SANDIA REPORT

SAND2004-8095
Unlimited Release
Printed February 2004

A Short Course on Measure and Probability Theories

Philippe P. Pébay

Prepared by
Sandia National Laboratories
Albuquerque, New Mexico 87185 and Livermore, California 94550

Sandia is a multiprogram laboratory operated by Sandia Corporation,
a Lockheed Martin Company, for the United States Department of Energy's
National Nuclear Security Administration under Contract DE-AC04-94-AL85000.

Approved for public release; further dissemination unlimited.

Sandia National Laboratories

Issued by Sandia National Laboratories, operated for the United States Department of Energy by Sandia Corporation.

NOTICE: This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government, nor any agency thereof, nor any of their employees, nor any of their contractors, subcontractors, or their employees, make any warranty, express or implied, or assume any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represent that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government, any agency thereof, or any of their contractors or subcontractors. The views and opinions expressed herein do not necessarily state or reflect those of the United States Government, any agency thereof, or any of their contractors.

Printed in the United States of America. This report has been reproduced directly from the best available copy.

Available to DOE and DOE contractors from
U.S. Department of Energy
Office of Scientific and Technical Information
P.O. Box 62
Oak Ridge, TN 37831

Telephone: (865) 576-8401
Facsimile: (865) 576-5728
E-Mail: reports@adonis.osti.gov
Online ordering: http://www.doe.gov/bridge

Available to the public from
U.S. Department of Commerce
National Technical Information Service
5285 Port Royal Rd
Springfield, VA 22161

Telephone: (800) 553-6847
Facsimile: (703) 605-6900
E-Mail: orders@ntis.fedworld.gov
Online ordering: http://www.ntis.gov/help/ordermethods.asp?loc=7-4-0#online

SAND2004-8095
Unlimited Release
Printed February 2004

A Short Course on Measure and Probability Theories

Philippe P. Pébay

Sandia National Laboratories
M.S. 9051, P.O. Box 969
Livermore, CA 94550, U.S.A.

pppebay@ca.sandia.gov

This page intentionally left blank

A Short Course on Measure and Probability Theories

Contents

Introduction 7

Acknowledgements 7

Typographic Conventions 8

Vocabulary Conventions 8

I Background: Sets, Functions and Structures 9
- I.1 Sets and Functions 9
  - I.1.1 Sets 9
  - I.1.2 Functions 10
- I.2 Algebraic Structures 12
  - I.2.1 Semigroups and Groups 12
  - I.2.2 Rings and Fields 14
  - I.2.3 Vector Spaces 15

II Measures and Probabilities 17
- II.1 Measure Spaces 17
  - II.1.1 σ-Algebras 17
  - II.1.2 Measures 18
  - II.1.3 Probability Spaces 20
- II.2 Measurable Functions and Random Variables 23
  - II.2.1 Measurable Functions 23
  - II.2.2 Simple Functions 25
  - II.2.3 Random Variables 27
  - II.2.4 Introduction to MARKOV Chains 28

III The LEBESGUE Measure 31
- III.1 Preamble 31
  - III.1.1 Fountainhead 31
  - III.1.2 Constructive Approach to the LEBESGUE Measure 33
- III.2 Measures on Boolean Algebras 34
  - III.2.1 Interval Measures 34
  - III.2.2 Boolean Algebras as Closures 35
- III.3 CARATHÉODORY's Extension 37
  - III.3.1 Outer-Measures 37
  - III.3.2 μ^*-measurability 38
  - III.3.3 σ-Algebra Associated with an Outer-Measure 41
- III.4 Measure Completion 42
  - III.4.1 Complete Measures 42
  - III.4.2 BOREL-STIELTJES Measures 44

IV L^p Spaces **47**
IV.1 Integrals of Positive Functions 47
IV.1.1 From Measures to Integrals 47
IV.1.2 Monotonic Convergence Theorem 49
IV.1.3 Additivity 51
IV.1.4 Density 53
IV.2 Probability Density Functions 55
IV.3 Integrability 56
IV.3.1 Integrable Functions 56
IV.3.2 Almost Everywhere 58
IV.3.3 $\mathscr{L}^p$ Spaces 60
IV.4 From $\mathscr{L}^p$ to L^p 62
IV.4.1 Quotientization 62
IV.4.2 Approximation in BANACH Spaces 63

V HILBERT Spaces **65**
V.1 Pre-HILBERT Spaces 65
V.1.1 Inner Products 65
V.1.2 Norm induced by an Inner Product 67
V.1.3 Orthogonality 70
V.2 Completeness-Based Projectors 71
V.2.1 Convex Minimization 71
V.2.2 Best Approximation Projector 72
V.2.3 Orthogonal Projectors 73
V.3 Approximation in HILBERT Spaces 75
V.3.1 HILBERT Bases 75
V.3.2 Orthogonal Polynomials 77
V.4 Application: WIENER Chaoses 80

Index **85**

References **85**

List of Figures

1 Functional and nonfunctional graphs 10
2 Construction of elementary algebraic structures 13
3 A monotonic sequence of simple functions 26
4 Two-state uniform MARKOV chain 29
5 Best approximation on a complete convex subset 73

List of Tables

1 CARATHÉODORY's Extension 33

Introduction

This brief Introduction to Measure Theory, and its applications to Probabilities, corresponds to the lecture notes of a seminar series given at Sandia National Laboratories in Livermore, during the spring of 2003.

The goal of these seminars was to provide a minimal background to Computational Combustion scientists interested in using more advanced stochastic concepts and methods, *e.g.*, in the context of uncertainty quantification. Indeed, most mechanical engineering *curricula* do not provide students with formal training in the field of probability, and even in less in measure theory. However, stochastic methods have been used more and more extensively in the past decade, and have provided more successful computational tools.

Scientists at the Combustion Reseach Facility of Sandia National Laboratories have been using computational stochastic methods for years. Addressing more and more complex applications, and facing difficult problems that arose in applications showed the need for a better understanding of theoretical foundations. This is why the seminar series was launched, and these notes summarize most of the concepts which have been discussed.

The goal of the seminars was to bring a group of mechanical engineers and computational combustion scientists to a full understanding of N. WIENER's polynomial chaos theory. Therefore, these lectures notes are built along those lines, and are not intended to be exhaustive. In particular, the author welcomes any comments or criticisms.

Acknowledgements

The author was supported by the United States Department of Energy, Office of Defense Programs. Thanks to the attendees of the seminar series, for their patience, comments and feedbacks; in particular, H.N. NAJM, B.J. DEBUSSCHERE, J.C. LEE, P.T. BOGGS and K.R. LONG, and to Pr J. POUSIN (INSA, Lyon) for his careful proofreading. Special thanks to Ms J. MATTO for having edited most of the manuscript.

Typographic Conventions

Since several types of structures will potentially be built on any given set, the following typographic conventions will be used in order to facilitate the reading:

- for a set with no particular structure on it, a slanted font: S;
- for a boolean algebra on S, a script font: $\mathcal{S}$;
- for a σ-algebra on S, a Gothic font: $\mathfrak{S}$.

Exceptions will be made for some standard and well-established notations, such as BOREL ($\mathfrak{B}^d$) and LEBESGUE ($\mathfrak{L}^d$) σ-algebras on $\mathbb{R}^d$.

Vocabulary Conventions

In order to avoid confusion, the following conventions regarding positiveness and negativeness will be used:

- *positive* means stricly greater than 0, denoted > 0
- *negative* means stricly smaller than 0, denoted < 0
- *nonpositive* means smaller than or equal to 0, denoted ≤ 0
- *nonnegative* means larger than or equal to 0, denoted ≥ 0.

In particular, 0 is neither positive nor negative, but it is both nonpositive and nonnegative.

I Background: Sets, Functions and Structures

Die ganzen Zahlen hat der liebe Gott gemacht, alles andere ist Menschenwerk.

L. KRONECKER[1]

The goal of this chapter is certainly not to provid an exhaustive coverage, but rather an overview of the very minimal background required for the understanding of the core text. The interested reader will find in-depth presentations of the matters discussed here in, *e.g.*, [Jec02] and [Lan02].

In all that follows, d will denote a positive integer.

I.1 Sets and Functions

I.1.1 Sets

It is largely beyond the scope of the present report to address the question *"what is a set?"*, however the naive definition of a *set* constituted of *elements* will be sufficient for the matter discussed here. In particular, $x \in S$ means that x is an element of a set S. For the interested audience, we mention that we place ourselves in the context of the VON NEUMANN[2]-BERNAYS[3]-GÖDEL[4] set theory (NBG), which is a generalization of the ZERMELO[5]-FRAENKEL[6] (ZF) set theory. In particular the sets of natural ($\mathbb{N}$), integer ($\mathbb{Z}$), rational ($\mathbb{Q}$), nonnegative real ($[0,\infty[$), real ($\mathbb{R}$) and complex ($\mathbb{C}$) numbers will be used. This being said, a few Set Theory concepts will be needed:

Definition I.1. A set is a *singleton* if it contains exactly one element. Denoting as x this element, the singleton is written $\{x\}$.

Definition I.2. A set B is a *subset of* another set A if any element of B is an element of A. It is denoted $B \subset A$.

Definition I.3. The *intersection of* two sets, A and B, is the set of elements common to A and B. It is denoted as $A \cap B$. The intersection of sets A_1 to A_n is written $\bigcap_{i=1}^{i=n} A_i$.

Definition I.4. The *union of* two sets A and B is the set obtained by combining the elements of A and B. It is denoted as $A \cup B$. The union of sets A_1 to A_n is written $\bigcup_{i=1}^{i=n} A_i$.

Definition I.5. The sets $[0,\infty[\cup\{\infty\}$ and $\mathbb{R}\cup\{-\infty,\infty\}$ are denoted respectively $[0,\infty]$ and $[-\infty,\infty]$.

Definition I.6. The set containing no elements is called the *empty set*. It is denoted $\varnothing$.

Definition I.7. The *set difference of* two sets A and B is the set obtained by removing all the elements from A which also belong to B. It is denoted $A \setminus B$ and also called *complement of B in A*.

[1] Leopold KRONECKER, Prussian mathematician, 1823–1891.
[2] John VON NEUMANN, Hungarian then American mathematician, 1903–1957.
[3] Paul BERNAYS, Swiss mathematician, 1888-1977.
[4] Kurt GÖDEL, Austrian then American mathematician, 1906–1978.
[5] Ernst ZERMELO, German mathematician, 1871–1953.
[6] Adolf FRAENKEL, German then Israeli mathematician, 1891–1965.

I.1.2 Functions

Definition I.8. If A and B are two sets, a *correspondence* from the *origin set* A to the *target set* B is a triple $f = (G,A,B)$ where G is a subset of $A \times B$ called a *graph*, and the set

$$\mathscr{D}_f = \{x \in A : (\exists y \in B)\ (x,y) \in G\}$$

is called the *definition set of* f. The graph G is said to be *functional* if, for all x in A, the set $\{y \in B : (x,y) \in G\}$ is either empty or a singleton.

Definition I.9. A correspondence $f = (G,A,B)$ is a *mapping*, or *map*, or *function* from A to B if G is functional and $\mathscr{D}_f = A$. It is generally symbolized as follows:

$$\begin{array}{rccc} f: & A & \longrightarrow & B \\ & x & \longmapsto & f(x) \end{array}$$

where $f(x)$ is the element of $\{y \in B : (x,y) \in G\}$. For all x in A, $f(x)$ is called the *image of* x *under* f. By extension, for any subset E of A,

$$f(E) = \{y \in B : (\exists x \in E)\ f(x) = y\}$$

is called the *image set of* E *under* f.

Remark I.10. The fact that $f(x)$ is actually well-defined is ensured by the fact that $\mathscr{D}_f = A$. Therefore, the set $\{y \in B : (x,y) \in G\}$ is not empty thus, by definition, it is necessarily a singleton.

Example I.11. If $A = \{a,b,c,d\}$ and $B = \{1,2,3\}$, then $G = \{(a,3),(c,1),(d,3)\}$ is a functional graph from A to B, defining a function from $\{a,c,d\}$ to G, illustrated by Figure 1, left. To the contrary, $H = \{(a,3),(c,1),(c,2),(d,3)\}$ is not a functional graph, since $\{y \in B : (c,y) \in H\}$ has two elements, 1 and 2 and in particular, it cannot define a function.

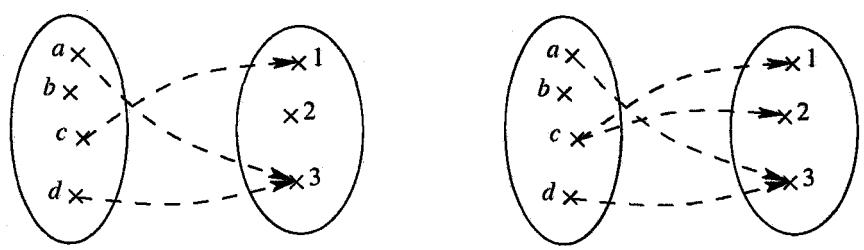

Figure 1. Functional (left) and nonfunctional (right) graphs.

Definition I.12. The set of mappings from a set A to a set B is denoted B^A.

Definition I.13. If A and B are two sets and f is a function in B^A, then for any y in B and any subset F of B, the sets

$$f^{<-1>}(y) = \{x \in A : f(x) = y\}$$
$$f^{<-1>}(F) = \{x \in A : (\exists y \in B)\ f(x) \in F\}$$

are called *inverse image of* x (resp. B) *under* f.

Example I.14. According to Definition I.12, $A^{\mathbb{N}}$ is the set of the mappings from $\mathbb{N}$ to A. In other words,

$$A^{\mathbb{N}} = \left\{ \begin{array}{lccc} u: & \mathbb{N} & \longrightarrow & A \\ & n & \longmapsto & u_n \end{array} \right\}$$

and such mappings are called *sequences of* elements of A,*sequence* denoted $(u_n)_{n\in\mathbb{N}}$. By extension, $(u_n)_{n\in E}$, where $E \subset \mathbb{N}$, denotes the elements of $(u_n)_{n\in\mathbb{N}}$ such that $n \in E$.

Definition I.15. Let A and B be two sets and f a function in B^A. If, for all y in B, $f^{<-1>}(y)$ contains at most, at least, or exactly one element, then f is respectively an *injection*, a *surjection*, or a *bijection*. It can also be said that f is, respectively, *injective*, *surjective*, or *bijective*.

Example I.16. Given any subset E of S, the *indicator function of* E is defined by

$$(\forall x \in S) \quad \mathbb{1}_E(x) = \left\{ \begin{array}{lll} 1 & \text{if} & x \in E \\ 0 & \text{if} & x \notin E \end{array} \right.$$

and is a surjection from S to $\{0,1\}$ if and only if $E \neq \varnothing$ and $E \neq S$ (otherwise, $\mathbb{1}_E{}^{<-1>}(0)$ or $\mathbb{1}_E{}^{<-1>}(1)$ is empty). If S has more than two elements, $\mathbb{1}_E$ cannot be injective, since either $\mathbb{1}_E{}^{<-1>}(0)$ or $\mathbb{1}_E{}^{<-1>}(1)$ has more than one element.

Definition I.17. A set is *denumerable* or *countably infinite* if it is bijective to $\mathbb{N}$.

Proposition I.18. *Let A and B be two sets and f a function in B^A.*

- *f is injective if and only if*

$$(\forall (x_1,x_2) \in A^2) \quad x_1 \neq x_2 \Longrightarrow f(x_1) \neq f(x_2).$$

- *f is surjective if and only if $f(A) = B$.*
- *f is bijective if and only if it is injective and surjective.*

Proof. Left to the reader as an exercise. □

Definition I.19. Given a set S, the *power set of* S is the set of all subsets of S and is denoted 2^S or $\mathscr{P}(S)$.

The following ZF axiom is essential in Measure Theory:

Axiom I.20. For any given set, its power set exists.

Remark I.21. The notation 2^S is justified by the fact that the notion of subset of S can be interpreted in an equivalentl way as being a map from S to $\{0,1\}$, formally denoted 2. More precisely, the following function:

$$\begin{array}{ccc} \mathscr{P}(S) & \longrightarrow & 2^S \\ E & \longmapsto & \mathbb{1}_E \end{array}$$

is obviously bijective, thus justifying the identification between $\mathscr{P}(S)$ and 2^S.

Definition I.22. Given a sequence $(u_n)_{n\in\mathbb{N}}$ in $[-\infty,\infty]^{\mathbb{N}}$, its *infimum limit* and *supremum limit* are respectively defined as follows:

$$\limsup_n u_n = \inf_{m\in\mathbb{N}} \sup_{n>m} u_n$$
$$\liminf_n u_n = \sup_{m\in\mathbb{N}} \inf_{n>m} u_n.$$

Example I.23. If, for all n in $\mathbb{N}$, $u_n = (-1)^n$, then $\liminf_n u_n = -1$ and $\limsup_n u_n = 1$.

Proposition I.24. *If $(u_n)_{n\in\mathbb{N}}$ is a sequence in $[-\infty,\infty]^{\mathbb{N}}$, then:*

$$\limsup_n u_n = \lim_m \sup_{n>m} u_n$$
$$\liminf_n u_n = \lim_m \inf_{n>m} u_n.$$

Proof. Exercise. □

Definition I.25. Given a set S and a sequence $(f_n)_{n\in\mathbb{N}}$ of functions in $[-\infty,\infty]^S$, the mappings $\mathrm{Inf}_{n\in\mathbb{N}} f_n$, $\mathrm{Sup}_{n\in\mathbb{N}} f_n$, $\liminf_n f_n$, and $\limsup_n f_n$ of $[-\infty,\infty]^S$ are defined as follows:

$$(\forall x \in S) \quad \left(\operatorname*{Sup}_{n\in\mathbb{N}} f_n\right)(x) = \sup_{n\in\mathbb{N}} f_n(x)$$
$$(\forall x \in S) \quad \left(\operatorname*{Inf}_{n\in\mathbb{N}} f_n\right)(x) = \inf_{n\in\mathbb{N}} f_n(x)$$
$$(\forall x \in S) \quad \left(\limsup_n f_n\right)(x) = \limsup_n f_n(x)$$
$$(\forall x \in S) \quad \left(\liminf_n f_n\right)(x) = \liminf_n f_n(x)$$

Remark I.26. For any k in $\mathbb{N}$, it is also possible to define the mappings $\mathrm{Inf}_{n\geqslant k} f_n$ and $\mathrm{Sup}_{n\geqslant k} f_n$ in a similar way.

I.2 Algebraic Structures

The axiomatic construction of semigroups, groups, rings, fields and vector spaces is now briefly recalled. These definitions are summarized in Figure 2.

I.2.1 Semigroups and Groups

Definition I.27. Let S be a set. A *binary operation on S* is a function with the form:

$$\begin{array}{ccc} S\times S & \longrightarrow & S \\ (x,y) & \longmapsto & x\star y. \end{array}$$

Remark I.28. Under the above assumptions, the binary operation is itself denoted $\star$.

Example I.29. Both $+$ and $\times$ are binary operations on $\mathbb{N}$, but $-$ is not, since the difference between two natural numbers is not necessarily a natural number. However, $-$ is a binary operation on, *e.g.*, $\mathbb{Z}$ which really allows to *count* (*zählen* in German, whence the symbol).

Definition I.30. A binary operation $\star$ on a set S is said to be *associative* if

$$(\forall (x,y,z) \in S^3) \quad x\star(y\star z) = (x\star y)\star z$$

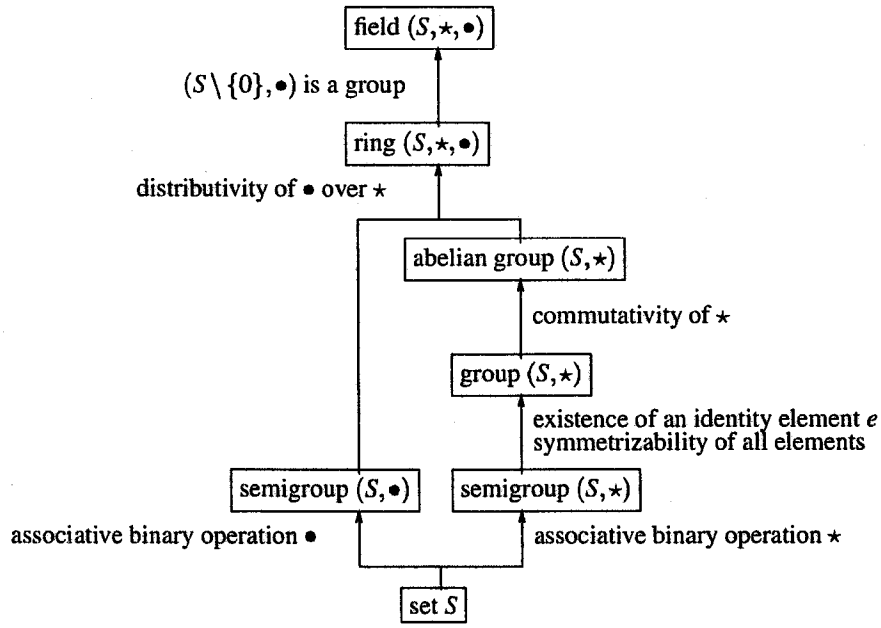

Figure 2. Construction of elementary algebraic structures.

and to be *commutative* if

$$(\forall(x,y)\in S^2)\quad x\star y = y\star x.$$

If $\star$ is an associative (resp. associative and commutative) binary operation on S, then $(S,\star)$ is said to be a *semigroup* (resp. *abelian semigroup*).

Remark I.31. In the case where a binary operation $\star$ is associative, and in this case only, the expression $x\star y\star z$ is meaningful; otherwise, it is ambiguous except is some *precedence* properties are explitly defined. The fact of being associative (resp. commutative) is referred to as the property of *associativity* (resp. *commutativity*). However, *in memoriam* of the pioneering work (and tragic fate) of N. ABEL[7] the adjective abelian is generally used in Group Theory, rather than commutative.

Example I.32. $(\mathbb{Z},-)$ is not a semigroup, since $x-(y-z)$ is not, in general, equal to $(x-y)-z$. To the contrary, both $(\mathbb{N},+)$ and $(\mathbb{N},\times)$ are semigroups, and, moreover, abelian semigroups.

Definition I.33. Let $(S,\star)$ be a semigroup. If

$$(\exists e\in S)\,(\forall x\in S)\quad x\star e = e\star x = x$$

then e is said to be *an identity element of* $(S,\star)$. In this case, any $x\in S$ such that

$$(\exists y\in S)\quad x\star y = y\star x = e$$

is said to be *symmetrizable in* $(S,\star)$ and y is called the *symmetric element of* x (and conversely). If each element of S is symmetrizable in $(S,\star)$, then $(S,\star)$ is said to be a *group*. In this case, if $\star$ is commutative, then $(S,\star)$ is said to be an *abelian group*.

[7]Niels ABEL, Norwegian mathematician, 1802–1829.

Proposition I.34. *A semigroup has at most one identity element. If such an identity element exists, then each element of the semigroup has at most one symmetric element.*

Proof. It is clear that a semigroup $(S,\star)$ does not necessarily have an identity element; *e.g.*, $(\mathbb{N}^*,+)$ exhibits such a case. Now, let $(S,\star)$ be a semigroup and assume it has two identity elements e_1 and e_2. Then, by definition of e_1,

$$e_1 \star e_2 = e_2 \star e_1 = e_1$$

and, by definition of e_2,

$$e_2 \star e_1 = e_1 \star e_2 = e_2$$

thus, necessarily, $e_1 = e_2$. Now, not all elements of a semigroup $(S,\star)$ with identity element e are necessarily symmetrizable: *e.g.*, 1 does not have a symmetric element in $(\mathbb{N},+)$, since one cannot find $x \in \mathbb{N}$ such that $x+1=0$. On the other hand, assume $x \in S$ has two symmetric elements y_1 and y_2 in $(S,\star)$. Then, by definition of y_1,

$$y_1 \star x = e$$

thus, using the associativity of $\star$,

$$(y_1 \star x) \star y_2 = y_1 \star (x \star y_2) = y_2$$

but, by definition of y_2, $x \star y_2 = e$, hence $y_1 = y_2$. □

Example I.35. Both $(\mathbb{N},+)$ and $(\mathbb{N},\times)$ have identity elements (respectively, 0 and 1), but none of them is a group: *e.g.*, 1 has no symmetric element in $(\mathbb{N},+)$, and neither does 0 in $(\mathbb{N},\times)$. In order to be able to symmetrize each element of $(\mathbb{N},+)$, one has to allow for negative integers. In other words, $(\mathbb{Z},+)$ is an abelian group. Now, if one wants to build a group from the semigroup $(\mathbb{N},\times)$, two problems must be solved: first, 0 must be eliminated because it cannot be symmetrized; second, all rational numbers must be added; namely, $(\mathbb{Q}^*,\times)$ is an abelian group.

To summarize: a group is a set equipped with an associative binary operation which has an identity element and such that each element is symmetrizable. One would now now like to have more than one operation, in order to generalize the usual arithmetics.

I.2.2 Rings and Fields

Definition I.36. If $\star$ and $\bullet$ are two binary operations on a set S such that

$$(\forall (x,y,z) \in S^3) \quad \left\{ \begin{array}{rcl} x \bullet (y \star z) & = & (x \bullet y) \star (x \bullet z) \\ (x \star y) \bullet z & = & (x \bullet z) \star (y \bullet z) \end{array} \right.$$

then it is said that $\bullet$ is *distributive over* $\star$. In this case, if $(S,\star)$ is an abelian group and $(S,\bullet)$ is a semigroup (resp. an abelian semigroup), then the triple $(S,\star,\bullet)$ is said to be a *ring* (resp. a *commutative ring*). If, in addition, $(S,\bullet)$ has a unit element, then $(S,\star,\bullet)$ is a *unit ring* (resp. a *commutative unit ring*). If $(S,\star,\bullet)$ is a ring, then the identity element of $(S,\star)$ is called zero of S, denoted 0_S or, if there is no risk of confusion, 0.

Example I.37. Historically, the notion of ring appeared to describe sets such as:

$$\mathbb{Z}[\sqrt{2}] = \left\{ a + b\sqrt{2}, (a,b) \in \mathbb{R}^2 \right\},$$

equipped with the usual addition and multiplication. Indeed, $(\mathbb{Z}[\sqrt{2}],+,\times)$ is a commutative unit ring, but the most intuitive example of such a structure is certainly $(\mathbb{Z},+,\times)$. A finite case can be easily constructed by considering $\mathbb{Z}_2=\{0,1\}$ equipped with the two following binary operations:

$+$	0	1
0	0	1
1	1	0

$\times$	0	1
0	0	0
1	0	1

and it is possible to check directly (by enumeration) that $(\mathbb{Z}_2,+,\times)$ is a commutative unit ring, with 0 and 1 as respective identity elements of $+$ and $\times$. In fact, one can easily view it as the two classes of integer numbers *modulo* 2 along with the usual arithmetics transported onto them (odd plus odd is even, odd times even is odd, *etc.*).

Definition I.38. A unit ring $(S,\star,\bullet)$ with at least two elements such that $(S\setminus\{0\},\bullet)$ is a group (resp. abelian group) is said to be a *field* (resp. a *commutative field*).

Remark I.39. The requirement of having at least two elements is intended to eliminate the undesireable case of the *nullring* $(0,\star,\bullet)$ with $0\star 0=0\bullet 0=0$.

Example I.40. Enforcing the extra requirements on $(\mathbb{Z},+,\times)$ leads to the construction of the field $(\mathbb{Q},+,\times)$: in fact, rational numbers precisely fill the need of being able to invert the multiplication of integer numbers.

In the present context, a few important fields will be needed:

Theorem I.41. *The sets $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, equipped with the usual addition and multiplication, are commutative fields.*

Proof. See, *e.g.*, [Lan02]. □

Remark I.42. For the sake of concision and although this is not perfectly rigorous, the field will be denoted as its underlying set when there is no risk of confusion; *e.g.*, ,"the field $\mathbb{R}$" will be used as a synonymous of "the field $(\mathbb{R},+,\times)$". In all this report, $\mathbb{K}$ will denote a field that might be either $\mathbb{R}$ or $\mathbb{C}$ (equipped with their respective usual additions and multiplications).

I.2.3 Vector Spaces

Only a very brief reminder is provided here, since even a simple introduction to the most important vector spaces properties would require an entire chapter on its own; *cf.* [Lan02] for more details.

Definition I.43. If $(E,\star)$ is an abelian group and $\cdot$ is a mapping in $E^{\mathbb{K}\times E}$ such that:

i. $(\forall x\in E)\quad 1\cdot x=x$;

ii. $(\forall(\lambda,\mu,x)\in\mathbb{K}^2\times E)\quad \lambda\cdot(\mu\cdot x)=(\lambda\times\mu)\cdot x$;

iii. $(\forall(\lambda,\mu,x)\in\mathbb{K}^2\times E)\quad (\lambda+\mu)\cdot x=(\lambda\cdot x)+(\mu\cdot x)$;

iv. $(\forall(\lambda,x,y)\in\mathbb{K}\times E^2)\quad \lambda\cdot(x+y)=(\lambda\cdot x)+(\lambda\cdot y)$,

then $(E,\star,\cdot)$ is said to be a $\mathbb{K}$-*vector space*. In this case, the elements of $\mathbb{K}$ and E are called, respectively, the *scalars* and the *vectors* of the vector space.

Example 1.44. The simplest examples of $\mathbb{K}$-vector spaces are the $\mathbb{K}^d$ sets equipped with their natural component-by-component addition and multiplication by a scalar. Another important class of $\mathbb{K}$-vector fields is that of *functional spaces*, *i.e.*, those whose vectors are functions. For example, the set of continuous functions over an interval can be naturally viewed as a $\mathbb{R}$-vector field.

In this report, two kinds of vector field will be dealt with, depending on the finiteness of their bases: finite-dimensional and infinite-dimensional vector spaces. Most undergraduate textbooks restrict themselves to the formers; the reader familiar with those but not with the latters should be extremely careful, since numerous essential finite-dimensional results do not extend to the infinite dimensions. Unfortunately, most of the vector spaces interesting for the purpose of this report are infinite-dimensional.

II Measures and Probabilities

In all that follows, S will denote a non-empty set. For any given subset $A \subset S$, the set $S \setminus A$ will be denoted A^c. For simplicity, it will be refered to as the *complement of* A (omitting "in S").

II.1 Measure Spaces

II.1.1 σ-Algebras

Definition II.1. Any $\mathfrak{S} \subset 2^S$ such that

i. $S \in \mathfrak{S}$

ii. if $A \in \mathfrak{S}$, then $A^c \in \mathfrak{S}$

iii. if $(A_n)_{n\in\mathbb{N}^*} \in \mathfrak{S}^{\mathbb{N}^*}$, then $\bigcup_{n\in\mathbb{N}^*} A_n \in \mathfrak{S}$

is called a *σ-algebra on* S. The elements of $\mathfrak{S}$ and the double $(S, \mathfrak{S})$ are called, respectively, *measurable sets* and *measurable space.*

Remark II.2. A σ-algebra necessarily contains $\varnothing$ and is closed for set difference and for denumerable intersection. In addition, closedness for denumerable union and intersection implies closedness for finite union and intersection.

Example II.3. $\{\varnothing, S\}$ and 2^S are two σ-algebras on S, respectively, the smallest and largest (for inclusion) possible σ-algebras on S. Using the same language as for other structures, a σ-algebra $\mathfrak{S}$ is *smaller* than a σ-algebra $\mathfrak{S}'$ if each element of the former belongs to the latter.

Example II.4. If $S = \{\Diamond, \Box, \triangle\}$ and $\mathfrak{S} = \{\varnothing, S, \{\Diamond\}, \{\Box, \triangle\}\}$, then $(S, \mathfrak{S})$ is a measurable space. To the contrary, if $\mathfrak{S}' = \{\varnothing, S, \{\Diamond\}, \{\Box\}, \{\triangle\}\}$ then $(S, \mathfrak{S}')$ is not a measurable space, since, *e.g.*, $\{\Box\}^c = \{\Diamond, \triangle\} \notin \mathfrak{S}'$.

Proposition II.5. *The intersection of two σ-algebras is a σ-algebra. Moreover, given any* $\mathscr{D} \subset 2^S$, *the intersection of all σ-algebras containing* $\mathscr{D}$ *is a σ-algebra.*

Proof. Exercise. □

Definition II.6. Under the hypothesis of Proposition II.5, the intersection of all σ-algebras containing $\mathscr{D}$ is called the *σ-algebra generated by* $\mathscr{D}$. In the case where $S = \mathbb{R}^d$, the σ-algebra generated by the set of all open sets (in the sense of the usual topology on $\mathbb{R}^d$) is called the *BOREL*[8]*σ-algebra* on $\mathbb{R}^d$ and denoted $\mathfrak{B}^d$; its elements are called *BOREL sets.*

Remark II.7. As a direct consequence of Definition II.6, the σ-algebra generated by any given $\mathscr{D} \subset 2^S$ is the smallest (for inclusion) σ-algebra on S that contains $\mathscr{D}$. It is also clear that, because any σ-algebra is closed for the complement, intersection and union operations, real singletons and intervals are BOREL sets.

Proposition II.8. *Any denumerable subset of* $\mathbb{R}$ *is a* BOREL *set.*

[8] Émile BOREL, French mathematician and politician, 1871–1956.

Proof. By definition, if $A \subset \mathbb{R}$ is denumerable, there exists a bijection $\varphi \in A^{\mathbb{N}}$, thus

$$A = \varphi(\mathbb{N}) = \{\varphi(n), n \in \mathbb{N}\} = \bigcup_{n \in \mathbb{N}} \{\varphi(n)\}.$$

Now, since real singletons are BOREL sets and σ-algebras are closed for denumerable union, A is a BOREL set. □

Example II.9. $\mathbb{N}$, $\mathbb{Z}$ and $\mathbb{Q}$ are BOREL sets.

Example II.10. A non-BOREL set: let $S = \mathbb{R}$ and E be the set of all real numbers with continued fraction expansion:

$$E = \left\{ x \in \mathbb{R} : \left(\exists (q_n)_{n \in \mathbb{N}} \in \mathbb{Q}^{\mathbb{N}} \right) x = q_0 + \frac{1}{q_1 + \frac{1}{q_2 + \cdots}} \right\}.$$

If F is defined as the subset of E such that there exists an increasing sequence $(r_n)_{n \in \mathbb{N}^*}$ in $\mathbb{Q}^{\mathbb{N}^*}$ such that q_{r_i} divides $q_{r_{i+1}}$ for all i in $\mathbb{N}^*$, then F is not a BOREL set. The main reason is that each particular choice of the r-sequence gives a BOREL set, and there are non-denumerably many different such sequences. More precisely, it is impossible to express F as a denumerable union of BOREL sets (*cf.* Lusin, *Fund. Math.* **10** (1927) p. 77).

II.1.2 Measures

Definition II.11. A *measure on* a measurable space $(S, \mathfrak{S})$ is a mapping μ in $[0, \infty]^{\mathfrak{S}}$ such that:

i. $\mu(\varnothing) = 0$;

ii. for any sequence $(A_n)_{n \in \mathbb{N}^*}$ of pairwise disjoined measurable sets,

$$\mu\left(\bigcup_{n \in \mathbb{N}^*} A_n \right) = \sum_{n \in \mathbb{N}^*} \mu(A_n).$$

Under these assumptions, the triple $(S, \mathfrak{S}, \mu)$ is a *measure space*. For any A in $\mathfrak{S}$, $\mu(A)$ is called the *measure of* A.

Example II.12. Given a measurable space $(S, \mathfrak{S})$, and x_0 in S, if μ is the mapping defined on $\mathfrak{S}$ by $\mu(A) = \mathbb{1}_A(x_0)$, then $(S, 2^S, \mu)$ is a measure space.

Example II.13. Denoting # as the cardinality of a set, $(\mathbb{N}, 2^{\mathbb{N}}, \#)$ is a measure space.

Example II.14. With $(S, \mathfrak{S})$ as defined in Example II.4, the mapping μ, defined by $\mu(\varnothing) = \mu(\{\square, \triangle\}) = 0$ and $\mu(S) = \mu(\{\Diamond\}) = 1$, is a measure on $(S, \mathfrak{S})$.

Proposition II.15. *Let $(X, \mathfrak{S}, \mu)$ be a measure space. If (A, B) in $\mathfrak{S}^2$ is such that $A \subset B$, then $\mu(A) \leqslant \mu(B)$.*

Proof. By definition, A and $B \setminus A$ are disjoined while $B = A \cup (B \setminus A)$. Therefore, it follows from the second axiom of Definition II.11 that

$$\mu(B) - \mu(A) = \mu(B \setminus A) \geqslant 0$$

since a measure is always positive. □

Proposition II.16. *Let $(X,\mathfrak{S},\mu)$ be a measure space. If $(A_n)_{n\in\mathbb{N}^*}$ is a sequence in $\mathfrak{S}$, then*

$$\mu\left(\bigcup_{n\in\mathbb{N}^*} A_n\right) \leqslant \sum_{n\in\mathbb{N}^*} \mu(A_n).$$

Proof. Assuming $B_1 = A_1$, define $B_{n+1} = A_{n+1} \setminus \bigcup_{i=1}^{n} A_i$ for all n in $\mathbb{N}^*$. These B_n are all measurable sets and, using the commutativity and associativity of the intersection,

$$\begin{aligned}\left(\forall (m,n) \in \mathbb{N}^{*^2}, m<n\right) \quad B_m \cap B_n &= \left(A_m \setminus \bigcup_{i=1}^{i=m-1} A_i\right) \cap \left(A_n \setminus \bigcup_{i=1}^{i=n-1} A_i\right) \\ &= \left(A_m \cap \bigcap_{i=1}^{i=m-1} A_i^c\right) \cap \left(A_n \cap \bigcap_{i=1}^{i=n-1} A_i^c\right) \\ &= \varnothing\end{aligned}$$

since $m \leqslant n-1$. Therefore, $\bigcap_{i=1}^{i=n-1} A_i^c$ contains A_m^c thus its intersection with A_m is empty. In other words, the B_n sets are pairwise disjoined and

$$\bigcup_{n\in\mathbb{N}^*} A_n = \bigcup_{n\in\mathbb{N}^*} B_n.$$

Now, by applying Proposition II.15 to the fact that, $B_n \subset A_n$ for all n in $\mathbb{N}^*$, it follows directly from the definition of a measure that

$$\mu\left(\bigcup_{n\in\mathbb{N}^*} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}^*} B_n\right) = \sum_{n\in\mathbb{N}^*} \mu(B_n) \leqslant \sum_{n\in\mathbb{N}^*} \mu(A_n).$$

□

Definition II.17. The property of μ exhibited by Proposition II.16 is called *countable subadditivity*.

Remark II.18. Countable subadditivity can be intutitively seen as corresponding (in the case of measure on σ-algebras) to the triangle inequality in the context of vector space norms.

Proposition II.19. *Let $(X,\mathfrak{S},\mu)$ be a measure space. If $(A_n)_{n\in\mathbb{N}^*}$ is a sequence in $\mathfrak{S}$ such that, for all n in $\mathbb{N}^*$, $A_n \subset A_{n+1}$, then*

$$\mu\left(\bigcup_{n\in\mathbb{N}^*} A_n\right) = \lim_n \mu(A_n).$$

Proof. Assuming $A_0 = \varnothing$, define $B_n = A_n \setminus A_{n-1}$ for all n in $\mathbb{N}^*$. Each B_n is a measurable set and, using the commutativity and associativity of the intersection,

$$\begin{aligned}\left(\forall (m,n) \in \mathbb{N}^{*^2}, m<n\right) \quad B_m \cap B_n &= (A_m \setminus A_{m-1}) \cap (A_n \setminus A_{n-1}) \\ &= \left(A_m \cap A_{m-1}^c\right) \cap \left(A_n \cap A_{n-1}^c\right) \\ &= \varnothing\end{aligned}$$

since $m \leqslant n-1$. Hence, $A_m \subset A_{n-1}$ (in particular, $A_m \subset A_m$) thus $A_m \cap A_{n-1}$ is empty. In other words, the B_n sets are pairwise disjoined and

$$\bigcup_{n\in\mathbb{N}^*} A_n = \bigcup_{n\in\mathbb{N}^*} B_n.$$

Therefore, on the one hand,

$$\mu\left(\bigcup_{n\in\mathbb{N}^*} A_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}^*} B_n\right) = \sum_{n\in\mathbb{N}^*} \mu(B_n). \tag{II.1}$$

On the other hand,

$$(\forall n \in N^*) \quad A_n = \bigcup_{i=1}^{n} B_i$$

from which follows, since the B_n are pairwise disjoined,

$$(\forall n \in N^*) \quad \mu(A_n) = \sum_{i=1}^{n} \mu(B_i)$$

thus

$$\lim_n \mu(A_n) = \sum_{n\in\mathbb{N}^*} \mu(B_n) \tag{II.2}$$

and, finally, the result arises by combining (II.1) and (II.2). □

Proposition II.20. *Let $(X,\mathfrak{S},\mu)$ be a measure space. If $(A_n)_{n\in\mathbb{N}^*}$ is a sequence in $\mathfrak{S}$ such that $\mu(A_1)$ is finite and, for all n in $\mathbb{N}^*$, $A_{n+1} \subset A_n$, then*

$$\mu\left(\bigcap_{n\in\mathbb{N}^*} A_n\right) = \lim_n \mu(A_n).$$

Proof. Similar to the the previous one. □

EXERCISE II.21. Given a measure space $(X,\mathfrak{S},\mu)$ and Y in $\mathfrak{S}$, prove that:

i. $\mathfrak{S}' = \{A \in \mathfrak{S} : A \subset Y\}$ is a σ-algebra on Y;

ii. $\mu_{|\mathfrak{S}'}$, restriction of μ to $\mathfrak{S}'$, is a measure on $(Y,\mathfrak{S}')$.

II.1.3 Probability Spaces

Definition II.22. A *probability space* $(S,\mathfrak{S},P)$ is a measure space such that $P(S) = 1$. In this case, any A in $\mathfrak{S}$, P and $P(A)$ are called, respectively, *event*, *probability measure on* $(S,\mathfrak{S})$ and *probability of* A.

Proposition II.23. *If $(S,\mathfrak{S},P)$ is a probability space,*

$$(\forall A \in \mathfrak{S}) \quad P(A^c) = 1 - P(A).$$

Proof. Immediate, since for any event A, $A \cap A^c = \varnothing$ and $A \cup A^c = S$. □

Example II.24. Let S be a finite subset of $\mathbb{N}$ and define the mapping P by

$$(\forall A \in 2^S) \quad P(A) = \frac{\#A}{\#S}.$$

In this setting, P is a probability measure on $(S,2^S)$, called the *uniform probability measure*. In this case, measuring a probability amounts to a combinatorial problem.

Remark II.25. When no confusion for $\mathfrak{S}$ is possible, it is also possible to speak of a probability function on S. P is often referred to simply as a "probability," instead of probability measure; however, such an approximation is discouraged since it leads to potential confusion between various concepts.

Now, using the framework of measure spaces, the concept of conditional probability arises naturally as follows:

Proposition II.26. *Let $(S,\mathfrak{S},P)$ be a probability space. For any B in $\mathfrak{S}$ such that $P(B) \neq 0$, the mapping defined as follows:*

$$(\forall A \in \mathfrak{S}) \quad P_B(A) = \frac{P(A \cap B)}{P(B)}$$

is a probability measure on $(S,\mathfrak{S})$. In addition, for any A such that $P(A) \neq 0$,

$$P(A)P_A(B) = P(B)P_B(A) = P(A \cap B).$$

Proof. For any B in $\mathfrak{S}$,

i. P is a measure, thus $P(\varnothing) = 0$ and therefore $P_B(\varnothing) = \frac{P(\varnothing)}{P(B)} = 0$.

ii. For any sequence $(A_k)_{k\in\mathbb{N}^*}$ of pairwise disjoined sets of $\mathfrak{S}$,

$$\begin{aligned} P_B\left(\bigcup_{k\in\mathbb{N}^*} A_k\right) &= \frac{1}{P(B)} P\left(\bigcup_{k\in\mathbb{N}^*} (A_k \cap B)\right) = \frac{1}{P(B)} \sum_{k\in\mathbb{N}^*} P(A_k \cap B) \\ &= \sum_{k\in\mathbb{N}^*} P_B(A_k). \end{aligned}$$

iii. $P_B(S) = \frac{P(B)}{P(B)} = 1$.

Thus, $(S,\mathfrak{S},P_B)$ is a probability space. Now, with the additional condition that $P(A) \neq 0$, $(S,\mathfrak{S},P_A)$ is also a probability space for the same reasons; the final equality then directly results from the definition. □

Definition II.27. Under the assumptions of Proposition II.26, P_B is called *conditional probability measure assuming B* (on $(S,\mathfrak{S})$). For any A in $\mathfrak{S}$, $P_B(A)$ can also be written $P(A|B)$ and called the *probability of A assuming B*. A and B are said to be *mutually independent* if $P(A|B) = P(A)$.

Remark II.28. The mutual independence of A and B can be expressed equivalently as $P(A \cap B) = P(A)P(B)$, since $P(A|B)P(B) = P(A \cap B)$.

The following result will prove very useful when dealing with events with unit probability, but which are not necessarily equal to S:

Lemma II.29. *Let $(S,\mathfrak{S},P)$ be a probability space. If A is an event such that $P(A) = 1$, then for any event B, $P(A \cap B) = P(B)$.*

Proof. By hypothesis, $P(A^c) = 0$; in addition, $B \cap A^c \subset A^c$ thus $P(B \cap A^c) = 0$. Now, since A and A^c are disjoined, so are $B \cap A$ and $B \cap A^c$, therefore

$$P(B) = P(B \cap (A \cup A^c)) = P((B \cap A) \cup (B \cap A^c)) = P(B \cap A) + P(B \cap A^c)$$

which leads to the conclusion. □

Theorem II.30 (Total Probability and BAYES[9] Formulas). *Let $(S,\mathfrak{S},P)$ be a probability space and $(A_n)_{n\in\mathbb{N}^*}$ be a sequence of pairwise disjoined events such that*

$$\begin{cases} (\forall n\in\mathbb{N}^*) \quad P(A_n) & \neq & 0 \\ \qquad\qquad \sum\limits_{n\in\mathbb{N}^*} P(A_n) & = & 1. \end{cases}$$

For any B in $\mathfrak{S}$,

$$P(B) = \sum_{n\in\mathbb{N}^*} P(B|A_n)P(A_n) \tag{II.3}$$

and, if $P(B)\neq 0$,

$$(\forall k\in\mathbb{N}^*) \quad P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{n\in\mathbb{N}^*} P(B|A_n)P(A_n)}. \tag{II.4}$$

Proof. If $P(B)=0$ then, given that $B\cap A_n\subset B$ for all n in $\mathbb{N}^*$, $P(B|A_n)=P(B\cap A_n)=0$ which proves the first result. If $P(B)\neq 0$ then assuming $A=\bigcup_{n\in\mathbb{N}^*}A_n$, $P(A)=1$ because of the hypothesis. Therefore, combining Lemma II.29 with the disjunction of the A_n shows that

$$\sum_{n\in\mathbb{N}^*} P_{A_n}(B)P(A_n) = \sum_{n\in\mathbb{N}^*} P(B\cap A_n) = P\left(\bigcup_{n\in\mathbb{N}^*}(B\cap A_n)\right) = P(B\cap A) = P(B).$$

Finally, combining (II.3) with Proposition II.26 establishes (II.4). □

Example II.31. An individual has been purchasing candy at the vending machine daily over an extended period of time, and intends to quit doing so. Based on his experience over the past months, he estimates that each day his behavior is governed only by the outcome of the previous day, and:

i. chances that he will use the vending machine are α in $]0,1[$ if he did not the previous day;

ii. chances that he will not use the vending machine are β in $]0,1[$ if he did the previous day.

Starting with a day away from the vending machine, what are the chances that the individual will permanently quit? First, a proper measurable space $(S,\mathfrak{S})$ must be defined. Formally, the outcome of each day can be expressed as 0 (failure) or 1 (success), thus leading to boolean sequences. In other words, $S=2^{\mathbb{N}^*}$ and, in this context, the event "not hitting the vending machine on day k," with k in $\mathbb{N}^*$ can be formulated as follows:

$$A_k = \left\{(u_n)_{n\in\mathbb{N}^*}\in 2^{\mathbb{N}^*} : u_k = 1\right\}$$

and its contrary as

$$A_k^c = \left\{(u_n)_{n\in\mathbb{N}^*}\in 2^{\mathbb{N}^*} : u_k = 0\right\}.$$

The notations are consistent because $A_k\cap A_k^c=\varnothing$ and $A_k\cup A_k^c=2^{\mathbb{N}^*}$. All such events are *a priori* acceptable, thus it is natural to choose $\mathfrak{S}=2^S=2^{2^{\mathbb{N}^*}}$ as the σ-algebra, since $(S,2^S)$ is the largest measurable space which can be built on S. Now, the hypothesis assumes the existence of a probability measure P over $(S,\mathfrak{S})$, satisfying the conditions (i) and (ii) above, which can be translated as follows:

$$(\forall k\in\mathbb{N}^*) \quad \begin{cases} P(A_{k+1}^c|A_k) & = & \alpha \\ P(A_{k+1}|A_k^c) & = & \beta \end{cases}$$

[9] Thomas BAYES, English minister and mathematician, 1702–1761.

thus, using the Total Probability Formula of Theorem II.30,

$$(\forall k \in \mathbb{N}^*) \quad \begin{cases} P(A_{k+1}) &= (1-\alpha-\beta)P(A_k)+\beta \\ P(A_{k+1}^c) &= (1-\alpha-\beta)P(A_k^c)+\alpha \end{cases}$$

and, using the classical fixed-point sequence explicitation technique, along with the hypothesis that $P(A_1)=1$ and $P(A_1^c)=0$, it follows that

$$(\forall k \in \mathbb{N}^*) \quad \begin{cases} P(A_k) &= \dfrac{\beta}{\alpha+\beta}+\dfrac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^{k-1} \\ P(A_k^c) &= \dfrac{\alpha}{\alpha+\beta}-\dfrac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^{k-1} \end{cases}$$

Now, using the hypothesis about conditionality, the probability that the individual manages to stay away from the vending machine during the n first days (starting when he decided to quit), given $n \in \mathbb{N}^*$, is:

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P\left(A_n \middle| \bigcap_{k=1}^{n-1} A_k\right) P\left(\bigcap_{k=1}^{n-1} A_k\right) = P(A_n|A_{n-1})P\left(\bigcap_{k=1}^{n-1} A_k\right)$$

whence, by recurrence,

$$P\left(\bigcap_{k=1}^{n} A_k\right) = P(A_1)\prod_{k=1}^{n-1} P(A_{k+1}|A_k) = (1-\alpha)^{n-1}.$$

Finally, the probability that the individual permanently quits his unhealthy (and expensive) habit can be determined using Proposition II.20 as follows:

$$P\left(\bigcap_{n\in\mathbb{N}^*} A_k\right) = \lim_n (1-\alpha)^{n-1} = 0,$$

which is a sad conclusion.

II.2 Measurable Functions and Random Variables

In this Section, $(S, \mathfrak{S})$ will denote a measurable space.

II.2.1 Measurable Functions

Definition II.32. A function f in $[-\infty,\infty]^S$ is said to be *measurable* if, for any interval I in $[-\infty,\infty]$, $f^{<-1>}(I)$ is measurable. A complex function over S is said to be measurable if both its real and imaginary parts are measurable.

Example II.33. Given any E in $\mathfrak{S}$, $\mathbb{1}_E$ is measurable.

Remark II.34. In the case of a BOREL σ-algebra, the measurable functions are called *BOREL functions*.

Example II.35. Any f in $\mathscr{C}^0(\mathbb{R})$ is a BOREL function: for all open intervals $I \subset \mathbb{R}$, $f^{<-1>}(I)$ is an open, thus BOREL, set.

Example II.36. $\mathbb{1}_{\mathbb{Q}}$ is a BOREL function, since $\mathbb{Q}$ is a BOREL set.

Proposition II.37. *The following properties are equivalent:*

i. $f \in [-\infty,\infty]^S$ *is measurable.*

ii. $(\forall a \in \mathbb{R})\quad f^{<-1>}(]a,\infty])$ *is measurable.*

iii. $(\forall a \in \mathbb{R})\quad f^{<-1>}([a,\infty])$ *is measurable.*

iv. $(\forall a \in \mathbb{R})\quad f^{<-1>}([-\infty,a[)$ *is measurable.*

v. $(\forall a \in \mathbb{R})\quad f^{<-1>}([-\infty,a])$ *is measurable.*

Proof. (i) clearly implies each of the four other statements. Conversely, assume (ii) is true. Then, for all a in $\mathbb{R}$ and all n in $\mathbb{N}^*$, $f^{<-1>}(]a-\frac{1}{n},\infty])$ is measurable; therefore,

$$\begin{aligned} f^{<-1>}([a,\infty]) &= \{x \in S : f(x) \geqslant a\} = \bigcap_{n\in\mathbb{N}^*} \{x \in S : f(x) > a - \tfrac{1}{n}\} \\ &= \bigcap_{n\in\mathbb{N}^*} f^{<-1>}(]a-\tfrac{1}{n},\infty]) \end{aligned}$$

is measurable. Using the complement shows that both $f^{<-1>}([-\infty,a])$ and $f^{<-1>}([-\infty,a[)$ are measurable thus, by intersection, any interval of $[-\infty,\infty]$ is measurable. Therefore, (ii) implies (i) and, because it has already been proved that (i) implies the other four statements, the equivalences are deduced immediately. □

Proposition II.38. *If F is a function of $\mathscr{C}^0(\mathbb{R}^2,\mathbb{R})$ and f and g are two measurable functions in $[-\infty,\infty]^S$, then $h = F(f,g)$ is measurable.*

Proof. For any a in $\mathbb{R}$, let G_a denote $\{(u,v) \in \mathbb{R}^2 : F(u,v) > a\}$. G_a is an open subset of $\mathbb{R}^2$, *i.e.*, it is a denumerable union of open rectangles:

$$\left(\exists\, (R_n =]a_n,b_n[\times]c_n,d_n[)_{n\in\mathbb{N}^*}\right) \quad G_a = \bigcup_{n\in\mathbb{N}^*} R_n.$$

Now,

$$\begin{aligned} \{x \in S : h(x) > a\} &= \{x \in S : (f(x),g(x)) \in G_a\} \\ &= \bigcup_{n\in\mathbb{N}^*} \{x \in S : (f(x),g(x)) \in R_n\}. \end{aligned}$$

In addition, for each $n \in \mathbb{N}^*$,

$$\begin{aligned} \{x \in S : (f(x),g(x)) \in R_n\} = \{x \in S : a_n < f(x) < b_n\} \\ \bigcap \{x \in S : c_n < g(x) < d_n\} \end{aligned}$$

is a measurable subset of S. Therefore, the denumerable union of all such sets is measurable, which proves that h is a measurable function. □

From this result the following corollary can be immediately deduced:

Corollary II.39.

i. If f and g are two measurable functions in $[-\infty,\infty]^S$, then $f+g$ and fg are measurable.

ii. If f is a measurable function in $[-\infty,\infty]^S$, then, for any p in $\mathbb{N}^*$, $|f|^p$ is measurable.

The important cases of the *infimum* and the *supremum* are now discussed:

Proposition II.40. *If $(f_n)_{n\in\mathbb{N}^*}$ is a sequence of measurable functions in $[-\infty,\infty]^S$, then for all k in $\mathbb{N}$, $\mathrm{Inf}_{n>k} f_n$ and $\mathrm{Sup}_{n>k} f_n$ are measurable.*

Proof. This results immediately from the fact that, for all (a,k) in $\mathbb{R}\times\mathbb{N}$,

$$\left\{x\in S : \sup_{n>k} f_n(x) > a\right\} = \bigcup_{n>k}\{x\in S : f_n(x) > a\}.$$

□

Corollary II.41. *Let $(f_n)_{n\in\mathbb{N}^*}$ be a sequence of measurable functions in $[-\infty,\infty]^S$.*

i. $\limsup_n f_n$ and $\liminf_n f_n$ are measurable.

ii. If, for all x in S, the sequence $(f_n(x))_{n\in\mathbb{N}^*}$ has a limit $f(x)$, then f is measurable.

Proof.

i. For all m in $\mathbb{N}$, define $g_m = \mathrm{Sup}_{n>m} f_n$. According to Proposition II.40, g_m is measurable and so is $\limsup_n f_n$ since

$$\limsup_n f_n = \inf_{m\in\mathbb{N}} g_m.$$

The proof is similar for $\liminf_n f_n$.

ii. If for all x in S, $(f_n(x))_{n\in\mathbb{N}^*}$ tends to $f(x)$, then

$$(\forall x\in S)\quad \limsup_n(f_n(x)) = \liminf_n(f_n(x)) = \lim_n f_n(x) = f(x).$$

Hence, $f = \limsup_n f_n$ is measurable.

□

II.2.2 Simple Functions

In the general case, a function is said to be *simple* if it takes a finite number of values. In the context of measurable functions, the definition is specialized as follows:

Definition II.42. A function f in $\mathbb{R}^S$ is said to be *simple* if $f(S)$ is a finite subset of $\mathbb{R}$.

Remark II.43. It is important to notice that a simple function is not allowed to take on infinite values. An equivalent expression is *piecewise constant*.

Proposition II.44. *If f is a simple function in $\mathbb{R}^S$, then there exists $n \in \mathbb{N}^*$, n real numbers $(a_i)_{1 \leqslant i \leqslant n}$ and n pairwise disjoined sets $(A_i)_{1 \leqslant i \leqslant n}$ such that*

$$\begin{cases} f &= \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \\ S &= \bigcup_{i=1}^{n} A_i. \end{cases}$$

Proof. Direct consequence of Definition II.42. □

Proposition II.45. *A simple function f in $\mathbb{R}^S$ is measurable if and only if all the sets A_i, as defined in Proposition II.44, are measurable.*

Proof. By definition of f, for all a in $\mathbb{R}$ the set $f^{<-1>}([a,\infty])$ is a union of k sets A_i, $0 \leqslant k \leqslant n$. More precisely, k is the cardinality of $[a,\infty] \cap (a_i)_{1 \leqslant i \leqslant n}$. Therefore, if all these A_i are measurable, so is any k-union thereof, and in particular $f^{<-1>}([a,\infty])$, thus f is measurable.
Conversely, if one of the A_i sets, say A_{i_0}, is not measurable, then let I be a an interval such as $I \cap (a_i)_{1 \leqslant i \leqslant n}$ contains only a_{i_0}. Such an interval exists, since $(a_i)_{1 \leqslant i \leqslant n}$ is a finite subset of $\mathbb{R}$, which is separable. Then, $f^{<-1>}(I)$ is not measurable thus neither is f. □

Theorem II.46. *Let f be a measurable function in $[0,\infty]^S$. There exists a sequence $(u_n)_{n \in \mathbb{N}^*}$ of simple measurable functions in $[0,\infty[^S$, such that:*

$$\begin{cases} (\forall n \in \mathbb{N}^*) & u_n \leqslant u_{n+1} \\ (\forall x \in S) & \lim_n u_n(x) = f(x) \end{cases}$$

Proof. For all n in $\mathbb{N}^*$ and all $1 \leqslant i \leqslant n2^n$, define the following sets and functions:

$$\begin{cases} E_{n,i} &= f^{<-1>}\left(\left[\frac{i-1}{2^n}, \frac{i}{2^n}\right[\right) \\ F_n &= f^{<-1>}([n,\infty]) \\ u_n &= n\mathbb{1}_{F_n} + \sum_{i=1}^{i=n2^n} \frac{i-1}{2^n} \mathbb{1}_{E_{n,i}}. \end{cases}$$

All these $E_{n,i}$ and F_n sets are measurable, thus the u_i are measurable simple functions. The cases $n = 1$ and $n = 2$ are illustrated by Figure 3.

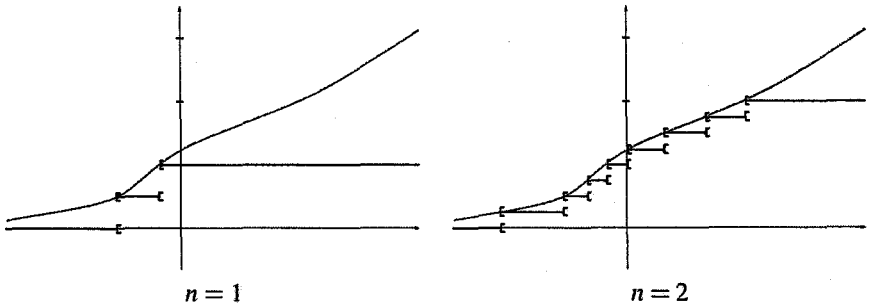

Figure 3. A monotonic sequence of simple functions: $n=1$ and $n=2$.

When n is incremented, F_n and each $E_{n,i}$ are split; more precisely:

$$E_{n,i} = E_{n+1,2i-1} \cup E_{n+1,2i}$$

while

$$\begin{aligned}
(\forall x \in E_{n,i}) \quad & u_n(x) = \tfrac{i-1}{2^n} \\
(\forall x \in E_{n+1,2i-1}) \quad & u_{n+1}(x) = \tfrac{2i-2}{2^{n+1}} = \tfrac{i-1}{2^n} \\
(\forall x \in E_{n+1,2i}) \quad & u_{n+1}(x) = \tfrac{2i-1}{2^{n+1}} > \tfrac{i-1}{2^n}.
\end{aligned}$$

On the other hand:

$$F_n = F_{n+1} \bigcup \left(\bigcup_{i=n2^{n+1}+1}^{i=(n+1)2^{n+1}} E_{n+1,i} \right)$$

while

$$\begin{aligned}
(\forall x \in F_n) \quad & u_n(x) = n \\
(\forall x \in Fn+1) \quad & u_{n+1}(x) = n+1 > n \\
(\forall x \in E_{n+1,i})(\forall i \geqslant n2^{n+1}+1) \quad & u_{n+1}(x) = \tfrac{i-1}{2^{n+1}} \geqslant n.
\end{aligned}$$

Therefore, since

$$S = F_n \bigcup \left(\bigcup_{i=1}^{i=n2^n} E_{n,i} \right),$$

it follows that

$$(\forall n \in \mathbb{N}^*)(\forall x \in S) \quad u_n(x) \leqslant u_{n+1}(x).$$

Concerning $\lim_n u_n$, two cases arise:

- if $f(x)$ is finite, then, for all $n > f(x)$,

$$|f(x) - u_n(x)| \leqslant \tfrac{1}{2^n} \to 0$$

- if $f(x) = \infty$, then $u_n(x) = n \to \infty$.

□

Remark II.47. If, in addition, f is bounded, then the convergence is uniform (*i.e.*, $\|u_n - f\|_\infty \to 0$).

II.2.3 Random Variables

Let $(T, \mathfrak{N})$ denote a measurable space; practically, the domain of interest considered here is limited to the cases where $(T, \mathfrak{N})$ is either a denumerable set along with its power set or $(\mathbb{R}^d, \mathfrak{B}^d)$.

Definition II.48. Given a probability space $(S, \mathfrak{S}, P)$, a *random variable* from S to T is a measurable function in T^S. $X(T)$ is called the *state space* and its elements the *states* of the random variable. If T is denumerable (resp. real, complex), the random variable is said to be *discrete* (resp. *real, complex*). The function $P_X = P \circ X^{<-1>}$ is called the *probability distribution of* X.

Proposition II.49. *Under the previous assumptions, $(T,\mathfrak{N},P_X)$ is a probability space.*

Proof. By hypothesis, $(T,\mathfrak{N})$ is a measurable space. In addition, because $(S,\mathfrak{S},P)$ is probability space,

i. $P_X(\varnothing) = P \circ X^{<-1>}(\varnothing) = P(\varnothing) = 0$;

ii. for any sequence $(A_n)_{n\in\mathbb{N}^*}$ measurable sets (with respect to $(T,\mathfrak{N})$,

$$P_X\left(\bigcup_{n\in\mathbb{N}^*} A_n\right) = P\left(\bigcup_{n\in\mathbb{N}^*} X(A_n)\right) = \sum_{n\in\mathbb{N}^*} P(X(A_n)) = \sum_{n\in\mathbb{N}^*} P_X(A_n);$$

iii. $P_X(T) = P \circ X^{<-1>}(T) \leqslant 1$. In addition, $S \subset X^{<-1>}(T)$ thus

$$1 = P(S) \leqslant P \circ X^{<-1>}(T) = P_X(T)$$

and therefore $P_X(T) = 1$.

□

Remark II.50. In general, the probability space $(S,\mathfrak{S},P)$ is not defined explicitly, but rather implicitly by the means of a random variable along with its probability distribution, *e.g.*, by chosing $X^{<-1>}(T)$ for S, the σ-algebra generated by $X^{<-1>}(\mathfrak{N})$ in 2^S for $\mathfrak{S}$ and $P_X \circ X$ for P. Technically, this allows to hide the probability space behind the random variable and its probability distribution.

II.2.4 Introduction to MARKOV Chains

The concept of MARKOV[10] chain is introduced here in an informal, inductive way; it will be formalized later.

Example II.51. Using the probability space of Example II.31, for all k in $\mathbb{N}^*$ the mapping X_k is defined on $2^{\mathbb{N}^*}$ by

$$\left(\forall (u_n)_{n\in\mathbb{N}^*} \in 2^{\mathbb{N}^*}\right) \quad X_k\left((u_n)_{n\in\mathbb{N}^*}\right) = u_k.$$

Each X_k is a discrete random variable with two states, 0 and 1. Its probability distribution is given by

$$\begin{cases} P_{X_k}(1) &= \dfrac{\beta}{\alpha+\beta} + \dfrac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^{k-1} \\ P_{X_k}(0) &= \dfrac{\alpha}{\alpha+\beta} - \dfrac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^{k-1} \end{cases}$$

In fact, the sequence $(X_k)_{k\in\mathbb{N}^*}$ presents the simplest case of a *uniform MARKOV chain*, with 2 states. Its behavior can be summarized as in Figure 4.

[10] Andrei MARKOV, Russian mathematician, 1856-1922.

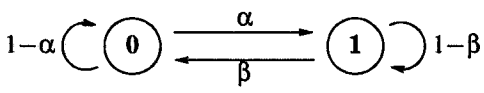

Figure 4. Two-state uniform MARKOV chain.

II Measures and Probabilities

III The LEBESGUE Measure

[...] une généralisation faite non pour le vain plaisir de généraliser, mais pour résoudre des problèmes antérieurement posés, est toujours une généralisation féconde.

H. LEBESGUE[11], in *Revue de métaphysique et de morale*, **34:2** (1927)

In this chapter, S will denote a nonempty set.

Now that an abstract measure framework has been defined, one would like to apply it so that it retrieves the results known in the particular case of the RIEMANN[12] integral, when it exists. In particular, simply by considering the case of constant functions, it appears clearly that such a measure must satisfy the following axioms:

Axiom III.1. The measure of any real interval is the difference between its upper and lower bounds.

Axiom III.2. The measure is invariant by translation.

Although only necessary, these axioms will prove restrictive enough so that it is impossible to make the entire $2^{\mathbb{R}}$ measurable while complying with them. The goal of section III.1 will precisely be to establish this impossibility, and the rest of this chapter will aim at defining a measure on as much of $2^{\mathbb{R}}$ as possible, while maintaining compliance with Axioms III.1 and III.2. At this point, several different approaches can be chosen; in particular, [Rud86] proposes an inductive method based on URYSOHN's[13] Lemma. On the other hand, the method chosen in the present chapter is a constructive one that provides a direct way to construct measures on $\mathbb{R}$ (and in particular the LEBESGUE measure) from right-continuous non decreasing functions (thus in particular from cumulative distribution functions). This angle of attack is especially appropriate here, because it is directly connected to applications in probability theory.

III.1 Preamble

The aim of this section is, first, to explain why there is no way to define a measure complying with Axioms III.1 and III.2 on the entire $2^{\mathbb{R}}$. To this goal, a counter-example will be exhibited. Then, in order to provide a clear overview of this somewhat lengthy process, the construction of the LEBESGUE measure by the means of CARATHÉODORY's[14] Extension is summarized.

III.1.1 Fountainhead

The fountainhead counter-example is based on the following axiom, which is part of the NBG Set Theory.

[11] Henri LEBESGUE, French mathematician, 1875–1941.
[12] Bernhard RIEMANN, German mathematician, 1826-1866.
[13] Pavel URYSOHN, Russian mathematician, 1898–1924.
[14] Constantin CARATHÉODORY, German mathematician, 1873–1950.

Axiom III.3 (Axiom of choice). Given any set of pairwise disjoined nonempty sets, there exists at least one set that contains exactly one element of each of the nonempty sets.

Lemma III.4. *There exists a subset of* $]0,1[$, *which is a set of class representatives of* $\mathbb{R}/\mathbb{Q}$.

Proof. First, notice that $(\mathbb{R},+)$ is abelian, thus each of its subgroups, and in particular $\mathbb{Q}$, is normal. In addition, recall that because $\mathbb{Q}$ is a normal subgroup of $(\mathbb{R},+)$, saying that two real numbers, x and y, are in relation with each other if and only if $x-y$ is rational defines an equivalence relation on $(\mathbb{R},+)$, thus leading to equivalence classes over $\mathbb{R}$; the collection of these equivalence classes constitutes $\mathbb{R}/\mathbb{Q}$, by definition. Moreover, the axiom of choice guarantees the existence of a set E of class representatives, *i.e.*, a set formed by taking exactly one element in each equivalence class. Let denote E such a set and, for any x in E, $[x]$ its integer part; $x-[x]$ is equivalent to x and belongs to $]0,1[$ if $x \notin \mathbb{Z}$, thus $x-[x]$ provides another class representative which complies with the requirements. If $x \in \mathbb{Z}$, it suffices to take $\frac{1}{2}$ as a class representative. □

Prior to exhibiting the counter-example, the following definition will be useful:

Definition III.5. For any subset A of $\mathbb{R}$ and r in $\mathbb{R}$, the set $\{x+r,\, x \in A\}$ is denoted $A+r$.

Proposition III.6. *There is no measure on* $(\mathbb{R},2^{\mathbb{R}})$ *which complies with Axioms III.1 and III.2.*

Proof. Let μ be a measure on $(\mathbb{R},2^{\mathbb{R}})$ which satisfies Axioms III.1 and III.2; this implies, in particular, that the set E defined in the proof of Lemma III.4 is measurable. Let also denote

$$F = \bigcup_{r\in\mathbb{Q}\cap]-1,1[} (E+r).$$

In particular, F is a subset of $]-1,2[$ thus, because of Axiom III.1,

$$\mu(F) \leqslant \mu(]-1,2[) = 3. \tag{III.1}$$

Moreover, for all x in $]0,1[$, there is an equivalence class in $\mathbb{R}/\mathbb{Q}$ to which x belongs; in other words, by definition of E, there exists y in E such that $x-y=r$ is rational, thus $x=y+r$ belongs to $E+r$. Therefore, $]0,1[\subset F$ and

$$\mu(F) \geqslant \mu(]0,1[) = 1. \tag{III.2}$$

Now, let r and s be two distinct rational numbers; any x which belongs to $E+r$ and $E+s$ must satisfy

$$(\exists (y,z) \in E^2) \quad x = y+r = z+s$$

and thus, $y-z = s-r \neq 0$. Hence, y and z are distinct and belong to the same equivalence class of $\mathbb{R}/\mathbb{Q}$, which contradicts the construction of E. Therefore, $E+r$ and $E+s$ are disjoined and, using Axiom III.2,

$$\mu(F) = \sum_{r\in\mathbb{Q}\cap]-1,1[} \mu(E+r) = \sum_{r\in\mathbb{Q}\cap]-1,1[} \mu(E) = \begin{cases} 0 & \text{if} \quad \mu(E)=0 \\ \infty & \text{if} \quad \mu(E)\neq 0, \end{cases} \tag{III.3}$$

which contradicts the fact that $1 \leqslant \mu(F) \leqslant 3$. Therefore, the hypothesis is inconsistent, *i.e.*, there cannot be a measure on $(\mathbb{R},2^{\mathbb{R}})$ that satisifies Axioms III.1 and III.2. □

A direct consequence of Proposition III.6 is that it is impossible to extend the RIEMANN integral to arbitrary sets of reals. The goal of this chapter will therefore be to extend of the classical notions of length and area to more complicated sets than those acceptable in the framework of RIEMANN integral; although it is not possible to do so for the entire $2^{\mathbb{R}}$, it will be shown that a large class of such sets can be addressed, in particular all BOREL sets.

Map	$\underset{\text{measure}}{\mu} \overset{e}{\Longrightarrow} \underset{\text{outer-measure}}{\mu^*} \overset{r}{\longrightarrow} \underset{\text{complete measure}}{\mu'} \overset{r}{\longrightarrow} \underset{\text{measure}}{\bar{\mu}}$
Definition set	$\underset{\text{boolean alg.}}{\mathcal{S}} \subset \underset{\sigma\text{-algebra}}{2^S} \supset \underset{\sigma\text{-algebra}}{\mathfrak{S}(\mu^*)} \supset \underset{\sigma\text{-algebra}}{\mathfrak{S}(\mathcal{S})}$

Table 1. Summary of CARATHÉODORY's Extension: $\overset{e}{\Longrightarrow}$ and $\overset{r}{\longrightarrow}$ denote, respectively, an extension and a restriction. Note: $\mathcal{S} \subset \mathfrak{S}(\mathcal{S})$.

III.1.2 Constructive Approach to the LEBESGUE Measure

Due to the complexity of the LEBESGUE measure construction, it is decomposed into several steps to make it easier to understand. It shall be kept in mind from the beginning that, as it often happens with constructive approaches, the whole picture will emerge only at the end.

The construction can be summarized as follows:

Step 1: The set I, formed of all intervals $]a,b]$ with $a < b$ in $\mathbb{R}$, is introduced. Although I does not have any structural properties, in particular is not closed for set operations, it is shown that from any nondecreasing right-continuous function F of $\mathbb{R}^{\mathbb{R}}$, a function μ_F can be defined over I with properties similar to those of measures. In particular, μ_F is σ-additive over I.

Step 2: The closure $\mathcal{J}$ of I for the usual set operations turns out to be the set of all possible finite unions of intervals with the form $]-\infty,a]$, $]b,c]$ or $]d,\infty[$ with $a, b < c$ and d in $\mathbb{R}$. In addition, any interval $]-\infty,a]$ or $]d,\infty[$ can be expressed as a denumerable union of pairwise disjoined elements of I. Therefore, the σ-additivity of μ_F is used to extend it to a function μ defined on $\mathcal{J}$.

Step 3: The triple $(\mathbb{R},\mathcal{J},\mu)$ has most of the properties of measure spaces, but still lacks a key one: $\mathcal{J}$ is not closed for denumerable union (nor for intersection); in fact, $\mathcal{J}$ is a *boolean algebra* and, for the sake of generality, this concept will be formalized. For instance, $\mathfrak{B} \not\subset \mathcal{J}$, *e.g.*, a closed interval is not measurable and neither is a finite set. The approach chosen here is to use a technique called CARATHÉODORY's Extension; this approach is generic, and thus not limited to the case of $\mathbb{R}$. Summarized in Table 1, this constructive approach proceeds as follows:

i. from a measure on a boolean algebra $\mathcal{S}$ on S, is built an *outer-measure* μ^*, defined on the the entire 2^S. Such objects are more general and in particular, lack σ-additivity. However, the relative roughness of μ^* allows it to remain compatible with Axioms III.1 and III.2 on the entire 2^S;

ii. then, the outer-measure is restricted to μ', defined over a particular subset $\mathfrak{S}(\mu^*)$ of S; this subset is chosen in such a way so that $(S,\mathfrak{S}(\mu^*),\mu')$ forms a measure space;

iii. now, the most striking aspect of this process is that $\mathfrak{S}(\mu^*)$ is not only large enough to be interesting, but that, provided an additional condition called σ-*finiteness*, it actually contains the initial boolean algebra $\mathcal{S}$. Since this implies that $\mathfrak{S}(\mu^*)$ also contains the σ-algebra $\mathfrak{S}(\mathcal{S})$ generated by $\mathcal{S}$, μ' is finally restricted to it, thus providing a measure $\bar{\mu}$ on the desired σ-algebra.

Step 4: This process is specialized, in the case where S is $\mathbb{R}$; this gives birth to a measure on the BOREL σ-algebra introduced in chapter II. However, it is established that the resulting measure space is not

complete measure, in a newly-defined sense (and intuitively close to that of completeness of metric spaces).

Step 5: A new meaning is given for the concept of *completeness* in the context of probability spaces. It is then explained how a measure space can be *completed* with respect to this definition. Finally, applying this process to the BOREL σ-algebra and measure concludes the construction of the LEBESGUE measure.

III.2 Measures on Boolean Algebras

The goal of this section is to construct, on a structure that is less restrictive than a σ-algebra on $\mathbb{R}$, a measure complying with Axioms III.1 and III.2.

III.2.1 Interval Measures

First, it is shown that with any right-continuous real function, there can be associated a set function defined on the class over left-open, right-closed intervals. Moreover, this set function will comply with Axiom III.2 and, for a particular choice of the "generating" function, with Axiom III.1.

Definition III.7. Let I denote the set of all intervals of the form $]a,b]$, where $(a,b) \in \mathbb{R}^2$ with $a \leqslant b$.

Remark III.8. I is indeed a set of subsets of $\mathbb{R}$; in other words, $I \subset 2^{\mathbb{R}}$.

Proposition III.9. *For any right-continuous and nondecreasing $F \in \mathbb{R}^{\mathbb{R}}$, let μ_F be the function defined over I as follows:*

$$(\forall (a,b) \in \mathbb{R}^2, a < b) \quad \mu_F(]a,b]) = F(b) - F(a).$$

Then μ_F satisfies the following properties:

i. *μ_F is nonnegative and $\mu_F(\varnothing) = 0$;*

ii. *for all $a \in \mathbb{R}$, $\mu_F(]a,b]) \downarrow 0$ when $b \downarrow a$;*

iii. *μ_F is monotonic over I;*

iv. *for all $(a,b,c) \in \mathbb{R}^3$, $a \leqslant b \leqslant c$ implies $\mu_F(]a,c]) = \mu_F(]a,b]) + \mu_F(]b,c])$;*

v. *μ_F is σ-additive over I.*

Proof. Only the proof of the last point is detailed here since the others are trivial and left to the reader. Let $(]a_k,b_k])_{k\in\mathbb{N}^*}$ be a sequence of pairwise disjoined intervals of I, such that their denumerable union belongs to I and thus can be written as $]a,b]$. For all $n \in \mathbb{N}^*$, it is always possible to re-index the n first intervals of the sequence such that $b_k \leqslant a_{k+1}$ for each $1 \leqslant k \leqslant n-1$, which also implies $a \leqslant a_1$ and $b_n \leqslant b$; then

$$\sum_{k=1}^{k=n} \mu_F(]a_k,b_k]) \leqslant \sum_{k=1}^{k=n} \mu_F(]a_k,b_k]) + \sum_{k=1}^{k=n-1} \mu_F(]b_k,a_{k+1}]) = \mu_F(]a_1,b_n]) \leqslant \mu_F(]a,b])$$

thus, taking the limit,

$$\sum_{n\in\mathbb{N}^*} \mu_F(]a_n,b_n]) \leqslant \mu_F(]a,b]).$$

In order to demonstrate the inverse inequality, it shall first be acknowledged that if $a = b$, then the result is immediate. If $a < b$, then let ε be in $]0, b-a[$. F is right-continuous, therefore

$$(\forall n \in \mathbb{N}^*)(\exists \varepsilon_n \in [0,\infty[^*) \quad \mu_F(]b_n, b_n+\varepsilon_n]) = F(b_n+\varepsilon_n) - F(b_n) < \frac{\varepsilon}{2^n}.$$

Now,

$$(\forall n \in \mathbb{N}^*) \quad [a+\varepsilon, b] \subset]a,b] = \bigcup_{n\in\mathbb{N}^*}]a_n, b_n] \subset \bigcup_{n\in\mathbb{N}^*}]a_n, b_n+\varepsilon_n[,$$

but because of HEINE[15]-BOREL-LEBESGUE's Theorem[16],

$$(\exists m \in \mathbb{N}^*) \quad [a+\varepsilon, b] \subset \bigcup_{k=1}^{k=m}]a_k, b_k+\varepsilon_k[$$

thus, by recurrence on k,

$$\begin{aligned}\mu_F([a+\varepsilon,b]) &\leqslant \sum_{k=1}^{k=m} \mu_F(]a_k, b_k+\varepsilon_k[) = \sum_{k=1}^{k=m} \mu_F(]a_k,b_k]) + \sum_{k=1}^{k=m} \mu_F(]b_k, b_k+\varepsilon_k[) \\ &\leqslant \sum_{k=1}^{k=m} \mu_F(]a_k,b_k]) + \sum_{k=1}^{k=m} \frac{\varepsilon}{2^k} \leqslant \sum_{n\in\mathbb{N}^*} \mu_F(]a_n,b_n]) + \varepsilon.\end{aligned}$$

Thus, because F is left-continuous, the conclusion arises by taking the limits when ε tends to 0. □

III.2.2 Boolean Algebras as Closures

First, a structure slightly more general than σ-algebras is introduced:

Definition III.10. Any $\mathcal{S} \subset 2^S$ such that:

i. $S \in \mathcal{S}$

ii. if $A \in \mathcal{S}$, then $A^c \in \mathcal{S}$

iii. if $(A,B) \in \mathcal{S}^2$, then $A \cup B \in \mathcal{S}$

is called a *boolean algebra on S*.

Example III.11. Any σ-algebra on S is a boolean algebra on S.

Example III.12. The set of all finite unions of real intervals is a boolean algebra on $\mathbb{R}$.

Remark III.13. A boolean algebra on S necessarily contains $\varnothing$ and is closed for set difference, finite union and finite intersection. In fact, the name is somewhat misleading, since a boolean algebra is *not* an algebra. It is in fact a *semialgebra*, when $\mathcal{S}$ is equipped with the $\cap$, $\cup$ and $\setminus$ binary operators.

Proposition III.14. *The boolean algebra $\mathcal{I}$ generated by I is the collection of all finite unions of intervals with the form $]-\infty, a]$, $]b,c]$ or $]d,\infty[$, a, $b < c$ and d in $\mathbb{R}$.*

[15] Heinrich HEINE, German mathematician, 1821–1881.

[16] From any open cover of a closed bounded domain can be extracted a finite open cover.

Proof. Exercise. Check the condition between b and c. □

The notion of measure is now extended, *via* a slight modification, to the case of boolean algebras:

Definition III.15. Let $\mathcal{S}$ be a boolean algebra on S. A *measure on* $(S, \mathcal{S})$ is a mapping μ in $[0,\infty]^{\mathcal{S}}$ such that:

i. $\mu(\varnothing) = 0$;

ii. for any sequence $(A_n)_{n\in\mathbb{N}^*}$ of pairwise disjoined sets in $\mathcal{S}$ such that their denumerable union belongs to $\mathcal{S}$,

$$\mu\left(\bigcup_{n\in\mathbb{N}^*} A_n\right) = \sum_{n\in\mathbb{N}^*} \mu(A_n).$$

For any A in $\mathcal{S}$, $\mu(A)$ is called the *measure of* A.

Proposition III.16. *A measure on a boolean algebra is countably subadditive.*

Proof. The proof is very similar to that of Proposition II.16, using the sets defined, for all n in $\mathbb{N}^*$, by $B_n = A \cap A_n$ and $C_{n+1} = B_{n+1} \setminus \bigcup_{i=1}^{n} B_i$ with $C_1 = B_1$. □

Proposition III.17. *Under the hypothesis of Proposition III.9, there exists a unique measure μ on $\mathfrak{I}$ such that $\mu_{|I} = \mu_F$.*

Proof. First, any interval with the form $]-\infty, a]$ or $]a, \infty[$, where $a \in \mathbb{R}$, can be written as a denumerable union of pairwise disjoined intervals in I. Therefore, in order to extend μ_F to $\mathfrak{I}$, it suffices to show that μ is completely determined by

$$\mu(A) = \sum_{n\in\mathbb{N}^*} \mu_F(I_n),$$

for all $(I_n)_{n\in\mathbb{N}^*} \in I^{\mathbb{N}^*}$ with denumerable union A. In fact, if A is also the denumerable union of the sequence with pairwise disjoined terms $(J_n)_{n\in\mathbb{N}^*} \in I^{\mathbb{N}^*}$,

$$(\forall n \in \mathbb{N}^*) \quad I_n = A \cap I_n = \bigcup_{m\in\mathbb{N}^*} (J_m \cap I_n)$$

and

$$(\forall m \in \mathbb{N}^*) \quad J_m = A \cap J_m = \bigcup_{n\in\mathbb{N}^*} (I_n \cap J_m).$$

Now, I is closed for $\cap$, thus each $I_n \cap J_m$ belongs to I, and because of the σ-additivity of I,

$$\sum_{n\in\mathbb{N}^*} \mu_F(I_n) = \sum_{n\in\mathbb{N}^*} \sum_{m\in\mathbb{N}^*} \mu_F(J_m \cap I_n) = \sum_{m\in\mathbb{N}^*} \sum_{n\in\mathbb{N}^*} \mu_F(I_n \cap J_m) \sum_{m\in\mathbb{N}^*} \mu_F(J_m).$$

□

Remark III.18. An immediate consequence of the latter result is that, in the particular case where F is the identity function over $\mathbb{R}$ then, if the $(I_k)_{1\leqslant k\leqslant n}$, $n \in \mathbb{N}^*$, are pairwise disjoined intervals with respective lower and upper bounds a_k and b_k,

$$\mu\left(\bigcup_{k=1}^{k=n} I_k\right) = \sum_{k=1}^{k=n} (b_k - a_k)$$

which clearly complies with Axioms III.1 and III.2.

III.3 CARATHÉODORY's Extension

Now that the first "brick" constituted by the boolean algebra $\mathcal{J}$ and its measure μ has been planted, CARATHÉODORY's Extension can built.

III.3.1 Outer-Measures

The goal here is to introduce a new class of maps on 2^S. Albeit too crude to be measures, these maps have enough properties to prepare the construction of measure spaces, starting from parts of 2^S that are not necessarily σ-algebras.

Definition III.19. An *outer-measure on S* is a mapping μ^* from 2^S to $[0,\infty]$ such that

i. $\mu^*(\varnothing) = 0$;

ii. for all subsets A and B of S, $A \subset B$ implies $\mu^*(A) \leqslant \mu^*(B)$;

iii. μ^* is countably subadditive.

Example III.20. Any measure on $(S, 2^S)$ is also an outer-measure on S, since the axioms of Definition II.11 imply those of Definition III.19.

Remark III.21. Outer-measures are more general objects than measures because they do not require σ-additivity: this allows them to address larger classes of objects without encountering the kind of contradictions met before. For instance, in the the proof of Lemma III.4, if an outer-measure μ^* was used instead of μ, then (III.1) as well as (III.2) would still hold, but (III.3) would not, precisely because of the loss of σ-additivity. Instead, (III.3) would become

$$\mu^*(F) \leqslant \sum_{r\in\mathbb{Q}\cap]-1,1[} \mu^*(E+r) = \sum_{r\in\mathbb{Q}\cap]-1,1[} \mu^*(E) = \begin{cases} 0 & \text{if} \quad \mu^*(E) = 0 \\ \infty & \text{if} \quad \mu^*(E) \neq 0, \end{cases}$$

which is not contradictory with the fact that $\mu^*(F)$ is in $[1,3]$. This only implies that $\mu^*(E) \neq 0$.

In the case where a measure on a boolean algebra is already known, there is a very convenient and general way of deducing an outer-measure from it. This technique is used, in particular, to demonstrate CARATHÉODORY's Extension.

Definition III.22. Let $\mathcal{S}$ be a either a boolean algebra or a σ-algebra on S. For any given $A \in \mathcal{S}$, an *$\mathcal{S}$-cover of A* is a sequence $(A_n)_{n\in\mathbb{N}^*} \in \mathcal{S}^{\mathbb{N}^*}$ such that

$$A \subset \bigcup_{n\in\mathbb{N}^*} A_n.$$

The set of all $\mathcal{S}$-covers of A is denoted as $\mathrm{Cov}_{\mathcal{S}}(A)$.

Proposition III.23. *Let $\mathcal{S}$ be a boolean algebra on S and μ a measure on $(S,\mathcal{S})$. The mapping μ^* defined on 2^S by*

$$(\forall A \in 2^S) \quad \mu^*(A) = \inf_{\mathrm{Cov}_{\mathcal{S}}(A)} \sum_{n\in\mathbb{N}^*} \mu(A_n)$$

is an outer-measure on S such that $\mu^*_{|\mathcal{S}} = \mu$.

Proof. With the usual convention that the *infimum* of any function over $\varnothing$ is ∞, μ^* is well defined on 2^S and is nonnegative by construction. In addition,

i. $\mu^*(\varnothing) = 0$ since $(\varnothing)_{\mathbb{N}^*} \in \mathcal{S}^{\mathbb{N}^*} = \mathrm{Cov}_{\mathcal{S}}(\varnothing)$;

ii. if $A \subset B \subset S$, then
$$\left(\forall (A_n)_{n\in\mathbb{N}^*} \in \mathcal{S}^{\mathbb{N}^*}\right) \quad B \subset \bigcup_{n\in\mathbb{N}^*} A_n \Longrightarrow A \subset \bigcup_{n\in\mathbb{N}^*} A_n,$$
thus, $\mathrm{Cov}_{\mathcal{S}}(A) \subset \mathrm{Cov}_{\mathcal{S}}(B)$ and, therefore, $\mu^*(A) \leqslant \mu^*(B)$;

iii. in order to establish the countable subadditivity for μ^*, let $(A_n)_{n\in\mathbb{N}^*}$ be a sequence of subsets of S, and consider the two following cases:

- either $\mu^*(A_n) = \infty$ for at least one $n \in \mathbb{N}^*$ and the conclusion is immediate;
- or, for all $n \in \mathbb{N}^*$, $\mu^*(A_n)$ is finite; by definition of μ^*, this means that, given any $\varepsilon \in [0,\infty[^*$,
$$(\forall n \in \mathbb{N}^*)(\exists (A_n^p)_{p\in\mathbb{N}^*} \in \mathrm{Cov}_{\mathcal{S}}(A_n)) \quad \sum_{p\in\mathbb{N}^*} \mu(A_n^p) < \mu^*(A_n) + \frac{\varepsilon}{2^n}$$
therefore,
$$\mu^*(A_n) \leqslant \sum_{n\in\mathbb{N}^*}\sum_{p\in\mathbb{N}^*} \mu(A_n^p) \leqslant \sum_{n\in\mathbb{N}^*} \mu(A_n) + \varepsilon,$$
and since this is true for all $\varepsilon \in [0,\infty[$, the countable subadditivity arises.

Therefore, μ^* is an outer-measure on S. Finally, $\mu^*_{|\mathcal{S}} = \mu$ since for any $A \in \mathcal{S}$, $(A, \varnothing, \cdots) \in \mathrm{Cov}_{\mathcal{S}}(A)$ thus $\mu^*(A) \leqslant \mu(A)$; conversely, for all $\mathfrak{S}$-cover $(A_n)_{n\in\mathbb{N}^*}$ of A, considering the $(B_n)_{n\in\mathbb{N}^*}$ defined by $B_1 = A_1 \cap A$ and
$$(\forall n \in \mathbb{N}^*) \quad B_{n+1} = (A_{n+1} \cap A) \setminus B_n,$$
one has that $\mu(A) \leqslant \mu^*(A)$. Since this is true for all $A \in \mathcal{S}$, it follows that $\mu^*_{|\mathcal{S}} = \mu$. □

Nevertheless, in the goal of extending the classical properties of the RIEMANN-integral, one wants to maintain σ-additivity, therefore the outer-measure must be restricted to a specific class of sets on which it will turn out to be a measure.

III.3.2 μ^*-measurability

A striking property of outer-measures is that they permit the construction of measure spaces in a straightforward manner. First, a particular class of subsets of S is defined:

Definition III.24. Let μ^* be an outer-measure on S. Any $A \subset S$ such that
$$(\forall E \subset S) \quad \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c) \tag{III.4}$$
is said to be *μ^*-measurable.*

Now, the following result justifies the entire approach:

Theorem III.25. *For any outer-measure μ^* on S, let denote $\mathfrak{S}(\mu^*)$ the collection of all μ^*-measurable sets; then, $\left(S, \mathfrak{S}(\mu^*), \mu^*_{|\mathfrak{S}(\mu^*)}\right)$ is a measure space.*

Proof. First,

i. for all $E \subset S$, $\mu^*(E \cap \varnothing) + \mu^*(E \cap \varnothing^c) = \mu^*(E)$ thus $\varnothing \in \mathfrak{S}(\mu^*)$;

ii. for all $A \in \mathfrak{S}(\mu^*)$, $A^{cc} = A$. Thus, because of the symmetry of (III.4), $A^c \in \mathfrak{S}(\mu^*)$;

iii. establishing that $\mathfrak{S}(\mu^*)$ is closed for denumerable union is quite long, and is split in two steps for the sake of clarity: first, union, then denumerable union. Hence, let A and B be in $\mathfrak{S}(\mu^*)$ and S be any subset of E; in particular, $E \cap (A \cup B)^c = E \cap A^c \cap B^c$, thus,

$$E \cap (A \cup B) = (E \cap A \cap B) \cup (E \cap A \cap B^c) \cup (E \cap A^c \cap B).$$

Therefore,

$$\mu^*(E \cap (A \cup B)^c) = \mu^*(E \cap A^c \cap B^c), \tag{III.5}$$

and because of the subadditivity of μ^*,

$$\mu^*(E \cap (A \cup B)) \leqslant \mu^*(E \cap A \cap B) + \mu^*(E \cap A \cap B^c) + \mu^*(E \cap A^c \cap B). \tag{III.6}$$

Now, applying successively the definition of a μ^*-measurable set to A and B shows first that $\mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap A^c)$, then

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

thus, combining with (III.5) and (III.6),

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

Conversely, the subadditivity of μ^* shows directly that

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

whence

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

Thus, since this is true for all $E \subset S$, $A \cup B$ belongs to $\mathfrak{S}(\mu^*)$. In other words, $\mathfrak{S}(\mu^*)$ is closed for union, and since $A \cap B = (A^c \cup B^c)^c$, it is also closed for intersection, and furthermore, for set difference. Now, if $(A_n)_{n \in \mathbb{N}}$ is a sequence μ^*-measurable sets, assume $B_1 = A_1$ and define, for all n in $\mathbb{N}^*$, $B_{n+1} = A_{n+1} \setminus \cup_{i=1}^{i=n} A_i$. By definition, these B_n are pairwise disjoined; moreover, by recurrence on union closedness, they are μ^*-measurable and

$$(\forall n \in \mathbb{N}^*) \quad \bigcup_{i=1}^{i=n} B_i = \bigcup_{i=1}^{i=n} A_i \in \mathfrak{S}(\mu^*). \tag{III.7}$$

Therefore, establishing closedness for denumerable union amounts to establishing closedness for disjoined denumerable union: the question is that of the μ^*-measurability of

$$B = \bigcup_{n \in \mathbb{N}^*} B_n = \bigcup_{n \in \mathbb{N}^*} A_n.$$

On the one hand, for all $E \subset S$ and all n in $\mathbb{N}^*$, it follows from the μ^*-measurability of A_n that

$$\mu^*\left(E \cap \bigcup_{i=1}^{i=n} B_i\right) = \mu^*\left(E \cap B_n \cap \bigcup_{i=1}^{i=n} B_i\right) + \mu^*\left(E \cap B_n^c \cap \bigcup_{i=1}^{i=n} B_i\right)$$
$$= \mu^*(E \cap B_n) + \mu^*\left(E \cap \bigcup_{i=1}^{i=n-1} B_i\right)$$

since the B_i are pairwise disjoined. Therefore, by recurrence,

$$(\forall (E,n) \in 2^S \times \mathbb{N}^*) \quad \mu^*\left(E \cap \bigcup_{i=1}^{i=n} B_i\right) = \sum_{i=1}^{i=n} \mu^*(E \cap B_i). \tag{III.8}$$

Now, by definition of B,

$$(\forall (E,n) \in 2^S \times \mathbb{N}^*) \quad B^c \subset \left(\bigcup_{i=1}^{i=n} B_i\right)^c$$

thus,

$$(\forall (E,n) \in 2^S \times \mathbb{N}^*) \quad \mu^*(E \cap B^c) \leqslant \mu^*\left(E \cap \left(\bigcup_{i=1}^{i=n} B_i\right)^c\right),$$

which, combined with (III.7) and (III.8), leads to:

$$(\forall (E,n) \in 2^S \times \mathbb{N}^*) \quad \mu^*(E) = \mu^*\left(E \cap \bigcup_{i=1}^{i=n} B_i\right) + \mu^*\left(E \cap \left(\bigcup_{i=1}^{i=n} B_i\right)^c\right)$$
$$\geqslant \sum_{i=1}^{i=n} \mu^*(E \cap B_i) + \mu^*(E \cap B^c).$$

and since this is true for all n in $\mathbb{N}^*$, one can take the limit when n tends to ∞ while conserving inequalities:

$$(\forall E \subset S) \quad \mu^*(E) \geqslant \sum_{n \in \mathbb{N}^*} \mu^*(E \cap B_n) + \mu^*(E \cap B^c). \tag{III.9}$$

On the other hand, for all $E \subset S$, $E = (E \cap B) \cup (E \cap B^c)$. Thus, because of the subadditivity of μ^*, it follows that

$$(\forall E \subset S) \quad \mu^*(E) \leqslant \mu^*(E \cap B) + \mu^*(E \cap B^c)$$

while, by definition of B and because of the countable subadditivity of μ^*,

$$(\forall E \subset S) \quad \mu^*(E \cap B) = \mu^*\left(\bigcup_{n \in \mathbb{N}^*} (E \cap B_n)\right) \leqslant \sum_{n \in \mathbb{N}^*} \mu^*(E \cap B_n).$$

Hence,

$$(\forall E \subset S) \quad \mu^*(E) \leqslant \mu^*(E \cap B) + \mu^*(E \cap B^c) \leqslant \sum_{n \in \mathbb{N}^*} \mu^*(E \cap B_n) + \mu^*(E \cap B^c),$$

which, combined with (III.9), shows the two following equalities:

$$(\forall E \subset S) \quad \mu^*(E) = \sum_{n \in \mathbb{N}^*} \mu^*(E \cap B_n) + \mu^*(E \cap B^c) \tag{III.10}$$

and

$$(\forall E \subset S) \quad \mu^*(E) = \mu^*(E \cap B) + \mu^*(E \cap B^c),$$

reaching to the conclusion that $(S, \mathfrak{S}(\mu^*))$ is a measurable space.

iv. Now, $\mu^*_{|\mathfrak{S}(\mu^*)}$ being denoted for simplicity as μ', it follows directly from the definition of μ^* that $\mu'(\varnothing) = \mu^*(\varnothing) = 0$.

v. Finally, if $(A_n)_{n\in\mathbb{N}^*}$ is a sequence of pairwise disjoined μ^*-measurable sets, with denumerable union A, it follows directly from (III.10) that

$$\mu'(A) = \mu^*(A) = \sum_{n\in\mathbb{N}^*} \mu^*(A\cap A_n) + \mu^*(A\cap A^c) = \sum_{n\in\mathbb{N}^*} \mu'(A_n).$$

Therefore, $(S, \mathfrak{S}(\mu^*), \mu')$ is a measure space.

□

Remark III.26. The μ^*-measurable subsets of S are exactly the measurable sets of the measure space $(S, \mathfrak{S}(\mu^*), \mu')$.

III.3.3 σ-Algebra Associated with an Outer-Measure

Theorem III.25 concludes that there is always a straightforward way to construct a measure space from an outer-measure on a set. A definition summarizes this result:

Definition III.27. Under the hypothesis of Theorem III.25, $\mathfrak{S}(\mu^*)$ is called the *σ-algebra associated with* μ^*.

Now that all the preliminary work has been done, CARATHÉODORY's Extension arises naturally.

Definition III.28. Let $\mathcal{S}$ be a either a boolean algebra or a σ-algebra on S and μ a measure on $(S, \mathcal{S})$. If there exists a sequence $(A_n)_{n\in\mathbb{N}^*} \in \mathcal{S}^{\mathbb{N}^*}$ such that

$$\begin{cases} (\forall n \in \mathbb{N}^*) \quad \mu(A_n) < \infty \\ S = \bigcup_{n\in\mathbb{N}^*} A_n, \end{cases}$$

then μ is said to be *σ-finite*.

Example III.29. The counting measure # on $(\mathbb{N}, 2^{\mathbb{N}})$ is σ-finite since *e.g.*, $\mathbb{N}$ can be covered with singletons, each of those having, by definition, a unit thus finite counting measure.

Theorem III.30 (CARATHÉODORY's Extension). *Let $\mathcal{S}$ be a boolean algebra on S and $\mathfrak{S}(\mathcal{S})$ the σ-algebra on S generated by $\mathcal{S}$. Any σ-finite measure on $(S, \mathcal{S})$ can be uniquely extended to a σ-finite measure on $(S, \mathfrak{S}(\mathcal{S}))$.*

Proof. The proof results directly from Lemma III.23 and Theorem III.25; the only part which remains to be proved is the fact that $\mathfrak{S}(\mathcal{S}) \subset \mathfrak{S}(\mu^*)$. This is left to the reader as an exercise. □

Now, this technique permits to exhibit particular measures having the relevant properties.

Corollary III.31. *If $F \in \mathbb{R}^{\mathbb{R}}$ is a right-continuous nondecreasing function, then there exists a unique measure μ defined on $(\mathbb{R}, \mathfrak{B})$ such that*

$$(\forall]a,b] \subset \mathbb{R}) \quad \mu(]a,b]) = F(b) - F(a).$$

Proof. Direct application of Theorem III.30. □

Definition III.32. Under the hypothesis of Corollary III.31 and in the particular case where F is the identity over $\mathbb{R}^d$, μ is called the *BOREL measure on* $\mathbb{R}^d$.

Remark III.33. At this stage, it should be noted that the BOREL measure on $\mathbb{R}$ complies with Axioms III.1 and III.2 by construction. In particular, the measure of a real singleton (which is a BOREL set as mentioned in Remark II.7) is 0.

Are we there yet? No, since there are still plenty of non-BOREL sets which could be measured with very limited additional efforts (*e.g.*, Example II.10 exhibits such a set). But it is already possible to apply these powerful results to reformulate the concept of cumulative distribution functions in a very generic way:

Definition III.34. A *cumulative distribution function*, or *CDF*, *on* $\mathbb{R}$ is a right-continuous nondecreasing function $F \in [0,\infty[^{\mathbb{R}}$ such that $F(-\infty) = 0$ and $F(\infty) = 1$.

Using CARATHÉODORY's Extension, it immediately follows that:

Theorem III.35. *Let F be a CDF on $\mathbb{R}$. There exists a unique probability measure, also denoted F, on $(\mathbb{R},\mathfrak{B})$ such that*

$$(\forall\,]a,b] \subset \mathbb{R}) \quad F(]a,b]) = F(b) - F(a).$$

Proof. The proof is immediate, from Corollary III.31, except for the fact that F is indeed a probability measure on $(\mathbb{R},\mathfrak{B})$. This arises because $]-n,n] \uparrow \mathbb{R}$, thus,

$$F(\mathbb{R}) = \lim_n F(]-n,n]) = \lim_n (F(n) - F(-n)) = 1 - 0 = 1.$$

□

Corollary III.36. *Any CDF on $\mathbb{R}$ is the CDF of a real random variable.*

Proof. This results directly from the fact that, as explained in Remark II.50, on each probability space $(\mathbb{R},\mathfrak{B},P)$ can be defined a real random variable X with PDF $P_X = P \circ X$. □

Example III.37. If μ denotes the BOREL measure on $\mathbb{R}$, $\mathbb{Q}$ is a zero measure set in $(\mathbb{R},\mathfrak{B},\mu)$ since, for any bijection $\varphi \in \mathbb{Q}^{\mathbb{N}}$ (its existence is guaranteed by the denumerability of $\mathbb{Q}$),

$$\mu(\mathbb{Q}) = \mu(\varphi(\mathbb{N})) = \mu\left(\bigcup_{n\in\mathbb{N}} \{\varphi(n)\}\right) = \sum_{n\in\mathbb{N}} \mu(\{\varphi(n)\}) = 0.$$

III.4 Measure Completion

III.4.1 Complete Measures

Measure completion consists of enriching, at no cost, a σ-algebra with sets or set differences whose measure is equal to 0.

Definition III.38. Given a measure space, a *zero measure set* is a measurable set whose measure is equal to 0.

Example III.39. $\varnothing$ is a zero measure set of $(\mathbb{N}, 2^{\mathbb{N}}, \#)$. Moreover, it is the only zero measure set of this measure space.

It is clear that any measurable subset of a zero measure set is also a zero measure set. However, any subset of a zero measure set is not necessarily measurable, although this would be a practical situation. The following concept is therefore introduced:

Definition III.40. A measure space is said to be *complete* if every subset of a zero measure set is measurable.

Example III.41. $(\mathbb{N}, 2^{\mathbb{N}}, \#)$ is a complete measure space since its only zero measure set is $\varnothing$, which belongs to the σ-algebra.

Example III.42. The measure space $(S, \mathfrak{S}, \mu)$ defined in Example II.14 is not complete since $\mu(\{\square, \triangle\}) = 0$, but neither $\{\square\}$ nor $\{\triangle\}$ are measurable.

Remark III.43. Although the notion of completeness is defined for measure spaces, it is also accepted to use it for measures, when no confusion is possible: saying that μ is complete means that the underlying measure space $(S, \mathfrak{S}, \mu)$ is complete.

In the case where a measure space is not complete, a natural question arises: is it possible to complete[17] it? The following result addresses this question:

Proposition III.44. *If $(S, \mathfrak{S}, \mu)$ is a measure space and*

$$\mathfrak{S}^{\mu} = \left\{ E \subset S : \left(\exists (A, B) \in \mathfrak{S}^2\right) A \subset E \subset B,\ \mu(B \setminus A) = 0 \right\},$$

then $(S, \mathfrak{S}^{\mu})$ is a measurable space. In addition, under the above conditions regarding A, B and E, and defining the function $\widetilde{\mu}$ on $\mathfrak{S}^{\mu}$ by $\widetilde{\mu}(E) = \mu(A)$, $(S, \mathfrak{S}^{\mu}, \widetilde{\mu})$ is a complete measure space.

Proof. The first part of the proof concerns the measurability of $(S, \mathfrak{S}^{\mu})$:

i. by definition, $\mathfrak{S} \subset \mathfrak{S}^{\mu}$, thus, $S \in \mathfrak{S}^{\mu}$;

ii. on the one hand, if $A \subset E \subset B$, then $B^c \subset E^c \subset A^c$; on the other hand, $A^c \setminus B^c = B \setminus A$, thus, if $E \in \mathfrak{S}^{\mu}$, so does E^c;

iii. similarily, if for all n in $\mathbb{N}^*$ $A_n \subset E_n \subset B_n$, then

$$\bigcup_{n \in \mathbb{N}^*} A_n \subset \bigcup_{n \in \mathbb{N}^*} E_n \subset \bigcup_{n \in \mathbb{N}^*} B_n,$$

while

$$\bigcup_{n \in \mathbb{N}^*} B_n \setminus \bigcup_{n \in \mathbb{N}^*} A_n \subset \bigcup_{n \in \mathbb{N}^*} (B_n \setminus A_n).$$

Thus, if for all n in $\mathbb{N}^*$ $\mu(B_n \setminus A_n) = 0$, then

$$\mu\left(\bigcup_{n \in \mathbb{N}^*} B_n \setminus \bigcup_{n \in \mathbb{N}^*} A_n \right) = 0,$$

which proves that $\mathfrak{S}^{\mu}$ is closed for denumerable union.

[17] intuitively, as it is done for incomplete metric spaces.

Now, the second part of the proof consists of establishing that $\widetilde{\mu}$ is a measure on $(S,\mathfrak{S}^{\mu})$, based in particular on the fact that, by definition, $\widetilde{\mu}_{|\mathfrak{S}} = \mu$:

iv. first, it must be ensured that $\widetilde{\mu}$ actually is a function in $[0,\infty]^{\mathfrak{S}^{\mu}}$. Definition and target sets are obvious, but is must be checked that it is indeed a function, *i.e.*, each E in $\mathfrak{S}^{\mu}$ has one and only one image under $\widetilde{\mu}$. The existence is ensured by construction of $\mathfrak{S}^{\mu}$ and $\widetilde{\mu}$; concerning the uniqueness, assume that the sets A_1, B_1, A_2 and B_2 in $\mathfrak{S}$ are such that $A_1 \subset E \subset B_1$, $A_2 \subset E \subset B_2$, $\mu(B_1 \setminus A_1) = 0$ and $\mu(B_2 \setminus A_2) = 0$. In this case, $A_2 \setminus A_1 \subset B_1 \setminus A_1$, thus,

$$\mu(A_1) = \mu(A_2) + \mu(A_2 \setminus A_1) = \mu(A_2),$$

which proves the uniqueness of $\widetilde{\mu}(E) = \mu(A_1) = \mu(A_2)$;

v. $\varnothing \in \mathfrak{S}$ thus $\widetilde{\mu}(\varnothing) = \mu(\varnothing) = 0$;

vi. for any sequence $(E_n)_{n\in\mathbb{N}^*}$ of pairwise disjoined measurable (with respect to $(S,\mathfrak{S}^{\mu})$) sets, using the same notations as for (iii) above, the A_n are pairwise disjoinded, since for all n in $\mathbb{N}^*$ $A_n \subset E_n$. Hence,

$$\widetilde{\mu}\left(\bigcup_{n\in\mathbb{N}^*} E_n\right) = \mu\left(\bigcup_{n\in\mathbb{N}^*} A_n\right) = \sum_{n\in\mathbb{N}^*} \mu(A_n) = \sum_{n\in\mathbb{N}^*} \widetilde{\mu}(E_n),$$

since μ is a measure on $(S,\mathfrak{S})$.

The last part of the proof concerns the completeness of $(S,\mathfrak{S}^{\mu},\widetilde{\mu})$: given any zero measure set E in $\mathfrak{S}^{\mu}$, there exists, by definition, (A,B) in $\mathfrak{S}^2$ such that $A \subset E \subset B$ and $\mu(B \setminus A) = 0$. In particular, $\mu(A) = \widetilde{\mu}(A) = 0$, thus,

$$\mu(B) = \mu(A) + \mu(B \setminus A) = 0.$$

Now, any subset F of E clearly satisfies $\varnothing \subset F \subset B$ and $\widetilde{\mu}(B \setminus \varnothing) = 0$, thus, $F \in \mathfrak{S}^{\mu}$, and therefore $\widetilde{\mu}(F) = 0$. □

Definition III.45. Under the hypothesis of Proposition III.44, $(S,\mathfrak{S}^{\mu},\widetilde{\mu})$ is said to be the *completion of* $(S,\mathfrak{S},\mu)$.

EXERCISE III.46. Complete the measure space $(S,\mathfrak{S},\mu)$ defined in Example II.14.

Remark III.47. Caution must be used when handling the concept of measure space completion when S is also equipped with a metric: there might be an ambiguity with completion in the sense of convergence of CAUCHY[18] sequences.

III.4.2 BOREL-STIELTJES Measures

Theorem III.48. *Under the hypothesis of Theorem III.25,* $\left(S,\mathfrak{S}(\mu^*),\mu^*_{|\mathfrak{S}(\mu^*)}\right)$ *is a complete measure space.*

Proof. Exercise. □

Remark III.49. As a consequence of the restriction of μ' from $\mathfrak{S}(\mu^*)$ to $\mathfrak{S}(\mathcal{S})$ in CARATHÉODORY's Extension, $(S,\mathfrak{S}(\mathcal{S}),\mu')$ is not necessarily complete; for instance, this is the case for BOREL measure spaces.

[18] Augustin CAUCHY, French mathematician, 1789–1857

Definition III.50. Under the hypothesis of Corollary III.31, the complete measure $\widetilde{\mu}$ defined on the σ-algebra $\mathfrak{B}^{\mu}$ is called the *STIELTJES*[19]*-LEBESGUE measure on* $\mathbb{R}$ *induced by* F. In particular, if μ is the BOREL measure, then $\mathfrak{B}^{\mu}$ (resp. $\ell = \widetilde{\mu}$) is called the *LEBESGUE* σ*-algebra* (resp. *LEBESGUE measure*) *on* $\mathbb{R}$ and denoted $\mathfrak{L}$ (resp. ℓ).

Remark III.51. For the sake of simplicity, the STIELTJES-LEBESGUE measure induced by a right-continuous nondecreasing function $F \in \mathbb{R}^{\mathbb{R}}$ is often also denoted F. Provided basic caution, this notation is unambiguous while being very intuitive.

[19]Thomas STIELTJES, Dutch astronomer and mathematician, 1856–1894.

III The LEBESGUE Measure

IV L^p Spaces

This new integral of Lebesque is proving itself a wonderful tool. I might compare it with a modern Krupp gun, so easily does it penetrate barriers which were impregnable.

E.H. VAN VLECK[20], in *Bulletin of the American Mathematical Society*, **23** (1916)

In this Chapter, $(S, \mathfrak{S}, \mu)$ will denote a measure space.

IV.1 Integrals of Positive Functions

IV.1.1 From Measures to Integrals

In order to avoid handling special cases at ∞ in the definitions and results presented below, the usual arithmetic of $[0,\infty[$ is extended to $[0,\infty]$ by the means of the following conventions:

i. for all a in $[0,\infty]$, $a+\infty=\infty+a=\infty$;

ii. for all a in $[0,\infty]^*$, $a\times\infty=\infty\times a=\infty$;

iii. $0\times\infty=0$.

Remark IV.1. Note that the arithmetic of limits is consistent with the two first conventions, but not with the third one since $0\times\infty$, in the case of limits, is not necessarily 0.

Definition IV.2. Let u be a measurable simple function in $[0,\infty[^S$, thus, with the notations of Proposition II.44, u can be written as follows:

$$u=\sum_{i=1}^{n} a_i \mathbb{1}_{A_i}.$$

The *integral of* u is defined as:

$$\int u\,\mathrm{d}\mu=\sum_{i=1}^{n} a_i\mu(A_i).$$

Proposition IV.3. *Under the previous assumptions, the integral of u is independent of the representation chosen for u: it is a unique value in* $[0,\infty]$.

Proof. Exercise. □

Definition IV.4. Let f be a measurable function in $[0,\infty]^S$. The *integral of* f is defined as:

$$\int f\,\mathrm{d}\mu=\sup\left\{\int u\,\mathrm{d}\mu, u\in[0,\infty[^S, u \text{ measurable simple}, u\leqslant f\right\}.$$

[20]Edward B. VAN VLECK, American mathematician, 1863–1943.

Remark IV.5. In the case where μ is the STIELTJES-LEBESGUE measure ℓ, $\mathrm{d}\ell$ is often denoted by $\mathrm{d}x$, because ℓ is induced by $x \mapsto x$. Although not rigorously correct, this notation is acceptable when there is no risk of confusion. An advantage of this tolerance is that it permits retrieval of the classic RIEMANN-integral notation since $\int_{\mathbb{R}} f(x)\,\mathrm{d}x$ is equal $\int f\,\mathrm{d}\ell$ for any f that is RIEMANN-integrable over $\mathbb{R}$.

Definition IV.6. If f is a measurable function in $[0,\infty]^S$ and E belongs to $\mathfrak{S}$, define

$$\int_E f\,\mathrm{d}\mu = \int f\mathbb{1}_E\,\mathrm{d}\mu.$$

Remark IV.7. The fact that $f\mathbb{1}_E$ is measurable is guaranteed by Corollary II.39.

Proposition IV.8.

i. *If f and g are two measurable functions in $[0,\infty]^S$ such that $f \leqslant g$, then*

$$\int f\,\mathrm{d}\mu \leqslant \int g\,\mathrm{d}\mu.$$

ii. *If f is a measurable function in $[0,\infty]^S$ and c belongs to $[0,\infty[$, then*

$$\int cf\,\mathrm{d}\mu = c\int f\,\mathrm{d}\mu.$$

Proof.

i. For any measurable simple function u in $[0,\infty[^S$, $u \leqslant f$ implies $u \leqslant g$. Thus,

$$\begin{aligned}\left\{\int u\,\mathrm{d}\mu, u \in [0,\infty[^S, u \text{ measurable simple}, u \leqslant f\right\} \\ \subset \quad \left\{\int u\,\mathrm{d}\mu, u \in [0,\infty[^S, u \text{ measurable simple}, u \leqslant g\right\},\end{aligned}$$

whence the conclusion arises immediately.

ii. Clearly, u is simple if and only if cu is simple, thus

$$\begin{aligned}\int cf\,\mathrm{d}\mu &= \sup\left\{\int u\,\mathrm{d}\mu, u \in [0,\infty[^S, u \text{ measurable simple}, u \leqslant cf\right\} \\ &= \sup\left\{\int cu\,\mathrm{d}\mu, u \in [0,\infty[^S, u \text{ measurable simple}, u \leqslant f\right\} \\ &= c\sup\left\{\int u\,\mathrm{d}\mu, u \in [0,\infty[^S, u \text{ measurable simple}, u \leqslant f\right\} \\ &= c\int f\,\mathrm{d}\mu.\end{aligned}$$

□

Proposition IV.9. *If f is a measurable function in $[0,\infty]^S$ and E belongs to $\mathfrak{S}$, then*

$$\mu(E)=0 \quad \Longrightarrow \quad \int_E f\,\mathrm{d}\mu = 0.$$

Proof. By definition,

$$\int_E f\,\mathrm{d}\mu = \int f\mathbb{1}_E\,\mathrm{d}\mu = \sup\left\{\int u\,\mathrm{d}\mu, u\in[0,\infty[^S, u \text{ measurable simple}, u\leqslant f\mathbb{1}_E\right\}.$$

With the notations of Proposition II.44, any measurable simple function u in $[0,\infty[^S$ can be expressed as:

$$u=\sum_{i=1}^n a_i\mathbb{1}_{A_i}$$

where n is in $\mathbb{N}^*$ and the A_i are pairwise disjoined measurable sets. Now, if $u\leqslant f\mathbb{1}_E$, it is clear that $u=u\mathbb{1}_E$ (since u must be zero on E^c). Therefore,

$$u=u\mathbb{1}_E=\sum_{i=1}^n a_i\mathbb{1}_{A_i}\mathbb{1}_E=\sum_{i=1}^n a_i\mathbb{1}_{A_i\cap E},$$

thus,

$$\int u\,\mathrm{d}\mu=\sum_{i=1}^n a_i\mu(A_i\cap E)$$

but, combining the non-negativity of measures and Proposition II.15,

$$(\forall 1\leqslant i\leqslant n)\quad 0\leqslant \mu(A_i\cap E)\leqslant \mu(E)=0,$$

thus,

$$\int u\,\mathrm{d}\mu=\sum_{i=1}^n a_i\mu(A_i\cap E)=0.$$

Finally,

$$\sup\left\{\int u\,\mathrm{d}\mu, u\in[0,\infty[^S, u \text{ measurable simple}, u\leqslant f\mathbb{1}_E\right\}=\sup\{0\}=0.$$

□

Example IV.10. If f is a function in $[0,\infty]^{\mathbb{R}}$, then, using the fact that $\mathbb{Q}$ is a zero BOREL measure set as explained in Example III.37, it follows that $\int_{\mathbb{Q}} f\,\mathrm{d}\ell=0$. This illustrates one of the advantages of the Lebesgue integral over the RIEMANN one, since the latter does not permit to integrate over $\mathbb{Q}$.

IV.1.2 Monotonic Convergence Theorem

Lemma IV.11. *Let u be a measurable simple function in $[0,\infty[^S$. The mapping ν defined on $\mathfrak{S}$ by*

$$(\forall E\in\mathfrak{S})\quad \nu(E)=\int_E u\,\mathrm{d}\mu$$

is a measure on $(S,\mathfrak{S})$.

Proof. With the notations of Proposition II.15, u can be written as

$$u = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i},$$

where n is in $\mathbb{N}^*$ and the A_i are pairwise disjoined measurable sets. Therefore,

i. $\mu(\varnothing) = 0$ since, for all i, $a_i \mathbb{1}_{A_i} \mathbb{1}_{\varnothing} = 0$.

ii. For any sequence $(E_k)_{k\in\mathbb{N}^*}$ of pairwise disjoined measurable sets, their union being denoted E,

$$\begin{aligned}
\nu\left(\bigcup_{k\in\mathbb{N}^*} E_k\right) &= \nu(E) = \int u\mathbb{1}_E \,\mathrm{d}\mu = \int \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \mathbb{1}_E \,\mathrm{d}\mu \\
&= \sum_{i=1}^{n} a_i \mu(A_i \cap E) = \sum_{i=1}^{n} \sum_{k\in\mathbb{N}^*} a_i \mu(A_i \cap E_k) \\
&= \sum_{k\in\mathbb{N}^*} \sum_{i=1}^{n} a_i \mu(A_i \cap E_k) = \sum_{k\in\mathbb{N}^*} \int u\mathbb{1}_{E_k} \,\mathrm{d}\mu \\
&= \sum_{k\in\mathbb{N}^*} \nu(E_k).
\end{aligned}$$

□

Theorem IV.12 (Monotonic Convergence or Beppo LEVI's[21] Theorem). *If $(f_n)_{n\in\mathbb{N}^*}$ is a sequence of measurable function in $[0,\infty]^S$ such that*

$$\begin{cases} (\forall n \in \mathbb{N}^*) & f_n \leqslant f_{n+1} \\ (\forall x \in S) & \lim_n f_n(x) = f(x), \end{cases}$$

then f is a measurable function, and

$$\int f \,\mathrm{d}\mu = \lim_n \int f_n \,\mathrm{d}\mu.$$

Proof. The measurability of f results directly from Corollary II.41. Now, the sequence $(\int f_n \,\mathrm{d}\mu)_{n\in\mathbb{N}}$ monotonically increases, thus, it has a limit $L \in [0,\infty]$. Since $f_n \leqslant f$ for all n in $\mathbb{N}^*$, then $\int f_n \,\mathrm{d}\mu \leqslant \int f \,\mathrm{d}\mu$, and therefore $L \leqslant \int f \,\mathrm{d}\mu$. Let u be a simple function in $[0,\infty[^S$ such that $u \leqslant f$ (this is always possible, according to Theorem II.46), and for any given c in $]0,1[$, define

$$(\forall n \in \mathbb{N}^*) \quad E_n = \{x \in S : f_n(x) \geqslant c u_n(x)\}.$$

For all n in $\mathbb{N}^*$, E_n is measurable, $E_n \subset E_{n+1}$ and $E_n \subset S$. Moreover, for all x in S, $f_n(x) \to u(x) > cu(x)$, thus there exists $n \in \mathbb{N}^*$ such that $x \in E_n$; therefore,

$$\bigcup_{n\in\mathbb{N}^*} E_n = S.$$

Now,

$$(\forall n \in \mathbb{N}^*) \quad \int f_n \,\mathrm{d}\mu \geqslant \int_{E_n} f_n \,\mathrm{d}\mu \geqslant c \int_{E_n} u \,\mathrm{d}\mu$$

[21] Beppo LEVI, Italian mathematician, 1876–1961

hence, according to Lemma IV.11 and Proposition II.19,

$$L = \lim_n \int f_n \, \mathrm{d}\mu \geqslant c \lim_n \int_{E_n} u \, \mathrm{d}\mu = c \lim_n \int_{\bigcup_{n \in \mathbb{N}^*} E_n} u \, \mathrm{d}\mu = c \int u \, \mathrm{d}\mu$$

and this is true for all c in $]0,1[$ and all simple functions u in $[0,\infty[^S$ such that $u \leqslant f$, thus

$$L \geqslant \sup_{0<c<1} c \int u \, \mathrm{d}\mu = \int u \, \mathrm{d}\mu$$

and

$$L \geqslant \sup \left\{ \int u \, \mathrm{d}\mu, u \in [0,\infty[^S, u \text{ measurable simple}, u \leqslant f \right\} = \int f \, \mathrm{d}\mu$$

whence, finally, $L = \int f \, \mathrm{d}\mu$. □

Corollary IV.13 (FATOU's[22] Lemma). *If $(f_n)_{n \in \mathbb{N}^*}$ is a sequence of measurable function in $[0,\infty]^S$, then*

$$\int \liminf_n f_n \, \mathrm{d}\mu \leqslant \liminf_n \int f_n \, \mathrm{d}\mu.$$

Proof. For all k in $\mathbb{N}$, define $g_k = \operatorname{Inf}_{n>k} f_n$ and $a_k = \operatorname{Inf}_{n>k} \int f_n \, \mathrm{d}\mu$. By definition, the sequences $(g_k)_{k \in \mathbb{N}}$ and $(a_k)_{k \in \mathbb{N}}$ monotonically increase, thus tends to, respectively, $\liminf_n f_n$ and $\liminf_n \int f_n \, \mathrm{d}\mu$ (each of these limits can be infinite). Now, for all k in $\mathbb{N}$ and all $n > k$, $g_k \leqslant f_n$ thus

$$(\forall k \in \mathbb{N})(\forall n > k) \quad \int g_k \, \mathrm{d}\mu \leqslant \int f_n \, \mathrm{d}\mu$$

hence

$$(\forall k \in \mathbb{N}) \quad \int g_k \, \mathrm{d}\mu \leqslant \operatorname*{Inf}_{n>k} \int f_n \, \mathrm{d}\mu = a_k.$$

Now, according to Proposition II.40, the g_k are measurable functions, thus applying Theorem IV.12 to $(g_k)_{k \in \mathbb{N}}$ show that $\liminf_n f_n$ is measurable and that, in addition,

$$\int \liminf_n f_n \, \mathrm{d}\mu = \lim_k \int g_k \, \mathrm{d}\mu \leqslant \lim_k a_k = \liminf_n \int f_n \, \mathrm{d}\mu.$$

□

IV.1.3 Additivity

Theorem IV.14. *If f and g are two measurable functions in $[0,\infty]^S$, then*

$$\int (f+g) \, \mathrm{d}\mu = \int f \, \mathrm{d}\mu + \int g \, \mathrm{d}\mu.$$

Proof. This result is established, first, for simple functions, then generalized to all measurable functions.

[22] Pierre FATOU, French mathematician and astronomer, 1878–1929.

Let f and g be two measurable simple functions in $[0,\infty[^S$, *i.e.*, according to Proposition II.44, there exists If f is a simple function in $\mathbb{R}^S$, then there exists (n,p) in $\mathbb{N}^{*2}$, $(a_i)_{1\leqslant i\leqslant n}$ and $(b_j)_{1\leqslant j\leqslant p}$ in $[0,\infty[$, n pairwise disjoined measurable sets $(A_i)_{1\leqslant i\leqslant n}$ and p pairwise disjoined measurable sets $(b_i)_{1\leqslant i\leqslant p}$ such that:

$$\begin{cases} f &= \sum_{i=1}^n a_i \mathbb{1}_{A_i} \\ g &= \sum_{j=1}^p b_j \mathbb{1}_{B_j} \\ S &= \bigcup_{i=1}^n A_i = \bigcup_{j=1}^p B_j \end{cases}$$

from which it follows directly that $f+g$ takes at most np non-negative values (among all $a_i+b_j \geqslant 0$), thus is a simple function in $[0,\infty[$. In addition, Corollary II.39 shows that $f+g$ is measurable, hence, according to Lemma IV.11, the mapping defined over $\mathfrak{S}$ by $E \mapsto \int_E (f+g)\,\mathrm{d}\mu$ is a measure on $(S,\mathfrak{S})$. Now,

$$\bigcup_{i=1}^n \bigcup_{j=1}^p (A_i \cap B_j) = \bigcup_{i=1}^n A_i \cap \bigcup_{j=1}^p B_j = S$$

thus, since the $A_i \cap B_j$ are pairwise disjoined,

$$\begin{aligned} \int (f+g)\,\mathrm{d}\mu &= \int_{\bigcup_{i=1}^n \cup_{j=1}^p (A_i\cap B_j)} (f+g)\,\mathrm{d}\mu \\ &= \sum_{i=1}^n \sum_{j=1}^p \int_{A_i\cap B_j} (f+g)\,\mathrm{d}\mu \\ &= \sum_{i=1}^n \sum_{j=1}^p (a_i+b_j)\mu(A_i\cap B_j) \\ &= \sum_{i=1}^n \sum_{j=1}^p a_i\mu(A_i\cap B_j) + \sum_{i=1}^n \sum_{j=1}^p b_j\mu(A_i\cap B_j) \\ &= \sum_{i=1}^n \sum_{j=1}^p \int_{A_i\cap B_j} f\,\mathrm{d}\mu + \sum_{i=1}^n \sum_{j=1}^p \int_{A_i\cap B_j} g\,\mathrm{d}\mu \\ &= \int_{\bigcup_{i=1}^n \cup_{j=1}^p (A_i\cap B_j)} f\,\mathrm{d}\mu + \int_{\bigcup_{i=1}^n \cup_{j=1}^p (A_i\cap B_j)} g\,\mathrm{d}\mu \\ &= \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu. \end{aligned}$$

Now, if f and g are any two measurable functions in $[0,\infty]^S$, Theorem II.46 shows that there exists two sequences $(u_n)_{n\in\mathbb{N}^*}$ and $(v_n)_{n\in\mathbb{N}^*}$ of simple measurable functions in $[0,\infty[^S$, such that:

$$\begin{cases} (\forall n \in \mathbb{N}^*) & u_n \leqslant u_{n+1} \\ (\forall x \in S) & \lim_n u_n(x) = f(x) \\ (\forall n \in \mathbb{N}^*) & v_n \leqslant v_{n+1} \\ (\forall x \in S) & \lim_n v_n(x) = g(x) \end{cases}$$

which implies that

$$\begin{cases} (\forall n \in \mathbb{N}^*) & u_n + v_n \leqslant u_{n+1} + v_{n+1} \\ (\forall x \in S) & \lim_n u_n(x) + v_n(x) = f(x) + g(x) \end{cases}$$

thus, according Theorem IV.12 applied to the sequence$(u_n+v_n)_{n\in\mathbb{N}^*}$ of measurable functions in $[0,\infty[^S$, $f+g$ is measurable and

$$\int (f+g)\,\mathrm{d}\mu = \lim_n \int (u_n+v_n)\,\mathrm{d}\mu$$

but, since the additivity of $\int \mathrm{d}\mu$ is already proven in the case of simple measurable functions, it finally follows that

$$\int (f+g)\,\mathrm{d}\mu = \lim_n \int u_n\,\mathrm{d}\mu + \lim_n \int v_n\,\mathrm{d}\mu = \int f\,\mathrm{d}\mu + \int g\,\mathrm{d}\mu.$$

□

Theorem IV.15. *If $(f_n)_{n\in\mathbb{N}^*}$ is a sequence of measurable functions in $[0,\infty]^S$ and f is defined by*

$$(\forall x \in S) \quad f(x) = \sum_{n\in\mathbb{N}^*} f_n(x)$$

then f is measurable on $[0,\infty]^S$ and

$$\int f\,\mathrm{d}\mu = \sum_{n\in\mathbb{N}^*} \int f_n\,\mathrm{d}\mu.$$

Proof. For any k in $\mathbb{N}^*$, define

$$g_k = \sum_{n=1}^{k} f_n.$$

By definition, $(g_k)_{k\in\mathbb{N}^*}$ monotonically increases and, for all x in S, $\lim_k g_k(x) = f(x)$ while, for all k in $\mathbb{N}^*$, $g_k(x) \leqslant f(x)$. In addition, it follows from Theorem IV.14 that

$$(\forall k \in \mathbb{N}^*) \quad \int g_k\,\mathrm{d}\mu = \sum_{n=1}^{k} \int f_n\,\mathrm{d}\mu$$

thus, it follows from Theorem IV.12 that

$$\int f\,\mathrm{d}\mu = \lim_k \int g_k\,\mathrm{d}\mu = \sum_{n\in\mathbb{N}^*} \int f_n\,\mathrm{d}\mu.$$

□

Remark IV.16. The series $\sum_n f_n(x)$ might well be divergent, in this case $f(x) = \infty$ since all $f_n(x)$ are non-negative.

Example IV.17. In the case where the measure space is $(\mathbb{N}, 2^{\mathbb{N}}, \#)$, Theorem IV.15 shows directly that, if $a_{ij} \geqslant 0$ for all i and j in $\mathbb{N}^*$, then

$$\sum_{i\in\mathbb{N}^*} \sum_{j\in\mathbb{N}^*} a_{ij} = \sum_{j\in\mathbb{N}^*} \sum_{i\in\mathbb{N}^*} a_{ij}$$

in a much more straightforward way than the classical Theory of Series does. This result is no longer valid if the non-negativity does not hold.

IV.1.4 Density

Theorem IV.18. *Let f be a measurable function in $[0,\infty]^S$ and define the mapping ν over $\mathfrak{S}$ by*

$$(\forall E \in \mathfrak{S}) \quad \nu(E) = \int_E f\,\mathrm{d}\mu.$$

i. $(S, \mathfrak{S}, \nu)$ *is a measure space.*

ii. For any measurable function g in $[0,\infty]^S$,

$$\int g\,\mathrm{d}\nu = \int gf\,\mathrm{d}\mu. \tag{IV.1}$$

Proof.

i. It is clear that ν is non-negative and that $\nu(\varnothing) = 0$. Now, for any sequence $(A_n)_{n\in\mathbb{N}^*}$ of pairwise disjoined sets in $\mathfrak{S}^{\mathbb{N}}$ with denumerable union denoted A, one has

$$f\mathbb{1}_A = \sum_{n\in\mathbb{N}^*} f\mathbb{1}_{A_n}$$

thus, since all $f\mathbb{1}_{A_n}$ are measurable, it follows from Theorem IV.15 that $f\mathbb{1}_A$ is also measurable, with

$$\int f\mathbb{1}_A\,\mathrm{d}\mu = \sum_{n\in\mathbb{N}^*} \int f\mathbb{1}_{A_n}\,\mathrm{d}\mu,$$

or, equivalently,

$$\nu(A) = \sum_{n\in\mathbb{N}^*} \nu(A_n).$$

ii. If g is simple then, according to Proposition II.44, there exists n in $\mathbb{N}^*$ real numbers $(a_i)_{1\leqslant i\leqslant n}$ and n pairwise disjoined measurable sets $(A_i)_{1\leqslant i\leqslant n}$ such that:

$$\begin{cases} g &= \sum_{i=1}^n a_i\mathbb{1}_{A_i} \\ S &= \bigcup_{i=1}^n A_i \end{cases}$$

thus, using the definition of ν combined with Theorem IV.14,

$$\int g\,\mathrm{d}\nu = \sum_{i=1}^n a_i\nu(A_i) = \sum_{i=1}^n a_i \int_{A_i} f\,\mathrm{d}\mu = \int \sum_{i=1}^n a_i f\mathbb{1}_{A_i}\,\mathrm{d}\mu = \int gf\,\mathrm{d}\mu.$$

Now, if g is any measurable function in $[0,\infty]^S$, Theorem II.46 shows that there exists a sequence $(u_n)_{n\in\mathbb{N}^*}$ of simple measurable functions in $[0,\infty[^S$, such that:

$$\begin{cases} (\forall n\in\mathbb{N}^*) & u_n \leqslant u_{n+1} \\ (\forall x\in S) & \lim_n u_n(x) = g(x) \end{cases}$$

thus, in particular,

$$\begin{cases} (\forall n\in\mathbb{N}^*) & u_n f \leqslant u_{n+1} f \\ (\forall x\in S) & \lim_n u_n(x) f(x) = g(x) f(x) \end{cases}$$

whence, according to Theorem IV.12,

$$\int g\,\mathrm{d}\nu = \lim_n \int u_n\,\mathrm{d}\nu = \lim_n \int u_n f\,\mathrm{d}\mu = \int gf\,\mathrm{d}\mu$$

□

Definition IV.19. Under the assumptions of Theorem IV.18, it is said that the measure ν has the *density f with respect to μ*, written $\mathrm{d}\nu = f\mathrm{d}\mu$ or, if no confusion is possible, $\nu = f\mu$.

Remark IV.20. No specific meaning is granted to d in $\mathrm{d}\nu$ and $\mathrm{d}\mu$; it only means that (IV.1) is true for any measurable function g in $[0,\infty]$.

Proposition IV.21. *If $F \in \mathscr{C}^1(\mathbb{R})$ is non-decreasing, then $\mathrm{d}F = F'\mathrm{d}\ell$.*

Proof. Exercise. □

IV.2 Probability Density Functions

In this Section, $(T,\mathfrak{T},\mu)$ will denote a measure space, where $(T,\mathfrak{T})$ can be either a finite set along with its power set or $(\mathbb{R}^d,\mathfrak{B}^d)$.

Definition IV.22. Let $(S,\mathfrak{S},P)$ be a probability space, and X a random variable in T^S. If f is a measurable function in $[0,\infty[^T$ such that $\mathrm{d}P_X = f\mathrm{d}\mu$, then f is said to be the *probability density function*, or *PDF*, *of X with respect to* μ.

Remark IV.23. This denomination is consistent with the notion of density in the more general case of measure spaces and measurable functions. When no confusion is possible, the mention to μ can be omitted.

Example IV.24. Let X be a discrete random variable with two possible values, 1 and 0, with respective probabilities $p \in [0,1]$ and $1-p$. X can, *e.g.*, model the outcome of a coin toss. Hence, $(1-p)\mathbb{1}_{\{0\}} + p\mathbb{1}_{\{1\}}$ is the PDF of X, since:

$$\begin{cases} \int (1-p)\mathbb{1}_{\{0\}} + p\mathbb{1}_{\{1\}}\,\mathrm{d}\# = (1-p)\#\{0\} + p\#\{1\} = 1 = P_X(\{0,1\}) \\ \int_{\{0\}} (1-p)\mathbb{1}_{\{0\}} + p\mathbb{1}_{\{1\}}\,\mathrm{d}\# = (1-p)\#\{0\} = 1-p = P_X(0) \\ \int_{\{1\}} (1-p)\mathbb{1}_{\{0\}} + p\mathbb{1}_{\{1\}}\,\mathrm{d}\# = p\#\{1\} = p = P_X(1) \end{cases}$$

X is said to have the *BERNOULLI*[23] *distribution with parameter p*, denoted $\mathrm{B}(p)$.

Definition IV.25. Let $(S,\mathfrak{S},P)$ be a probability space, X a random variable in T^S with PDF f, and φ in $\mathbb{R}^T$ a measurable function. When the following integral exists and is finite:

$$\int \varphi f\,\mathrm{d}\mu = \int \varphi(X)\,\mathrm{d}P$$

it is called the *moment of* $\varphi(X)$, denoted $\mathbb{E}[\varphi(X)]$. For n in $\mathbb{N}^*$ and when they exist, $\mathbb{E}[X^n]$, $\mathbb{E}[|X|^n]$, $\mathbb{E}[(X-\mathbb{E}[X])^n]$ and $\mathbb{E}[|X-\mathbb{E}[X]|^n]$ are respectively called *n^{th} moment*, *n^{th} absolute moment*, *n^{th} centered moment* and *n^{th} absolute centered moment of X*. In particular, the first moment and the second absolute centered moment are respectively called *expectation* and *variance of X*. The latter is also denoted $\mathbb{V}[X]$.

Remark IV.26. According to Theorem IV.14, when $\mathbb{V}[X]$ exists it is equal to $\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$.

Example IV.27. If the discrete random variable X has:

- the BERNOULLI distribution $\mathrm{B}(p)$, then

$$\mathbb{E}[X] = \int \mathrm{Id} f\,\mathrm{d}\# = 0(1-p)\#\{0\} + 1p\#\{1\} = p$$
$$\mathbb{V}[X] = \int |(1-p)\mathrm{Id}|^2 f\,\mathrm{d}\# = p^2(1-p)\#\{0\} + (p-1)^2 p\#\{1\} = p(1-p);$$

- the PDF $n \mapsto \frac{6}{\pi^2(n^2+1)}$ with respect to the counting measure on $(\mathbb{N},2^{\mathbb{N}})$, then it has no expectation, since $\sum_{n\in\mathbb{N}} \frac{n}{n^2+1} = \infty$.

[23] Jacob BERNOULLI, Swiss mathematician, 1654–1705

Proposition IV.28 (MARKOV Inequality). *If X is a non-negative real random variable with probability distribution P_X and expectation $\mathbb{E}[X]$, then*

$$(\forall \varepsilon \in [0,\infty[^*) \quad P_X([\varepsilon,\infty[) \leqslant \frac{\mathbb{E}[X]}{\varepsilon}.$$

Proof. For all ε in $[0,\infty[^*$,

$$\mathbb{E}[X] = \int \mathrm{Id}\,\mathrm{d}P_X \geqslant \int_{[\varepsilon,\infty[} \mathrm{Id}\,\mathrm{d}P_X \geqslant \int_{[\varepsilon,\infty[} \varepsilon\,\mathrm{d}P_X = \varepsilon P_X([\varepsilon,\infty[).$$

□

Corollary IV.29. *If X is a real random variable with probability distribution P_X and expectation $\mathbb{E}[X]$, then for all n in $\mathbb{N}^*$ such that the nth absolute centered moment of X exists,*

$$(\forall \varepsilon \in [0,\infty[^*) \quad P_{|X-\mathbb{E}[X]|}([\varepsilon,\infty[) \leqslant \frac{\mathbb{E}[|X-\mathbb{E}[X]|^n]}{\varepsilon^n}.$$

Proof. It suffices to apply Proposition IV.28 to $|X-\mathbb{E}[X]|^n$. □

Remark IV.30. In the case $n=2$, Corollary IV.29 becomes the BIENAYMÉ[24]-ČEBIČEV[25] Inequality.

IV.3 Integrability

IV.3.1 Integrable Functions

Definition IV.31. For any given $f \in [-\infty,\infty]^S$, define the following functions of $[0,\infty]^S$:

$$\begin{aligned} f^+ &= \max(f,0) \\ f^- &= \max(-f,0). \end{aligned}$$

Remark IV.32. Clearly, for all $f \in [-\infty,\infty]^S$, $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

Definition IV.33. A function $f \in [-\infty,\infty]^S$ is said to be *μ-integrable* if it is measurable and if

$$\int |f|\,\mathrm{d}\mu < \infty.$$

In this case, the *μ-integral of f* is defined as follows:

$$\int f\,\mathrm{d}\mu = \int f^+\,\mathrm{d}\mu - \int f^-\,\mathrm{d}\mu.$$

Remark IV.34. If μ is a LEBESGUE measure, then the expressions *LEBESGUE-integrable* and *LEBESGUE-integral* can also be used. Another advantage of the LEBESGUE Integral appears here: Definition IV.33 does not make sense in the context of the RIEMANN Integral; it suffices *e.g.*, to consider the function $f = \mathbb{1}_{[0,1]\cap\mathbb{Q}} - \mathbb{1}_{[0,1]\setminus\mathbb{Q}}$, which is LEBESGUE-integrable with $\int |f|\,\mathrm{d}\ell = \int f\,\mathrm{d}\ell = 0$. Now, in the context of the RIEMANN Integral, $\int |f(x)|\,\mathrm{d}x = 1$ but $\int f(x)\,\mathrm{d}x$ does not exist.

[24] Irénée-Jules BIENAYMÉ, French civil servant and mathematician, 1796–1878.
[25] Pafnuti ČEBIČEV, Russian mathematician, 1821–1894.

Example IV.35. Let consider the function $\mathbb{1}_{[0,\pi]}\cos$, which belongs to $\mathbb{R}^{\mathbb{R}}$. One has:

$$\int |f|\,\mathrm{d}\ell = \int \mathbb{1}_{[0,\pi]}|\cos|\,\mathrm{d}\ell \leqslant \int \mathbb{1}_{[0,\pi]}\,\mathrm{d}\ell = \pi < \infty$$

and f is LEBESGUE-integrable. Now, $f^+ = \mathbb{1}_{[0,\frac{\pi}{2}]}\cos$ and $f^- = -\mathbb{1}_{[-\frac{\pi}{2},0]}\cos$, thus

$$\int f\,\mathrm{d}\ell = \int \mathbb{1}_{[0,\frac{\pi}{2}]}\cos\,\mathrm{d}\ell - \int -\mathbb{1}_{[-\frac{\pi}{2},0]}\cos\,\mathrm{d}\ell = \int \mathbb{1}_{[0,\frac{\pi}{2}]}\cos\,\mathrm{d}\ell - \int \mathbb{1}_{[0,\frac{\pi}{2}]}\cos\,\mathrm{d}\ell = 0$$

because of the symmetry of cos with respect to $\frac{\pi}{2}$.

The definition is now slightly modified for complex-valued functions; the two cases are not comprised in a single definition, since the complex case does not allow for infinite values (since no such topology has been defined):

Definition IV.36. A function $f \in \mathbb{C}^S$ is said to be *μ-integrable* if it is measurable and if

$$\int |f|\,\mathrm{d}\mu < \infty.$$

In this case, the *μ-integral of f* is defined as follows:

$$\int f\,\mathrm{d}\mu = \int \Re(f)\,\mathrm{d}\mu + \mathrm{i}\int \Im(f)\,\mathrm{d}\mu.$$

Example IV.37. If $f \in \mathbb{C}^{\mathbb{R}}$ is defined by $f(t) = e^{\mathrm{i}t}\mathbb{1}_{[0,\pi]}(t)$ for all $t \in \mathbb{R}$, then

$$\int |f|\,\mathrm{d}\ell = \int \mathbb{1}_{[0,\pi]}\,\mathrm{d}\ell = \pi,$$

therefore f is LEBESGUE-integrable. Moreover, $\Re(f) = \mathbb{1}_{[0,\pi]}\cos$, $\Im(f) = \mathbb{1}_{[0,\pi]}\sin$, thus, using the results of Example IV.34,

$$\int f\,\mathrm{d}\ell = \int \mathbb{1}_{[0,\pi]}\cos\,\mathrm{d}\ell + \mathrm{i}\int \mathbb{1}_{[0,\pi]}\sin\,\mathrm{d}\ell = \mathrm{i}\int \sin\mathbb{1}_{[0,\pi]}\,\mathrm{d}\ell = \mathrm{i}\int_0^{\pi}\sin x\,\mathrm{d}x = 2\mathrm{i}$$

since sin is continuous over $[0,\pi]$ thus RIEMANN-integrable.

Proposition IV.38. *If f is μ-integrable, then*

$$\left|\int f\,\mathrm{d}\mu\right| \leqslant \int |f|\,\mathrm{d}\mu.$$

Proof. Let $z = \int f\,\mathrm{d}\mu$; z is a complex number and thus there exists $\alpha \in \mathbb{C}$ such that $|\alpha| = 1$ and $|z| = \alpha z$. If $u = \Re(\alpha f)$, then $u \leqslant |f|$ and

$$\left|\int f\,\mathrm{d}\mu\right| = \alpha\int f\,\mathrm{d}\mu = \int \alpha f\,\mathrm{d}\mu = \int u\,\mathrm{d}\mu \leqslant \int |f|\,\mathrm{d}\mu.$$

□

IV.3.2 Almost Everywhere

By definition of a measure, a zero measure set is not necessarily empty; moreover, as explained in Subsection III.4.1, if the measure space is complete, all subsets of zero measure sets are, themselves, zero measure sets. This apparently benevolent aspect will actually prove to be one of the most convenient features of the LEBESGUE integral. First, it allows to define a more general form of equality, as follows:

Definition IV.39. A property $\mathcal{P}(x)$ defined for $x \in S$ is said to be true *μ-almost everywhere*, abbreviated $\mu-$a.e., if there exists a zero measure set A such that $\mathcal{P}(x)$ is true for all $x \in A^c$. In the context of a probability measure, the expression *almost surely* is generally used, instead of almost everywhere.

Remark IV.40. It is important to understand why the definition of truth almost everywhere is given in such an (apparently) obfuscated way: the set of all x for which the property is true (or false) is not necessarily measurable. An alternate way of understanding the definition is to consider the following equivalence:

$$A^c \subset \{x \in S : \mathcal{P}(x)\} \iff \{x \in S : \neg\mathcal{P}(x)\} \subset A$$

and to see truth almost everywhere as the fact that the set of all x for which the property is false is a subset of a zero measure space. Obviously, if $(S, \mathfrak{S}, \mu)$ is complete, this difficulty vanishes.

Example IV.41. If f and g are two functions defined over S,

$$f \overset{\text{a.e.}}{=} g \iff (\exists A \in \mathfrak{S}, \mu(A) = 0)(\forall x \in A^c) \quad f(x) = g(x)$$

and $\overset{\text{a.e.}}{=}$ is clearly an equivalence relation. In the context of probability measures, $\overset{\text{a.s.}}{=}$

Example IV.42. If $(f_n)_{n \in \mathbb{N}^*}$ is a sequence of functions defined over S,

$$f_n \xrightarrow{\text{a.e.}} f \iff (\exists A \in \mathfrak{S}, \mu(A) = 0)(\forall x \in A^c) \quad \lim_n f(x) = \lim_n g(x).$$

It is important to acknowledge that $\xrightarrow{\text{a.e.}}$ is a pointwise convergence (*i.e.*, in the target set, and not in the function space).

Remark IV.43. In fact, the principle of truth almost everywhere turns out to be one of the most fruitful of Measure Theory, since it allows to consider functions which are defined "only" almost everywhere. If f is such a function, *i.e.*, f is defined in the complement of a zero measure set A, then define (over the entire S) the function $\widehat{f} = f\mathbb{1}_{A^c}$; by definition, f shares the same properties (measurability, integrability and integral) as those of $\widehat{f}$. This convention is clearly consistent with everything done before, and extends the scope of measurable functions.

From now on, everything will have to be understood in this context.

Proposition IV.44. *Let f be a non-negative measurable function; then,*

$$\int f \, \mathrm{d}\mu = 0 \iff f \overset{\text{a.e.}}{=} 0.$$

Proof. Let $f \in [0, \infty]^S$ be measurable. First, if $f \overset{\text{a.e.}}{=} 0$,

$$(\exists A \in \mathfrak{S}, \mu(A) = 0)(\forall x \in A^c) \quad f(x) = 0$$

thus, by additivity and because of Proposition IV.9,

$$\int f \, \mathrm{d}\mu = \int f\mathbb{1}_{A^c} \, \mathrm{d}\mu + \int f\mathbb{1}_A \, \mathrm{d}\mu = \int f\mathbb{1}_{A^c} \, \mathrm{d}\mu = \int 0 \, \mathrm{d}\mu = 0.$$

Conversely, assume $\int f \, d\mu = 0$ and define

$$(\forall n \in \mathbb{N}^*) \quad A_n = \left\{ x \in S : f(x) \geqslant \frac{1}{n} \right\}.$$

Now, for all $n \in \mathbb{N}^*$, $\frac{1}{n} \mathbb{1}_{A_n} \leqslant f$ thus

$$(\forall n \in \mathbb{N}^*) \quad \mu(A_n) \leqslant n \int f \, d\mu = 0$$

and, because of Proposition II.19,

$$(\forall n \in \mathbb{N}^*) \quad \mu\left(\{x \in S : f(x) \neq 0\}\right) = \mu\left(\bigcup_{n \in \mathbb{N}^*} A_n \right) = \lim_n \mu(A_n) = 0.$$

□

The following theorem is, in many respects, one of the most powerful results of Measure Theory:

Theorem IV.45 (LEBESGUE Dominated Convergence Theorem). *Let $(f_n)_{n \in \mathbb{N}^*}$ a sequence of μ-integrable functions such that $f_n \xrightarrow{\text{a.e.}} f$ and:*

$$\left(\exists g \in [0, \infty]^S, \int |g| \, d\mu < \infty \right) (\forall n \in \mathbb{N}^*) \quad |f_n| \overset{\text{a.e.}}{\leqslant} g,$$

then f is μ-integrable a.e. and

i. $\lim_n \int |f_n - f| \, d\mu = 0$;

ii. $\lim_n \int f_n \, d\mu = \int f \, d\mu$.

Proof. By hypothesis, there exists a zero measure set A such that, for all $x \in A^c$, $\lim_n f_n(x) = f(x)$. Using the same conventions as in Remark IV.43, $\widehat{f}$, is the pointwise limit of the sequence of the μ-integrable, thus measurable, $\widehat{f}_n$ functions; hence, $\widehat{f}$ is measurable and, since $|\widehat{f}| \leqslant g$, $\widehat{f}$ is μ-integrable. For the sake of clarity, and because this does not change anything to the result, f is now used, instead of $\widehat{f}$. Applying FATOU Lemma (Corollary IV.13) to the sequence of functions defined by $h_n = 2g - |f_n - f|$ leads to

$$\int 2g \, d\mu \leqslant \int \liminf_n 2g - |f_n - f| \, d\mu = 2g - \int \liminf_n |f_n - f| \, d\mu$$

i.e.,

$$\int \liminf_n |f_n - f| \, d\mu \leqslant 0$$

whence

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

Finally,

$$\left| \int f_n \, d\mu - \int f \, d\mu \right| \leqslant \int |f_n - f| \, d\mu$$

from where the second equality arises. □

Corollary IV.46. *If $(f_n)_{n\in\mathbb{N}^*}$ is a sequence of measurable functions of $[-\infty,\infty]^S$, then*

$$\sum_{n\in\mathbb{N}^*}\int |f_n|\,\mathrm{d}\mu = \int \sum_{n\in\mathbb{N}^*}|f_n|\,\mathrm{d}\mu$$

and, if this number is finite, then each $|f_n|$ is μ-integrable, the series $\sum_{n\in\mathbb{N}^*}f_n$ pointwise converges a.e. and

$$\sum_{n\in\mathbb{N}^*}\int f_n\,\mathrm{d}\mu = \int \sum_{n\in\mathbb{N}^*}f_n\,\mathrm{d}\mu$$

Proof. Exercise. □

IV.3.3 $\mathscr{L}^p$ Spaces

First, a new kind of conjugation is introduced:

Definition IV.47. Any two positive reals p and q are said to be *conjugate exponents* if $\frac{1}{p}+\frac{1}{q}=1$. As a convention, 1 and ∞ are considered to pertain to this case.

Remark IV.48. Clearly, if p and q are conjugate exponents, then both of them are in $[1,\infty]$. Moreover, given $p\in[1,\infty]$, there is only one p', such that p and p' are conjugate exponents. If $p\in]1,\infty[$, $p'=\frac{p}{p-1}$.

Example IV.49. An important case is that of $p=p'=2$.

The following preliminary lemma will be very helpful:

Lemma IV.50. *Given any $(\alpha,\beta)\in[0,\infty[^2$ such that $\alpha+\beta=1$ then,*

$$(\forall(u,v)\in[0,\infty]^2)\quad u^\alpha v^\beta \leqslant \alpha u+\beta v$$

Proof. If u or v is 0 or ∞, the result is trivial. Otherwise, assuming $s=\ln u$ and $t=\ln v$, the convexity of exp ensures that

$$u^\alpha v^\beta = e^{\alpha s+\beta t} \leqslant \alpha e^s+\beta e^t = \alpha u+\beta v.$$

□

Theorem IV.51 (HÖLDER[26] Inequality for Integrals). *Given any p in $]1,\infty[$ then, for all measurable functions f and g of $[0,\infty]^S$,*

$$\int fg\,\mathrm{d}\mu \leqslant \left(\int f^p\,\mathrm{d}\mu\right)^{1/p}\left(\int g^{p'}\,\mathrm{d}\mu\right)^{1/p'}$$

Proof. For the sake of clarity, define $a=(\int f^p\,\mathrm{d}\mu)^{\frac{1}{p}}$ and $b=(\int g^{p'}\,\mathrm{d}\mu)^{\frac{1}{p'}}$. If $a=0$ or $b=0$, then $f\overset{\text{a.e.}}{=}0$ or $g\overset{\text{a.e.}}{=}0$ thus $\int fg\,\mathrm{d}\mu=0$, and if $a=\infty$ or $b=\infty$ the inequality is obvious. Therefore, assume that both a and b are in $]0,\infty[$ and define $F=\frac{f}{a}$ and $G=\frac{g}{b}$. According to Lemma IV.50, one has, $\mu-$a.e.,

$$F(x)G(x) \leqslant \frac{1}{p}F(x)^p+\frac{1}{q}G(x)^q$$

from which the result follows by integration. □

[26] Otto HÖLDER, German mathematician, 1859–1937.

Corollary IV.52 (CAUCHY-SCHWARZ[27] Inequality for Integrals). *For all measurable functions f and g of $[0,\infty]^S$,*

$$\int fg\,\mathrm{d}\mu \leqslant \left(\int f^2\,\mathrm{d}\mu\right)^{1/2}\left(\int g^2\,\mathrm{d}\mu\right)^{1/2}$$

Proof. Immediate, with $p = p' = 2$. □

Theorem IV.53 (MINKOWSKI[28] Inequality for Integrals). *Given any p in $]1,\infty[$ then, for all measurable functions f and g of $[0,\infty]^S$,*

$$\left(\int (f+g)^p\,\mathrm{d}\mu\right)^{1/p} \leqslant \left(\int f^p\,\mathrm{d}\mu\right)^{1/p} + \left(\int g^p\,\mathrm{d}\mu\right)^{1/p}.$$

Proof. Exercise: use the convexity of $t \mapsto t^p$ and apply HÖLDER Inequality to $f(f+g)^{p-1}$. □

Definition IV.54. Given any $p \in [1,\infty[$, the set of all measurable functions $f \in \mathbb{C}^S$ such that $|f|^p$ is μ-integrable is denoted $\mathscr{L}^p(S,\mathfrak{S},\mu)$, on which the function $\|\cdot\|_p$ is defined as follows:

$$(\forall f \in \mathscr{L}^p(S,\mathfrak{S},\mu)) \quad \|f\|_p = \left(\int |f|^p\,\mathrm{d}\mu\right)^{1/p}.$$

Definition IV.55. A measurable function $f \in \mathbb{C}^S$ such that

$$(\exists M \in [0,\infty[) \quad \mu\left(\{x \in S : |f(x)| > M\}\right) = 0$$

is said to be *essentially bounded* and the set of such functions is denoted $\ell^\infty(S,\mathfrak{S},\mu)$, on which the function $\|\cdot\|_\infty$ is defined as follows:

$$(\forall f \in \mathscr{L}^p(S,\mathfrak{S},\mu)) \quad \|f\|_\infty = \inf\left\{M \in [0,\infty[: \mu\left(\{x \in S : |f(x)| > M\}\right) = 0\right\}.$$

Remark IV.56. The nomenclature hars several exceptions; in particular:

- if no confusion is possible, $\mathscr{L}^p(S,\mathfrak{S},\mu)$ is often simply denoted $\mathscr{L}^p(\mu)$ or even $\mathscr{L}^p$;
- for any interval $I \subset \mathbb{R}$, although this is not consistent with the previous point, $\mathscr{L}^p(I,\mathfrak{L}(I),\ell_I)$ is generally denoted $\mathscr{L}^p(I)$ in the literature and, in particular, $\mathscr{L}^p(\mathbb{R})$ stands for $\mathscr{L}^p(\mathbb{R},\mathfrak{L},\ell)$.

At the expense of a new definition, the essential properties of the $\mathscr{L}^p$ spaces can be elegantly summarized.

Definition IV.57. A function $\|\cdot\|$ defined on a $\mathbb{K}$-vector field E is said to be a *seminorm on E* if:

i. $(\forall f \in E) \quad \|f\| \geq 0$;

ii. $(\forall (\alpha,f) \in \mathbb{K}\times E) \quad \|\alpha f\| = |\alpha|\,\|f\|$;

iii. $(\forall (f,g) \in E^2) \quad \|f+g\| \leq \|f\| + \|g\|$.

[27] Hermann SCHWARZ, German mathematician, 1843–1921.
[28] Hermann MINKOWSKI, Russian then German mathematician, 1864–1909.

Theorem IV.58. *For all $p \in [1,\infty]$,*

i. $\mathscr{L}^p(S,\mathfrak{S},\mu)$ *is a* $\mathbb{C}$*-vector field;*

ii. $\|\cdot\|_p$ *is a seminorm on* $\mathscr{L}^p(S,\mathfrak{S},\mu)$*;*

iii. for all $f \in \mathscr{L}^p(S,\mathfrak{S},\mu)$ *and* $g \in \mathscr{L}^{p'}(S,\mathfrak{S},\mu)$, $|\int fg\,\mathrm{d}\mu| \leqslant \|f\|_p \|g\|_{p'}$.

Proof. Directly from HÖLDER and MINKOWSKI Inequalities (Theorems IV.51 and IV.53). □

IV.4 From $\mathscr{L}^p$ to L^p

IV.4.1 Quotientization

First, it is useful to notice that, in the context of $\mathscr{L}^p$ spaces, the equivalence relation $\overset{\text{a.e.}}{=}$ can be reformulated in a more convenient way:

Proposition IV.59. *Given any $p \in [1,\infty]$, for all f and g in $\mathscr{L}^p(S,\mathfrak{S},\mu)$,*

$$f \overset{\text{a.e.}}{=} g \quad \Longleftrightarrow \quad \|f-g\|_p = 0.$$

Proof. Because the modulus is positive definite, $f(x) = g(x)$ if and only if $|f(x)-g(x)|^p = 0$. Therefore,

$$f \overset{\text{a.e.}}{=} g \quad \Longleftrightarrow \quad |f-g|^p \overset{\text{a.e.}}{=} 0 \quad \Longleftrightarrow \quad \int |f-g|^p \,\mathrm{d}\mu = 0$$

according to Proposition IV.44, and the result follows by taking p^{th} roots. □

Now, the following quotientization allows to get rid of "doublets", *i.e.*, functions which do not differ in terms of integration:

Definition IV.60. Given any $p \in [1,\infty]$, $L^p(S,\mathfrak{S},\mu)$ denotes the quotient space of $\mathscr{L}^p(S,\mathfrak{S},\mu)$ by the equivalence relation $\overset{\text{a.e.}}{=}$, or, in a more synthetic form,

$$L^p = \mathscr{L}^p / \overset{\text{a.e.}}{=}.$$

Remark IV.61. Lightweight notations similar to those described in Remark IV.56 are acceptable, when no confusion is possible, such as $L^p(\mu)$, $L^p(\mathbb{R})$ and $L^p(I)$.

Proposition IV.62. *For all $p \in [1,\infty]$, $(L^p(S,\mathfrak{S},\mu), \|\cdot\|_p)$ is a normed space.*

Proof. By construction and because of Theorem IV.58, $\|\cdot\|_p$ is a seminorm on $L^p(S,\mathfrak{S},\mu)$. In addition, if $\|f\|_p = 0$ then Proposition IV.44 shows that $|f|^p \overset{\text{a.e.}}{=} 0$ thus, clearly, $f \overset{\text{a.e.}}{=} 0$; in other words, $\|\cdot\|_p$ is definite and therefore is a norm. □

Proposition IV.63. *If I is an interval of $\mathbb{R}$ such that $\mu(I) < \infty$, then*

i. $(\forall p \in [1,\infty])\quad L^\infty(I) \subset L^p(I)$*;*

ii. $(\forall (p,q) \in [1,\infty]^2, p \leqslant q) \quad \mathrm{L}^q(I) \subset \mathrm{L}^p(I)$ *and the injection is continuous.*

Proof.

i. Given any $f \in \mathrm{L}^\infty(I)$ and any $p \in [1,\infty]$, one has

$$\int_I |f|^p \, \mathrm{d}\mu \leqslant \mu(I) \, \|f\|_\infty^p < \infty.$$

ii. Given any $(p,q) \in [1,\infty]^2$ such that $p < q$ and any $f \in \mathrm{L}^p(I)$, let denote $J = I \cap f^{<-1>}([1,\infty])$; by definition, over J $|f|^p \leqslant |f|^q$ thus

$$\int_I |f|^p \, \mathrm{d}\mu \leqslant \int_J |f|^q \, \mathrm{d}\mu + \int_{I \setminus J} |f|^p \, \mathrm{d}\mu \leqslant \int_I |f|^q \, \mathrm{d}\mu + \int_{I \setminus J} \mathrm{d}\mu \leqslant \|f\|_q^q + \mu(I) < \infty$$

thus $f \in \mathrm{L}^q(I)$. Concerning the continuity of the injection, let $h \in \mathrm{L}^q(I)$ and apply HÖLDER Inequality (Theorem IV.51) to the case where $f = |h|^p$, $g = 1$, $r = {}^q/_p$ and $s = \frac{r}{r-1}$:

$$\int_I |h|^p \, \mathrm{d}\mu \leqslant \left(\int_I |f|^r \, \mathrm{d}\mu \right)^{1/r} \left(\int_I \mathrm{d}\mu \right)^{1/s}$$

thus $\|h\|_p \leqslant C \, \|h\|_q$, where $C \in [0,\infty[$.

□

IV.4.2 Approximation in BANACH Spaces

First, a few reminders:

Definition IV.64. A seminorm $\|\cdot\|$ on a $\mathbb{K}$-vector field E is said to be a *norm on* E if:

$$(\forall f \in E) \quad \|f\| = 0 \implies f = 0_E.$$

In this case, the ordered pair $(E, \|\cdot\|)$ is said to be a *normed space*.

Remark IV.65. An equivalent definition is to say that a norm is *positive definite* seminorm. In fact, the first axiom of Definition IV.57 implies that the converse of (IV.64) holds.

Example IV.66. Defining the functions $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_\infty$ on $\mathbb{K}^d$ as follows:

$$\left(\forall (x_1, x_2, \ldots, x_n) \in \mathbb{K}^d\right) \quad \begin{cases} \|(x_1, x_2, \ldots, x_n)\|_1 & = & \displaystyle\sum_{i=1}^{i=d} |x_i| \\ \|(x_1, x_2, \ldots, x_n)\|_2 & = & \left(\displaystyle\sum_{i=1}^{i=d} |x_i|^2 \right)^{1/2} \\ \|(x_1, x_2, \ldots, x_n)\|_\infty & = & \displaystyle\max_{1 \leq i \leq d} |x_i|, \end{cases}$$

then $(\mathbb{K}^d, \|\cdot\|_1)$, $(\mathbb{K}^d, \|\cdot\|_2)$ and $(\mathbb{K}^d, \|\cdot\|_\infty)$ are normed spaces. $\|\cdot\|_2$ and $\|\cdot\|_\infty$ are often referred to as, respectively, the *euclidean norm* and the supremum *norm* on $\mathbb{K}^d$.

Any given normed space can be viewed as a metric space, by defining the distance between two vectors as the norm of their difference. Therefore, any given normed space can naturally be equipped with a topology, the one of its associated metric space. Except otherwise indicated, a normed space is always considered to be equipped with this topology. Now, the first question that arises when toplogizing a set is that of the completeness; it is not true in general that normed vector spaces are complete. Therefore, the following particular class of spaces is introduced:

Definition IV.67. A complete normed space is called a *BANACH*[29] *space.*

Example IV.68. All finite dimension normed spaces are BANACH spaces.

EXERCISE IV.69. Show that the converse statement is not true.

The following result is often seen as a generalization of the famous RIESZ[30]-FISCHER[31] Theorem for L^2 spaces:

Theorem IV.70. *For all $p \in [1,\infty]$, $(L^p(S,\mathfrak{S},\mu),\|\cdot\|_p)$ is a* BANACH *space.*

Proof. The case $p=\infty$ is easy to establish: given a CAUCHY sequence $(f_n)_{n\in\mathbb{N}^*}$ in $\ell^\infty(S,\mathfrak{S},\mu)$, it suffices to consider, for all $(n,m)\in\mathbb{N}^{*2}$,

$$A_{n,m}=\left\{x : |f_n(x)-f_m(x)| > \|f_n-f_m\|_\infty\right\}.$$

Using the fact $\mathscr{L}^p(S,\mathfrak{S},\mu)$ is a vector field, it follows that for all $(n,m)\in\mathbb{N}^{*2}$, $f_n-f_m\in\mathscr{L}^p(S,\mathfrak{S},\mu)$ thus, by definition of $\|\cdot\|_\infty$, $\mu(A_{n,m})=0$. Therefore, the denumerable union A of the $A_{n,m}$ is a zero measure set while

$$(\forall x\in A^c)\quad |f_n(x)-f_m(x)| \leqslant \|f_n-f_m\|_\infty$$

i.e., $(f_n)_{n\in\mathbb{N}^*}$ is uniformely CAUCHY over A^c thus converges uniformely towards a function f defined over A^c. Assuming, *e.g.*, $f=0$ over A, $(f_n)_{n\in\mathbb{N}^*}$ tends to f in $(L^p(S,\mathfrak{S},\mu),\|\cdot\|_\infty)$. □

Remark IV.71. To avoid obfuscation, the same notation f is used for the function of $\mathscr{L}^p(S,\mathfrak{S},\mu)$ and it class representative in $L^p(S,\mathfrak{S},\mu)$. In this sense f is uniquely defined in $L^p(S,\mathfrak{S},\mu)$ while, in $\mathscr{L}^p(S,\mathfrak{S},\mu)$, f is only uniquely defined on A^c.

Remark IV.72. When $\mu=\infty$, the results are no longer true; *i.e.*, $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ are not comparable.

Theorem IV.73. *If $I\subset\mathbb{R}$ is a bounded interval, then, for all $p\in[1,\infty[$, $\mathscr{C}^0_c(I)$ is dense in $L^p(I)$. In addition, $\mathscr{C}^0_0(I)$ is dense in $L^\infty(I)$.*

Proof. Exercise □

Remark IV.74. Focusing on the case where $I=\mathbb{R}$ and $p=1$ is particularly instructive; in this case, Theorem IV.73 can be interpreted as follows: if the "distance" between two compact support continuous functions f and g is henceforth defined as $\int_{\mathbb{R}}|f(t)-g(t)|\,\mathrm{d}t$, then the completion of the subsequent metric space is, exactly, $L^1(\mathbb{R})$ (through the previously mentioned canonical injection of $L^1(\mathbb{R})$ in $\mathscr{L}^1(\mathbb{R})$). This shows that the LEBESGUE Integral is, undoubtlessly, the "*good*" generalization of the RIEMANN Integral.

[29] Stefan BANACH, Polish mathematician, 1892–1945.
[30] Frigyes RIESZ, Hungarian mathematician, 1880-1956.
[31] Ernst FISCHER, German mathematician, 1875–1954.

V HILBERT Spaces

Une géométrie ne peut être plus vraie qu'une autre ;
elle peut seulement être plus commode.

H. POINCARÉ[32], [Poi02]

In this chapter, $\mathbb{K}$ will denote a field which can be either $\mathbb{R}$ or

$\mathbb{C}$. Due to the extent of the topic, this chapter only provides an overview of the essential definitions and properties; for a more in-depth understanding, [Yos95] is particularily recommended.

V.1 Pre-HILBERT Spaces

In this Section, E, F and G are $\mathbb{K}$-vector fields. The goal of this section is the construction of the pre-HILBERT[33] spaces by the formal definition of an inner product over a vector space. It is then explained how such an inner product can induce a norm and, therefore, a topology on the vector field.

V.1.1 Inner Products

Definition V.1. A function $f \in F^E$ is said to be *linear* if:

$$(\forall(\alpha,\beta) \in \mathbb{K}^2)\ (\forall(x,y) \in E^2) \quad f(\alpha x + \beta y) = \alpha f(x) + \beta f(y)$$

and *conjugate linear* if:

$$(\forall(\alpha,\beta) \in \mathbb{K}^2)\ (\forall(x,y) \in E^2) \quad f(\alpha x + \beta y) = \overline{\alpha} f(x) + \overline{\beta} f(y).$$

Remark V.2. If $\mathbb{K} = \mathbb{R}$, conjugate linearity and linearity coincide.

Example V.3. Given a and b distinct in $\mathbb{R}$, $N \in \mathscr{C}^0([a,b]^2)$, and $f \in \mathscr{C}^0([a,b])$, define

$$\begin{array}{rccl} S(f): & [a,b] & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & \int_a^b N(x,y) f(y)\,\mathrm{d}y. \end{array}$$

One can prove that $S(f) \in \mathscr{C}^0([a,b])$ thus that S is a function from $\mathscr{C}^0([a,b])$ onto itself. In addition, S is linear since for all α and β in $\mathbb{R}$ and all f and g in $\mathscr{C}^0([a,b])$,

$$\begin{aligned} S(\alpha f + \beta g) &= \int_a^b N(x,y)(\alpha f(y) + \beta g(y))\,\mathrm{d}y \\ &= \alpha \int_a^b N(x,y) f(y)\,\mathrm{d}y + \beta \int_a^b N(x,y) g(y)\,\mathrm{d}y \\ &= \alpha S(f) + \beta S(g). \end{aligned}$$

[32] Henri POINCARÉ, French mathematician, 1854–1912.
[33] David HILBERT, German mathematician, 1862–1943.

Example V.4. The operator $\frac{\mathrm{d}}{\mathrm{d}x}$ can be seen as a linear mapping from $\mathscr{C}^1(\mathbb{R})$ onto $\mathscr{C}^0(\mathbb{R})$ since

$$(\forall(\alpha,\beta)\in\mathbb{K}^2))(\forall(f,g)\in\mathscr{C}^1(\mathbb{R})^2)) \quad \frac{\mathrm{d}}{\mathrm{d}x}(\alpha f+\beta g)=\alpha\frac{\mathrm{d}}{\mathrm{d}x}f+\beta\frac{\mathrm{d}}{\mathrm{d}x}g.$$

Definition V.5. A function in $G^{E\times F}$ is said to be *bilinear* if it is linear with respect two both variables, and *sesquilinear* if it is linear (resp. conjugate linear) with respect to the first (resp. second) variable

Remark V.6. It is equivalent to say that $f\in G^{E\times F}$ is bilinear if:

$$\begin{cases} (\forall(x_1,x_2,y)\in E^2\times F) & f(x_1+x_2,y)=f(x_1,y)+f(x_2,y) \\ (\forall(x,y_1,y_2)\in E\times F^2) & f(x,y_1+y_2)=f(x,y_1)+f(x,y_2) \\ (\forall(x,y)\in E\times F)(\forall\alpha\in\mathbb{K}) & f(\alpha x,y)=\alpha f(x,y)=f(x,\alpha y) \end{cases}$$

and sesquilinear if:

$$\begin{cases} (\forall(x_1,x_2,y)\in E^2\times F) & f(x_1+x_2,y)=f(x_1,y)+f(x_2,y) \\ (\forall(x,y_1,y_2)\in E\times F^2) & f(x,y_1+y_2)=f(x,y_1)+f(x,y_2) \\ (\forall(x,y)\in E\times F)(\forall\alpha\in\mathbb{K}) & f(\alpha x,y)=\alpha f(x,y) \\ (\forall(x,y)\in E\times F)(\forall\alpha\in\mathbb{K}) & f(x,\alpha y)=\overline{\alpha} f(x,y) \end{cases}$$

In particular, if $\mathbb{K}=\mathbb{R}$ then sesquilinearity and bilinearity coincide.

Definition V.7. A linear (resp. conjugate linear, sesquilinear, bilinear) function into $\mathbb{K}$ is called a *linear* (resp. *conjugate linear, sesquilinear, bilinear) form over* E. The set of all continuous linear forms over E is called the *topological dual of* E and denoted E'.

Remark V.8. In this context, the target set $\mathbb{K}$ is seen as a vector space over itself: each element of $\mathbb{K}$ is at the same time a scalar and a vector. It is immediate that the dual is also a $\mathbb{K}$-vector field. When no confusion with the algebraic dual is possible, E' will simply be called the *dual of* E.

Example V.9. The trace is a linear form over $\mathfrak{M}_n(\mathbb{K})$, for all $n\in\mathbb{N}^*$.

Example V.10. The following mapping:

$$\begin{array}{rccc} \langle\cdot,\cdot\rangle: & E'\times E & \longrightarrow & \mathbb{K} \\ & (f,x) & \longmapsto & f(x) \end{array}$$

is a bilinear form called the *duality pairing over* E.

Definition V.11. A function in $\mathbb{K}^{E\times E}$ is said to be *symmetric* if

$$(\forall(x,y)\in E^2) \quad f(x,y)=f(y,x)$$

and *conjugate symmetric* if

$$(\forall(x,y)\in E^2) \quad f(x,y)=\overline{f(y,x)}. \tag{V.1}$$

Remark V.12. If $\mathbb{K}=\mathbb{R}$, then any conjugate symmetric mapping is symmetric, and conversely.

Definition V.13. A *hermitian form* is a conjugate symmetric sesquilinear form.

Remark V.14. In order to establish that a form is hermitian, it thus suffices to prove that it is left-linear and satisfies (V.1). If $\mathbb{K}=\mathbb{R}$, a hermitian form is a symmetrical bilinear form (also called *euclidean form*).

Definition V.15. A bilinear or sesquilinear form f over E^2 is said to be *degenerate* if

$$(\exists x \in E \setminus \{0_E\}) \quad f(x,x) = 0$$

and *non-degenerate* otherwise.

Definition V.16. A bilinear or sesquilinear form f over E^2 is said to be *positive definite* if

$$(\forall x \in E \setminus \{0_E\}) \quad f(x,x) > 0.$$

Remark V.17. A positive definite bilinear or sesquilinear form is necessarily non-degenerate. The converse is clearly untrue: *e.g.*, the bilinear form defined over $\mathbb{R}^2$ by $(x,y) \mapsto -xy$ is non-degenerate, but is not positive definite.

Definition V.18. An *inner product over* E, generally denoted $(\cdot|\cdot)$, is a positive definite hermitian form over E^2. In this case, the ordered pair $(E,(\cdot|\cdot))$ is said to be a *pre-HILBERT space.*

Example V.19. Given two distinct real numbers a and b, E denotes here the $\mathbb{K}$-vector field formed by $\mathscr{C}^0([a,b],\mathbb{K})$ along with its usual operators (sum of functions, multiplication of a function by a scalar). The mapping $(\cdot|\cdot)$ is defined over E^2 as follows:

$$(\forall (f,g) \in E^2) \quad (f|g) = \int_a^b f(x)\overline{g(x)}\,\mathrm{d}x.$$

Clearly, for all $(f,g,h) \in E^3$,

i. $(f|f) \in \mathbb{K}$;

ii. $(f|f) \geqslant 0$ and $(f|f) = 0$ if and only if $f = 0$;

iii. $(\forall(\alpha,\beta) \in \mathbb{K}^2) \quad ((\alpha f + \beta g)|h) = \alpha(f|h) + \beta(g|h)$;

iv. $(f|g) = \overline{(g|f)}$

which proves that $(\cdot|\cdot)$ is an inner product over E.

V.1.2 Norm induced by an Inner Product

Provided the following important results, it is possible to make a normed space from any given pre-HILBERT space, and therefore to derive an associated topolgy.

Proposition V.20 (CAUCHY-SCHWARZ Inequality for Inner Products). *If* $(E,(\cdot|\cdot))$ *is a pre-*HILBERT *space, then*

$$(\forall (x,y) \in E^2) \quad |(x|y)| \leqslant (x|x)^{\frac{1}{2}}(y|y)^{\frac{1}{2}}.$$

Proof. The following three cases are considered:

- if $(y|y) = 0$, *i.e.*, $y = 0_E$, then the inequality is trivially satisfied;

- if $(y|y) = 1$, then

$$\begin{aligned}(x-(x|y)\,y|x-(x|y)\,y) &= (x|x)-(x|y)\,\overline{(x|y)}-(x|y)\,\overline{(x|y)}+(x|y)\,\overline{(x|y)}\,(y|y)\\ &= (x|x)-|\,(x|y)\,|^2\end{aligned}$$

  whence, since $(x-(x|y)\,y|x-(x|y)\,y) \geqslant 0$, the inequality arises;

- if $(y|y) \neq 0$ then, assuming $z = \frac{y}{(y|y)^{\frac{1}{2}}}$, one retrieves the latter case for x et z since $(z|z) = 1$; hence,

$$\frac{|\,(x|y)\,|}{(y|y)^{\frac{1}{2}}} = |\,(x|z)\,| \leqslant (x|x)^{\frac{1}{2}}$$

  and the result follows.

□

Proposition V.21 (MINKOWSKI Inequality for Inner Products). *If $(E,(\cdot|\cdot))$ is a pre-HILBERT space, then*

$$(x+y|x+y)^{\frac{1}{2}} \leqslant (x|x)^{\frac{1}{2}} + (y|y)^{\frac{1}{2}}.$$

Proof. By definition of an inner product, for all $(x,y) \in E^2$,

$$\begin{aligned}(x+y|x+y) &= (x|x)+(x|y)+(y|x)+(y|y)\\ &= (x|x)+(y|y)+2\Re((x|y))\\ &\leqslant (x|x)+(y|y)+2|\,(x|y)\,|\\ &\leqslant (x|x)+(y|y)+2(x|x)^{\frac{1}{2}}(y|y)^{\frac{1}{2}}\end{aligned}$$

according to CAUCHY-SCHWARZ Inequality (Proposition V.20). Hence,

$$(x+y|x+y) \leqslant \left((x|x)^{\frac{1}{2}} + (y|y)^{\frac{1}{2}}\right)^2$$

and since both sides are non-negative, one can take their square roots while preserving the inequality. □

Provided this framework, the construction of a topology naturally associated to a pre-HILBERT space now arises naturally:

Proposition V.22. *If $(E,(\cdot|\cdot))$ is a pre-HILBERT space, then the following function:*

$$\begin{array}{rccc}\|\cdot\|: & E & \longrightarrow & \mathbb{K}\\ & x & \longmapsto & (x|x)^{\frac{1}{2}}\end{array}$$

is a norm on E.

Proof. It results immediately from the definition of an inner product that $(\cdot|\cdot)^{\frac{1}{2}}$ is positive definite. Positive homogeneity is trivial and the triangle inequality is just another way to formulate MINKOWSKI Inequality (Proposition V.21). □

Definition V.23. Under the hypothesis of Proposition V.22, $\|\cdot\|$ is called the *norm induced by* $(\cdot|\cdot)$.

Remark V.24. Except otherwise indicated, a pre-HILBERT space is always considered to be normed by its induced norm.

The existence of the induced norm allows to extend many topological concepts, such as that of continuity.

Proposition V.25. *If $(E,(\cdot|\cdot))$ is a finite-dimensional pre-*HILBERT *space, then its induced norm continuous. In addition, for any given $y \in E$, the mappings over E by $x \mapsto (x|y)$ and $x \mapsto (y|x)$ are also continuous.*

Proof. In fact, these mappings are even uniformly continuous; for $\|\cdot\|$, this directly arises from the Triangle Inequality. Concerning the two other functions, it suffices to apply CAUCHY-SCHWARZ Inequality. □

As it is possible to associate a norm to each inner product, one naturally wonders whether the converse is true. The following result provides a convenient and intuitive necessary condition:

Proposition V.26 (Parallelogram Law). *If $(E,(\cdot|\cdot))$ is a pre-*HILBERT *space, then the norm $\|\cdot\|$ induced by $(\cdot|\cdot)$ satisfies*

$$(\forall (x,y) \in E^2) \quad \|x+y\|^2 + \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right).$$

Proof. By definition of $\|\cdot\|$, for all $(x,y) \in E^2$ one has

$$\begin{aligned}\|x+y\|^2 + \|x-y\|^2 &= (x+y|x+y) + (x-y|x-y)\\ &= 2\left(\|x\|^2 + \|y\|^2\right) + (x|y) + (y|x) - (x|y) - (y|x)\end{aligned}$$

whence the conclusion. □

Example V.27. Consider, for $n \in \mathbb{N}^*$, the vectors x and y in $\mathbb{R}^n$ usual, such that all their components are 0, except $x_1 = y_2 = 1$. One has, on the one hand,

$$2\left(\|x\|_1^2 + \|y\|_1^2\right) = 2\left(\|x\|_2^2 + \|y\|_2^2\right) = 2\left(\|x\|_\infty^2 + \|y\|_\infty^2\right) = 2(1+1) = 4,$$

while, on the other hand,

$$\|x+y\|_1^2 + \|x-y\|_1^2 = (1+1)^2 + (1+1)^2 = 8,$$

$$\|x+y\|_2^2 + \|x-y\|_2^2 = 1^2 + 1^2 + 1^2 + 1^2 = 4$$

and

$$\|x+y\|_\infty^2 + \|x-y\|_\infty^2 = 1 + 1 = 2.$$

This suffices to conclude, because of Proposition V.26, that neither $\|\cdot\|_1$ nor $\|\cdot\|_\infty$ can be derived from an inner product on $\mathbb{R}^n$. To the contrary, the euclidean norm satisfies the Parallelogram Law in this particular case (in fact, it does in all cases, since it is indeed derived from the usual inner product on $\mathbb{R}^n$).

V.1.3 Orthogonality

The abstract definition of an inner product allows the generalization of some classical Euclidean Geometry concepts and results to less intuitive geometries such as those of functional spaces. In particular, orthogonality is extended as follows:

Definition V.28. Let E_1 and E_2 be two $\mathbb{K}$-vector fields, f be a bilinear form over $E_1 \times E_2$. If $(x,y) \in E_1 \times E_2$ is such that $f(x,y)=0$, then it is said that x and y are *orthogonal with respect to* f.

Remark V.29. When there is no risk of confusion, the reference to the bilinear form is omitted; in this case, two vectors x and y will simply be said to be orthogonal, denoted $x \perp y$. In particular, in the context of a pre-HILBERT space and except otherwise indicated, orthogonality will be understood with respect to its inner product.

Example V.30. Given any bilinear form over E^2, 0_E is always orthogonal to all others vectors of E, including itself; the latter property is generally formulated by saying that 0_E is *isotropic*.

Definition V.31. Let $(E,(\cdot|\cdot))$ be a pre-HILBERT space. For any $A \subset E$, the set

$$A^{\perp} = \{f \in E' : (\forall x \in A)\ \langle f,x\rangle = 0\}$$

is called the *orthogonal space of* A.

Proposition V.32. *Because of the left-linearity of* $\langle\cdot,\cdot\rangle$, $A^{\perp}$ *is a vector subspace of* E', *even if* A *is not a subspace.*

Proof. Immediate consequence of the left-linearity of $\langle\cdot,\cdot\rangle$. It shall be acknowledged that this property holds even if A is not a vector subspace of E. □

Remark V.33. In the case where E' is identified to E, and in particular in the finite-dimensional case, the identification of $\langle\cdot,\cdot\rangle$ to $(\cdot|\cdot)$ allows to view $A^{\cdot}$ as a subspace of E and, therefore, to retrieve the euclidean point of view of orthogonality between vectors.

Proposition V.34. *Let* $(E,(\cdot|\cdot))$ *be a pre-*HILBERT *space. For any* $n \in \mathbb{N}^*$ *and any family* $\{v_i\}_{1\leqslant i\leqslant n} \in E^n$ *of pairwise orthogonal vectors, then*

$$(\forall\{\lambda_i\}_{1\leqslant i\leqslant n} \in \mathbb{K}^n) \quad \left\|\sum_{i=1}^{i=n}\lambda_i v_i\right\|^2 = \sum_{i=1}^{i=n}|\lambda_i|^2\,\|v_i\|^2.$$

Proof. It suffices to develop the left-hand side as follows:

$$\left\|\sum_{i=1}^{i=n}\lambda_i v_i\right\|^2 = \left(\sum_{i=1}^{i=n}\lambda_i v_i \middle| \sum_{j=1}^{j=n}\lambda_j v_j\right) = \sum_{i=1}^{i=n}\lambda_i\overline{\lambda_i}\,(v_i|v_i) + \sum_{1\leqslant i\neq j\leqslant n}\lambda_i\overline{\lambda_j}\,(v_i|v_j)$$

and the result arises because of the pairwise orthogonality of the v_i. □

Corollary V.35 (PYTHAGORAS[34] Theorem). *Let* $(E,(\cdot|\cdot))$ *be a pre-*HILBERT *space. The following statement holds:*

$$(\forall(x,y) \in E^2) \quad x \perp y \quad \Longrightarrow \quad \|x+y\|^2 = \|x\|^2 + \|y\|^2.$$

[34] PYTHAGORAS OF SAMOS, Greek philosopher, mathematician and astronomer, *ca.*569BC–*ca.*475BC

Proof. Immediate from Proposition V.34, with $n = 2$ and $\lambda_1 = \lambda_2 = 1$. □

Remark V.36. In the case where $\mathbb{K} = \mathbb{R}$, the converse of PYTHAGORAS Theorem is true, since the proof of MINKOWSKI Inequality (Proposition V.21) then implies $(x|y) = 0$. However, this is no longer true if $\mathbb{K} = \mathbb{C}$, precisely because, in this case, $\Re((x|y)) = 0$ does not imply $(x|y)) = 0$; a simple counter-example can be exhibited with $E = \mathbb{C}^2$ equipped with its usual $\mathbb{K}$-vector field structure, over which the inner product is defined by

$$(\forall (x_1,x_2,y_1,y_2) \in \mathbb{C}^4) \quad ((x_1,x_2)|(y_1,y_2)) = x_1\overline{y_1} + x_2\overline{y_2}$$

and taking $x = (1,\mathrm{i})$, $y = (-\mathrm{i},1)$.

V.2 Completeness-Based Projectors

By adding a topological requirement, namely completeness, it will now be seen that many important results classicaly viewed as pertaining to Analysis can be interpreted or established from a geometric standpoint.

V.2.1 Convex Minimization

Definition V.37. A pre-HILBERT space, complete with respect to its induced norm, is said to be a ***HILBERT** space.*

Example V.38. $\mathrm{L}^2(\mathbb{R},\mathfrak{L},\ell)$, equipped with the usual L^2 inner product defined as follows:

$$(\forall (f,g) \in \mathrm{L}^2(\mathbb{R},\mathfrak{L},\ell)^2) \quad (f|g)_{\mathrm{L}^2} = \int f\overline{g}\,\mathrm{d}\ell$$

is a HILBERT space.

Remark V.39. When there is no risk of confusion, the identification between a HILBERT space $(H,(\cdot|\cdot))$ and H will be made, in order to simplify notations. This is consistent with similar mathematical conventions aiming at concision; *e.g.*, a vector field is a triple $(E,+,\cdot)$, but the usage is to refer to it as, simply, E.

Example V.40. For all $n \in \mathbb{N}^*$, $\mathbb{C}^n$ equipped with $(\cdot|\cdot)$ defined as follows:

$$(\forall x = (x_i)_{1\leqslant i\leqslant n} \in \mathbb{C}^n)\,(\forall y = (y_i)_{1\leqslant i\leqslant n} \in \mathbb{C}^n) \quad (x|y) = \sum_{i=1}^{i=n} x_i\overline{y_i}$$

is a HILBERT space, by transport of the completeness of $\mathbb{C}$.

Example V.41. The set $\ell^2(\mathbb{C})$ equipped with the inner product defined as follows:

$$\left(\forall u = (u_n)_{n\in\mathbb{N}^*} \in \mathbb{C}^{\mathbb{N}^*}\right)\left(\forall v = (v_n)_{n\in\mathbb{N}^*} \in \mathbb{C}^{\mathbb{N}^*}\right) \quad (u|v) = \sum_{n\in\mathbb{N}^*} u_n\overline{v_n}$$

is a HILBERT space.

Remark V.42. By definition, a HILBERT space is also a BANACH space. The converse is not true, as exhibited by Example V.27.

It is now possible to establish one of the most important Optimization results:

Theorem V.43 (F. RIESZ Theorem). *Let $(E,(\cdot|\cdot))$ be a pre-*HILBERT *space and $\|\cdot\|$ its induced norm. If $A \neq \varnothing$ is a complete convex subset of E, then $\min_A \|\cdot\|$ exists and is reached only once.*

Proof. The norm $\|\cdot\|$ satisfies the Parallelogram Law (Proposition V.26), which can be equivalently written as follows:

$$(\forall (x,y) \in E^2) \quad \|x-y\|^2 = 2\left(\|x\|^2 + \|y\|^2\right) - 4\left\|\frac{x+y}{2}\right\|^2$$

thus, defining $\delta = \inf_{x\in A} \|x\|$, it follows that

$$(\forall (x,y) \in A^2) \quad \|x-y\|^2 \leqslant 2\left(\|x\|^2 + \|y\|^2\right) - 4\delta^2 \tag{V.2}$$

since the convexity of A ensures that $\frac{x+y}{2} \in A$. The proof is now split into existence and uniqueness aspects:
EXISTENCE: Let $(x_n)_{n\in\mathbb{N}^*} \in A^{\mathbb{N}^*}$ be such that $\lim_n \|x_n\| = \delta$, which is always possible since the sequence of norms is a real sequence and $\mathbb{R}$ is complete. According to (V.2), one has

$$(\forall (n,m) \in \mathbb{N}^{*2}) \quad \|x_n - x_m\|^2 \leqslant 2\left(\|x_n\|^2 - \delta^2 + \|x_m\|^2 - \delta^2\right)$$

thus $(x_n)_{n\in\mathbb{N}^*}$ is a CAUCHY sequence in A. By hypothesis, A is complete with respect to $\|\cdot\|$ and therefore this sequence converges towards a limit in A; necessarily, $\lim_n x_n = \delta$.
UNIQUENESS: If δ is reached in x and y, then it follows directly from (V.2) that $\|x-y\|^2 = 0$ thus $x = y$, since a norm is positive definite. □

In the context of HILBERT spaces, the theorem can be expressed in a somewhat lighter formulation:

Corollary V.44. *If $A \neq \varnothing$ is a convex subset of a* HILBERT *space with induced norm $\|\cdot\|$, then $\min_A \|\cdot\|$ exists and is reached only once.*

Proof. Trivial: if H is a HILBERT space, then it is complete and therefore A, being a closed subset of H, is also complete and the conditions of Theorem V.43 are satisfied. □

V.2.2 Best Approximation Projector

An important consequence of F.RIESZ Theorem concerns the best approximation among a complete set:

Corollary V.45 (Best Approximation Theorem). *Let $(E,(\cdot|\cdot))$ be a pre-*HILBERT *space and $A \neq \varnothing$ be a complete convex subset of E. For all $x \in E$, the two following properties are equivalent:*

i. $P_A x \in A$ is such that $\|x - P_A x\| = \min_{y\in A} \|x-y\|$;

ii. $P_A x \in A$ is such that, for all $y \in A$, $\Re((x - P_A x | y - P_A x)) \leqslant 0$.

In addition, for all $x \in E$, there exists one and only one $P_A x \in A$ which satisfies these properties.

Proof. Exercise. □

Remark V.46. As for Theorem V.43, if E is a HILBERT space, it suffices to know that $A \neq \varnothing$ is a closed convex.

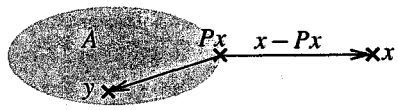

Figure 5. Best approximation on a complete convex subset.

The Best Approximation Theorem is illustrated by Figure 5, and formalized by a new definition:

Definition V.47. Under the assumptions of Theorem V.43, the mapping P from E to A defined by $P_A : x \mapsto P_A x$ is called the *best approximation projector from E onto A.*

And, in particular,

Proposition V.48. *The best approximation projector from a pre-*HILBERT *space onto any of its convex complete non-empty subsets is an idempotent contraction.*

Proof. Given any best approximation projector P from a pre-HILBERT $(E,(\cdot|\cdot))$ onto a convex complete non-empty $A \subset E$, the idempotence is immediate since $P(E) \subset A$ and, for all $y \in A$, $P(y) = y$. In addition, according to Theorem V.43,

$$(\forall (x,y) \in E^2) \quad \begin{cases} \Re((x - P_A x | P_A y - P_A x)) \leqslant 0 \\ \Re((y - P_A y | P_A x - P_A y)) \leqslant 0 \end{cases}$$

thus, by right-bilinearity,

$$(\forall (x,y) \in E^2) \quad \Re((x - y - (P_A x - P_A y) | P_A x - P_A y)) \geqslant 0$$

whence

$$(\forall (x,y) \in E^2) \quad \Re((x - y | P_A x - P_A y)) \geqslant \Re(\|P_A x - P_A y\|^2) = \|P_A x - P_A y\|^2 ;$$

now, according to CAUCHY-SCHWARZ Inequality (Proposition V.20),

$$(\forall (x,y) \in E^2) \quad \Re((x - y | P_A x - P_A y)) \leqslant \|x - y\| \, \|P_A x - P_A y\| ;$$

therefore,

$$(\forall (x,y) \in E^2) \quad \|P_A x - P_A y\|^2 \leqslant \|x - y\| \, \|P_A x - P_A y\|$$

which proves the contractivity of P_A. □

V.2.3 Orthogonal Projectors

An other important consequence of the Best Approximation Theorem arises when the projection is made onto some particular vector subspace; the vocabulary is first specialized, in order to retrieve and extend a classical geometric concept:

Definition V.49. If $(E,(\cdot|\cdot))$ is a pre-HILBERT space and $A \neq \varnothing$ is a complete vector subspace of E, then the best approximation projector of E onto A is called the *orthogonal projector of E* onto A.

Remark V.50. As before, it shall be acknowledged that this definition holds in particular when E is a HILBERT space of which A is a closed vector subspace.

This extension of a well-known concept of Euclidean Geometry makes plenty of sense when understanding it in the following context:

Corollary V.51. *If $(E,(\cdot|\cdot))$ is a pre-*HILBERT *space then, for all complete closed vector subspaces $A \neq \varnothing$ of E, $P_A \in L(E,A)$ and:*

$$(\forall x \in E) \quad \Big(y = P_A x \iff (\forall z \in A)\ (x-y|z) = 0\Big)$$

Proof. First, the fact that, for all $x \in E$, $P_A x$ exists, is unique and belongs to A is guaranteed by Corollary V.45, since a vector subspace is, in particular, convex. This result also shows that

$$(\forall x \in E) \quad \|x - P_A x\| = \min_{y \in A} \|x - y\|$$

thus, for all $(x,z) \in E \times A$, the mapping

$$\begin{array}{rccc} f: & \mathbb{K} & \longrightarrow & [0,\infty[\\ & \lambda & \longmapsto & \|x - P_A x - \lambda z\|^2 \end{array}$$

is minimal for $\lambda = 0$, since $P_A x - \lambda z \in A$. Now,

$$\begin{aligned} (\forall (x,z,\lambda) \in E \times A \times \mathbb{C}) \quad f(\lambda) &= (x - P_A x - \lambda z | x - P_A x - \lambda z) \\ &= |\lambda|^2 \|z\|^2 - 2\Re((x - P_A x|z)\lambda) + \|x - P_A x\|^2 \end{aligned}$$

thus f is a polynomial function and, because it reaches a *minimum* in 0, its derivative there vanishes, implying that $\Re((x - P_A x|z)) = 0$. If $\mathbb{K} = \mathbb{R}$, then it immediately follows that

$$(\forall (x,z) \in E \times A) \quad (x - P_A x|z) = 0. \tag{V.3}$$

If $\mathbb{K} = \mathbb{C}$, then applying the same reasoning to $\mathrm{i}z$ which also belongs to F leads to $\Im((x - P_A x|z)) = 0$ and therefore to (V.3). Conversely, given $x \in E$, assume that $y \in A$ is such that, for all $z \in A$, $(x - y|z) = 0$. It then follows immediately from PYTHAGORAS Theorem that

$$(\forall z \in A) \quad \|x - y - z\|^2 = \|x - y\|^2 + \|z\|^2 \geqslant \|x - y\|^2$$

and, therefore, $\|x - y\| = \min_{u \in A} \|x - u\|$, *i.e.*, $y = P_A x$ according to Corollary V.45. Concerning the linearity of P_A, it then suffices to acknowledge that, on the one hand,

$$(\forall (x,y,z) \in E^2 \times A) \quad \begin{cases} (x - P_A x|z) = 0 \\ (y - P_A y|z) = 0 \end{cases}$$

hence, by left-linearity,

$$(\forall (x,y,z) \in E^2 \times A) \quad (x + y - P_A x - P_A y|z) = 0$$

thus, for all $(x,y) \in E^2$, $P_A x + P_A y = P_A(x+y)$ because of the equivalence established above. On the other hand,

$$(\forall (\lambda,x,y) \in \mathbb{K} \times E \times A) \quad \lambda(x - P_A x|y) = (\lambda x - \lambda P_A x|y) = 0$$

thus, for all $(\lambda,x,y) \in \mathbb{K} \times E$, $P_A(\lambda x) = \lambda P_A x$. □

Remark V.52. As usual, if E is a HILBERT space, it suffices to require that $\overline{A}^E = A$.

EXERCISE V.53. Show that Corollary V.51 still holds if $A \neq \varnothing$ is only assumed to be a finite-dimensional vector subspace of a HILBERT space.

Finally, the results above lead to:

Corollary V.54. *If $(E, (\cdot|\cdot))$ is a pre-*HILBERT *space and A is a complete vector subspace of E, then :*

$$E = A \oplus A^{\perp}.$$

Proof. If P is the orthogonal projector on A, then any $x \in H$ can be written as follows:

$$x = P_A x + (x - P_A x),$$

where, by definition, $P_A x \in A$ and $x - P_A x \in A^{\perp}$; in addition, this decomposition is unique because of the uniqueness of $P_A x$. □

Remark V.55. Here, A can be empty: in this case, $A^{\perp} = E$. As usual, if E is a HILBERT space, it suffices to require $\overline{A}^E = A$.

Now, it has already been observed that, for all x in a pre-HILBERT space, $y \mapsto \langle x|y\rangle$ is a continuous linear form. Theorem V.43 allows to address the converse statement:

Theorem V.56 (RIESZ Representation Theorem). *If $(E, (\cdot|\cdot))$ is a pre-*HILBERT *space then, for any $\varphi \in E'$, there exists a unique $y \in E$ such that*

$$(\forall x \in E) \quad \varphi(x) = (x|y).$$

In addition, $\|\varphi\|_{E'} = \|y\|_E$.

Proof. Exercise. □

V.3 Approximation in HILBERT Spaces

In this Section, I will denote a non-empty set.

V.3.1 HILBERT Bases

Definition V.57. Let $(E, (\cdot|\cdot))$ be a pre-HILBERT space. Any family $\{e_i\}_{i \in I}$ of vectors of E is said to be *orthogonal* if

$$(\forall (i,j) \in I^2, i \neq j) \quad (e_i|e_j) = 0$$

and *orthonormal* if

$$(\forall (i,j) \in I^2) \quad (e_i|e_j) = \delta_{ij}.$$

Example V.58. In $\ell^2(\mathbb{N})$ as defined in Example V.41, the family of vectors $\{u_n\}_{n \in \mathbb{N}}$ defined as follows:

$$(\forall (n,m) \in \mathbb{N}^2) \quad u_{m_n} = \delta_{mn}$$

is orthonormal.

Definition V.59. Let $\mathscr{F} = \{e_i\}_{i\in I}$ be an orthogonal family of a pre-HILBERT space $(E, (\cdot|\cdot))$. For all $x \in E$, the scalars $(\widehat{x}_i)_{i\in I}$ defined by

$$(\forall i \in I) \quad \widehat{x}_i = \frac{(x|e_i)}{\|e_i\|^2}$$

are called the *FOURIER*[35] *coefficients of x with respect to $\mathscr{F}$.*

Example V.60. In the case where $E = \mathrm{L}^2_p(]0, 2\pi[)$, the "standard" definition of FOURIER coefficients is retrieved: $(\cdot|\cdot)$ is the L^2 inner product, for which the family $\{x \mapsto e^{inx}\}_{n\in\mathbb{Z}}$ is, indeed, orthogonal.

In the case where I is denumerable, and this will be the case from now on (*i.e.*, the study is restricted to that of *separable* HILBERT spaces), this allows to generalize immediately the classic BESSEL[36]'s Inequality.

Theorem V.61 (BESSEL's Inequality). *If $\mathscr{F} = \{e_i\}_{i\in I}$ is an orthogonal family of a pre-*HILBERT *space $(E, (\cdot|\cdot))$, then*

$$(\forall x \in E) \quad \sum_{i\in I} \left|(x|e_i)\right|^2 \leqslant \|x\|^2$$

Proof. Let J be a finite subset of I, and denote $(x|e_i) = x_i$ for all $i \in I$. Using the definition of an orthogonal family, one has for all $x \in E$:

$$\begin{aligned} 0 \leqslant \left\| x - \sum_{i\in J} (x|e_i)\, e_i \right\|^2 &= \left(x - \sum_{i\in J} x_i e_i \middle| x - \sum_{i\in J} x_i e_i \right) \\ &= \|x\|^2 - \sum_{i\in J} \overline{x}_i\, (x|e_i) - \sum_{i\in J} x_i\, (e_i|x) + \sum_{i\in J} x_i \overline{x}_i \\ &= \|x\|^2 - \sum_{i\in J} |x_i|^2 \end{aligned}$$

whence the conclusion. □

Remark V.62. An equivalent formulation of BESSEL's Inequality, using the notations of Definition V.59, is:

$$(\forall x \in E) \quad \sum_{i\in I} |\widehat{x}_i|^2\, \|e_i\|^2 \leqslant \|x\|^2$$

which is especially interesting when the family is orthonormal since, in this case, the $\|e_i\|^2$ are all equal to 1.

Now, in the context of HILBERT spaces, these results become more accurate.

Definition V.63. A subset A of a HILBERT space H is said to be *total in H* if $A^\perp = \{0_H\}$.

Definition V.64. Let H be a HILBERT space. An orthonormal family $\mathscr{F} = \{e_i\}_{i\in I}$ of a H is said to be a *HILBERT basis of H* if

$$(\forall x \in H) \quad x = \sum_{i\in I} \widehat{x}_i e_i$$

where the $\widehat{x}_i$ denote the FOURIER coefficients of x with respect to $\mathscr{F}$.

Provided these definitions, one can characterize the HILBERT bases:

[35] Joseph FOURIER, French mathematician, 1768–1830.
[36] Friedrich BESSEL, German astronomer and mathematician, 1784–1846.

Theorem V.65. *Let $\mathscr{F} = \{e_i\}_{i\in I}$ be an orthonormal family of a* HILBERT *space H. The following statements are equivalent:*

i. *$\mathscr{F}$ is a* HILBERT *basis of H;*

ii. *$\mathscr{F}$ is total in H;*

iii. *the linear span of $\mathscr{F}$ is dense in H;*

iv. *for all $(x,y) \in H^2$, $(x|y) = \sum_{i\in I} \widehat{x}_i \overline{\widehat{y}_j}$*

Proof. Exercise. □

In addition, in the case of HILBERT bases, BESSEL's Inequality becomes:

Corollary V.66 (PARSEVAL[37]'s Theorem). *Let H be a* HILBERT *space. $\mathscr{F} = \{e_i\}_{i\in I}$ is a* HILBERT *basis of H if and only if*

$$(\forall x \in H) \quad \|x\|^2 = \sum_{i\in I} |\widehat{x}_i|^2 \, \|e_i\|^2 \tag{V.4}$$

Proof. Assume (V.4) is true, and consider $x \in \mathscr{F}^{\perp}$. For all $i \in I$, $(x|e_i) = 0$ thus $\|x\| = 0$ whence $x = 0_H$. Hence, by Theorem V.65, $\mathscr{F}$ is a HILBERT basis of H. The converse statement is immediate, by using the same proof as for BESSEL's inequality combined with the density of the linear span of $\mathscr{F}$. □

V.3.2 Orthogonal Polynomials

Orthogonal polynomials are a particularly useful application of HILBERT spaces, in numerical analysis as well as in probability theory. The general idea is quite simple: let f be a continuous positive function over a given interval Isuch that, for all $n \in \mathbb{N}$, the $x \mapsto x^n$ monomials belong to the HILBERT space $H = \mathrm{L}^2(I, \mathfrak{B}, \mu)$, where μ is the measure with density q with respect to the LEBESGUE measure. In other words,

$$(\forall n \in \mathbb{N}) \quad \int_I |x|^n q(x)\,\mathrm{d}x < \infty.$$

By orthogonalizing this sequence, a family $(p_n)_{n\in\mathbb{N}}$ of pairwise orthogonal (in the sense of H) polynomials is obtained:

$$(\forall (n,m) \in \mathbb{N}^2, n \neq m) \quad \int_I p_n(x) p_m(x) q(x)\,\mathrm{d}x = 0.$$

If, in addition, the family $\mathscr{F} = (p_n)_{n\in\mathbb{N}}$ is total, then it is a HILBERT basis and one can express any function of H in terms of its FOURIER coefficients with respect to $\mathscr{F}$. In the context of numerical analysis, this means that any function of this space can be approached as closely as desired by means of linear combinations of p_n polynomials.

The study will be restricted here to that of HERMITE[38] polynomials, because they are directly related to WIENER[39] spaces. In fact, the mathematical literature uses two different and not equivalent definitions of HERMITE polynomials; probabilists generally work with the following ones:

[37] Marc-Antoine PARSEVAL DES CHÊNES, French mathematician, 1755–1836.
[38] Charles HERMITE, French mathematician, 1822–1901.
[39] Norbert WIENER, American mathematician, 1894–1964.

Definition V.67. For all $n \in \mathbb{N}^*$, the function H_n defined over $\mathbb{R}$ by

$$(\forall x \in \mathbb{R}) \quad H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-\frac{x^2}{2}}$$

is called the nth *HERMITE polynomial.* In addition, H_0 is defined as being identically equal to 1.

Example V.68. One can easily check that, for all $x \in \mathbb{R}$,

$$\begin{aligned}
H_0(x) &= 1 \\
H_1(x) &= x \\
H_2(x) &= x^2 - 1 \\
H_3(x) &= x^3 - 3x \\
H_4(x) &= x^4 - 6x^2 + 3 \\
H_5(x) &= x^5 - 10x^3 + 15x \\
H_6(x) &= x^6 - 15x^4 + 45x^2 - 15.
\end{aligned}$$

Remark V.69. HERMITE polynomials are also often defined by

$$(\forall (n,x) \in \mathbb{N} \times \mathbb{R}) \quad h_n(x) = (-1)^n e^{x^2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2}$$

which leads to a completely different set of polynomials.

The H_n satisfy the following properties:

Proposition V.70. *If H_n denotes the nth* HERMITE *polynomial, then*

$$(\forall (n,x) \in \mathbb{N} \times \mathbb{R}) \quad \left\{ \begin{array}{rcl} H_n(-x) & = & (-1)^n H_n(x) \\ H_{n+1}(x) & = & xH_n(x) - nH_{n-1}(x) \\ H_n'(x) & = & nH_{n-1}(x) \end{array} \right.$$

and in particular each H_n is a polynomial with degree n.

Proof. These identities result directly from the fact that the $H_n(x)$ are the coefficients of the expansion in powers of t of $\exp(tx - \frac{1}{2}t^2)$. A recurrence on the second identity combined with the fact that $H_0 = 1$ yields the last property. □

Remark V.71. A direct consequence of Propositon V.70 is that, for all $n \in \mathbb{N}$, $\frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2}$ is the product of e^{-x^2} by a polynomial function. This implies that each $\frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2}$ tends to 0 at $\pm\infty$.

Proposition V.72. *If H_n denotes the nth* HERMITE *polynomial, then the family $\left\{ \frac{1}{\sqrt{n!}} H_n \right\}_{n \in \mathbb{N}}$ is a* HILBERT *basis of $H = \mathrm{L}^2\left(\mathbb{R}, \mathfrak{L}, \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \mathrm{d}x\right)$.*

Proof. One obtains the following equality with n integrations by parts and using Remark V.71:

$$(\forall n \in \mathbb{N}^*) \quad \int_{-\infty}^{\infty} H_n(x) x^n e^{-\frac{x^2}{2}} \, \mathrm{d}x = n! \int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} x^n \, \mathrm{d}x = n! \sqrt{2\pi}.$$

Similarily, k integrations by parts yield:

$$(\forall (k,n) \in \mathbb{N}^2, k < n) \quad \int_{-\infty}^{\infty} H_n(x) x^k e^{-\frac{x^2}{2}} \, \mathrm{d}x = (-1)^n \int_{-\infty}^{\infty} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-\frac{x^2}{2}} x^k \, \mathrm{d}x = 0.$$

Therefore, combining the linearity of the integral with the fact that each H_n has degree n,

$$(\forall n \in \mathbb{N}) \quad \|H_n\|_H^2 = \int_{-\infty}^{\infty} H_n(x)^2 e^{-\frac{x^2}{2}} \, \mathrm{d}x = n! \sqrt{2\pi},$$

and

$$(\forall (m,n) \in \mathbb{N}^2, m < n) \quad (H_n | H_m)_H = \int_{-\infty}^{\infty} H_n(x) H_m(x) e^{-\frac{x^2}{2}} \, \mathrm{d}x = 0.$$

The orthogonality for all $m \neq n$ arises by switching the roles of m and n. The proof that the family is indeed total is left as an exercise. □

The following property of HERMITE polynomials is especially useful for stochastic applications:

Proposition V.73. *If X is a gaussian random variable with $\mathbb{V}[X] = 1$, then*

$$\mathbb{E}[H_n(X)] = \mathbb{E}[X]^n.$$

Proof. If $n = 0$ the result is immediate. Otherwise, if $n \in \mathbb{N}^*$, let $\mathbb{E}[X] = \mu$; then,

$$\begin{aligned}
\mathbb{E}[H_n(X)] &= \frac{(-1)^n}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{x^2}{2}} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-\frac{x^2}{2}} e^{-\frac{(x-\mu)^2}{2}} \, \mathrm{d}x \\
&= \frac{(-1)^n}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2}} \int_{-\infty}^{\infty} e^{\mu x} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-\frac{x^2}{2}} \, \mathrm{d}x \\
&= \frac{(-1)^n}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2}} \left(\left[e^{\mu x} \frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}} e^{-\frac{x^2}{2}} \right]_{-\infty}^{\infty} - \mu \int_{-\infty}^{\infty} e^{\mu x} \frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}} e^{-\frac{x^2}{2}} \, \mathrm{d}x \right),
\end{aligned}$$

thanks to an integration by parts. According to Remark V.71, the bracketed term vanishes, and thus

$$\mathbb{E}[H_n(X)] = \frac{(-1)^{n-1}}{\sqrt{2\pi}} \mu e^{-\frac{\mu^2}{2}} \int_{-\infty}^{\infty} e^{\mu x} \frac{\mathrm{d}^{n-1}}{\mathrm{d}x^{n-1}} e^{-\frac{x^2}{2}} \, \mathrm{d}x.$$

By recurrence, it then follows that

$$\mathbb{E}[H_n(X)] = \frac{\mu^n}{\sqrt{2\pi}} e^{-\frac{\mu^2}{2}} \int_{-\infty}^{\infty} e^{\mu x} e^{-\frac{x^2}{2}} \, \mathrm{d}x = \frac{\mu^n}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu)^2}{2}} \, \mathrm{d}x = \mu^n.$$

□

Remark V.74. It is also interesting to notice, and left to the reader as an exercise, that

$$(\forall (s,x) \in \mathbb{R}^2) \quad e^{sx - \frac{1}{2}s^2} = \sum_{k=0}^{\infty} \frac{s^n}{n!} H_n(x).$$

For any fixed x, the radius of convergence of this power series in t is ∞. In particular, differentiation can be made term by term, and thus

$$(\forall (n,x) \in \mathbb{N} \times \mathbb{R}) \quad \frac{\mathrm{d}^n}{\mathrm{d}s^n} e^{sx - \frac{1}{2}s^2} |_{s=0} = H_n(x).$$

V.4 Application: WIENER Chaoses

In this Section, H will denote a real separable HILBERT space and $(S, \mathfrak{S}, P)$ a complete probability space.

Definition V.75. A stochastic process $W = \{W(h)\}_{h \in H}$ defined in $(S, \mathfrak{S}, P)$ is said to be an *isonormal gaussian process on H* if W is a family of gaussian centered random variables such that:

$$(\forall (h,g) \in H^2) \quad \mathbb{E}[W(h)W(g)] = (h|g)_H.$$

Proposition V.76. *Let $W = \{W(h)\}_{h \in H}$ be an isonormal gaussian process on H. The map $h \mapsto W(h)$ is a linear isometry from H to a closed vector subspace of $\mathrm{L}^2(S, \mathfrak{S}, P)$, denoted $\mathscr{H}_1$.*

Proof. Each $W(h)$ is a random variable on S and, according to the hypothesis,

$$(\forall h \in H) \quad \mathbb{E}\left[W(h)^2\right] = \|h\|_H^2 < \infty. \tag{V.5}$$

Therefore, the image set of $h \mapsto W(h)$ is in $\mathrm{L}^2(S, \mathfrak{S}, P)$, and

$$(\forall (\lambda, \mu, f, g) \in \mathbb{R}^2 \times H^2) \quad \mathbb{E}\left[(W(\lambda g + \mu h) - \lambda W(g) - \mu W(h))^2\right] = 0,$$

whence

$$(\forall (\lambda, \mu, f, g) \in \mathbb{R}^2 \times H^2) \quad W(\lambda g + \mu h) - \lambda W(g) - \mu W(h) \overset{\text{a.s.}}{=} 0,$$

which proves the linearity, in the sense of $\mathrm{L}^2(S, \mathfrak{S}, P)$. In addition, it follows directly from (V.5) that the mapping is isometric since, by definition,

$$\mathbb{E}\left[W(h)^2\right] = \int W(h)^2 \, \mathrm{d}P = \|W(h)\|^2_{\mathrm{L}^2(S, \mathfrak{S}, P)}.$$

Now, the image by a linear isometry of a complete set being also complete (this result, easy to establish, is left to the reader as an exercise), the image set $\mathscr{H}_1$ of H by $h \mapsto W(h)$ is a complete vector subspace of $\mathrm{L}^2(S, \mathfrak{S}, P)$, which implies that is closed. □

Remark V.77. By definition, each element of $\mathscr{H}_1$ is a centered gaussian random variable; however, the converse is not true.

Definition V.78. Let $W = \{W(h)\}_{h \in H}$ be an isonormal gaussian process on H. For all $n \in \mathbb{N}^*$, the nth *WIENER chaos* $\mathscr{H}_n$ is defined as being the closed vector subspace of $\mathrm{L}^2(S, \mathfrak{S}, P)$ generated by the family of random variables $\left\{H_n(W(h)), h \in H, \|h\|_H = 1\right\}$. The set of constants of $\mathrm{L}^2(S, \mathfrak{S}, P)$ is denoted as $\mathscr{H}_0$.

Remark V.79. This is indeed consistent with the previous definition of $\mathscr{H}_1$, because of the linearity of $h \mapsto W(h)$:

$$(\forall h \in H, h \neq 0) \quad W(h) = \|h\|_H \, W\left(\frac{h}{\|h\|_H}\right)$$

which belongs to the closed vector subspace of $\mathrm{L}^2(S, \mathfrak{S}, P)$ generated by $\left\{W(h), h \in H, \|h\|_H = 1\right\}$.

Example V.80. In the case where $H = \mathbb{R}$ along with its standard HILBERT space structure, then the only $h \in H$ such that $\|h\|_H = 1$ are ± 1. Combining Proposition V.70 and the linearity of $h \mapsto W(h)$ yields

$$(\forall n \in \mathbb{N}) \quad H_n(W(-1)) = H_n(-W(1)) = (-1)^n H_n(W(1)),$$

and thus each $\mathscr{H}_n$ is a one-dimensional subspace of $\mathrm{L}^2(S, \mathfrak{S}, P)$ generated by $H_n(W(1))$. For instance, since $W(1)$ is by definition the standard normal random variable on $\mathrm{L}^2(S, \mathfrak{S}, P)$, the elements of $\mathscr{H}_1$ are the centered gaussian random variables on this space.

Lemma V.81. *If X and Y are two random variables with joint gaussian normal centered distribution and H_n denotes the nth* HERMITE *polynomial, then*

$$(\forall (m,n) \in \mathbb{N}^2) \quad \mathbb{E}[H_n(X)H_m(Y)] = \begin{cases} 0 & \text{if} \quad n \neq m \\ n!(\mathbb{E}[XY])^n & \text{if} \quad n = m \end{cases}.$$

Proof. By definition, the random vector (X,Y) is centered, gaussian, with the following variance-covariance matrix:

$$\Gamma = \begin{pmatrix} 1 & \mathbb{E}[XY] \\ \mathbb{E}[XY] & 1 \end{pmatrix}.$$

Therefore, for each $(s,t) \in \mathbb{R}^2$, the linear combination of $sX+tY$ is a centered gaussian random variable with variance

$$\mathbb{V}[sX+tY] = \begin{pmatrix} s & t \end{pmatrix} \begin{pmatrix} 1 & \mathbb{E}[XY] \\ \mathbb{E}[XY] & 1 \end{pmatrix} \begin{pmatrix} s \\ t \end{pmatrix} = s^2 + 2st\mathbb{E}[XY] + t^2,$$

whence

$$(\forall (s,t) \in \mathbb{R}^2) \quad \mathbb{E}\left[e^{sX+tY}\right] = e^{\frac{1}{2}(s^2+2st\mathbb{E}[XY]+t^2)},$$

and thus, multiplying both sides by $e^{-\frac{1}{2}(s^2+t^2)}$,

$$(\forall (s,t) \in \mathbb{R}^2) \quad \mathbb{E}\left[e^{sX-\frac{1}{2}s^2} e^{tY-\frac{1}{2}t^2}\right] = e^{st\mathbb{E}[XY]}.$$

On the one hand,

$$\begin{aligned}
(\forall (s,t) \in \mathbb{R}^2) \quad \frac{\partial^{n+m}}{\partial s^n \partial t^m} e^{st\mathbb{E}[XY]} &= \mathbb{E}[XY]^n \frac{\partial^m}{\partial t^m}\left(t^n e^{st\mathbb{E}[XY]}\right) \\
&= \mathbb{E}[XY]^n \sum_{k=0}^{k=m} \binom{m}{k} \frac{\partial^k}{\partial t^k} t^n \frac{\partial^{m-k}}{\partial t^{m-k}} e^{st\mathbb{E}[XY]} \\
&= \mathbb{E}[XY]^n e^{st\mathbb{E}[XY]} \sum_{k=0}^{k=m} \binom{m}{k} \mathbb{E}[XY]^{m-k} s^{m-k} \frac{\partial^k}{\partial t^k} t^n.
\end{aligned}$$

Now, if $s=t=0$, each $s^{m-k}\frac{\partial^k}{\partial t^k} t^n$ vanishes, except if $k=n=m$. Therefore,

$$\frac{\partial^{n+m}}{\partial s^n \partial t^m} e^{st\mathbb{E}[XY]}\Big|_{(s,t)=(0,0)} = \begin{cases} 0 & \text{if} \quad n \neq m \\ n!(\mathbb{E}[XY])^n & \text{if} \quad n = m \end{cases}.$$

On the other hand, in a neighborhood of $(0,0)$,

$$\frac{\partial^{n+m}}{\partial s^n \partial t^m} \mathbb{E}\left[e^{sX-\frac{1}{2}s^2} e^{tY-\frac{1}{2}t^2}\right] = \mathbb{E}\left[\frac{\partial^n}{\partial s^n} e^{sX-\frac{1}{2}s^2} \frac{\partial^m}{\partial t^m} e^{tY-\frac{1}{2}t^2}\right],$$

and thus, according to Remark V.74,

$$\frac{\partial^{n+m}}{\partial s^n \partial t^m} \mathbb{E}\left[e^{sX-\frac{1}{2}s^2} e^{tY-\frac{1}{2}t^2}\right]\Big|_{(s,t)=(0,0)} = \mathbb{E}[H_n(X)H_m(Y)],$$

which yields the desired result. □

Theorem V.82. *For all $(n,m) \in \mathbb{N}^2$, $n \neq m$, one has $\mathscr{H}_n \perp \mathscr{H}_m$.*

Proof. This result is a direct consequence of Lemma V.81. □

Example V.83. With $H = \mathbb{R}$ along with its standard HILBERT space structure, let X, Y and Z belong to, respectively, $\mathscr{H}_0$, $\mathscr{H}_1$, and $\mathscr{H}_2$. According to Example V.80, there exists a unique $(a,b,c) \in \mathbb{R}^3$ such that $X = a$, $Y = bW(1)$ and $Z = cH_2(W(1))$, and one has

$$\begin{aligned}
\mathbb{E}[XY] &= ab\int W(1)\,\mathrm{d}P = 0,\\
\mathbb{E}[XZ] &= ac\int W(1)^2 - 1\,\mathrm{d}P = ac\left(\int W(1)^2\,\mathrm{d}P - \int \mathrm{d}P\right) = ac(1-1) = 0,\\
\mathbb{E}[YZ] &= bc\int W(1)(W(1)^2-1)\,\mathrm{d}P = bc\left(\int W(1)^3\,\mathrm{d}P - \int W(1)\,\mathrm{d}P\right) = bc(0-0) = 0.
\end{aligned}$$

Henceforth the σ-field generated by the random variables $\{W(h)\}_{h\in H}$, *i.e.*, the smallest σ-algebra on S such that these random variables are measurable, will be denoted $\mathfrak{S}(W)$.

Lemma V.84. *The family* $\left\{e^{W(h)}\right\}_{h\in H}$ *is a total subset of* $\mathrm{L}^2(S,\mathfrak{S},P)$.

Proof. Technical, *cf.* [Nua95]. □

The fundamental result for WIENER chaos decomposition can then be stated as follows:

Theorem V.85. *If a stochastic process W on $\mathrm{L}^2(S,\mathfrak{S},P)$ is an isonormal gaussian process on H, then*

$$\mathrm{L}^2(S,\mathfrak{S}(W),P) = \bigoplus_{n\in\mathbb{N}} \mathscr{H}_n$$

Proof. If $X \in \mathrm{L}^2(S,\mathfrak{S}(W),P)$ is such that $X \perp \mathscr{H}_n$ for all $n \in \mathbb{N}$, then

$$(\forall n \in \mathbb{N})(\forall h \in H, \|h\|_H = 1) \quad \mathbb{E}[XH_n(W(h))] = 0$$

It can be easily shown that each monomial with degree n can be expressed as a linear combination of HERMITE polynomials of order not greater than n; therefore,

$$(\forall n \in \mathbb{N})(\forall h \in H, \|h\|_H = 1) \quad \mathbb{E}[X(W(h))^n] = 0,$$

and thus

$$(\forall t \in \mathbb{R})(\forall h \in H, \|h\|_H = 1) \quad \mathbb{E}\left[Xe^{tW(h)}\right] = 0.$$

Using the linearity of $h \mapsto W(h)$, it then follows that

$$(\forall h \in H, h \neq 0) \quad \mathbb{E}\left[Xe^{W(h)}\right] = \mathbb{E}\left[X\exp\left(\|h\|_H W\left(\tfrac{1}{\|h\|_H}h\right)\right)\right] = 0,$$

whence, by Lemma V.84, $X = 0$. □

Example V.86. Let $(S,\mathfrak{S},P) = (\mathbb{R},\mathfrak{L},\nu)$, where ν has a standard normal density with respect to ℓ, $H = \mathbb{R}$ with its standard HILBERT space structure, and define W by

$$(\forall (h,x) \in \mathbb{R}^2) \quad W(h)(x) = hx.$$

Each $W(h)$ can thus been naturally seen as a real random variable on $(\mathbb{R}, \mathfrak{L}, \nu)$ and, by definition of ν, the CDF of $W(h)$ for any $h \neq 0$ is given by

$$(\forall x \in \mathbb{R}) \quad F_{W(h)}(x) = \nu\left(W(h)^{-1}(]-\infty, x])\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x}{h}} e^{-\frac{t^2}{2}} \, \mathrm{d}t = \frac{1}{\sqrt{2\pi}h} \int_{-\infty}^{x} e^{-\frac{u^2}{2h^2}} \, \mathrm{d}u$$

while $W(0)$ is identically zero. Therefore, each $W(h)$ has indeed a centered gaussian distribution. In addition, for all $W(g)$ and $W(h)$, the linearity of $h \mapsto W(h)$ ensures that any linear combination of them has also a centered gaussian distribution, and $\mathbb{E}[W(h)W(g)] = hg = (h|g)_H$. Hence, the hypotheses of Theorem V.85 are met, and since $W(1) : x \mapsto x$, Proposition V.72 is retrieved.

V HILBERT Spaces

References

[Jec02] T.J. Jech. *Set Theory*. Monographs in Mathematics. Springer Verlag, third millenium edition, 2002.

[Lan02] S. Lang. *Algebra*. Monographs in Mathematics. Springer Verlag, revised third edition, 2002.

[Nua95] D. Nualart. *The Malliavin Calculus and Related Topics*. Probability and Its Applications. Springer Verlag, 1995.

[Poi02] H. Poincaré. *La science et l'hypothèse*. Flammarion, 1902.

[Rud86] W. Rudin. *Real and Complex Analysis*. Series in Higher Mathematics. McGraw-Hill, third edition, 1986.

[Yos95] K. Yosida. *Functional Analysis*. Classics in Mathematics. Springer Verlag, sixth edition, 1995.

DISTRIBUTION:

2 Pr Jérôme Pousin
I.N.S.A., Center for Mathematics
Bâtiment Léonard de Vinci
69621 Villeurbanne cedex, France

2 Pr Mauro Valorani
University of Rome La Sapienza
Dpt di Meccanica & Aeronautica
Via Eudossiana, 18
00184 Roma, Italy

1 MS 9051
Bert J. Debusschere, 8351

1 MS 9051
Jeremiah C. Lee, 8351

1 MS 9051
Andy McIlroy, 8351

1 MS 9051
Habib N. Najm, 8351

1 MS 9051
Henri C. Ozarovsky, 8351

5 MS 9051
Philippe P. Pébay, 8351

1 MS 9915
Helgi Adalsteinsson, 8961

1 MS 9915
Curtis L. Janssen, 8961

1 MS 9915
Mitchel W. Sukalski, 8961

1 MS 9217
Paul T. Boggs, 8962

1 MS 9217
Kevin R. Long, 8962

1 MS 9051
Jaideep Ray, 8964

3 MS 9018
Central Technical Files, 8940-1

1 MS 0899
Technical Library, 9616

1 MS 9021
Classification Office, 8511, for Technical Library, MS 0899, 9616 DOE/OSTI via URL