,y) + d_{\mathbb{X}}(y,z)$.

where $B(x,\epsilon) = \{y \in \mathbb{X} : d_X(x,y) < \epsilon\}$. $\mathcal{O}_{\mathbb{X}}$ is the collection of open sets of $\mathbb{X}$. When $\mathbb{X} = \mathbb{R}^n$ an usual choice of metric is $d_{\mathbb{R}^n}(x,y) = \left(\sum_{i=1}^n (x_i - y_i)^2\right)^{1/2}$.

Theorem 1.4. *Let $S \subset \mathbb{X}$, $\mathcal{F} = \sigma(\mathcal{C})$ where $\mathcal{C}$ is a collection of subsets of $\mathbb{X}$. Define $\mathcal{C} \cap S = \{A \cap S : A \in \mathcal{C}\}$. Then,*

$$\sigma(\mathcal{C} \cap S) = \sigma(\mathcal{C}) \cap S \text{ is a } \sigma\text{-algebra associated with } S.$$

Proof. First, note that since $\mathcal{C} \subseteq \sigma(\mathcal{C})$ we have $\mathcal{C} \cap S \subseteq \sigma(\mathcal{C}) \cap S$. From Example 1.1.3, $\sigma(\mathcal{C}) \cap S$ is a σ-algebra associated with S. Then, it follows from Theorem 1.3 that $\sigma(\mathcal{C} \cap S) \subseteq \sigma(\mathcal{C}) \cap S$.

Now, we need only show that $\sigma(\mathcal{C} \cap S) \supseteq \sigma(\mathcal{C}) \cap S$ 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. $\mathcal{G} = \{A \subseteq \mathbb{X} : A \cap S \in \sigma(\mathcal{C} \cap S)\}$.

By construction, $\mathcal{C} \subset \mathcal{G}$ since $A \in \mathcal{C} \implies A \cap S \in \mathcal{C} \cap S \subseteq \sigma(\mathcal{C} \cap S)$. Thus, $A \in \mathcal{G}$. We will show that $\mathcal{G}$ is a σ-algebra. 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 $\sigma(\mathcal{C}) \cap S \subseteq \sigma(\mathcal{C} \cap S)$.

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 $A_n \in \mathcal{G}, n \in \mathbb{N}$ and note that

$$\left(\underset{n \in \mathbb{N}}{\cup} A_n\right) \cap S = \underset{n \in \mathbb{N}}{\cup} (A_n \cap S).$$

   Since, $A_n \cap S \in \sigma(\mathcal{C} \cap S)$, $\underset{n \in \mathbb{N}}{\cup} (A_n \cap S) \in \sigma(\mathcal{C} \cap S)$ and $\underset{n \in \mathbb{N}}{\cup} A_n \in \mathcal{G}$.

Thus, $\mathcal{G}$ is a σ-algebra. ■

1.2 Measure spaces

We start by defining a measure on a measurable space.

Definition 1.3. *Given a measurable space* $(\mathbb{X}, \mathcal{F})$ *a (positive) measure* $\mu : \mathcal{F} \to [0, \infty]$ *is a function having the following properties:*

1. $\mu(\emptyset) = 0$

2. *For all* $A_j \in \mathcal{F}$, $j = 1, 2, \ldots$ *with* $A_j \cap A_i = \emptyset$ *if* $i \neq j$,

$$\mu\left(\cup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(A_j).$$

The triple $(\mathbb{X}, \mathcal{F}, \mu)$ *is called a measure space.*

Remark 1.3.

1. *Property 2 in Definition 1.3 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* $A_1 \subseteq A_2 \subseteq \cdots$ *is said to be exhausting if* $\lim_{j \to \infty} A_j := \cup_{j=1}^{\infty} A_j = \mathbb{X}$. *A measure* μ *is called* σ*-finite if there is an exhausting sequence* $\{A_1, A_2, \cdots\} \in \mathcal{F}$ *such that* $\mu(A_j) < \infty$ *for all* j.

4. *If* μ *satisfies properties 1 and 2 but* $\mathcal{F}$ *is not a* σ*-algebra,* μ *is called a pre-measure. But in this case, property 2 requires that* $\cup_{j=1}^{\infty} A_j \in \mathcal{F}$.

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

Definition 1.4. *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 (probability) spaces can be difficult as we will soon discover. What follows are very simple examples of measure and 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 if F has finitely many elements. μ is called the counting measure.*

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.

3. Let $\Omega = \{\omega_1, \omega_2, \cdots\}$ and $p_i \in [0, 1]$ for $i = 1, 2, \cdots$ such that $\sum_{i=1}^{\infty} p_i = 1$. Let $(\Omega, 2^{\Omega})$ be a measurable space, then the set function

$$P(A) = \sum_{i:\omega_i \in A} p_i = \sum_{i=1}^{\infty} p_i \mu_{\omega_i}(A), \;\; A \subseteq \Omega$$

is a probability measure.

1.2.1 Some properties of measures

Theorem 1.5. *For $A, A_1, A_2, \cdots \in \mathcal{F}$ we have for any measure μ,*

1. $A \subseteq A_1 \implies \mu(A) \leq \mu(A_1)$ *(monotonicity)*

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

3. $\mu(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} \mu(A_i)$ *(subadditivity)*

Proof. 1. A and $A_1 - A$ are disjoint sets. Hence, $\mu(A \cup (A_1 - A)) = \mu(A) + \mu(A_1 - A) = \mu(A_1)$, which implies $\mu(A) \leq \mu(A_1)$.

2. $A \cup A_1 = A \cup (A_1 - A)$ and $A_1 = (A \cap A_1) \cup (A_1 - A)$. So, by the second equality, given

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

3. Let $B_1 = A_1$, $B_2 = A_2 - A_1$, $B_3 = A_3 - \cup_{j=1}^{2} A_j$, ... Hence, $\{B_j\}$ is a disjoint collection and $B_j \subseteq A_j$, $\mu\left(\cup_{j=1}^{\infty} B_j = \cup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} \mu(B_j) \leq \sum_{j=1}^{\infty} \mu(A_j)$. ■

An important property of any probability measure is continuity. But since probabilities are set functions, it is useful to define what we mean by limits of sets.

Definition 1.5. *Let $\{A_i\}_{i \in \mathbb{N}}$ be a collection of sets. We have,*

1. *If $A_1 \subseteq A_2 \subseteq A_3 \ldots$, $\lim\limits_{n \to \infty} A_n := \cup_{i=1}^{\infty} A_i$*

2. *If $A_1 \supseteq A_2 \supseteq A_3 \ldots$, $\lim\limits_{n \to \infty} A_n := \cap_{i=1}^{\infty} A_i$*

3. *If $A_1, A_2, \ldots$ is an arbitrary sequence of sets, let $B_n = \cap_{m=n}^{\infty} A_m$ $(B_1 \subseteq B_2 \subseteq \ldots)$ and $C_n = \cup_{m=n}^{\infty} A_m$ $(C_1 \supseteq C_2 \supseteq \ldots)$. Then, put $B = \lim\limits_{n \to \infty} B_n = \cup_{n=1}^{\infty} \cap_{m=n}^{\infty} A_m$ and $C = \lim\limits_{n \to \infty} C_n = \cap_{n=1}^{\infty} \cup_{m=n}^{\infty} A_m$. We say that $A = \lim\limits_{n \to \infty} A_n$ exists if $B = C$, and we write $A = B = C$. B is called the limit inferior of $\{A_1, A_2, \ldots\}$ or $\liminf\limits_{n \to \infty} A_n$ and C is called the limit superior of $\{A_1, A_2, \ldots\}$ or $\limsup\limits_{n \to \infty} A_n$.*

Theorem 1.6. *Let $(\mathbb{X}, \mathcal{F})$ be a measurable space. A function $\mu : \mathcal{F} \to [0, \infty]$ is a measure if, and only if,*

a) $\mu(\emptyset) = 0$

b) If $A_1, A_2 \in \mathcal{F}$ are disjoint $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$

c) If $A_1, A_2, \cdots \in \mathcal{F}$ and $A_1 \subseteq A_2 \subseteq \cdots$ with $\lim_{n \to \infty} A_n = A \in \mathcal{F}$ we have

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

Proof. ($\Longrightarrow$) If μ is a measure then properties a) and b) follow directly from properties 1) and 2) from the definition of measure. Now, for property c) define $B_1 := A_1$, $B_2 := A_2 - A_1, \cdots$. From Remark 1.1.5 we have that $B_j \in \mathcal{F}$ and the collection $\{B_j\}_{j=1}^{\infty}$ is pairwise disjoint. Put $A_n = \cup_{j=1}^{n} B_j$ and $\cup_{n=1}^{\infty} A_n = \cup_{j=1}^{\infty} B_j = A$. The last equality results from the definition of the limit of a sequence of non-decreasing sets. Then, by σ-additivity of μ

$$\begin{aligned}\mu(A) &= \mu\left(\cup_{j=1}^{\infty} B_j\right) = \sum_{j=1}^{\infty} \mu(B_j) = \lim_{n\to\infty} \sum_{j=1}^{n} \mu(B_j) = \lim_{n\to\infty} \mu(B_1 \cup B_2 \cup \cdots \cup B_n) \\ &= \lim_{n\to\infty} \mu(A_n).\end{aligned}$$

($\Longleftarrow$) Now, assume that $\mu : \mathcal{F} \to [0, \infty]$ satisfies a)-c). We will show that in this case μ will satisfy properties 1) and 2) from the definition of measure. Let $\{B_j\}_{j=1}^{\infty}$ be a pairwise disjoint sequence in $\mathcal{F}$ and define $A_n := \cup_{j=1}^{n} B_j$ and $A := \cup_{n=1}^{\infty} A_n = \cup_{j=1}^{\infty} B_j$. Clearly, $A_1 \subseteq A_2 \subseteq \cdots$ and $\lim_{n\to\infty} A_n = A$. Using b), we have

$$\mu(A_1) = \mu(B_1),\ \mu(A_2) = \mu(B_1) + \mu(B_2),\ \cdots,\ \mu(A_n) = \sum_{j=1}^{n} \mu(B_j).$$

From c) we conclude (second equality) that

$$\mu\left(\cup_{j=1}^{\infty} B_j\right) = \mu(A) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \left(\sum_{j=1}^{n} \mu(B_j) \right) = \sum_{j=1}^{\infty} \mu(B_j).$$

■

Remark 1.4. *If μ is a finite measure, c) in Theorem 1.6 can be replaced by either of these two equivalent conditions*

c') If $A_1, A_2, \cdots \in \mathcal{F}$ and $A_1 \supseteq A_2 \supseteq \cdots$ with $\lim_{n\to\infty} A_n = A \in \mathcal{F}$ we have

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

or

c") If $A_1, A_2, \cdots \in \mathcal{F}$ and $A_1 \supseteq A_2 \supseteq \cdots$ with $\lim_{n\to\infty} A_n = \emptyset \in \mathcal{F}$ we have

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

Theorem 1.7. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Then,*

1. $P(A^c) = 1 - P(A)$ *for all* $A \in \mathcal{F}$

2. $A, B \in \mathcal{F}$, $A \subseteq B \implies P(A) \leq P(B)$ *for all* $A, B \in \mathcal{F}$

3. *If* $A_i \in \mathcal{F}$, $i = 1, 2, \ldots,$ *then*

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{1\leq i_1<i_2\leq n} P(A_{i_1} \cap A_{i_2}) + \sum_{1\leq i_1<i_2<i_3\leq n} P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) + \cdots + (-1)^{n+1} P(\cap_{i=1}^n A_i)$$

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 1.5.1. 3. Let $n = 2$. Then, from Theorem 1.5.2 we have

$$P(A_1 \cup A_2) \;=\; P(A_1) + P(A_2) - P(A_1 \cap A_2).$$

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

$$P(B_{n-1}) = P(\cup_{i=1}^{n-1} A_i) = \sum_{i=1}^{n-1} P(A_i) - \sum_{1\leq i<j\leq n-1} P(A_i \cap A_j) + \sum_{1\leq i<j<k\leq n-1} P(A_i \cap A_j \cap A_k) + \cdots + (-1)^n P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}).$$

We will show that this representation holds for n. From the case where there are only two sets

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

But,

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

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

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

■

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

Theorem 1.8. *Suppose $\{A_n\}_{n=1}^{\infty} \in \mathcal{F}$, where $(\Omega, \mathcal{F}, P)$ is a probability space. Let $A = \lim_{n\to\infty} A_n$. Then, $A \in \mathcal{F}$ and $P(A_n) \to P(A)$ as $n \to \infty$.*

Proof. Since $\{A_n\}_{n=1,2,\ldots}$ has a limit, there exist $C_1 \supseteq C_2 \supseteq C_3 \supseteq \ldots$ and $B_1 \subseteq B_2 \subseteq B_3 \subseteq \ldots$ as in Definition 1.5, such that $B = \cup_{i=1}^{\infty} B_i = \cap_{i=1}^{\infty} C_i = C = A$. By construction, $B = B_1 \cup (B_2 - B_1) \cup (B_3 - B_2) \cup \cdots = \chi_1 \cup \chi_2 \cup \ldots$. The collection $\{\chi_1, \chi_2, \ldots\}$ is pairwise disjoint. By σ-additivity of measures we have $P(B) = \sum_{i=1}^{\infty} P(\chi_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(\chi_i)$. 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, $P(B) = \lim_{n\to\infty} P(B_n)$.

By De Morgan's Laws $C = \cap_{i=1}^{\infty} C_i = (\cup_{i=1}^{\infty} C_i^c)^c$. Therefore, $P(C) = 1 - P(\cup_{i=1}^{\infty} C_i^c)$. Now, $\cup_{i=1}^{\infty} C_i^c = C_1^c \cup (C_2^c - C_1^c) \cup (C_3^c - C_2^c) \cdots = \theta_1 \cup \theta_2 \cup \theta_3 \ldots$, where the collection $\{\theta_1, \theta_2, \ldots\}$ is pairwise disjoint. Hence, $P(\cup_{i=1}^{\infty} C_i^c) = \sum_{i=1}^{\infty} P(\theta_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(\theta_i)$. But $\sum_{i=1}^{n} P(\theta_i) = P(C_n^c)$ and $P(C_n^c) = 1 - P(C_n)$. Hence, $P(\cup_{i=1}^{\infty} C_i^c) = \lim_{n\to\infty}(1 - P(C_n)) = 1 - \lim_{n\to\infty} P(C_n)$. Consequently, $P(C) = 1 - (1 - \lim_{n\to\infty} P(C_n)) = \lim_{n\to\infty} P(C_n)$.

Finally, by construction, $B_n \subseteq A_n \subseteq C_n$, 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 $A = B = C$, $\lim_{n\to\infty} P(A_n) = P(A)$.

To see that $A \in \mathcal{F}$, just note that $A = C = \cap_{i=1}^{\infty} C_i$ where C_i's represent countable unions of events. Hence, C_i are events for all i and by De Morgan's Laws $\cap_{i=1}^{\infty} C_i$ are events. ∎

1.3 Independence of events

Definition 1.6. 1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Given any $B \in \mathcal{F}$ such that $P(B) \neq 0$, we define $P_B : \mathcal{F} \to [0, 1]$ where*

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

$P_B(A)$ is called the conditional probability of A given B.[2]

[2] We show below that this is indeed a probability measure.

2. *Any A, $B \in \mathcal{F}$ are said to be independent if $P(A \cap B) = P(A)P(B)$;*

3. *If $2 < n \in \mathbb{N}$ then $E_1, \cdots, E_n \in \mathcal{F}$ are independent if*

$$P\left(\underset{m\in I}{\cap} E_m\right) = \prod_{m\in I} P(E_m) \text{ for all } I \subset \{1, \cdots, n\}. \tag{1.1}$$

Remark 1.5. *Note that (1.1) represents $\sum_{i=2}^{n} \binom{n}{i} = 2^n - n - 1$ equations.*

Theorem 1.9. *Let $(\Omega, \mathcal{F}, P)$ and A, $B \in \mathcal{F}$ such that $P(B) \neq 0$.*

1. *A and B independent $\iff P_B(A) = P(A)$*

2. *$P_B : \mathcal{F} \to [0, 1]$ defines a new probability measure on $\mathcal{F}$.*

Proof. 1. Since A and B are independent $P(A \cap B) = P(A)P(B)$ and since $P_B(A) = \frac{P(A\cap B)}{P(B)}$ we have $P_B(A) = \frac{P(A)P(B)}{P(B)} = P(A)$. Now, $P_B(A) = P(A) \implies P(A \cap B)/P(B) = P(A)$ which implies $P(A \cap B) = P(A)P(B) \implies A$ and B are independent.
2. We must show that P_B is a probability measure on $\mathcal{F}$ for any B such that $P(B) \neq 0$. First, note that $P_B(\emptyset) = P(\emptyset \cap B)/P(B) = P(\emptyset)/P(B) = 0$ and $P_B(\Omega) = P(\Omega \cap B)/P(B) = P(B)/P(B) = 1$. Second, $P_B\left(\cup_{j=1}^{\infty} A_j\right) = P\left(\left(\cup_{j=1}^{\infty} A_j\right) \cap B\right)/P(B) = P((\cup_{j=1}^{\infty} A_j \cap B))/P(B) = \sum_{j=1}^{\infty} P_B(A_j)$. ■

Theorem 1.10. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. If $A, B \in \mathcal{F}$ are independent, then:*

1. *A and B^c are independent (or A^c and B are independent).*

2. *A^c and B^c are independent.*

Proof. 1. Recall that $A \cup B = B \cup (A \cap B^c)$ and $P(A \cup B) = P(B) + P(A \cap B^c)$. The last equality together with Theorem 1.5.2 gives $P(A) - P(A \cap B) = P(A \cap B^c)$. Now, by independence of A and B we have $P(A \cap B^c) = P(A) - P(A \cap B) = P(A) - P(A)P(B)$. Hence, $P(A \cap B^c) = P(A)(1 - P(B)) = P(A)P(B^c)$.

2. Note that

$$\begin{aligned}
A^c \cap B^c &= (A \cup B)^c \text{ by DeMorgan's Laws} \\
&= \Omega - (A \cup B) \\
P(A^c \cap B^c) &= 1 - P(A \cup B) \\
&= 1 - (P(A) + P(B) - P(A)P(B)) \text{ by independence of } A \text{ and } B \\
&= (1 - P(A))(1 - P(B)) = P(A^c)P(B^c).
\end{aligned}$$

■

Theorem 1.11. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $A_1, A_2, \cdots, A_n \in \mathcal{F}$. *If* $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$ *then* $P(A_1 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$.

Proof. We will use induction. For $n = 2$, we have that if $P(A_1) > 0$, $P(A_2|A_1) = P(A_1 \cap A_2)/P(A_1)$ which implies $P(A_1 \cap A_2) = P(A_1)P(A_2|A_1)$. Now, assume that

$$P(A_1 \cap \cdots \cap A_{n-1}) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_{n-1}|A_1 \cap A_2 \cap \cdots \cap A_{n-2})$$

and define $B_n = (A_1 \cap A_2 \ldots A_{n-1}) \cap A_n$ with $P(A_1 \cap \cdots \cap A_{n-1}) > 0$. Then,

$$\begin{aligned}
P(B_n) &= P(A_1 \cap \cdots \cap A_{n-1})P(A_n|A_1 \cap \cdots \cap A_{n-1}) \\
&= P(A_1)P(A_2|A_1) \ldots P(A_{n-1}|A_1 \cap A_2 \cap \cdots \cap A_{n-2})P(A_n|A_1 \cap \cdots \cap A_{n-1})
\end{aligned}$$

by the assumption in the induction argument. ■

Recall the definition of a partition for a set.

Definition 1.7. $\{E_1, E_2 \ldots\}$ *is a partition of* Ω *if* $\bigcup_{i \in \mathbb{N}} E_i = \Omega$ *and* $E_i \cap E_j = \emptyset$, *for all* $i \neq j$.

Theorem 1.12. *Let* $(\Omega, \mathcal{F}, P)$ *be a probability space and* $\{E_1, E_2, \ldots\}$ *be a partition of* Ω. *If* $A \in \mathcal{F}$,

$$P(A) = \sum_{i=1}^{\infty} P(A|E_i)P(E_i).$$