MEASURE THEORY AND LEBESGUE INTEGRATION

Lecture notes for the *"Tutorial Course on Measure Theory"*

Contents

INTRODUCTION

These lecture notes throw the foundations of basic measure theory which would allow for a systematic treatment of events and random observations in the framework of probability theory.
For an interval contained in the real line or a nice region in the plane, the length of the interval or the area of the region give an idea of the size. We want to extend the notion of size to as large a class of sets as possible. In the first section we discuss the class of sets that it is possible to measure. Section 2 describes measurable functions on measurable spaces while Section 3 introduce the definition of measures and the most important examples. Based on the notions of measure spaces and measurable maps, we introduce the integral of a measurable map with respect to a general measure. Integration theory is a cornerstone in a systematic theory of probability that allows for the definition and investigation of expected values and higher moments of random variables. In these notes, we define the integral by an approximation scheme with simple functions. Then we deduce basic statements such as Monotone Convergence Theorem. Other important convergence theorems are deduced from it. We

end these notes with two fundamental results: the Radon-Nikodym Theorem and Fubini Theorem. Both will have important consequences in probability theory. General references for the present notes are classical textbooks in real analysis. We refer in particular to [3] and follow also the presentation of [2, 5]. Other (free) sources of inspiration are [1, 4].

Notations. — For a given set S, we use standard notation for operators on subsets of S, namely, if $A, B \subset S$, then

$$A \cap B, \qquad A \cup B, \qquad A \setminus B$$

denote, respectively, the intersection of A and B, the union of A and B and the complement of B in A, i.e. the collection of all elements of A that do not belong to B. In particular, the *complement* of B is

$$B^c = S \setminus A.$$

When dealing with arbitrary collection $\{A_i,\ i \in I\}$ of subsets of S, we write

$$\bigcap_{i\in I} A_i, \qquad \bigcup_{i\in I} A_i$$

for the interesction and union of *all* the sets A_i, $i \in I$. We denote the empty set by $\varnothing$. We denote the $\mathcal{S}$ the collection of all the subsets of S. Any collection $\mathcal{C}$ of subset of S is called a class of subsets, i.e. $\mathcal{C}$ is a class of subsets of S if $\mathcal{C} \subseteq \mathcal{P}(S)$. Given a countable family of subsets $(A_n)_n \subset S$, we write $(A_n)_n \uparrow A$ whenever $A_n \subset A_{n+1}$ for any $n \in \mathbb{N}$ and $\bigcup_n A_n = A$. We write $(A_n)_n \downarrow A$ whenever $A_{n+1} \subset A_n$ for any $n \in \mathbb{N}$ and $\bigcap_n A_n = A$.

1. Measurable sets: σ-algebras, π-systems

1.1. Motivation and terminology. — In everyday language, we call random experiment any experiment which, if repeated under the same conditions, may lead to various possible results which can not be predicted in advance. The set of all the possible results is called the state space associated to the experiment and is classically denoted by S. A generic possible result of the experiment is classically denoted ω, so that $\omega \in S$. Then, a random event, associated to this experiment is a subset of S (i.e. a set of possible results) for which we can decide whether it has been realized or not. It appears quite intuitive that the set Σ of all random events should enjoy the following properties:

- the set of all the possible result has to be a random event, i.e. $S \in \Sigma$,
- if one can decide that a random event occurred, one has to be able to decide that it did not occurred. Thus, if $A \in \Sigma$, one must have $A^c \in \Sigma$,
- if one can decide that the event A has occurred and that B has occurred, one has to be able also to decide whether the event "A or B has occurred", i.e. if $A \in \Sigma$ and $B \in \Sigma$, one must have $A \cup B \in \Sigma$.

In other words, Σ should be at least stable by two operations: taking the complement and taking (finite) union. It turns out that, since one aims to repeat our experiment an arbitrary large number of times, one can also think of S to be an infinite set, *it appears important that Σ is also stable by countable union*.
These simple heuristic arguments lead to the following fundamental concept of σ-algebras and measurable spaces. Let S be a given set, $S \neq \varnothing$.

Definition 1.1. — *A collection Σ of subsets of S is called a σ-algebra on S if*

1. $S \in \Sigma$
2. $A \in \Sigma$ *implies* $A^c \in \Sigma$
3. $\{A_n\}_{n\in\mathbb{N}} \subset \Sigma$ *implies* $\bigcup_{n=1}^{\infty} A_n \in \Sigma$.

One can check very easily that if $(A_n)_n \subset \Sigma$ then $\cap_n A_n \in \Sigma$. In other words, a σ-algebra on S is a family of subsets of S closed under *any countable collections of set operations*.

Definition 1.2. — *If S is a given set and Σ is a σ-algebra on S, then the pair (S, Σ) is called a measurable space. Any element $A \in \Sigma$ is called a Σ-measurable subset of S.*

We will introduce later on the notion of *measure* over S, notice however that the notion of *measurable space* is independent of the notion of measure.

Remark 1.3. — *A very simple example of σ-algebra is the one made of all the possible subsets of S that we shall denote $\mathcal{P}(S)$ (power set σ-algebra). Another trivial example of σ-algebra on S is the one only made of $\{\varnothing, S\}$ (it is of course of very little interest).* //

Remark 1.4. — *For those who are familiar with the notion of topological space, the axiomatization of measurable sets is somehow reminiscent to the notion of open sets in topological space. Recall the definition of topological space (X, τ): a set X is called a topological space if there exists a collection τ of subsets of X, satisfying the following*

(i) The empty set and X are in τ.
(ii) The union of any collection of sets in τ is also in τ.
(iii) The intersection of any finite collection of sets in τ is also in τ.

The collection τ is called a topology on X and the elements of τ are called the open sets of τ. One sees that the definition of topology may look like the one of σ-algebra but differs drastically since, for instance, the complementary of an open set is usually not an open set. Quoting Williams [5]*: "In topology, one axiomatizes the notion of 'open set', insisting in particular that the union of* ***any*** *collection of open sets is open, and that the intersection of a* ***finite*** *collection of open sets is open. In measure theory, one axiomatizes the notion of 'measurable set', insisting that the union of a* ***countable*** *collection of measurable sets is measurable, and that the intersection of a countable collection of measurable sets is also measurable."* //

Example 1.5. — **(Interpretation)**. *Assume one tosses a coin twice and let S be the the set of all possible outcomes Heads or Tails:*

$$S = \{HH, HT, TT, TH\}.$$

Let

$$\Sigma = \{\varnothing; S; \{HT, HH\}; \{TH, TT\}\}.$$

Then, Σ is a σ-algebra on S. In probability, one should think as the σ-algebra on S as the collections of all the events that you "know". In this example, Σ corresponds to the case in which we know the outcome of the first toss.

Example 1.6. — *If $A \subset S$, then $\Sigma = \{\varnothing; S; A; A^c\}$ is a σ-algebra on S.*

One has the following properties:

Proposition 1.7. — *Let $\{\Sigma_i\}_{i\in I}$ be a family of σ-algebras. Then $\bigcap_i \Sigma_i$ is a σ-algebra .*

Démonstration: First, $S \in \Sigma_i$ for all i, hence $S \in \bigcap_i \Sigma_i$. Next, $A \in \bigcap_i \Sigma_i \implies A \in \Sigma_i$ for all i, hence $A^c \in \Sigma_i$ for all i, and so $A^c \in \bigcap_i \Sigma_i$. Finally, $\{A_n\} \in \bigcap_i \Sigma_i \implies \{A_n\} \in \Sigma_i$ for all i, hence $\bigcup_n A_n \in \Sigma_i$ for all i, and so $\bigcup_n A_n \in \bigcap_i \Sigma_i$. ■

The above result is no more true for unions of σ-algebras as illustrated by the following example:

Example 1.8. — *Let $S = \{a, b, c\}$ and $\Sigma_1 = \{\varnothing; S; \{a\}; \{b, c\}\}$ and $\Sigma_2 = \{\varnothing; S; \{b\}; \{a, c\}\}$. Then, Σ_1 and Σ_2 are σ-algebrason S but $\Sigma_1 \cup \Sigma_2$ is not. Indeed,*

$$\Sigma_1 \cup \Sigma_2 = \{\varnothing; S; \{a\}; \{b\}; \{a, c\}; \{b, c\}\}$$

and $\{a\} \in \Sigma_1 \cup \Sigma_2$, $\{b\} \in \Sigma_1 \cup \Sigma_2$ while $\{a\} \cup \{b\} = \{a, b\} \notin \Sigma_1 \cup \Sigma_2$. The collection $\Sigma_1 \cup \Sigma_2$ is not closed under (finite) union, it is not a σ-algebra on S.

It is usually difficult to describe the typical element of a given σ-algebra. For this reason, one usually prefers to deal with some explicit sets and build the smallest σ-algebra that contains them:

Definition 1.9. — *Let $\mathcal{C}$ be a class of subsets of S. Defining*

$$\sigma(\mathcal{C}) = \bigcap_i \{\Sigma_i : \Sigma_i \textit{ is a } \sigma\textit{-algebra} \quad \textit{with } \mathcal{C} \subset \Sigma_i\},$$

as the intersection of all σ-algebras containing $\mathcal{C}$, we call $\sigma(\mathcal{C})$ the σ-algebra generated by $\mathcal{C}$.

Since there is at least one σ-algebra containing $\mathcal{C}$ (namely, $\mathcal{P}(S)$) we are not taking the intersection over an empty class of σ-algebras. Moreover, from Proposition 1.7, $\sigma(\mathcal{C})$ is a σ-algebra on S. Actually, $\sigma(\mathcal{C})$ is the smallest σ-algebra containing $\mathcal{C}$, i.e. one has the following

Proposition 1.10. — *Let $\mathcal{C} \subseteq \mathcal{P}(S)$ be a class. Then, $\sigma(\mathcal{C})$ is characterized by the following two properties:*

1. *$\mathcal{C} \subseteq \sigma(\mathcal{C})$;*
2. *if Σ is a σ-algebra on S s.t. $\mathcal{C} \subseteq \Sigma$, then $\sigma(\mathcal{C}) \subseteq \Sigma$.*

Proof. — The proof is left as an Exercise. It consists in proving first that $\sigma(\mathcal{C})$ satisfies 1. & 2 and then that, if $\mathcal{G}$ is another σ-algebra satisfying

1. $\mathcal{C} \subseteq \mathcal{G}$;
2. if Σ is a σ-algebra s.t. $\mathcal{C} \subseteq \Sigma$, then $\mathcal{G} \subseteq \Sigma$;

then $\mathcal{G} = \sigma(\mathcal{C})$. □

***Example 1.11* (Fundamental example: Borel σ-algebra).** — *If S is a topological space (see Remark 1.4) and if τ is the collection of open subsets of S. Then, the σ-algebra generated by the family of open subsets of S is called the **Borel σ-algebra** on S and is denoted by $\mathcal{B}(S)$. In other words, if (S, τ) is a topological space then $\mathcal{B}(S) = \sigma(\tau)$.*

The Borel σ-algebra $\mathcal{B}(\mathbb{R})$ is the most important of all σ-algebras. It is very difficult to exhibit subset of $\mathbb{R}$ that does not belong to $\mathcal{B}(\mathbb{R})$ (it is however possible, see [5]). In other words, almost every subset of $\mathbb{R}$ you may think about belongs to $\mathcal{B}(\mathbb{R})$. Anyway, the actual structure of elements of $\mathcal{B}(\mathbb{R})$ may be very intricate and the following result will help us in getting a more precise idea of it.

Proposition 1.12. — *Let $\pi(\mathbb{R}) = \{(-\infty, x]\ ;\ x \in \mathbb{R}\}$, then $\mathcal{B}(\mathbb{R}) = \sigma(\pi(\mathbb{R}))$.*

Proof. — Recall first that any open set of $\mathbb{R}$ (for the usual topology we are dealing with) is a countable union of open intervals. In particular, for any $y \in \mathbb{R}$, $(-\infty, y)$ is an open subset of $\mathbb{R}$. We first claim that $\pi(\mathbb{R}) \subset \mathcal{B}(\mathbb{R})$. Indeed, for any $x \in \mathbb{R}$, since

$$(-\infty, x] = \bigcap_{n \geqslant 1} (-\infty, x + 1/n)$$

is the countable intersection of open subsets of $\mathbb{R}$, one has $(-\infty, x] \subset \mathcal{B}(\mathbb{R})$ for any x and this proves our claim. In particular, $\sigma(\pi(\mathbb{R})) \subset \mathcal{B}(\mathbb{R})$. To prove the converse inclusion, we only have to prove that any open subset A of $\mathbb{R}$ belongs to $\sigma(\pi(\mathbb{R}))$. Again, since A is the countable union of open intervals and $\sigma(\pi(\mathbb{R}))$ is a σ-algebra , it is enough to prove that

$$(a, b) \in \sigma(\pi(\mathbb{R})) \qquad \forall a, b \in \mathbb{R},\ a < b.$$

For any $x \in \mathbb{R}$ with $x > a$ one has

$$(a, x] = (-\infty, x] \cap (-\infty, a]^c \in \sigma(\pi(\mathbb{R})).$$

Moreover, setting $\delta = \frac{b-a}{2} > 0$, one has clearly

$$(a, b) = \bigcup_n (a, b - \delta/n].$$

Thus, $(a, b) \in \sigma(\pi(\mathbb{R}))$ as the countable union of elements of $\sigma(\pi(\mathbb{R}))$. This completes the proof. □

From the above proof, one easily deduces the following:

Proposition 1.13. — *For all $i = 1, \ldots, 6$, it holds*

$$\mathcal{B}(\mathbb{R}) = \sigma(\mathcal{C}_i)$$

with

$$\mathcal{C}_1 = \{(a,b) : a,b \in \mathbb{R}\}; \qquad \mathcal{C}_2 = \{[a,b] : a,b \in \mathbb{R}\}; \qquad \mathcal{C}_3 = \{(-\infty,b) : b \in \mathbb{R}\}$$
$$\mathcal{C}_4 = \{(-\infty,b] : b \in \mathbb{R}\}; \qquad \mathcal{C}_5 = \{(a,\infty) : a \in \mathbb{R}\}; \qquad \mathcal{C}_6 = \{[a,\infty) : a \in \mathbb{R}\}.$$

Remark 1.14. — *Any singleton $\{x\}$ $(x \in \mathbb{R})$ is contained in $\mathcal{B}(\mathbb{R})$. Indeed, given $x \in \mathbb{R}$, one has*

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x + \frac{1}{n}\right)$$

with $\left(x - \frac{1}{n}, x + \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R})$ for any $n \geqslant 1$. Thus, $\{x\}$ is the countable intersection of elements in $\mathcal{B}(\mathbb{R})$ and therefore $\{x\} \in \mathcal{B}(\mathbb{R})$. As a consequence, since $\mathbb{Q}$ is countable, one has

$$\mathbb{Q} = \bigcup_{q \in \mathbb{Q}} \{q\}$$

is the countable union of elements of $\mathcal{B}(\mathbb{R})$ so $\mathbb{Q} \in \mathcal{B}(\mathbb{R})$. More generally, any countable subset of $\mathbb{R}$ belongs to $\mathcal{B}(\mathbb{R})$.

1.2. π-systems and d-systems. — The above set $\pi(\mathbb{R})$ is a typical example of a more general class of sets:

Definition 1.15. — *Let S be a set. A collection $\mathscr{I}$ of subsets of S is called a π-system (or p-system) if $\mathscr{I}$ is closed under finite intersections, i.e.*

$$A, B \in \mathscr{I} \implies A \cap B \in \mathscr{I}.$$

A collection $\mathscr{D}$ of subsets of S is called a d-system (or Dynkyn*-system) on S if*

(i) $S \in \mathscr{D}$;
(ii) *if $A, B \in \mathscr{D}$ and $B \subset A$ then $A \setminus B \in \mathscr{D}$;*
(iii) *if $(A_n)_n$ is a sequence of elements of $\mathscr{D}$ with $(A_n)_n \uparrow A$ then $A \in \mathscr{D}$.*

One has the following characterization of σ-algebras in terms of π-systems and d-systems:

Proposition 1.16. — *A collection Σ of subsets of S is a σ-algebra if and only if Σ is both a π-system and a d-system.*

Proof. — It is quite clear that a σ-algebra is both a π-system and a d-system. Assume now that Σ is both a π-system and a d-system. Let $A, B \in \Sigma$ and let $(F_n)_n \subset \Sigma$. Then, $A^c = S \setminus A$ belongs to Σ by virtue of the second property (ii) of π-systems here above. In the same way, $B^c \in \Sigma$ and, since Σ is a π-system one has $A \cap B \in \Sigma$ so that $A \cup B = S \setminus (A^c \cup B^c) \in \Sigma$. Now, by induction, one proves easily that $A_n := F_1 \cup \ldots \cup F_n \in \Sigma$ for any $n \geqslant 1$. Moreover, $A_n \subset A_{n+1}$ and $\bigcup_n A_n = \bigcup_k F_k$. Therefore, from the third property of d-systems(point (iii) in the above definition), one deduces that $\bigcup_k F_k \in \Sigma$. This proves that Σ is a σ-algebra. □

As for σ-algebrasone can prove the following

Lemma 1.17. — *If $\{\mathcal{D}_i\}_{i\in I}$ is a family of d-system on S then $\bigcap_{i\in I}\mathcal{D}_i$ is a d-system on S.*

With this, given a class $\mathcal{C}$ of subsets of S, one can define $d(\mathcal{C})$ as the smallest d-system containing $\mathcal{C}$ as we did for $\sigma(\mathcal{C})$, i.e.

$$d(\mathcal{C}) = \bigcap_i \{\mathcal{D}_i \ ; \ \mathcal{D}_i \ d\text{-system containing } \mathcal{C}\}\,.$$

It is clear then that $d(\mathcal{C}) \subset \sigma(\mathcal{C})$ and one can prove the following:

***Proposition 1.18* (Dynkyn's Lemma)**. — *If $\mathscr{I}$ is a π-system, then $d(\mathscr{I}) = \sigma(\mathscr{I})$.*

Proof. — From the previous Proposition, we only have to prove that $d(\mathscr{I})$ is a π-system (since it will then be of course the smallest σ-algebra containing $\mathscr{I}$). First, let

$$\mathcal{D}_1 = \{B \in d(\mathscr{I}) \ \text{ such that } \ B \cap C \in d(\mathscr{I})\ \forall C \in \mathscr{I}\}.$$

Because $\mathscr{I}$ is a π-system, one has $\mathscr{I} \subset \mathcal{D}_1$. Moreover, it is not difficult to check as an exercise that $\mathcal{D}_1$ is a d-system because $d(\mathscr{I})$ is. In particular, $\mathcal{D}_1$ is a d-system containing $\mathscr{I}$ so that $\mathcal{D}_1 \supset d(\mathscr{I})$. In other words, $\mathcal{D}_1 = d(\mathscr{I})$. Now, let

$$\mathcal{D}_2 = \{A \in d(\mathscr{I}) \ \text{ such that } \ A \cap B \in d(\mathscr{I})\ \forall B \in d(\mathscr{I})\}.$$

Since $\mathcal{D}_1 = d(\mathscr{I})$ one has $\mathscr{I} \subset \mathcal{D}_2$. But, here again, one proves easily that $\mathcal{D}_2$ is a d-system because $d(\mathscr{I})$ is. Consequently, $\mathcal{D}_2 = d(\mathscr{I})$. This exactly means that $d(\mathscr{I})$ is a π-system and achieves the proof. □

***Theorem 1.19* (*Monotone class Theorem*)**. — *Let $\mathcal{D}$ be a d-system and let $\mathcal{C}$ be a π-system with $\mathcal{C} \subset \mathcal{D}$. Then $\sigma(\mathcal{C}) \subset \mathcal{D}$.*

Proof. — Introduce $d(\mathcal{C})$ as the d-system generated by $\mathcal{C}$. From Dynkin's lemma $d(\mathcal{C}) = \sigma(\mathcal{C})$. But since $d(\mathcal{C})$ is the smallest d-system containing $\mathcal{C}$ one has $d(\mathcal{C}) \subset \mathcal{D}$. This achieves the proof. □

1.3. Restriction of σ-algebras. — Once a measurable space (S, Σ) is given, it can be useful to define a σ-algebra on a smaller subset of S which inherits the properties of Σ. This can be done thanks to the following

Proposition 1.20. — *Let (S, Σ) be a measurable space and let $S_0 \subseteq S$. Then the class*

$$\Sigma \cap S_0 = \{A \cap S_0 : \ A \in \Sigma\}$$

is a σ-algebra on S_0. This is called the σ-algebra induced by Σ on S_0.

Proof. — The proof is an easy and useful exercise. □

Example 1.21. — *Given $I \subset \mathbb{R}$, we can define the Borel σ-algebra $\mathcal{B}(I)$ on I as the σ-algebra induced by $\mathcal{B}(\mathbb{R})$ on I.*

Proposition 1.22. — *Given a non-empty set S and a class $\mathcal{C} \subseteq \mathcal{P}(S)$ one has*

$$\sigma(\mathcal{C} \cap S_0) = \sigma(\mathcal{C}) \cap S_0 \qquad \forall S_0 \subset S.$$

In other words, given $S_0 \subset S$, the σ-algebra on S_0 induced by $\sigma(\mathcal{C})$ is just the σ-algebra generated by $\mathcal{C} \cap S_0$.

Example 1.23. — *Thanks to the above Proposition, one sees that, for $I \subset \mathbb{R}$, the previous definition of $\mathcal{B}(I)$ as the σ-algebra induced by $\mathcal{B}(\mathbb{R})$ on I exactly coincides with the one given in Definition 1.11 as the σ-algebra generated by the open subsets of I.*

2. Measurable functions

Once the concept of measurable space introduced, it becomes important to understand the various properties of functions between such spaces:

Definition 2.1. — *Let (S_1, Σ_1) and (S_2, Σ_2) be two measurable spaces and let $h : S_1 \to S_2$ be a mapping from S_1 to S_2. Then, h is called Σ_1/Σ_2-measurable if, for any $A \in \Sigma_2$, one has $h^{-1}(A) \in \Sigma_1$ where the inverse image $h^{-1}(A)$ is defined as*

$$h^{-1}(A) = \{s \in S_1 \, : \, h(s) \in A\}.$$

If there is no ambiguity on which σ-algebras we are dealing with, i.e. if Σ_1 and Σ_2 are well-understood, we shall only say that h is measurable.

Remark 2.2. — *The definition is the exact analogue of the definition of continuous mapping between two topological spaces. Here again, the role played by the open sets in topological spaces is played now by measurable sets.* //

Remark 2.3. — *Notice that the definition of the inverse image $h^{-1}(A)$ does not imply that the mapping h in invertible. Moreover, one checks easily from the definition of inverse image that, given a collection (non necessarily countable) of sets $(A_\alpha)_\alpha$:*

$$h^{-1}\left(\bigcup_\alpha A_\alpha\right) = \bigcup_\alpha h^{-1}(A_\alpha) \qquad \textit{while} \qquad h^{-1}(A^c) = (h^{-1}(A))^c, \qquad \forall A.$$

In particular, one can also check that $h^{-1}\left(\bigcap_\alpha A_\alpha\right) = \bigcap_\alpha h^{-1}(A_\alpha)$. //

One deduces from the above definition the immediate property of measurable mappings:

Proposition 2.4. — *Let (S_1, Σ_1), (S_2, Σ_2) and (S_3, Σ_3) be measurable spaces and let h_1 be measurable from (S_1, Σ_1) and h_2 be measurable from (S_2, Σ_2) to (S_3, Σ_3), then $h_2 \circ h_1$ is measurable from (S_1, Σ_1) to (S_3, Σ_3).*

One also has the following useful result:

Lemma 2.5. — *Let (S_1,Σ_1) and (S_2,Σ_2) be two measurable spaces and let $h : S_1 \to S_2$. Assume that there is some class $\mathcal{C} \subset \Sigma_2$ such that $\Sigma_2 = \sigma(\mathcal{C})$. If $h^{-1}(A) \in \Sigma_1$ for any $A \in \mathcal{C}$, then h is Σ_1/Σ_2-measurable.*

Proof. — Let $\mathscr{E} = \{A \in \Sigma_2 \text{ such that } h^{-1}(A) \in \Sigma_1\}$. From the property of inverse image recalled in Remark 2.3, it is easily deduced that $\mathscr{E}$ is a σ-algebra on S_2. By assumption, $\mathcal{C} \subset \mathscr{E}$ so that $\sigma(\mathcal{C}) \subset \mathscr{E}$. This means that $\mathscr{E} = \Sigma_2$. □

For real valued mapping, one has the following specific definition

Definition 2.6. — *Let (S,Σ) be a measurable space and let $h : S \to \mathbb{R}$. One shall only say that h is Σ-measurable whenever h is $\Sigma/\mathcal{B}$-measurable. We shall denote by $\mathcal{M}(\Sigma)$ the set of Σ-measurable mappings, by $\mathcal{M}(\Sigma)^+$ the set of nonnegative Σ-measurable mappings and by $\mathcal{M}_b(\Sigma)$ the set of bounded Σ-measurable mappings.*

Remark 2.7. — *Because* $\limsup$ *of sequences of finite-valued functions maybe infinite, it is convenient to extend the above definition to functions h taking values in $[-\infty,\infty]$ in the following way: h will be called Σ-measurable if h is $\Sigma/\mathcal{B}([-\infty,\infty])$-measurable.* //

Proposition 2.8. — *Let (S,Σ) be a measurable space and let $h : S \to \mathbb{R}$. Then, h is Σ-measurable if and only if $\{h \leqslant c\} \in \Sigma$ for any $c \in \mathbb{R}$ where $\{h \leqslant c\} = \{s \in S\,;\, h(s) \leqslant c\}$.*

Proof. — It suffices to apply the above Lemma 2.5 to the class $\mathcal{C} = \pi(\mathbb{R})$ which is such that $\mathcal{B} = \sigma(\mathcal{C})$ according to Prop. 1.12. □

Remark 2.9. — *It is clear that the conclusion of the above Prop. still hold true if one replaces $\{h \leqslant c\}$ by $\{h < c\}$, $\{h > c\}$ or $\{h \geqslant c\}$.* //

Whenever S is a topological space and $\Sigma = \mathcal{B}(S)$, one has a specific definition

Definition 2.10. — *Let S be a topological space and $h : S \to \mathbb{R}$. We say that h is a **Borel function** if h is $\mathcal{B}(S)$-measurable.*

Remark 2.11. — *It is clear from Prop. 2.8 that, if S is a topological space and $h : S \to \mathbb{R}$ is continuous, then h is a Borel function. However, the class of Borel functions is much larger than the one of continuous functions. For instance, monotone functions are Borel function (Check it).* //

Lemma 2.12. — *Sums and products of measurable $\mathbb{R}$-valued functions are measurable, i.e. given a measurable space (S,Σ), if $\lambda \in \mathbb{R}$ and if $h_1, h_2 \in \mathcal{M}(\Sigma)$, then $h_1+\lambda h_2 \in \mathcal{M}(\Sigma)$ and $h_1\, h_2 \in \mathcal{M}(\Sigma)$.*

Proof. — The proof is left as an (easy) exercise. □

Lemma 2.13. — *Let (S,Σ) be a measurable space and let $(h_n)_n$ be a sequence of mappings of $\mathcal{M}(\Sigma)$. Then,*

$$(i)\ \inf_n h_n \qquad (ii) \qquad \liminf_n h_n \quad \textit{and} \quad (iii) \quad \limsup_n h_n$$

are Σ-measurable (in the sense of Remark 2.7). Moreover,

$$\{s \in S\,;\, \lim_n h_n(s) \textit{ exists }\} \in \Sigma.$$

Proof. — (i) For any $c \in \mathbb{R}$, one has

$$\{\inf_n h_n \geqslant c\} = \bigcap_n \{h_n \geqslant c\}$$

so that $\{\inf_n h_n \geqslant c\} \in \Sigma$ for any $c \in \mathbb{R}$ and $\inf_n h_n$ is measurable according to Prop. 2.8. Now, for any $n \in \mathbb{N}$ and any $s \in S$, set

$$f_n(s) = \inf\{h_k(s)\,;\, k \geqslant n\}.$$

From (i), $f_n \in \mathcal{M}(\Sigma)$. Moreover,

$$f(s) := \limsup_n f_n(s) = \lim_n f_n(s) = \sup_n f_n(s)$$

so that $\{f \leqslant c\} = \bigcap_n \{f_n \leqslant c\}$ which implies $\{f \leqslant c\} \in \Sigma$. This proves that $f \in \mathcal{M}(\Sigma)$ and (ii) is proven. Now, the same reasoning yields (iii). Finally, since

$$A := \{s \in S\,;\, \lim_n h_n(s) \text{ exists }\} = \{\liminf_n h_n < \infty\} \cap \{\limsup_n h_n < \infty\} \cap g^{-1}(\{0\})$$

where $g(s) = \limsup_n h_n(s) - \liminf_n h_n(s)$, we see that $\{s \in S\,;\, \lim_n h_n(s) \text{ exists }\} \in \Sigma$. □

Given a real-valued mapping $h : S \to \mathbb{R}$ where (S,Σ) is a measurable space. One can define the positive and negative part of h as

$$h^+(s) = \max(h(s),0) \qquad h^-(s) = -\min(h(s),0), \qquad s \in S.$$

Notice that both h^+ and h^- are *nonnegative* mapping. Moreover,

$$h(s) = h^+(s) - h^-(s) \qquad \forall s \in S.$$

One calls h^+ the positive part of h while h^- is its negative part. One has the following which is easily deduced from the above Lemma:

Lemma 2.14. — *Given a real-valued mapping $h : S \to \mathbb{R}$ where (S,Σ) is a measurable space. Then, h is S-measurable if and only if both $h^{\pm}$ are S-measurable.*

Example 2.15 **(*Fundamental Example:* Indicator)**. — *Let $A \subset S$, the* indicator *function of A is defined as follows*

$$\mathbb{1}_A(s) = \begin{cases} 1 & \textit{if } s \in A, \\ 0 & \textit{if } s \notin A. \end{cases} \tag{2.1}$$

Then, given a σ-algebra Σ on S, $\mathbb{1}_A$ is S-measurable if and only if $A \in \Sigma$.

***Definition 2.16* (Simple functions).** — *A function $f : S \to \mathbb{R}$ is said to be a* simple function *if it takes only a finite number of values. In particular, it can be expressed as a finite linear combination of indicator functions, i.e.*

$$f = \sum_{i=1}^{N} \alpha_i \mathbb{1}_{A_i}$$

with $\alpha_i \in \mathbb{R}$ and $A_i \subseteq S$ for all i.

***Remark 2.17*.** — *The above representation is not unique. It however can be made unique if we take $\alpha_1 > \alpha_2 > \ldots$ and $A_i = f^{-1}(\{\alpha_i\})$. In such case we say that f is written in* standard *or* canonical form.

***Theorem 2.18*.** — *Let (S, Σ) be a measurable space, and let $f : S \to [-\infty, +\infty]$. Then*

1. *f is measurable if and only if there exists a sequence $\{\varphi_n\}$ of simple measurable functions s.t. $\varphi_n \to f$;*
2. *if f is bounded from below, then $\{\varphi_n\}$ can be chosen s.t. $\varphi_n \uparrow f$;*

Remark 2.19*.** — *Since nonnegative functions are bounded from below, the above theorem asserts that* ***any nonnegative measurable function is the (pointwise) limit of an increasing sequence of simple functions. *It will be mainly in this form that the above Theorem will be applied.*

Given a collection of maps, it is convenient (notably in probability theory) to exhibit a common σ-algebra which makes each of them measurable:

***Definition 2.20*.** — *Let (S, Σ) be a measurable space and let $(h_\lambda)_{\lambda \in \Lambda}$ be a collection of maps from S to $\mathbb{R}$, i.e. $h_\lambda : S \to \mathbb{R}$ for any $\lambda \in \Lambda$. One denotes*

$$\mathscr{H} = \sigma(h_\lambda \,;\, \lambda \in \Lambda)$$

the smallest σ-algebra $\mathscr{H}$ on S such that each map $h_\lambda \in \mathcal{M}(\mathscr{H})$ for any $\lambda \in \Lambda$. The σ-algebra $\mathscr{H}$ is called the σ-algebra generated by the family $(h_\lambda)_{\lambda \in \Lambda}$.

In practical situation, the meaning and usefulness of the above definition stands in the following: given a family $(h_\lambda)_{\lambda \in \Lambda}$ of measurable applications from $S \to \mathbb{R}$. Then, $\mathscr{H}$ consists precisely of those subsets $A \in \Sigma$ such that, for each and every $x \in S$, you can decide whether or not $x \in A$ only by knowing $h_\lambda(x)$ for *any* $\lambda \in \Lambda$. We will see a probabilistic interpretation of this in the next chapter. We end this section with a general fundamental result which will be required in several proofs in the sequel:

***Theorem 2.21* (Monotone class theorem).** — *Let $\mathcal{H}$ be a class of bounded functions from a set S into $\mathbb{R}$ satisfying the three conditions:*

1. *$\mathcal{H}$ is a vector space over $\mathbb{R}$,*
2. *$\mathcal{H}$ contains the constant functions,*
3. *if $(f_n)_n$ is a sequence of nonnegative functions in $\mathcal{H}$ such that $f_n \uparrow f$ where f is a bounded function on S, then $f \in \mathcal{H}$,*

4. there exists some π-system $\mathscr{I}$ on S such that $\mathbb{1}_A \in \mathcal{H}$ for any $A \in \mathscr{I}$.

Then, if $\mathcal{H}$ contains every $\sigma(\mathscr{I})$-measurable function on S.

We refer to [**5**, Chapter A3] for a proof of this result. One of the main applications of the Monotone Class Theorem is that of showing that certain property is satisfied by all sets in an σ-algebra , generally starting by the fact that the field generating the σ-algebra satisfies such property and that the sets that satisfies it constitutes a monotone class. Typically, it plays, for measurable functions, the role that Proposition 3.12 is playing for measures and it allows to deduce results about general measurable functions only from results about indicator functions of elements of π-systems.

3. MEASURES

3.1. Definition and main properties. — We begin with the following definition

Definition 3.1. — *Let (S, Σ) be a measurable space. A measure on (S, Σ) is a mapping $\mu : \Sigma \to [0, \infty]$ such that*

1. *$\mu(\varnothing) = 0$;*
2. *for any sequence $(A_n)_n$ of disjoint subsets of Σ one has*

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n).$$

The second property is called σ-additivity. Given $A \in \Sigma$, we call $\mu(A)$ the measure of A. A measure space is a triple (S, Σ, μ) where (S, Σ) is a measurable space and μ a measure over (S, Σ).

***Example 3.2* (Dirac measure)**. — *Let (S, Σ) be a measurable space and let $x \in S$ be given. Introduce, for any $A \in \Sigma$:*

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

Then, δ_x is a measure on (S, Σ) called the Dirac measure *at $x \in S$.*

***Example 3.3* (Counting measure)**. — *Let (S, Σ) be a measurable space and let J be a countable subset of S. We introduce the counting measure as*

$$\nu = \sum_{j \in J} \delta_j$$

i.e. $\nu(A) = \sum_{j \in J} \delta_j(A)$ for all $A \in \Sigma$. It counts the number of elements of D belonging to A

***Example 3.4* (Discrete measure)**. — *Let (S, Σ) be a measurable space and let J be a countable subset of S. For any $x \in J$, let $m_x > 0$ be given. Define then*

$$\mu = \sum_{x \in J} m_x \delta_x.$$

Then, μ is a measure over (S, Σ) and it is said to be discrete (it can take only a countable number of nonnegative values).

Definition 3.5. — *Let (S, Σ, μ) be a measure space. It is called* ***finite*** *if $\mu(S) < \infty$. It is called* ***σ-finite*** *if there is a sequence $(A_n)_n \subset \Sigma$ such that $\mu(A_n) < \infty$ for any $n \geqslant 1$ and $\bigcup_n A_n = S$.*

The most important case of finite measure spaces is the one of probability triple

Definition 3.6. — *Let (S, Σ, μ) be a measure space. The measure μ is called a* ***probability measure*** *if $\mu(S) = 1$ and then (S, Σ, μ) is a probability triple.*

We establish several properties of measures.

Proposition 3.7. — *Let (S, Σ, μ) be a measure space. Then,*

1. *$\mu(B) = \mu(A) + \mu(B \setminus A)$ for any $A, B \in \Sigma$ with $A \subset B$. In particular, $\mu(A) \leqslant \mu(B)$.*
2. *$\mu(A \cup B) \leqslant \mu(A) + \mu(B)$ for any $A, B \in \Sigma$.*
3. *$\mu\left(\cup_{i=1}^n A_i\right) \leqslant \sum_{i=1}^n \mu(A_i)$ for any $(A_i)_{i=1,\ldots,n} \subset \Sigma$.*

If moreover $\mu(S) < \infty$ then

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B) \quad \textit{for any} \quad A, B \in \Sigma.$$

Proof. — The proof of the first point is clear from the additivity of μ. For the second point, one notices that

$$A \cup B = A \bigcup_{\text{disjoint}} \left(B \setminus (A \cap B)\right).$$

Therefore, $\mu(A \cup B) = \mu(A) + \mu\left(B \setminus (A \cap B)\right)$. Moreover,

$$B = (A \cap B) \bigcup_{\text{disjoint}} \left(B \setminus (A \cap B)\right)$$

so that $\mu(B) = \mu(A \cap B) + \mu\left(B \setminus (A \cap B)\right) \geqslant \mu\left(B \setminus (A \cap B)\right)$. Putting altogether yields

$$\mu(A \cup B) \leqslant \mu(A) + \mu(B).$$

The third point is proven by induction from the second one. For the last point, one still uses the above reasoning:

$$\mu(B) = \mu(A \cap B) + \mu\left(B \setminus (A \cap B)\right)$$

and, since $\mu(A \cap B) < \infty$, this last identity actually reads

$$\mu\left(B \setminus (A \cap B)\right) = \mu(B) - \mu(A \cap B)$$

which, combined with the above identity $\mu(A \cup B) = \mu(A) + \mu\left(B \setminus (A \cap B)\right)$ provides the conclusion. □

Remark 3.8. — *It is very important to notice that, as we pointed out in the above proof, the identity in the first point is equivalent to $\mu(B \setminus A) = \mu(B) - \mu(A)$ only whenever $\mu(A) < \infty$.* //

The following result provides the ***monotone convergence properties of measures***

Proposition 3.9. — *Let (S,Σ,μ) be a measure space and let $(A_n)_n$ be a sequence of elements of Σ.*

1. *If $A_n\subset A_{n+1}$ for any $n\geqslant 1$, then*
$$\mu\left(\bigcup_n A_n\right)=\sup_n \mu(A_n)=\lim_n \mu(A_n).$$
In other words, if $(A_n)\uparrow A$ then $\mu(A)=\lim_n\mu(A_n)$.
2. *If $A_n\supset A_{n+1}$ for any $n\geqslant 1$ and $\mu(A_1)<\infty$ then*
$$\mu\left(\bigcap_n A_n\right)=\inf_n \mu(A_n)=\lim_n \mu(A_n).$$
In other words, if $(A_n)_n\downarrow A$ then $\mu(A)=\lim_n\mu(A_n)$.

Proof. — Define $B_1:=A_1$ and $B_n:=A_n\setminus A_{n-1}$ for any $n\geqslant 2$. Then, the sets $(B_n)_n$ are disjoints and, for any $n\geqslant 1$, $A_n=B_1\cup\ldots\cup B_n$ so that
$$\mu(A_n)=\sum_{k=1}^n\mu(B_k)\qquad\forall n\geqslant 1.$$
Since all the terms of the series are nonnegative, one sees that
$$\sup_n\mu(A_n)=\lim_n\mu(A_n)=\sum_{k=1}^\infty\mu(B_k).$$
Since $\bigcup_k B_k=\bigcup_n A_n=:A$ one gets that $\sum_{k=1}^\infty\mu(B_k)=\mu\left(A\right)$ which is the desired result.
To prove the second property, one simply uses the first one with $B_n=S\setminus A_n$. $\square$

Definition 3.10. — *Let (S,Σ,μ) be a given measure space. An element $A\in\Sigma$ is called a μ-null set if $\mu(A)=0$.*

One deduces from the above Theorem the following:

Corollary 3.11. — *The union of a countable number of μ-null sets is a μ-null set.*

3.2. Construction of measure spaces. — Let us now explain in a more concrete way how to construct explicitly measures over some measurable space (S,Σ). Since it is difficult to have a clear representation of what a given σ-algebra is, it is uneasy to construct set functions over such mathematical object. The main idea to overcome this difficulty is to use the notion of π-system. This can be done thanks to the following

Proposition 3.12. — *Let $\mathscr{I}$ be a π-system on S and let $\Sigma=\sigma(\mathscr{I})$. Suppose that μ_1 and μ_2 are two measures on (S,Σ) with $\mu_1(S)=\mu_2(S)<\infty$ and $\mu_1=\mu_2$ on $\mathscr{I}$. Then, $\mu_1=\mu_2$ on Σ.*

Proof. — Set
$$\mathcal{D} = \{F \in \Sigma \text{ such that } \mu_1(F) = \mu_2(F)\}.$$
It is easy to see that $\mathcal{D}$ is a d-system containing $\mathscr{I}$ so that $d(\mathscr{I}) \subset \mathcal{D}$. Therefore, from Dynkyn's Lemma (Prop. 1.18), $\Sigma \subset \mathcal{D}$ which is the desired result. □

Corollary 3.13. — *If two probability measures agree on a π-system, then they agree on the σ-algebra generated by that π-system.*

The above Proposition shows that, at least for finite measure spaces, it is enough to know the measure on some suitable π-system . Construction of measure on a measurable space (S, Σ) is usually a difficult task but the above result indicates that a way to do it would be to assign the value of the measure on a suitable (and rich enough) class $\mathcal{C}$ such that $\sigma(\mathcal{C}) = \Sigma$ and then to extend such a "measure" to Σ. The main path to do this is given by the following method. Introduce first some definition:

Definition 3.14. — *A collection Σ_0 of subsets of S is called an algebra on S if*

1. *$S \in \Sigma_0$;*
2. *If $A \in \Sigma_0$ then $A^c \in \Sigma_0$ where $A^c = S \setminus A$ is the complementary of A;*
3. *If $A, B \in \Sigma_0$ then $A \cup B \in \Sigma_0$.*

In other words, an algebra on S is a class containing S, stable by finite union and complement.

Definition 3.15. — *Let Σ_0 be an algebra on S (but not necessarily a σ-algebra). A mapping $\ell : \Sigma_0 \to [0, \infty]$ is said to be a* pre-measure *on Σ_0 if $\ell(\varnothing) = 0$ and for any pairwise disjoints $\{A_n\}_n \subset \Sigma_0$ with $\bigcup_n A_n \in \Sigma_0$ it holds*
$$\ell\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \ell(A_n).$$
Moreover, a pre-measure ℓ is said to be σ-finite on Σ_0 if there exists a sequence $\{A_n\}_n \subset \Sigma_0$ with $\bigcup_n A_n = S$ and $\ell(A_n) < \infty$ for any $n \in \mathbb{N}$.

The definition of pre-measure is exactly the one of measure we already saw except that it is defined on an algebra (which explains why the fact that $\bigcup_n A_n \in \Sigma_0$ is an additional assumption (obviously met if Σ_0 is a σ-algebra). In particular, if Σ_0 is a σ-algebra , then any measure on Σ_0 is a pre-measure.

Remark 3.16. — *As for measure, it is very easy to check that pre-measure is monotone: if ℓ is a pre-measure on Σ_0 and $A, B \in \Sigma_0$ are given with $A \subset B$ then $\ell(A) \leqslant \ell(B)$.*

The following shows also that the knowledge of some additive set function over some algebra yield naturally to the construction of some measure:

Theorem 3.17 **(Caratheodory's extension Theorem).** — *Let S be a given set and let Σ_0 be an algebra on S and $\Sigma = \sigma(\Sigma_0)$. If $\ell : \Sigma_0 \to [0, \infty]$ is a pre-measure on (S, Σ_0) then there exists a measure μ on (S, Σ) such that*
$$\mu(A) = \ell(A) \qquad \forall A \in \Sigma_0.$$

If moreover ℓ is a σ-finite pre-measure on Σ_0, then such a measure μ on (S, Σ) is unique *and σ-finite.*

Remark 3.18. — *We refer to the Appendix for a proof of the above fundamental result (in a slightly different form). Though it is one of the principal result in measure theory since it allows to construct measures well-adapted to practical situations, once such measures are constructed, Caratheodory's theorem becomes almost useless.* //

3.3. The fundamental case of the Lebesgue measure. — The fundamental example is of course the one of the Lebesgue measure over $(\mathbb{R}, \mathcal{B}(\mathbb{R})$. Let $S = \mathbb{R}$. We define

$$\mathcal{C}_0 = \{[a, b) \; ; \; -\infty \leqslant a \leqslant b \leqslant \infty \in \mathbb{R}\}$$

and let

$$\Sigma_0 = \left\{ \bigcup_{j=1}^{N} I_j : \; I_j \in \mathcal{C}_0 \;\; \forall j, \; I_i \cap I_j = \varnothing \text{ if } i \neq j, \; N \in \mathbb{N} \right\}.$$

One proves without major difficulty that Σ_0 is an algebra on $\mathbb{R}$. Let us define a pre-measure on Σ_0 by setting

1. $\ell([a, b)) = b - a$ for any $b \geqslant a$;
2. $\ell((-\infty, b)) = \ell((a, \infty)) = \ell(\mathbb{R}) = +\infty$;
3. $\ell\left(\bigcup_{j=1}^{N} I_j\right) = \sum_{j=1}^{N} \ell(I_j)$ if $\{I_j\}_{j=1,\ldots,N} \subset \mathcal{C}_0$ are pairwise disjoints.

Then, according to Carathéodory extension theorem and since $\sigma(\Sigma_0) = \sigma(\mathcal{C}_0) = \mathcal{B}(\mathbb{R})$ is the Borel σ-algebra , we get the following:

Theorem 3.19. — *The exists a unique measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ that we denote λ (or $\mathfrak{m}$) and such that*

$$\lambda([a, b)) = b - a \qquad \forall a < b.$$

We call this measure the Lebesgue measure *on $(\mathbb{R}, \mathcal{B}(\mathbb{R})$.*

Remark 3.20. — *We can define in the same way the Lebesgue measure on $(I, \mathcal{B}(I))$ for all $I \subset \mathbb{R}$.*

Remark 3.21. — *The measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ is σ-finite since $([-n, n))_n \uparrow \mathbb{R}$ but is not finite since*

$$\lambda(\mathbb{R}) = \lim_n \lambda([-n, n)) = \lim_n 2n = \infty.$$

Main properties of the Lebesgue measure. — Let us now list some important properties of the Lebesgue measure. First,

$$\lambda(\{x\}) = 0 \qquad \forall x \in \mathbb{R}.$$

Indeed, setting $A_n = \left[x - \frac{1}{n}, x + \frac{1}{n}\right)$ for any $n \in \mathbb{N}$, one has $(A_n)_n \downarrow \{x\}$ and

$$\lambda(A_n) = \frac{2}{n} \qquad \forall n.$$

In particular, $\lambda(A_1) < \infty$ and, using the continuity from above, we get

$$\lambda(\{x\}) = \lim_n \lambda(A_n) = 0.$$

The fact that singletons have zero Lebesgue measure shows that λ is not a counting measure. Moreover, we also have that

$$\lambda([a,b]) = \lambda((a,b] \cup \{a\}) = \lambda((a,b]) + \lambda(\{a\}) = \lambda((a,b]) \quad \forall b \geqslant a.$$

In the same way

$$\lambda((a,b)) = \lambda([a,b]) = \lambda((a,b]) = \lambda([a,b)) = b-a.$$

Moreover, one has

$$\lambda(A) = 0 \qquad \forall A \subset \mathbb{R}\ ;\ A \text{ countable}.$$

Indeed,

$$A = \bigcup_{a\in A} \{a\}$$

and, the union being countable and disjoint, $\lambda(A) = \sum_{a\in A} \lambda(\{a\}) = 0$.
Notice however that there exist subsets of $\mathcal{B}(\mathbb{R})$ which have zero Lebesgue measure and are not countable.

3.4. Atoms, purely atomic measures, diffuse measures, image measure. — Let us introduce the following definition:

Definition 3.22. — *Let (S,Σ,μ) be a given measure space and assume that*

$$\{x\} \in \Sigma \qquad \forall x \in S.$$

Given $x_0 \in S$, we say that x_0 is an atom *of (S,Σ,μ) if*

$$\mu(\{x_0\}) > 0.$$

Introducing D_μ as the set of atoms of μ, we say that μ is diffuse if $D_\mu = \varnothing$ while μ is said to be purely atomic *if D_μ is countable and $\mu(D_\mu^c) = 0$.*

Example 3.23. — *As we saw it, $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ is such that $\{x\} \in \mathcal{B}(\mathbb{R})$ for all $x \in \mathbb{R}$ and if λ is the Lebesgue measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$ then λ is diffuse on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. On the contrary, given $a \in \mathbb{R}$, δ_a is purely atomic with only one atom. More generally, discrete measures are purely atomic.*

One has the following decomposition theorem:

Theorem 3.24. — *Let (S,Σ,μ) be a σ-finite measure space with $\{x\} \in \Sigma$ for all $x \in S$. Then, there exists a diffuse measure μ_1 and a purely atomic measure μ_2 on (S,Σ) such that*

$$\mu = \mu_1 + \mu_2,$$

i.e. $\mu(A) = \mu_1(A) + \mu_2(A)$ for all $A \in \Sigma$.

3.5. Negliglible sets and completeness. — Let us begin with the following definition

Definition 3.25. — *Let (S, Σ, μ) be a given measure space. A set $A \subset S$ is said to be negligible if there exists $B \in \Sigma$ such that*

$$A \subset B \qquad \textit{and} \qquad \mu(B) = 0.$$

In other words, a set is negligible if it is contained in a μ-null set. Notice however that a negligible set is not necessarily measurable (we did not ask A to belong to Σ). Of course, μ-null sets are negligible and one has the following definition

Definition 3.26. — *A measure space (S, Σ, μ) is said to be complete if every negligible set is measurable.*

Example 3.27. — *One can show that the measure space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ (where λ is the Lebesgue measure), is not complete.*

If a measure space (S, Σ, μ) is not complete, we can always enlarge it and extend the measure μ to make it complete as illustrated by the following In general we can manually complete a measure by adding the subsets of null-measure sets to the σ-algebra on which the measure is defined.

Theorem 3.28. — *Let (S, Σ, μ) be a given measure space. Define the set of all negigible sets $\mathcal{N}$ of (S, Σ, μ), i.e.*

$$\mathcal{N} = \{N \in \mathcal{P}(S) : \; N \subseteq B \textit{ for some } B \in \Sigma \; \mu(B) = 0\}\,.$$

Introduce then

$$\overline{\Sigma} = \{A \cup N : \; A \in \Sigma, \; N \in \mathcal{N}\}\,.$$

Then $\overline{\Sigma}$ is a σ-algebra on S and the function

$$\bar{\mu}(A \cup N) := \mu(A) \qquad \forall A \in \overline{\Sigma}$$

defines a unique extension of μ that is a complete measure on $\overline{\Sigma}$.
We call $\overline{\Sigma}$ completion of Σ w.r.t. μ.

Example 3.29. — *Considering $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$, the completion of the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$ is called the Lebesgue σ-algebra (its elements are called the Lebesgue measurable sets) and the associated measure $\overline{\lambda}$ is still called the Lebesgue measure (sometimes denoted by* Leb*).*

Let us introduce the following definition

Definition 3.30. — *Let (S, Σ, μ) be a measure space. Let $P(s)$ be a statement which may or may not hold for $s \in S$. If the set $\{s \in S \; ; \; P(s) \textit{ is false } \}$ is a μ-null set then we say that P holds μ – almost everywhere.*

4. LEBESGUE INTEGRATION THEORY

The concept of Riemann integral (familiar from basic analysis and calculus) applies only to continuous functions (or else to functions having very peculiar point of discontinuity). In particular, it makes no sense in full generality to compute the integral of some measurable function f over $\mathbb{R}$. Indeed, f being 'only' measurable, it can be almost everywhere discontinuous. Moreover, if one wants to deal with measurable function over some abstract space (S, Σ), it becomes also meaningless to speak about the continuity of f and, *a fortiori*, to compute its Riemann integral. For such functions, it has been necessary to develop another notion (due primarily to H. Lebesgue) of integral with respect to some abstract measure.

In all this section, (S, Σ, μ) will denote a given measure space.

4.1. Construction of the integral and first properties. — We are interested in defining, for suitable elements f of $\mathcal{M}(\Sigma)$ the integral of f with respect to the measure μ denoted $\int f d\mu$. We introduce first the following notation: we say that $f \in \mathcal{S}(\Sigma)$ if f is a simple and measurable function and $f \in \mathcal{S}^+(\Sigma)$ if f is a Σ-measurable and nonnegative simple function. We begin with the case of simple functions:

***Definition 4.1* (Integral of simple functions). —** *Let $f \in \mathcal{S}^+(\Sigma)$ given by*

$$f = \sum_{k=1}^{m} \alpha_k \mathbb{1}_{A_k} \tag{4.2}$$

for some $\alpha_k \in [0, \infty]$ and $A_k \in \Sigma$ for any $k = 1, \ldots, m$. We then define the integral of f with respect to μ as

$$I_\mu(f) = \sum_{k=1}^{m} \alpha_k \mu(A_k) \tag{4.3}$$

with the convention $\infty \times 0 = 0 \times \infty = 0$.

***Remark 4.2*. —** *We decided here to use different notation for integral of simple function and integral of general measurable (and nonnegative) function. This is just to emphasize the fact that the second is a generalization of the first one. Of course, in practice, we shall only use the usual notation $\int_S f d\mu$, the notation $I_\mu(f)$ is needed only for the construction of the integral.*

Since the decomposition of simple function as a combination of indicator function is not unique, it is important to check that the above definition makes sense and does not depend on the choice of the decomposition:

***Proposition 4.3*. —** *Let $f \in \mathcal{S}^+(\Sigma)$ be given by*

$$f = \sum_{j=1}^{M} \beta_j \mathbb{1}_{B_j} = \sum_{i=1}^{N} \alpha_i \mathbb{1}_{A_i}, \qquad (B_j)_j \,;\, (A_i)_i \subset \Sigma$$

Then

$$\sum_{i=1}^{N}\alpha_i\mu(A_i)=\sum_{j=1}^{M}\beta_j\mu(B_j).$$

In other words, the value of $I_\mu(f)$ does not depend on the choice of the representation of f.

The basic properties of the integral of nonnegative simple functions are listed in the following:

Lemma 4.4. — *Let f, g be two nonnegative simple functions.*

1. *For any $\alpha,\beta\geqslant 0$, $\alpha\, f+\beta\, g$ is a nonnegative simple function and*
$$I_\mu(\alpha\, f+\beta\, g)=\alpha\, I_\mu(f)+\beta\, I_\mu(g).$$
2. *If $f(s)\leqslant g(s)$ for μ-almost every $s\in S$, then $I_\mu(f)\leqslant I_\mu(g)$.*
3. *If $\mu\left(\{f\neq g\}\right)=0$ then $I_\mu(f)=I_\mu(g)$.*

Remark 4.5. — *Notice that, if $f,g\in\mathcal{S}^+(\Sigma)$ then $\max(f,g)$ and $\min(f,g)$ are nonnegative simple functions.*

Proof. — 1) Let $f=\sum_{i=1}^N a_i\mathbb{1}_{A_i}$ and $g=\sum_{j=1}^M b_j\mathbb{1}_{B_j}$ be given. We assume that $A_i=f^{-1}\left(\{a_i\}\right)$ so that $\{A_i\}$ are pairwise disjoints and $S=\bigcup_{i=1}^N A_i$. In the same way, we set $B_j=g^{-1}\left(\{b_j\}\right)$. Then

$$\begin{aligned}I_\mu(\alpha\, f+\beta\, g)=&\ I_\mu\Bigg(\sum_{i=1}^{N}\alpha a_i\mathbb{1}_{A_i}+\sum_{j=1}^{M}\beta b_j\mathbb{1}_{B_j}\Bigg)\\ =&\ I_\mu\Bigg(\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\alpha a_i+\beta b_j\right)\mathbb{1}_{A_i\cap B_j}\Bigg)\end{aligned}$$

By definition we get then

$$\begin{aligned}I_\mu(\alpha\, f+\beta\, g)&=\sum_{i=1}^{N}\sum_{j=1}^{M}\left(\alpha a_i+\beta b_j\right)\mu(A_i\cap B_j)\\ &=\sum_{i=1}^{N}\sum_{j=1}^{M}\alpha a_i\mu(A_i\cap B_j)+\sum_{i=1}^{N}\sum_{j=1}^{M}\beta b_j\mu(A_i\cap B_j)\end{aligned}$$

and, since $\{A_i\}_i$ and $\{B_j\}_j$ are partitions of S, we get

$$I_\mu(\alpha\, f+\beta\, g)=\alpha\sum_i a_i\mu(A_i)+\beta\sum_j b_j\mu(B_j)=\alpha\, I_\mu(f)+\beta\, I_\mu(g).$$

2) Obviously $g-f$ is simple and nonnegative. Write its canonical representation as $g-f=$

$\sum_{i=1}^N \alpha_i \mathbb{1}_{A_i}$. Then, since $g - f \geqslant 0$, each $\alpha_i \geqslant 0$. From linearity

$$I_\mu(g) - I_\mu(f) = I_\mu(g-f) = \sum_{i=1}^N \alpha_i \mu(A_i) \geqslant 0$$

which proves the desired monotonicity.

3) If $\mu(\{f \neq g\}) = 0$, it implies that $f \leqslant g$ μ-a. e. which, from the previous point implies that $I_\mu(g) \geqslant I_\mu(f)$. But of course, since $g \leqslant f$ μ-a.e. one also has $I_\mu(f) \geqslant I_\mu(g)$. □

With the above definition, we are in position to define the integral of any nonnegative Σ-measurable function.

Definition 4.6. — *For any $f \in \mathcal{M}(\Sigma)^+$, we define*

$$\mu(f) = \int_S f d\mu := \sup_{\substack{h \leqslant f \\ h \in \mathcal{S}^+(\Sigma)}} I_\mu(h) \leqslant \infty.$$

Remark 4.7. — *Here, two different notations are introduced, the notation $\mu(f)$ for the integral of f is common in probability theory. Notice that, for $f = \mathbb{1}_A$, $\mu(f) = \mu(A)$.*

A first important property is that if f is a nonnegative simple function, then

$$I_\mu(f) = \int_S f d\mu.$$

Again, one can check that the definition makes sense. One has the following important lemma:

Lemma 4.8. — *If $f \in \mathcal{M}(\Sigma)^+$ and $\int_S f d\mu = 0$ then $\mu(\{f > 0\}) = 0$*

Proof. — For any $n \in \mathbb{N}$, set $F_n = \{f > 1/n\}$. Then, $F_n \subset F_{n+1}$ and $\bigcup_n F_n = \{f > 0\}$. Now, Prop. 3.9 allows us to prove the result arguing by contradiction. Indeed, if $\mu(\{f > 0\}) > 0$ then there exists $n_0 \in \mathbb{N}$ such that $\mu(F_{n_0}) > 0$. In this case, defining

$$h = \frac{1}{n_0} \mathbb{1}_{F_{n_0}}$$

one sees that h is a nonnegative simple function with $h \leqslant f$. Therefore,

$$\int_S f d\mu \geqslant I_\mu(h) = \frac{1}{n_0} \mu(F_{n_0}) > 0$$

which contradicts the assumption. □

From now on, we adopt the following **notation**:

if $(f_n)_n$ is a sequence of functions, we write $f_n \uparrow f$ if, for any $s \in S$, $(f_n(s))_n$ is a nondecreasing sequence that converge to some limit denoted $f(s)$.

One can now state the main convergence result

***Theorem 4.9* (Monotone convergence Theorem). —** *Let $(f_n)_n$ be a sequence of functions in $\mathcal{M}(\Sigma)^+$ such that $f_n \uparrow f$ μ-a.e. on S. Then*

$$\lim_n \int_S f_n d\mu = \int_S f d\mu$$

where the sequence $\left(\int_S f_n d\mu\right)_n$ is a nondecreasing sequence.

Proof. — Observe that $f \in \mathcal{M}(\Sigma)^+$ since it is a limit of non-negative measurable functions, and by monotonicity of integral, the sequence $\left\{\int f_n \, d\mu\right\}$ is non increasing. Define then

$$\alpha = \lim_n \int_S f_n \, d\mu.$$

Since $f_n \leqslant f$ for all n,

$$\alpha \leqslant \int_X f d\mu.$$

We show the reverse inequality holds. Take $c \in (0,1)$, $\varphi \in \mathcal{S}^+(\Sigma)$ s.t. $\varphi \leqslant f$, $\varphi = \sum_{i=1}^k a_i \mathbb{1}_{A_i}$ and define

$$E_n = \{s \in S\,; :\; c\varphi(s) \leqslant f_n(s)\} \in \Sigma$$

where $E_n \in \Sigma$ since $c\varphi - f_n$ is a linear combination of measurable functions. It is clear that $E_n \subset E_{n+1}$ for any n. Moreover

$$E_n \uparrow S.$$

Indeed, for all $s \in S$

1. $f(s) = 0$ implies $\varphi(s) \leqslant f(s) = 0$ and $f_n(s) \leqslant f(s) = 0$. Thus, $f_n(s) = \varphi(s) = 0$, so $s \in E_n$ for all n since $c\varphi(s) \leqslant f_n(s)$ (i.e. $0 \leqslant 0$).
2. $f(s) > 0$ implies

   $$c\varphi(s) < f(s) \qquad \text{since } \varphi \leqslant f \text{ and } c \in (0,1)$$

   so since $f_n \uparrow f$

   $$\exists\, n_0 \in \mathbb{N} \text{ s.t. } \quad c\varphi(s) < f_{n_0}(s) \quad \Rightarrow \quad s \in E_{n_0}$$

   implying

   $$\forall\, s \in S \quad \exists\, n \in \mathbb{N} \quad \text{s.t.} \quad s \in E_n \quad \Rightarrow \quad S = \bigcup_{n=1}^{\infty} E_n.$$

Summarizing, for every $s \in S$ there exists at least an n such that $s \in E_n$. Then we have

$$\int_S f_n \, d\mu = \int_S f_n \mathbb{1}_{E_n} + f_n \mathbb{1}_{E_n^c} \, d\mu \geqslant \int_S f_n \mathbb{1}_{E_n} \, d\mu \geqslant \int_S c\varphi \mathbb{1}_{E_n}, d\mu = c \int_{E_n} \varphi \, d\mu$$

i.e.

$$\int_S f_n \, d\mu \geqslant c \sum_{i=1}^k a_i \mu(E_n \cap A_i).$$

Using the lower continuity of the measure μ, we have $\lim_n \mu(E_n \cap A_i) = \mu(A_i)$ so that, at the limit

$$\alpha = \lim_n \int_S f_n \, d\mu \geqslant c \sum_{i=1}^{k} a_i \mu(A_i) = c \int_S \varphi \, d\mu.$$

The last inequality holds for every $c \in (0, 1)$ so

$$\alpha \geqslant \int_S \varphi \, d\mu$$

for every $\varphi \in \mathcal{S}^+(\Sigma)$ s.t. $\varphi \leqslant f$. Taking the supremum over the set of such φ we get

$$\alpha \geqslant \int_S f \, d\mu$$

which is the desired result. □

A simple Corollary of the above result is the following (equivalent) definition of the integral

Corollary 4.10. — *Let $f \in \mathcal{M}(\Sigma)^+$ be given and let $(f_n)_n$ be a sequence of Σ-measurable and nonnegative simple function such that $f_n \uparrow f$ (see Theorem 2.18). Then*

$$\int_S f d\mu = \lim_n I_\mu(f_n)$$

where the sequence $(I_\mu(f_n))_n$ is nondecreasing.

Proof. — The proof is immediate. □

Such a result has a lot of important consequences, in particular, it implies that the integral is linear:

Lemma 4.11. — *Let $\alpha, \beta \geqslant 0$ and let $f, g \in \mathcal{M}(\Sigma)^+$. Then*

$$\int_S (\alpha f + \beta g) d\mu = \alpha \int_S f d\mu + \beta \int_S g d\mu.$$

Proof. — From the above Corollary 4.10 and Theorem 2.18, there are two sequences of nonnegative simple functions $(f_n)_n$ and $(g_n)_n$ such that $f_n \uparrow f$ and $f_n \uparrow g$. Then,

$$\int_S f d\mu = \lim_n I_\mu(f_n) \qquad \text{and} \qquad \int_S g \, d\mu = \lim_n I_\mu(g_n).$$

One get then the conclusion since I_μ is linear (see Lemma 4.4). □

Until now, we only defined the integral of nonnegative function. One can extend this definition by linearity to any measurable function, using the positive and negative part of function.

Definition 4.12. — *For any $f \in \mathcal{M}(\Sigma)$, we say that f is μ-integrable and write $f \in \mathcal{L}^1(S, \Sigma, \mu)$ if*

$$\int_S |f|\, d\mu = \int_S f^+\, d\mu + \int_S f^-\, d\mu < \infty.$$

We define then the integral of f with respect to μ as

$$\int_S f\, d\mu = \int_S f^+ d\mu - \int_S f^-\, d\mu.$$

Remark 4.13. — *Notice that, by construction, $\left|\int_S f\, d\mu\right| \leqslant \int_S |f| d\mu$ (Explain why).* //

Here again, one prove without difficulty that the integral is linear:

Lemma 4.14. — *Let $f, g \in \mathcal{L}^1(S, \Sigma, \mu)$ and let $\alpha, \beta \in \mathbb{R}$, then $\alpha f + \beta g \in \mathcal{L}^1(S, \Sigma, \mu)$ and*

$$\int_S (\alpha f + \beta g)\, d\mu = \alpha \int_S f\, d\mu + \beta \int_S g\, d\mu.$$

4.2. Convergence results. — All the key convergence results on integration are consequences of the above Theorem 4.9:

***Lemma 4.15* (Fatou's Lemma).** — *For any sequence $(f_n)_n \in \mathcal{M}(\Sigma)^+$ one has*

$$\int_S \liminf_n f_n d\mu \leqslant \liminf_n \int_S f_n d\mu.$$

Proof. — For any $k \in \mathbb{N}$, set $g_k = \inf_{n \geqslant k} f_k$ so that $g_k \uparrow \liminf_n f_n$ and, from Theorem 4.9:

$$\lim_k \int_S g_k d\mu = \int_S \liminf_n f_n d\mu. \tag{4.4}$$

Now, for any $n \geqslant k$ one has $f_n \geqslant g_k$ so that $\int_S f_n d\mu \geqslant \int_S g_k d\mu$, i.e.

$$\int_S g_k d\mu \leqslant \inf_{n \geqslant k} \int_S f_n d\mu.$$

Letting now $k \to \infty$ we get

$$\lim_k \int_S g_k d\mu \leqslant \liminf_n \int_S f_n d\mu$$

which, combined with (4.4) yields the conclusion. □

***Lemma 4.16* (Reverse Fatou's Lemma).** — *Let $(f_n)_n \subset \mathcal{M}(\Sigma)^+$ be such that there exists $f \in \mathcal{M}(\Sigma)^+$ such that $f_n \leqslant f$ for any $n \in \mathbb{N}$ and $\int_S f d\mu < \infty$. Then*

$$\int_S \limsup_n f_n d\mu \geqslant \limsup_n \int_S f_n d\mu.$$

Proof. — The proof consists only in applying Fatou's Lemma to $g_n = f - f_n$. □

***Theorem 4.17* (Dominated convergence Theorem).** — *Let $(f_n)_n \subset \mathcal{M}(\Sigma)$ be such that there exists $f \in \mathcal{M}(\Sigma)$ such that*

$$\lim_n f_n(s) = f(s) \qquad \text{for } \mu\text{-almost every } s \in S.$$

Assume moreover that there exists $g \in \mathcal{L}^1(S,\Sigma,\mu)$ such that $|f_n(s)| \leqslant g(s)$ for any $n \in \mathbb{N}$ and μ-almost every $s \in S$. Then, $f, f_n \in \mathcal{L}^1(S,\Sigma,\mu)$ for any $n \in \mathbb{N}$ with

$$f_n \to f \qquad \text{in } \mathcal{L}^1(S,\Sigma,\mu)$$

i.e. $\lim_n \int_S |f_n - f|\, d\mu = 0$. *In particular,*

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

Proof. — For any $n \in \mathbb{N}$ one has $|f_n - f| < 2g$ with $2g \in \mathcal{L}^1(S,\Sigma,\mu)$. By the reverse Fatou's Lemma

$$\limsup_n \int_S |f_n - f| d\mu \leqslant \int_S \liminf_n |f_n - f| d\mu = \int_S 0\, d\mu = 0.$$

This proves the result. □

One has a kind of characterization of the convergence in $\mathcal{L}^1(S,\Sigma,\mu)$:

***Lemma 4.18* (Scheffé's Lemma).** — *Assume that $f_n, f \in \mathcal{L}^1(S,\Sigma,\mu)$ and that $f_n \to f$ μ-a.e. Then*

$$\lim_n \int_S |f_n - f|\, d\mu = 0 \Longleftrightarrow \lim_n \int_S |f_n| d\mu = \int_S |f|\, d\mu.$$

The proof of Scheffé's Lemma is left as an ***Exercise***.

4.3. Examples – Riemann vs Lebesgue. — Let us explicit here some computations of the integral with respect to given measure.

***Proposition 4.19*.** — *Let (S,Σ) be a given measurable space, let $x \in S$ and consider the measure $\mu = \delta_x$ on (S,Σ). Then, for any $f \in \mathcal{M}(\Sigma)^+$ one has*

$$\int_S f d\delta_x = f(x). \tag{4.5}$$

More generally, $f \in \mathcal{L}^1(S,\Sigma,\delta_x)$ if and only if $f^\pm(x) < \infty$ and in this case (4.5) holds true.

Proof. — The proof is very simple but it provides the general procedure to compute integral. It consists in the following steps:

1. *First step:* Assume that f is an indicator function, i.e $f = \mathbb{1}_A$ for some $A \in \Sigma$. Then, by definition
$$\int_S f d\delta_x = I_{\delta_x}(f) = \delta_x(A) = \mathbb{1}_A(x) = f(x).$$
This proves (4.5) for indicator function.

2. *Second step:* Assume now that f is a nonnegative simple function, $f \in \mathcal{S}^+(\Sigma)$. Then, $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$ for $a_i \geqslant 0$ and $A_i \in \Sigma$ for $i = 1, \ldots, n$. Then, using again the definition of I_{δ_x} (or equivalently, using the first step and the linearity of I_{δ_x} we get
$$\int_S f d\delta_x = I_{\delta_x}(f) = \sum_{i=1}^n a_i \delta_x(A_i) = \sum_{i=1}^n a_i \mathbb{1}_{A_i}(x) = f(x)$$
which proves (4.5) for $f \in \mathcal{S}^+(\Sigma)$.
3. *Third step:* Let now $f \in \mathcal{M}(\Sigma)^+$. Then, from Corollary 4.10 and Theorem 2.18, there is a sequence $(f_n)_n \subset \mathcal{S}^+(\Sigma)$ such that $f_n \uparrow f$ and
$$\int_S f d\delta_x = \lim_n I_{\delta_x}(f_n) = \lim_n f_n(x)$$
where we used the second step for the last identity. Since $\lim_n f_n(x) = f(x)$, we have (4.5).

The second part of the result is obvious. □

The above can be extended to general discrete measure

Proposition 4.20. — *Let (S, Σ) be a given measure space. Let $J \subset S$ be countable and let $(m_x)_{x \in J}$ be given with $m_x > 0$ for all $x \in J$. Set then $\mu = \sum_{x \in J} m_x \delta_x$. Then, for any $f \in \mathcal{M}^+(\Sigma)$ one has*
$$\int_S f d\mu = \sum_{x \in J} m_x f(x).$$

It is important now to understand the meaning of the integral with respect to the Lebesgue measure. One has the following :

Proposition 4.21. — *Let $([a, b], \mathcal{L}([a, b]), \lambda)$, where λ is Lebesgue measure and $\mathcal{L}([a, b])$ the Lebesgue σ-algebra . Then if $f : [a, b] \to \mathbb{R}$ is Riemann integrable it is also Lebesgue-integrable w.r.t. λ and*
$$\int_{[a,b]} f d\lambda = \int_a^b f(x) dx.$$

Remark 4.22. — *In practice, Riemann integrable functions are Lebesgue integrable but the converse is not true since $\mathbb{1}_{\mathbb{Q}}$ is not Riemann integrable while it is Lebesgue integrable with*
$$\int_{[a,b]} \mathbb{1}_{\mathbb{Q}} d\lambda = I_\lambda(\mathbb{1}_{\mathbb{Q} \cap [a,b]}) = \lambda(\mathbb{Q} \cap [a, b]) = 0.$$

4.4. Image measure and integral transform. — We begin with the following

Proposition 4.23. — *Let (S, Σ, μ) be a measure space and $(X, \mathcal{F})$ a measurable space. Let $F : S \to X$ be a $\Sigma/\mathcal{F}$-measurable function. Then the function defined for all $A \in \mathcal{F}$ by*
$$(\mu \circ F^{-1})(A) := \mu\left(F^{-1}(A)\right) = \mu\left(\{x \in X : F(x) \in A\}\right)$$

is a measure on $(X, \mathcal{F})$*. It is called the image measure of* μ *under* F *(or the push-forward measure of* μ *by* F*).*

Proof. — Exercise. □

Remark 4.24. — *If* μ *is* σ*-finite measure on* (S, Σ) *then so is* $\mu \circ F^{-1}$ *on* $(X, \mathcal{F})$*.*

One has then the following which allows to compute integral with respect to the image measure:

***Theorem 4.25* (Change of variable formula).** — *Let* (S, Σ, μ) *be a measure space and* $(X, \mathcal{F})$ *a measurable space. Let* $F : S \to X$ *be a* $\Sigma/\mathcal{F}$*-measurable function and let* $\nu = \mu \circ F^{-1}$ *denote the image measure of* μ *under* F*. Let* $f \in \mathcal{M}^+(\mathcal{F})$ *be given. Then*

$$\int_X f d\nu = \int_S f \circ F d\mu. \tag{4.6}$$

More generally, $f \in \mathcal{L}^1(X, \mathcal{F}, \nu)$ *if and only if* $f \circ F \in \mathcal{L}^1(S, \Sigma, \mu)$ *and in this case* (4.6) *holds.*

Proof. — The proof follows the same step as the one of Propositon 4.19, i.e prove first the result for indicator functions, use then the linearity to show the result for simple function and pass to the limit to get the result for general nonnegative and measurable functions. □

5. Integrals over subsets and the Radon-Nikodym Theorem

We begin by the definition of integral over some measurable set:

Definition 5.1. — *For any* $f \in \mathcal{M}(\Sigma)^+$ *and any* $A \in \Sigma$*, we set*

$$\int_A f\, d\mu = \int_S (f \mathbb{1}_A)\, d\mu$$

and we define the measure $\mu_f := f\mu$ *as*

$$\mu_f(A) = \int_A f\, d\mu.$$

It is a very simple exercise to show that μ_f is a measure on (S, Σ). More generally, one can prove the following

Proposition 5.2. — *Let* $f \in \mathcal{M}(\Sigma)^+$ *and* $h \in \mathcal{M}(\Sigma)$*. Then,* $h \in \mathcal{L}^1(S, \Sigma, \mu_f)$ *if and only if* $fh \in \mathcal{L}^1(S, \Sigma, \mu)$ *and then*

$$\int_S fh\, d\mu = \int_S h\, d\mu_f.$$

Notice that, if $f \in \mathcal{M}(\Sigma)^+$ and $A \in \Sigma$ then

$$\mu(A) = 0 \Longrightarrow \mu_f(A) = 0.$$

One may ask if there are other types of measures ν on (S, Σ) that satisfy the above property. The Radon-Nikodym Theorem provides an answer to this question for σ-finite measures.

***Theorem 5.3* (Radon-Nikodym Theorem).** — *Let (S,Σ) be a measurable space and let ν and μ be two σ-finite measures on (S,Σ). Then, the following are equivalent:*

(a) for any $A \in \Sigma$, $\mu(A) = 0$ implies that $\nu(A) = 0$;
(b) $\nu = \mu_f = f\mu$ for some $f \in \mathcal{M}(\Sigma)^+$.

In this case, f is defined uniquely (modulo μ-null sets).

We refer to [5] or [3] for a complete proof.

***Definition 5.4*.** — *Under the assumptions of the above Theorem, assuming (a) and (b) to hold, we say that ν is absolutely continuous with respect to μ and write*

$$\nu \ll \mu.$$

We shall say that the function f is a version of the density of ν with respect to μ or equivalently, that f is the Radon-Nikodym derivative of ν with respect to μ and denote

$$f = \frac{d\nu}{d\mu} \qquad \mu - a.\ e.$$

6. Product structure: Fubini’s Theorem

In all this section, we will assume that (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2) are two measure spaces. We will denote

$$S = S_1 \times S_2 = \{(s_1, s_2),\ s_i \in S_i, i = 1, 2\}$$

the cartesian product of S_1 and S_2. The scope is to build a σ-algebra Σ on S and a measure on (S,Σ) such that, given $A_i \in \Sigma_i$ $(i = 1, 2)$ the cartesian product $A_1 \times A_2 \in \Sigma$ and

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2). \tag{6.7}$$

Let us begin with the construction of the σ-algebra on X.

6.1. Product σ-algebra. — If $A_1 \in \Sigma_1$ and $A_2 \in \Sigma_2$ we shall call measurable rectangle the cartesian product

$$A = A_1 \times A_2.$$

We denote by $\mathcal{C}_0$ the collection of all measurable rectangles in S

***Definition 6.1*.** — *We denote by $\Sigma = \Sigma_1 \otimes \Sigma_2$ the σ-algebra on S generated by $\mathcal{C}_0$, i.e.*

$$\Sigma_1 \otimes \Sigma_2 = \sigma(\mathcal{C}_0) = \sigma\left(\{A \times B\ ;\ A \in \Sigma_1,\ B \in \Sigma_2\}\right).$$

We call it the product σ-algebra of (Σ_1, Σ_2) and the measurable space $(S,\Sigma) = (S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$ is called the product spaces of the measurable spaces (S_1, Σ_1) and (S_2, Σ_2).

If $E \subset S$, for any $s_1 \in S_1$ we denote the s_1-section of E the set

$$\tau_{s_1}(E) = \{s_2 \in S_2 \,:\, (s_1, s_2) \in E\} \subset S_2$$

and similarly, for any $s_2 \in S_2$ we define the s_2-section as

$$T_{s_2}(E) = \{s_1 \in S_1 \,:\, (s_1, s_2) \in E\} \subset S_1.$$

Given a function $f \,:\, S \to \mathbb{R}$, for each $s_i \in S_i$ $(i = 1, 2)$ we define

$$f(s_1, \cdot) \,:\, S_2 \to \mathbb{R}$$

the mapping $s_2 \in S_2 \mapsto f(s_1, S_2)$ and

$$f(\cdot, s_2) \,:\, S_1 \to \mathbb{R}$$

the mapping $s_1 \in S_1 \mapsto f(s_1, s_2)$
One has the following

Lemma 6.2. — *Let $s_i \in S_i$ $(i = 1, 2)$ be given.*

1. *If $E \in \Sigma_1 \otimes \Sigma_2$ then $\tau_{s_1}(E) \in \Sigma_2$ and $T_{s_2}(E) \in \Sigma_1$;*
2. *If $f \,:\, S \to \mathbb{R}$ is $(\Sigma_1 \otimes \Sigma_2)$-measurable, then $f(s_1, \cdot)$ is Σ_2-measurable and $f(\cdot, s_2)$ is Σ_1-measurable.*

Remark 6.3. — *Notice that, actually, $\Sigma_1 \otimes \Sigma_2$ is the smallest of all the σ-algebras $\mathcal{A}$ on $S = S_1 \times S_2$ such that the projection mappings:*

$$p_i \,:\, \boldsymbol{s} = (s_1, s_2) \in S_1 \times S_2 \mapsto p_i(\boldsymbol{s}) = s_i \in S_i$$

are $\mathcal{A}/\Sigma_i$-measurable $(i = 1, 2)$.

Example 6.4. — *If $S_1 = S_2 = \mathbb{R}$ and $\Sigma_1 = \Sigma_2 = \mathcal{B}(\mathbb{R})$, then the product σ-algebra $\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ is exactly the Borel σ-algebra $\mathcal{B}(\mathbb{R}^2)$ on $\mathbb{R}^2$, i.e*

$$\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) = \sigma\left(\left\{ \text{ open subset of } \mathbb{R}^2 \right\}\right).$$

In particular, any continuous function $f \,:\, \mathbb{R}^2 \to \mathbb{R}$ is $\mathcal{B}(\mathbb{R}^2)$-measurable.

Of course the above procedure can be extended, by a simple induction, to the product of any finite number of measurable spaces:

Definition 6.5. — *Let $(S_i, \Sigma_i)_{i=1,\ldots,d}$ is a finite family of measurable spaces. On $S = S_1 \times S_2 \ldots \times S_d$ one can construct the product σ-algebra*

$$\Sigma_1 \otimes \Sigma_2 \otimes \ldots \otimes \Sigma_d = (\Sigma_1 \otimes \ldots \Sigma_{d-1}) \otimes \Sigma_d$$

where $\Sigma_1 \otimes \ldots \Sigma_{d-1}$ is the product σ-algebra on $S_1 \times \ldots \times S_{d-1}$.

Remark 6.6. — *An important property of the product of σ-algebrasis its associativity, i.e. if $(S_i, \Sigma_i)_{i=1,\ldots,3}$ are measurable spaces then*

$$\Sigma_1 \otimes \Sigma_2 \otimes \Sigma_3 = (\Sigma_1 \otimes \Sigma_2) \otimes \Sigma_3 = \Sigma_1 \otimes (\Sigma_2 \otimes \Sigma_3)$$

where $\Sigma_1 \otimes \Sigma_2$ is the product σ-algebra on $S_1 \times S_2$ while $\Sigma_2 \otimes \Sigma_3$ is the product σ-algebra on $S_2 \times S_3$.

Thanks to the above procedure, one can define, for any $d \geqslant 2$, product σ-algebra

$$\underbrace{\mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \otimes \ldots \otimes \mathcal{B}(\mathbb{R})}_{d \text{ times}}$$

and prove that it exactly coincides with the Borel σ-algebra $\mathcal{B}(\mathbb{R}^d)$ on $\mathbb{R}^d$ generated by the open subsets of $\mathbb{R}^d$. In particular,

$$\mathcal{B}(\mathbb{R}^d) = \mathcal{B}(\mathbb{R}^k) \otimes \mathcal{B}(\mathbb{R}^n)$$

for any $k, n \in \mathbb{N}$ with $k + n = d$.

6.2. Fubini Theorem. — Among all the measures that can be constructed over $(X, \Sigma_1 \times \Sigma_2)$ there is one which satisfies (6.7) whenever μ_1, μ_2 are both σ-finite:

Proposition 6.7. — *Assume that (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2) are both σ-finite measure spaces. Then, there exists a unique measure μ on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$ such that*

$$\mu(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2) \qquad \forall A_1 \in \Sigma_1 \ ; \ \ A_2 \in \Sigma_2.$$

Moreover, μ is σ-finite on $(S_1 \times S_2, \Sigma_1 \otimes \Sigma_2)$.
Such a measure is usually denoted $\mu = \mu_1 \otimes \mu_2$ and is called the product measure of μ_1 and μ_2.

Example 6.8. — *Applying the above result to $S_1 = S_2 = \mathbb{R}$ and $\Sigma_1 = \Sigma_2 = \mathcal{B}(\mathbb{R})$, $\mu_1 = \mu_2 = \lambda$, one obtains that*

$$\lambda_2 = \lambda \otimes \lambda$$

is a σ-finite measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}))$ such that

$$\lambda_2(A \times B) = \lambda(A)\lambda(B) \qquad \forall A, B \in \mathcal{B}(\mathbb{R}).$$

The measure λ_2 is called the Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$ where we recall that $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}^2$.

The computation of integral with respect to $\mu_1 \otimes \mu_2$ is given by the two following results

Proposition 6.9 (**Tonelli Theorem**). — *Assume that (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2) are both σ-finite measure spaces and let $\mu = \mu_1 \otimes \mu_2$ be the product measure on $(X, \Sigma_1 \otimes \Sigma_2)$ with $S = S_1 \times S_2$. Let $f : S \to \mathbb{R}^+$ be $\Sigma_1 \otimes \Sigma_2$-measurable. Then,*

1. *the mapping $F_1 : s_1 \in S_1 \mapsto \displaystyle\int_{S_2} f(s_1, \cdot) d\mu_2$ is Σ_1-measurable,*
2. *the mapping $F_2 : s_2 \in S_2 \mapsto \displaystyle\int_{S_1} f(\cdot, s_2) d\mu_1$ is Σ_2-measurable,*

3. *The integral of f over S with respect to μ is given by*

$$\int_S f d\mu = \int_{S_1} F_1(s_1) d\mu_1(s_1) = \int_{S_2} F_2(s_2) d\mu_2(s_2). \tag{6.8}$$

i.e.

$$\begin{aligned}\int_{S_1\times S_2} f(s_1,s_2)\, d\left(\mu_1\otimes\mu_2\right)(s_1,s_2) &= \int_{S_1}\left(\int_{S_2} f(s_1,s_2)\, d\mu_2(s_2)\right) d\mu_1(s_1)\\ &= \int_{S_2}\left(\int_{S_1} f(s_1,s_2)\, d\mu_1(s_1)\right) d\mu(s_2).\end{aligned}$$

The above result covers the case of nonnegative measurable mappings on the product space. If one deals with more general mappings, under some additional integrability assumption, one has

Proposition 6.10 **(FUBINI THEOREM)**. — *Assume that (S_1, Σ_1, μ_1) and (S_2, Σ_2, μ_2) are both σ-finite measure spaces and let $\mu = \mu_1 \otimes \mu_2$ be the product measure on $(S, \Sigma_1 \otimes \Sigma_2)$ with $S = S_1 \times S_2$. Let $f : S \to \mathbb{R}$ be μ-integrable, i.e.*

$$\int_{S_1\times S_2} |f|\, d(\mu_1\otimes\mu_2) < \infty.$$

Then,

1. *the mapping $F_2 : s_2 \in S_2 \mapsto \displaystyle\int_{S_1} f(s_1,s_2)\, d\mu_1(s_1)$ is well-defined and μ_2-integrable;*
2. *the mapping $F_1 : s_1 \in S_1 \mapsto \displaystyle\int_{S_2} f(s_1,s_2)\, d\mu_2(s_2)$ is well-defined and μ_1-integrable*
3. *The integral of f over S with respect to μ is given by*

$$\int_S f\, d\mu = \int_{S_1} F_1(s_1)\, d\mu_1(s_1) = \int_{S_2} F_2(s_2)\, d\mu_2(s_2).$$

Remark 6.11. — *Using Tonelli Theorem, one actually sees that f is μ-integrable if and only if*

$$\int_{S_1}\left(\int_{S_2} |f(s_1,s_2)|\, d\mu_2(s_2)\right) d\mu_1(s_1) < \infty$$

which is equivalent to

$$\int_{S_2}\left(\int_{S_1} |f(s_1,Ss_2)|\, d\mu_1(s_1)\right) d\mu_2(s_2) < \infty$$

since $|f|$ is nonnegative.

Exercise 1. — *Let $S_1 = S_2 = \mathbb{N}$ and let $\mu_1 = \mu_2$ be the counting measure on $\mathbb{N}$. Define $f : \mathbb{N}\times\mathbb{N} \to \mathbb{R}$ by*

$$f(k,n) = \begin{cases} 1 & \textit{if } k = n \\ -1 & \textit{if } k = n+1\ . \\ 0 & \textit{otherwise}. \end{cases}$$

Show that

$$\int_{\mathbb{N}}\left(\int_{\mathbb{N}} f(k,n)\,d\mu(k)\right)\,d\mu(n) \neq \int_{\mathbb{N}}\left(\int_{\mathbb{N}} f(k,n)\,d\mu(n)\right)\,d\mu(k).$$

Why is this not a contradiction to the Fubini theorem ?

Corollary 6.12. — *Let (X,Σ,μ) be a given σ-finite measure space and let $f : X \to \mathbb{R}^+$ be Σ-measurable. Then*

$$\int_X f\,d\mu = \int_{\mathbb{R}^+} \mu\left(\{x \in X \;;\; f(x) > t\}\right)\,d\lambda(t)$$

where λ is the Lebesgue measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$.

Proof. — Consider the product space $Y = X \times \mathbb{R}^+$ endowed with the σ-algebra $\Sigma \otimes \mathcal{B}(\mathbb{R})$ and the measure $\nu = \mu \otimes \lambda$. Let $\Phi : Y \to \mathbb{R}$ be defined by

$$\Phi(x,t) = \mathbb{1}_{[0,f(x))}(t) = \mathbb{1}_{\{f>t\}}(x).$$

Clearly

$$f(x) = \int_{\mathbb{R}^+} \Phi(x,t)\,d\lambda(t) \qquad \forall x \in X$$

and therefore

$$\int_X f\,d\mu = \int_X \left(\int_{\mathbb{R}^+} \Phi(x,t)\,d\lambda(t)\right)\,d\mu(x)$$

which, according to Tonelli theorem, reads

$$\int_X f\,d\mu = \int_{\mathbb{R}^+} \left(\int_X \Phi(x,t)\,d\mu(x)\right)\,d\lambda(t).$$

Since, for any $t \in \mathbb{R}^+$

$$\int_X \Phi(x,t)\,d\mu(x) = \mu\left(\{f > t\}\right)$$

we get the conclusion. □

Example 6.13. — *Let $D = \{(x,y) \in \mathbb{R}^2 \;;\; x^2+y^2 \leqslant 1\}$ be the unit disc of $\mathbb{R}^2$. One has $D \in \mathcal{B}(\mathbb{R}^2)$ since $D = F^{-1}([0,1])$ with $F(x,y) = x^2+y^2$ is a continuous and thus $(\mathcal{B}(\mathbb{R}^2))$-measurable function. Let us compute $\lambda_2(D)$. One has,*

$$\lambda_2(D) = \int_{\mathbb{R}^2} \mathbb{1}_D(x,y)\,d\lambda_2(x,y)$$

and, according to Tonelli Theorem, since $\lambda_2 = \lambda \otimes \lambda$

$$\lambda_2(D) = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbb{1}_D(x,y)\,d\lambda(x)\right)\,d\lambda(y).$$

For a given $y \in \mathbb{R}$, let us compute

$$F(y) = \int_{\mathbb{R}} \mathbb{1}_D(x,y)\,d\lambda(x).$$

Notice that $F(y) = \lambda(T_y(D))$ *with the notations used already since, given* $y \in \mathbb{R}$, $(x, y) \in D$ *if and only if* $x \in T_y(D)$ *which amounts to* $x^2 \leqslant 1 - y^2$. *Thus,* $T_y(D) = \varnothing$ *if* $y^2 > 1$ *i.e.* $F(y) = 0$ *if* $y \notin [-1, 1]$. *Moreover, for* $y \in [-1, 1]$,

$$x \in T_y(D) \Longleftrightarrow x \in [-\sqrt{1-y^2}, \sqrt{1-y^2}]$$

i.e.

$$F(y) = \lambda(T_y(D)) = \lambda\left([-\sqrt{1-y^2}, \sqrt{1-y^2}]\right) = 2\sqrt{1-y^2}.$$

To summarize

$$F(y) = 2\sqrt{1-y^2}\mathbb{1}_{[-1,1]}(y)$$

and

$$\lambda_2(D) = \int_{\mathbb{R}} F(y)\, d\lambda(y) = 2\int_{[-1,1]} \sqrt{1-y^2}\, d\lambda(y).$$

Since $y \mapsto \sqrt{1-y^2}$ *is continuous, the Lebesgue integral coincide with the Riemann integral and*

$$\lambda_2(D) = 2\int_{-1}^{1} \sqrt{1-y^2}dy.$$

Setting $y = \sin\theta$, $dy = \cos\theta\, d\theta$ *and*

$$\lambda_2(D) = 2\int_{-\pi/2}^{\pi/2} \sqrt{1-\sin^2\theta}\cos\theta\, d\theta = 2\int_{-\pi/2}^{\pi/2} \cos^2\theta\, d\theta.$$

One can use then integration by parts or remember that $2\cos^2\theta = 1 + \cos 2\theta$ *to get*

$$\lambda_2(D) = \int_{-\pi/2}^{\pi/2} (1 + \cos 2\theta)\, d\theta = (\theta + \sin 2\theta)\Big|_{-\pi/2}^{\pi/2} = \pi$$

which is indeed the area of the unit disc.
If we consider now the cylinder $A = \{(x, y, z) \in \mathbb{R}^3\ ;\ x^2 + y^2 \leqslant 1 \quad z \in [0, h]\}$ *for* $h > 0$, *we see that*

$$A = D \times [0, h]$$

where $D \subset \mathbb{R}^2$ *and* $[0, h] \subset \mathbb{R}$. *Thus,* A *is a measurable rectangle of* $\mathbb{R}^3 = \mathbb{R}^2 \times \mathbb{R}$ *endowed with* $\mathcal{B}(\mathbb{R}^2) \otimes \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^3)$. *Moreover, by construction*

$$\lambda_3(A) = \lambda_2(D)\lambda([0, h]) = \pi\, h$$

which is indeed the volume of a cylinder.

7. APPENDIX: CONSTRUCTION OF MEASURES

In Section 3, we have provided the definition of a measure as a nonnegative real-valued set function, together with some of its properties. We also briefly explained in Section 3.2 the main steps in showing the existence and constructing a measure on a sufficiently rich class of subsets thanks to Caratheodory extension theorem (see Theorem 3.17). We give complete proof of this result here in this Appendix. Notations may slightly differ with respect to the core of these notes.

Definition 7.1. — *A function $\gamma : \mathcal{P}(X) \to [0, \infty]$ is called* outer measure *if*

1. $\gamma(\varnothing) = 0$
2. $A \subseteq B$ *implies* $\gamma(A) \leqslant \gamma(B)$
3. $A = \bigcup_{n=1}^{\infty} A_n$ *implies* $\gamma(A) \leqslant \sum_{n=1}^{\infty} \gamma(A_n)$

namely it is a monotonic, σ-sub-additive set function.

A common way to generate outer measures is as follows.

Proposition 7.2. — *Suppose $\mathcal{C}$ is a collection of subsets of X such that $\varnothing \in \mathcal{C}$ and there exist $C_1, C_2, \ldots$ belonging to $\mathcal{C}$ such that $X = \bigcup_{i=1}^{\infty} C_i$. Suppose $\ell : \mathcal{C} \to [0, \infty]$ is given with $\ell(\varnothing) = 0$. Define*

$$\mu^*(A) = \inf_{\{B_n\}} \left\{ \sum_{n=1}^{\infty} \ell(B_n) : \ \{B_n\} \subset \mathcal{C}, \ A \subseteq \bigcup_{n=1}^{\infty} B_n \right\} \text{(1)}$$

Then, $\mu^\star$ is an outer measure.

Proof. — First, for $A = \varnothing$ take $\{B_n\} = \{\varnothing, \varnothing, \ldots\}$, so $\mu^*(A) = \sum_{n=1}^{\infty} \ell(\varnothing) = 0$.
Second, let $A \subseteq B$. Then the collection of countable covers of A contains the corresponding collection for B, i.e.

$$\left\{ \{A_n\} \in \mathcal{C} : \ A \subseteq \bigcup_{n=1}^{\infty} A_n \right\} \supseteq \left\{ \{B_n\} \in \mathcal{C} : \ A \subseteq B \subseteq \bigcup_{n} B_n \right\}$$

(verify this) from which taking the infimum on a larger space yields

$$\mu^*(A) = \inf_{\{A_n\}} \sum_{n=1}^{\infty} \ell(A_n) \leqslant \inf_{\{B_n\}} \sum_{n=1}^{\infty} \ell(B_n) = \mu^*(B)$$

It remains then to prove the σ-sub-additivity. Let $A_1, A_2, \ldots$ be subsets of X and let $\varepsilon > 0$ be given. For each i, there exist $C_{i1}, C_{i2}, \ldots$ belonging to $\mathcal{C}$ such that $A_i \subset \bigcup_{j=1}^{\infty} C_{ij}$ and

$$\sum_{j=1}^{\infty} \ell(C_{i,j}) \leqslant \mu^*(A_i) + \frac{\varepsilon}{2^i}.$$

(1) Here the sequences $\{B_n\}$'s (i.e. the sequences $\{B_n\} \in \mathcal{C}$ s.t. $A \subseteq \bigcup_{n=1}^{\infty} B_n$) are called *countable covers of A*.

Since $A = \bigcup_i A_i \subseteq \bigcup_i \bigcup_j C_{i,j}$, the family $\{C_{i,j}\}_{i,j}$ is a countable cover of A in $\mathcal{C}$ and we have

$$\mu^*(A) = \mu^*\left(\bigcup_i A_i\right) \leqslant \sum_i \sum_j \ell(C_{i,j}) \leqslant \sum_i \left(\mu^*(A_i) + \frac{\varepsilon}{2^i}\right) = \sum_i \mu^*(A_i) + \varepsilon$$

and since ε is arbitrary $\mu^* \left(\bigcup_i A_i\right) \leqslant \sum_i \mu^*(A_i)$. □

Example 7.3. — *Let $X = \mathbb{R}$ and let $\mathcal{C}$ be the collection of intervals of the form $(a, b]$, (i.e with the notations introduced earlier $\mathcal{C} = \mathcal{C}_0$). Let*

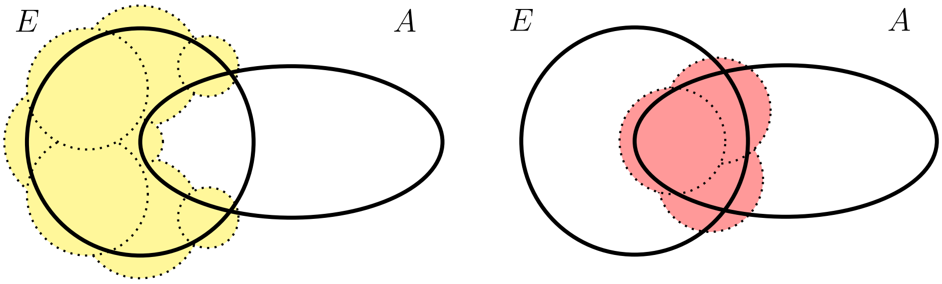
$$\ell((a, b]) = b - a.$$

Define the associated μ^* as above. Then, μ^* is an outer measure, but we will see later on that μ^* is not a measure on $\mathcal{P}(X)$.

Observe that for all $A, E \in \mathcal{P}(X)$ we have

$$\mu^*(E) = \mu^*\Big((E \cap A)\bigcup(E \cap A^c)\Big) \leqslant \mu^*(E \cap A) + \mu^*(E \cap A^c) \tag{7.9}$$

by sub-additivity of μ^*. That is, in general the outer measure of E is less or equal than the sum of the outer measure of $E \cap A$ (the infimum of the red region in figure) and of the outer measure of $E \cap A^c$ (the infimum of the yellow region in figure).



Definition 7.4. — *Let $\mu^\star$ be an outer measure. A set $A \in \mathcal{P}(X)$ is said to be μ^**-measurable *if*

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c), \qquad \forall E \in \mathcal{P}(X),$$

and define

$$\mathcal{G} = \{A \in \mathcal{P}(X) : \ A \text{ is } \mu^*\text{-measurable}\}.$$

A motivation for the notion of μ^*-measurability is given by the following

***Theorem 7.5* (Caratheodory's Theorem).** —
$\mathcal{G}$ is a σ-algebra and $\mu := \mu^*|_{\mathcal{G}}$ is a measure on $\mathcal{G}$.

Proof. — *First step: $\mathcal{G}$ is an algebra.*

– $X \in \mathcal{G}$ since

$$\mu^*(E) = \mu^*(E \cap X) + \mu^*(E \cap \varnothing)$$

– $A \in \mathcal{G}$ implies $A^c \in \mathcal{G}$ since

$$\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) + \mu^*(E \cap \underbrace{(A^c)^c}_{A})$$

– Let $A, B \in \mathcal{G}$,let us prove that $A \bigcup B \in \mathcal{G}$: For all $E \in \mathcal{P}(X)$

$$\mu^*(E) = \mu^*(E \cap B) + \mu^*(E \cap B^c);$$

substitute E (which is arbitrary) with $E \cap A$ and $E \cap A^c$ to get,

$$\begin{aligned}\mu^*(E \cap A) &= \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c)\\ \mu^*(E \cap A^c) &= \mu^*(E \cap A^c \cap B) + \mu^*(E \cap A^c \cap B^c)\end{aligned}$$

and sum side by side to get, $\forall E \in \mathcal{P}(X)$,

$$\underbrace{\mu^*(E)}_{\text{since } A \in \mathcal{G}} = \underbrace{\mu^*(E \cap A \cap B)}_{1} + \underbrace{\mu^*(E \cap A \cap B^c)}_{2} + \underbrace{\mu^*(E \cap A^c \cap B)}_{3} + \mu^*(E \cap \underbrace{A^c \cap B^c}_{(A \bigcup B)^c}). \tag{7.10}$$

Substitute now E with $E \cap (A \bigcup B)$ in the last expression to get, $\forall E \in \mathcal{P}(X)$,

$$\begin{aligned}\mu^*(E \cap (A \cup B)) &= \mu^*(E \cap \underbrace{(A \cup B) \cap A \cap B}_{A \cap B}) + \mu^*(E \cap \underbrace{(A \cup B) \cap A \cap B^c}_{A \cap B^c})\\ &\quad + \mu^*(E \cap \underbrace{(A \cup B) \cap A^c \cap B}_{A^c \cap B}) + \mu^*(E \cap \underbrace{(A \cup B) \cap (A \cup B)^c}_{\varnothing})\\ &= \underbrace{\mu^*(E \cap A \cap B)}_{1} + \underbrace{\mu^*(E \cap A \cap B^c)}_{2} + \underbrace{\mu^*(E \cap A^c \cap B)}_{3}\end{aligned} \tag{7.11}$$

so from (7.10) we have that, $\forall E \in \mathcal{P}(X)$,

$$\mu^*(E) = \mu^*(E \cap (A \cup B)) + \mu^*(E \cap (A \cup B)^c)$$

so that $A \bigcup B \in \mathcal{G}$. So $\mathcal{G}$ is an algebra.

Second step: $\mathcal{G}$ is a σ-algebra : Let $\{A_n\}$ be pairwise disjoint sets in $\mathcal{G}$ and let $B_n = \bigcup_{i=1}^n A_i$ and $A = \bigcup_i A_i$. Given $E \in \mathcal{P}(X)$ we have

$$\mu^*(E \cap B_n) = \mu^*(E \cap B_n \cap A_n) + \mu^*(E \cap B_n \cap A_n^c) = \mu^*(E \cap A_n) + \mu^*(E \cap B_{n-1}).$$

In the same way $\mu^*(E \cap B_{n-1}) = \mu^*(E \cap A_{n-1}) + \mu^*(E \cap B_{n-2})$ so that, iterating,

$$\mu^*(E \cap B_n) \geqslant \sum_{k=1}^n \mu^*(E \cap A_k).$$

Since $B_n \in \mathcal{G}$ for any n (since $\mathcal{G}$ is an algebra) we have

$$\mu^*(E) = \mu^*(E \cap B_n) + \mu^*(E \cap B_n^c) \geqslant \sum_{k=1}^n \mu^*(E \cap A_k) + \mu^*(E \cap A^c)$$

where we used the monotonicity of the outer measure and the fact that $A^c \subset B_n^c$. Let then $n \to \infty$ and using the countable σ-sub-additivity

$$\begin{aligned} \mu^*(E) &\geqslant \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \cap A^c) \\ &\geqslant \mu^*\left(\bigcup_{k=1}^{\infty} E \cap A_k\right) + \mu^*(E \cap A^c) \\ &\geqslant \mu^*(E \cap A) + \mu^*(E \cap A^c). \end{aligned} \tag{7.12}$$

Since clearly $\mu^*(E \cap A) + \mu^*(E \cap A^c) \geqslant \mu^*(E)$ (finite sub-additivity) we get

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

i.e. $B \in \mathcal{G}$. We prove that $\mathcal{G}$ is closed under the countable union of pairwise disjoints sets. If $\{A_n\}_n \subset \mathcal{G}$ are given, one construct

$$A_1' = A_1, \quad A_2' = A_2 \setminus A_1, \qquad A_n' = A_n \setminus \bigcup_{i=1}^{n-1} A_i'$$

so that $\{A_n'\}$ are pairwise disjoints and $\bigcup_n A_n' = \bigcup_n A_n$. Since $\mathcal{G}$ is an algebra, $A_n' \in \mathcal{G}$ for any n so that $\bigcup_n A_n \in \mathcal{G}$. This proves that $\mathcal{G}$ is σ-algebra .

Third step: $\mu^*|_{\mathcal{G}}$ is a measure. We need to show that $\mu^*|_{\mathcal{G}}$ is σ-additive, i.e. if $\{A_n\}$ are pairwise disjoint elements of $\mathcal{G}$ s.t. $A = \bigcup_{n=1}^{\infty} A_n$, then $\mu^*(A) = \sum_{n=1}^{\infty} \mu^*(A_n)$.
Notice that, since we proved that $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$, one has that all the inequalities in (7.12) are actually identities, i.e.

$$\mu^*(E) = \sum_{k=1}^{\infty} \mu^*(E \cap A_k) + \mu^*(E \cap A^c) \qquad \forall E \in \mathcal{P}(X).$$

In particular, for $E = A$ we get

$$\mu^*(A) = \sum_{k=1}^{\infty} \mu^*(A \cap A_k) = \sum_{k=1}^{\infty} \mu^*(A_k)$$

which proves the result. □

Remark 7.6. — *With the above notations, if $A \in \mathcal{P}(X)$ is such that $\mu^*(A) = 0$ then $A \in \mathcal{G}$ (i.e. $\mathcal{G}$ contains the μ^*-null sets). Indeed, given $E \in \mathcal{P}(X)$, we have*

$$\mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E \cap A^c) \leqslant \mu^*(E)$$

and, since the other inequality always holds, we get that $\mu^*(E \cap A) + \mu^*(E \cap A^c) = \mu^*(E)$, i.e. $A \in \mathcal{G}$.

The above Proposition allows to construct easily, from a given outer measure μ^* a measure μ on the σ-algebra of μ^*-measurable sets. It would somehow be preferable to construct a measure on a given σ-algebra rather than the one associated to μ^*. To do so, we describe now a general construction. We have the following where $\sigma(\Sigma_0)$ is the σ-algebra generated by Σ_0:

Theorem 7.7 (**Carathéodory Extension Theorem**). — *Let Σ_0 be an algebra on X and $\ell : \Sigma_0 \to [0, \infty]$ be a premeasure on Σ_0. Define*

$$\mu^*(E) = \inf\left\{ \sum_{n=1}^{\infty} \ell(A_n) \ : \ A_n \in \Sigma_0, \ E \subset \bigcup_{n=1}^{\infty} A_n \right\} \qquad \forall E \in \mathcal{P}(X).$$

Then

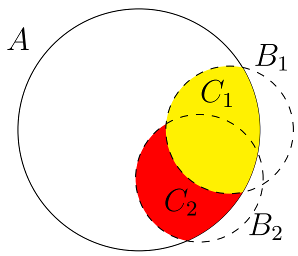
(1) μ^* is an outer measure on X; we denote by $\mathcal{G}$ the associated σ-algebra of μ^*-measurable subsets;
(2) $\mu^*(A) = \ell(A)$ if $A \in \Sigma_0$;
(3) every set in Σ_0 and every μ^*-null set is μ^*-measurable; in particular $\sigma(\Sigma_0) \subset \mathcal{G}$.
(4) if ℓ is σ-finite, then there is a unique extension to $\sigma(\Sigma_0)$.

Proof. — The fact that μ^* is an outer measure is exactly Proposition 7.2. We denote by $\mathcal{G}$ the σ-algebra of μ^*-measurable sets.
Let us prove point (2), i.e. $\mu^*|_{\Sigma_0} = \ell$: let $A \in \Sigma_0$;

- let $\{B_n\} = \{A, \emptyset, \emptyset, \ldots\}$, and

$$\underbrace{\mu^*(A) \leqslant \sum_n \ell(B_n)}_{\text{from def. of } \mu^* \text{ without inf}} = \ell(A)$$

- for every $\{B_n\}$ s.t. $A \subset \bigcup_n B_n$, let $C_1 = B_1 \cap A, \ldots, C_n = (B_n - \bigcup_{j=1}^{n-1} B_n) \cap A$



so that the $C_n \in \Sigma_0$ (since $A, B_1, B_2, \ldots$ do and Σ_0 is an algebra), are disjoint, are s.t. $C_n \subset B_n$ and $A = \bigcup_n C_n$. Hence by monotonicity

$$\ell(A) = \sum_{n=1}^{\infty} \ell(C_n) \leqslant \sum_{n=1}^{\infty} \ell(B_n) \qquad \text{for every such } \{B_n\}$$

and taking the infimum yields $\ell(A) \leqslant \mu^*(A)$. This proves that $\mu^*(A) = \ell(A)$.

Let us now prove (3). Suppose $A \in \Sigma_0$. Let $\varepsilon > 0$ and let $E \in \mathcal{P}(X)$. Pick $\{B_n\} \subset \Sigma_0$ such that $E \subset \bigcup_n B_n$ and $\sum_n \ell(B_n) \leqslant \mu^*(E) + \varepsilon$. Then

$$\mu^*(E) + \varepsilon \geqslant \sum_n \ell(B_n \cap A) + \ell(B_n \cap A^c) \geqslant \mu^*(E \cap A) + \mu^*(E \cap A^c)$$

where, for the last inequality, we used that $\{B_n \cap A\}$ is a covering of $E \cap A$ and $\{B_n \cap A^c\}$ a covering of $E \cap A^c$. Since ε is arbitrary, we have $\mu^*(E) \geqslant \mu^*(E \cap A) + \mu^*(E \cap A^c)$ and, as already saw, this means that A is μ^*-measurable. The fact that μ^*-null sets are μ^*-measurable comes from Remark 7.6.
It remains to prove the uniqueness, i.e. if ℓ is σ-finite and ν is another measure on $\sigma(\Sigma_0)$ such that $\mu|_{\Sigma_0} = \nu|_{\Sigma_0} = \ell$ then $\mu = \nu$. We first assume ℓ to be finite. The proof follows then easily from Proposition 3.12 once we notice that the algebra Σ_0 is actually a π-system.
It remains to consider the case when ℓ is σ-finite. Write $X = \bigcup_i S_i$, where $S_i \uparrow X$ and $\ell(S_i) < \infty$ for each i. By the preceding point, we have uniqueness of the extension of the premeasure ℓ_i defined by $\ell_i(A) = \ell(A \cap S_i)$. If μ and ν are two extensions of ℓ and $A \in \sigma(\Sigma_0)$, then

$$\mu(A) = \lim_k \mu(A \cap S_k) = \lim_k \nu(A \cap S_k) = \nu(A)$$

which proves the result. □

To summarize what we have shown so far: given a premeasure ℓ on an algebra Σ_0, one can associate to μ an outer measure μ^* such that

1. $\mathcal{G}$, the class of μ^*-measurable subsets of X, is a σ-algebra containing Σ_0, from which $\sigma(\Sigma_0) \subseteq \mathcal{G}$
2. $\mu^*|_{\mathcal{G}} = \mu$ is a measure on $\mathcal{G}$, therefore $\mu^*|_{\sigma(\Sigma_0)}$ is a measure on $\sigma(\Sigma_0)$
3. $\mu^*|_{\Sigma_0} = \mu|_{\Sigma_0} = \ell$;

that is we have found that there exists a measure μ on $\sigma(\Sigma_0)$ that extends the premeasure ℓ on Σ_0, and this is given by $\mu := \mu^*|_{\sigma(\Sigma_0)}$.

Moreover, besides the existence result, the above point (4) ensures also the uniqueness of the construction if ℓ is σ-finite.

References

[1] BASS, R. F. — *Real analysis for graduate students*, (2014) http://bass.math.uconn.edu/real.html

[2] CINLAR, E. — **Probability and Stochastics**, Springer-Verlag, 2011.

[3] DUDLEY, R. M. — **Real Analysis and Probability**, Cambridge University Press, 2002.

[4] HUNTER, J. K. — *Measure Theory*, Lecture notes (2011) https://www.math.ucdavis.edu/~hunter/measure_theory/measure_notes.pdf

[5] WILLIAMS, D. — **Probability with Martingales** Cambridge University Press, 1991.

B. Lods, Sept. 2018

LECTURE NOTES FOR THE *“Tutorial Course on Measure Theory”*