LECTURE NOTES

IN MEASURE THEORY

Christer Borell
Matematik
Chalmers och Göteborgs universitet
412 96 Göteborg
(Version: Feb 05)

PREFACE

These are lecture notes on integration theory for a ten-week course at the Chalmers University of Technology and the Göteborg University. The parts defining the course essentially lead to the same results as the first three chapters in the Folland book [F] , which is used as a text book on the course. The proofs in the lecture notes sometimes differ from those given in [F] . Here is a brief description of the differences to simplify for the reader.

In Chapter 1 we introduce so called π- and λ-systems, which are substitutes for monotone classes of sets [F]. Besides we prefer to emphasize metric outer measures instead of so called premeasures. Throughout the course, a variety of important measures are obtained as image measures of the linear measure on the real line. In Section 1.6 positive measures in $\mathbf{R}$ induced by increasing right continuous mappings are constructed in this way.

Chapter 2 deals with integration and is very similar to [F] and most other texts.

Chapter 3 starts with some standard facts about metric spaces and relates the concepts to measure theory. For example Ulam's Theorem is included. The existence of product measures is based on properties of π- and λ-systems.

Chapter 4 deals with different modes of convergence and is mostly close to [F] . Here we include a section about orthogonality since many students have seen parts of this theory before.

The Lebesgue Decomposition Theorem and Radon-Nikodym Theorem in Chapter 5 are proved using the von Neumann beautiful L^2-proof.

To illustrate the power of abstract integration these notes contain several sections, which do not belong to the course but may help the student to a better understanding of measure theory. The corresponding parts are set between the symbols

$$\downarrow\downarrow\downarrow$$

and

$$\uparrow\uparrow\uparrow$$

respectively.

Finally I would like to express my deep gratitude to the students in my classes for suggesting a variety of improvements and a special thank

to Jonatan Vasilis who has provided numerous comments and corrections in my original text.

Göteborg 2006
Christer Borell

CONTENT

CHAPTER 1

MEASURES

Introduction

The Riemann integral, dealt with in calculus courses, is well suited for computations but less suited for dealing with limit processes. In this course we will introduce the so called Lebesgue integral, which keeps the advantages of the Riemann integral and eliminates its drawbacks. At the same time we will develop a general measure theory which serves as the basis of contemporary analysis and probability.

In this introductory chapter we set forth some basic concepts of measure theory, which will open for abstract Lebesgue integration.

1.1. σ-Algebras and Measures

Throughout this course

$\mathbf{N} = \{0, 1, 2, ...\}$ (the set of natural numbers)
$\mathbf{Z} = \{..., -2, -1, 0, 1, , 2, ...\}$ (the set of integers)
$\mathbf{Q}$ = the set of rational numbers
$\mathbf{R}$ = the set of real numbers
$\mathbf{C}$ = the set of complex numbers.

If $A \subseteq \mathbf{R}$, A_+ is the set of all strictly positive elements in A.

If f is a function from a set A into a set B, this means that to every $x \in A$ there corresponds a point $f(x) \in B$ and we write $f : A \to B$. A function is often called a map or a mapping. The function f is injective if

$$(x \neq y) \Rightarrow (f(x) \neq f(y))$$

and surjective if to each $y \in B$, there exists an $x \in A$ such that $f(x) = y$. An injective and surjective function is said to be bijective.

A set A is finite if either A is empty or there exist an $n \in \mathbf{N}_+$ and a bijection $f : \{1, ..., n\} \to A$. The empty set is denoted by ϕ. A set A is said to be denumerable if there exists a bijection $f : \mathbf{N}_+ \to A$. A subset of a denumerable set is said to be at most denumerable.

Let X be a set. For any $A \subseteq X$, the indicator function χ_A of A relative to X is defined by the equation

$$\chi_A(x) = \begin{cases} 1 \text{ if } x \in A \\ 0 \text{ if } x \in A^c. \end{cases}$$

The indicator function χ_A is sometimes written 1_A. We have the following relations:

$$\chi_{A^c} = 1 - \chi_A$$

$$\chi_{A \cap B} = \min(\chi_A, \chi_B) = \chi_A \chi_B$$

and

$$\chi_{A \cup B} = \max(\chi_A, \chi_B) = \chi_A + \chi_B - \chi_A \chi_B.$$

Definition 1.1.1. Let X be a set.

a) A collection $\mathcal{A}$ of subsets of X is said to be an algebra in X if $\mathcal{A}$ has the following properties:

(i) $X \in \mathcal{A}$.
(ii) $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$, where A^c is the complement of A relative to X.
(iii) If $A, B \in \mathcal{A}$ then $A \cup B \in \mathcal{A}$.

(b) A collection $\mathcal{M}$ of subsets of X is said to be a σ-algebra in X if $\mathcal{M}$ is an algebra with the following property:

If $A_n \in \mathcal{M}$ for all $n \in \mathbf{N}_+$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{M}$.

If $\mathcal{M}$ is a σ-algebra in X, $(X, \mathcal{M})$ is called a measurable space and the members of $\mathcal{M}$ are called measurable sets. The so called power set $\mathcal{P}(X)$, that is the collection of all subsets of X, is a σ-algebra in X. It is simple to prove that the intersection of any family of σ-algebras in X is a σ-algebra. It follows that if $\mathcal{E}$ is any subset of $\mathcal{P}(X)$, there is a unique smallest σ-algebra $\sigma(\mathcal{E})$ containing $\mathcal{E}$, namely the intersection of all σ-algebras containing $\mathcal{E}$.

The σ-algebra $\sigma(\mathcal{E})$ is called the σ-algebra generated by $\mathcal{E}$. The σ-algebra generated by all open intervals in $\mathbf{R}$ is denoted by $\mathcal{R}$. It is readily seen that the σ-algebra $\mathcal{R}$ contains every subinterval of $\mathbf{R}$. Before we proceed, recall that a subset E of $\mathbf{R}$ is open if to each $x \in E$ there exists an open subinterval of $\mathbf{R}$ contained in E and containing x; the complement of an open set is said to be closed. We claim that $\mathcal{R}$ contains every open U subset of $\mathbf{R}$. To see this suppose $x \in U$ and let $x \in]a, b[\subseteq U$, where $-\infty < a < b < \infty$. Now pick $r, s \in \mathbf{Q}$ such that $a < r < x < s < b$. Then $x \in]r, s[\subseteq U$ and it follows that U is the union of all bounded open intervals with rational boundary points contained in U. Since this family of intervals is at most denumberable we conclude that $U \in \mathcal{R}$. In addition, any closed set belongs to $\mathcal{R}$ since its complements is open. It is by no means simple to grasp the definition of $\mathcal{R}$ at this stage but the reader will successively see that the σ-algebra $\mathcal{R}$ has very nice properties. At the very end of Section 1.3, using the so called Axiom of Choice, we will exemplify a subset of the real line which does not belong to $\mathcal{R}$. In fact, an example of this type can be constructed without the Axiom of Choice (see Dudley's book $[D]$).

In measure theory, inevitably one encounters ∞. For example the real line has infinite length. Below $[0, \infty] = [0, \infty[\cup \{\infty\}$. The inequalities $x \leq y$ and $x < y$ have their usual meanings if $x, y \in [0, \infty[$. Furthermore, $x \leq \infty$ if $x \in [0, \infty]$ and $x < \infty$ if $x \in [0, \infty[$. We define $x + \infty = \infty + x = \infty$ if $x, y \in [0, \infty]$, and

$$x \cdot \infty = \infty \cdot x = \begin{cases} 0 & \text{if } x = 0 \\ \infty & \text{if } 0 < x \leq \infty. \end{cases}$$

Sums and multiplications of real numbers are defined in the usual way.

If $A_n \subseteq X$, $n \in \mathbf{N}_+$, and $A_k \cap A_n = \phi$ if $k \neq n$, the sequence $(A_n)_{n \in \mathbf{N}_+}$ is called a disjoint denumerable collection. If $(X, \mathcal{M})$ is a measurable space, the collection is called a denumerable measurable partition of A if $A = \cup_{n=1}^{\infty} A_n$ and $A_n \in \mathcal{M}$ for every $n \in \mathbf{N}_+$. Some authors call a denumerable collection of sets a countable collection of sets.

Definition 1.1.2. (a) Let $\mathcal{A}$ be an algebra of subsets of X. A function $\mu : \mathcal{A} \to [0, \infty]$ is called a content if

(i) $\mu(\phi) = 0$
(ii) $\mu(A \cup B) = \mu(A) + \mu(B)$ if $A, B \in \mathcal{A}$ and $A \cap B = \phi$.

(b) If $(X, \mathcal{M})$ is a measurable space a content μ defined on the σ-algebra $\mathcal{M}$ is called a positive measure if it has the following property:

For any disjoint denumerable collection $(A_n)_{n \in \mathbf{N}_+}$ of members of $\mathcal{M}$

$$\mu(\cup_{n=1}^{\infty} A_n) = \Sigma_{n=1}^{\infty} \mu(A_n).$$

If $(X, \mathcal{M})$ is a measurable space and the function $\mu : \mathcal{M} \to [0, \infty]$ is a positive measure, $(X, \mathcal{M}, \mu)$ is called a positive measure space. The quantity $\mu(A)$ is called the μ-measure of A or simply the measure of A if there is no ambiguity. Here $(X, \mathcal{M}, \mu)$ is called a probability space if $\mu(X) = 1$, a finite positive measure space if $\mu(X) < \infty$, and a σ-finite positive measure space if X is a denumerable union of measurable sets with finite μ-measure. The measure μ is called a probability measure, finite measure, and σ-finite measure, if $(X, \mathcal{M}, \mu)$ is a probability space, a finite positive measure space, and a σ-finite positive measure space, respectively. A probability space is often denoted by $(\Omega, \mathcal{F}, P)$. A member A of $\mathcal{F}$ is called an event.

As soon as we have a positive measure space $(X, \mathcal{M}, \mu)$, it turns out to be a fairly simple task to define a so called μ-integral

$$\int_X f(x) d\mu(x)$$

as will be seen in Chapter 2.

The class of all finite unions of subintervals of $\mathbf{R}$ is an algebra which is denoted by $\mathcal{R}_0$. If $A \in \mathcal{R}_0$ we denote by $l(A)$ the Riemann integral

$$\int_{-\infty}^{\infty} \chi_A(x)dx$$

and it follows from courses in calculus that the function $l : \mathcal{R}_0 \to [0, \infty]$ is a content. The algebra $\mathcal{R}_0$ is called the Riemann algebra and l the Riemann content. If I is a subinterval of $\mathbf{R}$, $l(I)$ is called the length of I. Below we follow the convention that the empty set is an interval.

If $A \in \mathcal{P}(X)$, $c_X(A)$ equals the number of elements in A, when A is a finite set, and $c_X(A) = \infty$ otherwise. Clearly, c_X is a positive measure. The measure c_X is called the counting measure on X.

Given $a \in X$, the probability measure δ_a defined by the equation $\delta_a(A) = \chi_A(a)$, if $A \in \mathcal{P}(X)$, is called the Dirac measure at the point a. Sometimes we write $\delta_a = \delta_{X,a}$ to emphasize the set X.

If μ and ν are positive measures defined on the same σ-algebra $\mathcal{M}$, the sum $\mu + \nu$ is a positive measure on $\mathcal{M}$. More generally, $\alpha\mu + \beta\nu$ is a positive measure for all real α, $\beta \geq 0$. Furthermore, if $E \in \mathcal{M}$, the function $\lambda(A) = \mu(A \cap E)$, $A \in \mathcal{M}$, is a positive measure. Below this measure λ will be denoted by μ^E and we say that μ^E is concentrated on E. If $E \in \mathcal{M}$, the class $\mathcal{M}_E = \{A \in \mathcal{M};\ A \subseteq E\}$ is a σ-algebra of subsets of E and the function $\theta(A) = \mu(A)$, $A \in \mathcal{M}_E$, is a positive measure. Below this measure θ will be denoted by $\mu_{|E}$ and is called the restriction of μ to $\mathcal{M}_E$.

Let $I_1, ..., I_n$ be subintervals of the real line. The set

$$I_1 \times ... \times I_n = \{(x_1, ..., x_n) \in \mathbf{R}^n;\ x_k \in I_k,\ k = 1, ..., n\}$$

is called an n-cell in $\mathbf{R}^n$; its volume $\mathrm{vol}(I_1 \times ... \times I_n)$ is, by definition, equal to

$$\mathrm{vol}(I_1 \times ... \times I_n) = \Pi_{k=1}^n l(I_k).$$

If $I_1, ..., I_n$ are open subintervals of the real line, the n-cell $I_1 \times ... \times I_n$ is called an open n-cell. The σ-algebra generated by all open n-cells in $\mathbf{R}^n$ is denoted by $\mathcal{R}_n$. In particular, $\mathcal{R}_1 = \mathcal{R}$. A basic theorem in measure theory states that there exists a unique positive measure v_n defined on $\mathcal{R}_n$ such that the measure of any n-cell is equal to its volume. The measure v_n is called the volume measure on $\mathcal{R}_n$ or the volume measure on $\mathbf{R}^n$. Clearly, v_n is σ-finite. The measure v_2 is called the area measure on $\mathbf{R}^2$ and v_1 the linear measure on $\mathbf{R}$.

Theorem 1.1.1. *The volume measure on* $\mathbf{R}^n$ *exists.*

Theorem 1.1.1 will be proved in Section 1.5 in the special case $n = 1$. The general case then follows from the existence of product measures in Section 3.4. An alternative proof of Theorem 1.1.1 will be given in Section 3.2. As soon as the existence of volume measure is established a variety of interesting measures can be introduced.

Next we prove some results of general interest for positive measures.

Theorem 1.1.2. *Let* $\mathcal{A}$ *be an algebra of subsets of* X *and* μ *a content defined on* $\mathcal{A}$. *Then,*

(a) μ *is finitely additive, that is*

$$\mu(A_1 \cup ... \cup A_n) = \mu(A_1) + ... + \mu(A_n)$$

if $A_1, ..., A_n$ *are pairwise disjoint members of* $\mathcal{A}$.

(b) *if* $A, B \in \mathcal{A}$,

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

Moreover, if $\mu(A \cap B) < \infty$, *then*

$$\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$$

(c) $A \subseteq B$ *implies* $\mu(A) \leq \mu(B)$ *if* $A, B \in \mathcal{A}$.

(d) μ *finitely sub-additive, that is*

$$\mu(A_1 \cup ... \cup A_n) \leq \mu(A_1) + ... + \mu(A_n)$$

if $A_1, ..., A_n$ *are members of* $\mathcal{A}$.

If $(X, \mathcal{M}, \mu)$ *is a positive measure space*

(e) $\mu(A_n) \to \mu(A)$ *if* $A = \cup_{n \in \mathbf{N}_+} A_n$, $A_n \in \mathcal{M}$, *and*

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq ... \ .$$

(f) $\mu(A_n) \to \mu(A)$ *if* $A = \cap_{n \in \mathbf{N}_+} A_n$, $A_n \in \mathcal{M}$,

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq ...$$

and $\mu(A_1) < \infty$.
(g) μ *is sub-additive, that is for any denumerable collection* $(A_n)_{n \in \mathbf{N}_+}$ *of members of* $\mathcal{M}$,

$$\mu(\cup_{n=1}^{\infty} A_n) \leq \Sigma_{n=1}^{\infty} \mu(A_n).$$

PROOF (a) If $A_1, ..., A_n$ are pairwise disjoint members of $\mathcal{A}$,

$$\mu(\cup_{k=1}^{n} A_k) = \mu(A_1 \cup (\cup_{k=2}^{n} A_k))$$

$$= \mu(A_1) + \mu(\cup_{k=2}^{n} A_k)$$

and, by induction, we conclude that μ is finitely additive.

(b) Recall that

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

Now $A = (A \setminus B) \cup (A \cap B)$ and we get

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

Moreover, since $A \cup B = (A \setminus B) \cup B$,

$$\mu(A \cup B) = \mu(A \setminus B) + \mu(B)$$

and, if $\mu(A \cap B) < \infty$, we have

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

(c) Part (b) yields $\mu(B) = \mu(B \setminus A) + \mu(A \cap B) = \mu(B \setminus A) + \mu(A)$, where the last member does not fall below $\mu(A)$.

(d) If $(A_i)_{i=1}^n$is a sequence of members of $\mathcal{A}$ define the so called disjunction $(B_k)_{k=1}^n$ of the sequence $(A_i)_{i=1}^n$ as

$$B_1 = A_1 \text{ and } B_k = A_k \setminus \cup_{i=1}^{k-1} A_i \text{ for } 2 \leq k \leq n.$$

Then $B_k \subseteq A_k$, $\cup_{i=1}^k A_i = \cup_{i=1}^k B_i$, $k = 1, .., n$, and $B_i \cap B_j = \phi$ if $i \neq j$. Hence, by Parts (a) and (c),

$$\mu(\cup_{k=1}^n A_k) = \Sigma_{k=1}^n \mu(B_k) \leq \Sigma_{k=1}^n \mu(A_k).$$

(e) Set $B_1 = A_1$ and $B_n = A_n \setminus A_{n-1}$ for $n \geq 2$. Then $A_n = B_1 \cup .... \cup B_n$, $B_i \cap B_j = \phi$ if $i \neq j$ and $A = \cup_{k=1}^\infty B_k$. Hence

$$\mu(A_n) = \Sigma_{k=1}^n \mu(B_k)$$

and

$$\mu(A) = \Sigma_{k=1}^\infty \mu(B_k).$$

Now e) follows, by the definition of the sum of an infinite series.

(f) Put $C_n = A_1 \setminus A_n$, $n \geq 1$. Then $C_1 \subseteq C_2 \subseteq C_3 \subseteq ...$,

$$A_1 \setminus A = \cup_{n=1}^\infty C_n$$

and $\mu(A) \leq \mu(A_n) \leq \mu(A_1) < \infty$. Thus

$$\mu(C_n) = \mu(A_1) - \mu(A_n)$$

and Part (e) shows that

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

This proves (f).

(g) The result follows from Parts d) and e).
    This completes the proof of Theorem 1.1.2.

The hypothesis "$\mu(A_1) < \infty$ " in Theorem 1.1.2 (f) is not superfluous. If $c_{\mathbf{N}_+}$ is the counting measure on $\mathbf{N}_+$ and $A_n = \{n, n+1, ...\}$, then $c_{\mathbf{N}_+}(A_n) = \infty$ for all n but $A_1 \supseteq A_2 \supseteq ....$ and $c_{\mathbf{N}_+}(\cap_{n=1}^{\infty} A_n) = 0$ since $\cap_{n=1}^{\infty} A_n = \phi$.

If $A, B \subseteq X$, the symmetric difference $A\Delta B$ is defined by the equation

$$A\Delta B =_{def} (A \setminus B) \cup (B \setminus A).$$

Note that

$$\chi_{A\Delta B} = \mid \chi_A - \chi_B \mid .$$

Moreover, we have

$$A\Delta B = A^c \Delta B^c$$

and

$$(\cup_{i=1}^{\infty} A_i)\Delta(\cup_{i=1}^{\infty} B_i) \subseteq \cup_{i=1}^{\infty}(A_i \Delta B_i).$$

Example 1.1.1. Let μ be a finite positive measure on $\mathcal{R}$. We claim that to each set $E \in \mathcal{R}$ and $\varepsilon > 0$, there exists a set A, which is finite union of intervals (that is, A belongs to the Riemann algebra $\mathcal{R}_0$), such that

$$\mu(E\Delta A) < \varepsilon.$$

To see this let $\mathcal{S}$ be the class of all sets $E \in \mathcal{R}$ for which the conclusion is true. Clearly $\phi \in \mathcal{S}$ and, moreover, $\mathcal{R}_0 \subseteq \mathcal{S}$. If $A \in \mathcal{R}_0$, $A^c \in \mathcal{R}_0$ and therefore $E^c \in \mathcal{S}$ if $E \in \mathcal{S}$. Now suppose $E_i \in \mathcal{S}$, $i \in \mathbf{N}_+$. Then to each $\varepsilon > 0$ and i there is a set $A_i \in \mathcal{R}_0$ such that $\mu(E_i \Delta A_i) < 2^{-i}\varepsilon$. If we set

$$E = \cup_{i=1}^{\infty} E_i$$

then

$$\mu(E\Delta(\cup_{i=1}^{\infty} A_i)) \leq \Sigma_{i=1}^{\infty} \mu(E_i \Delta A_i) < \varepsilon.$$

Here

$$E\Delta(\cup_{i=1}^{\infty} A_i) = \{E \cap (\cap_{i=1}^{\infty} A_i^c)\} \cup \{E^c \cap (\cup_{i=1}^{\infty} A_i)\}$$

and Theorem 1.1.2 (f) gives that

$$\mu(\{E \cap (\cap_{i=1}^{n} A_i^c)\} \cup \{(E^c \cap (\cup_{i=1}^{\infty} A_i)\}) < \varepsilon$$

if n is large enough (hint: $\cap_{i\in I}(D_i \cup F) = (\cap_{i\in I} D_i) \cup F$). But then

$$\mu(E\Delta \cup_{i=1}^{n} A_i) = \mu(\{E \cap (\cap_{i=1}^{n} A_i^c)\} \cup \{E^c \cap (\cup_{i=1}^{n} A_i)\}) < \varepsilon$$

if n is large enough we conclude that the set $E \in \mathcal{S}$. Thus $\mathcal{S}$ is a σ-algebra and since $\mathcal{R}_0 \subseteq \mathcal{S} \subseteq \mathcal{R}$ it follows that $\mathcal{S} = \mathcal{R}$.

Exercises

1. Prove that the sets $\mathbf{N} \times \mathbf{N} = \{(i, j);\ i, j \in \mathbf{N}\}$ and $\mathbf{Q}$ are denumerable.

2. Suppose $\mathcal{A}$ is an algebra of subsets of X and μ and ν two contents on $\mathcal{A}$ such that $\mu \leq \nu$ and $\mu(X) = \nu(X) < \infty$. Prove that $\mu = \nu$.

3. Suppose $\mathcal{A}$ is an algebra of subsets of X and μ a content on $\mathcal{A}$ with $\mu(X) < \infty$. Show that

$$\mu(A \cup B \cup C) = \mu(A) + \mu(B) + \mu(C)$$

$$-\mu(A \cap B) - \mu(A \cap C) - \mu(B \cap C) + \mu(A \cap B \cap C).$$

4. A collection $\mathcal{C}$ of subsets of X is an algebra with the following property: If $A_n \in \mathcal{C}$, $n \in \mathbf{N}_+$ and $A_k \cap A_n = \phi$ if $k \neq n$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{C}$. Prove that $\mathcal{C}$ is a σ-algebra.

5. Let $(X, \mathcal{M})$ be a measurable space and $(\mu_k)_{k=1}^{\infty}$ a sequence of positive measures on $\mathcal{M}$ such that $\mu_1 \leq \mu_2 \leq \mu_3 \leq ...$. Prove that the set function

$$\mu(A) = \lim_{k \to \infty} \mu_k(A),\ A \in \mathcal{M}$$

is a positive measure.

6. Let $(X, \mathcal{M}, \mu)$ be a positive measure space. Show that

$$\mu(\cap_{k=1}^{n} A_k) \leq \sqrt[n]{\Pi_{k=1}^{n} \mu(A_k)}$$

for all $A_1, ..., A_n \in \mathcal{M}$.

7. Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space with $\mu(X) = \infty$. Show that for any $r \in [0, \infty[$ there is some $A \in \mathcal{M}$ with $r < \mu(A) < \infty$.

8. Show that the symmetric difference of sets is associative:

$$A\Delta(B\Delta C) = (A\Delta B)\Delta C.$$

9. $(X, \mathcal{M}, \mu)$ is a finite positive measure space. Prove that

$$\mid \mu(A) - \mu(B) \mid \leq \mu(A\Delta B).$$

10. Let $E = 2\mathbf{N}$. Prove that

$$c_{\mathbf{N}}(E\Delta A) = \infty$$

if A is a finite union of intervals.

11. Suppose $(X, \mathcal{P}(X), \mu)$ is a finite positive measure space such that $\mu(\{x\}) > 0$ for every $x \in X$. Set

$$d(A, B) = \mu(A\Delta B), \; A, B \in \mathcal{P}(X).$$

Prove that

$$d(A, B) = 0 \;\Leftrightarrow A = B,$$
$$d(A, B) = d(B, A)$$

and

$$d(A, B) \leq d(A, C) + d(C, B).$$

12. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space. Prove that

$$\mu(\cup_{i=1}^n A_i) \geq \Sigma_{i=1}^n \mu(A_i) - \Sigma_{1 \leq i < j \leq n} \mu(A_i \cap A_j)$$

for all $A_1, ..., A_n \in \mathcal{M}$ and integers $n \geq 2$.

1.2. Measure Determining Classes

Suppose μ and ν are probability measures defined on the same σ-algebra $\mathcal{M}$, which is generated by a class $\mathcal{E}$. If μ and ν agree on $\mathcal{E}$, is it then true that μ and ν agree on $\mathcal{M}$? The answer is in general no. To show this, let

$$X = \{1, 2, 3, 4\}$$

and

$$\mathcal{E} = \{\{1, 2\}, \{1, 3\}\}.$$

Then $\sigma(\mathcal{E}) = \mathcal{P}(X)$. If $\mu = \frac{1}{4}c_X$ and

$$\nu = \frac{1}{6}\delta_{X,1} + \frac{1}{3}\delta_{X,2} + \frac{1}{3}\delta_{X,3} + \frac{1}{6}\delta_{X,4}$$

then $\mu = \nu$ on $\mathcal{E}$ and $\mu \neq \nu$.

In this section we will prove a basic result on measure determining classes for σ-finite measures. In this context we will introduce so called π-systems and λ-systems, which will also be of great value later in connection with the construction of so called product measures in Chapter 3.

Definition 1.2.1. A class $\mathcal{G}$ of subsets of X is a π-system if $A \cap B \in \mathcal{G}$ for all $A, B \in \mathcal{G}$.

The class of all open n-cells in $\mathbf{R}^n$ is a π-system.

Definition 1.2.2. A class $\mathcal{D}$ of subsets of X is a λ-system if the following properties hold:

(a) $X \in \mathcal{D}$.

(b) If $A, B \in \mathcal{D}$ and $A \subseteq B$, then $B \setminus A \in \mathcal{D}$.

(c) If $(A_n)_{n\in\mathbf{N}_+}$ is a disjoint denumerable collection of members of the class $\mathcal{D}$, then $\cup_{n=1}^{\infty} A_n \in \mathcal{D}$.

Theorem 1.2.1. *If a class $\mathcal{M}$ is both a π-system and λ-system, then $\mathcal{M}$ is a σ-algebra.*

PROOF. If $A \in \mathcal{M}$, then $A^c = X \setminus A \in \mathcal{M}$ since $X \in \mathcal{M}$ and $\mathcal{M}$ is a λ-system. Moreover, if $(A_n)_{n\in\mathbf{N}_+}$ is a denumerable collection of members of $\mathcal{M}$,

$$A_1 \cup ... \cup A_n = (A_1^c \cap ... \cap A_n^c)^c \in \mathcal{M}$$

for each n, since $\mathcal{M}$ is a λ-system and a π-system. Let $(B_n)_{n=1}^{\infty}$ be the disjungation of $(A_n)_{n=1}^{\infty}$. Then $(B_n)_{n\in\mathbf{N}_+}$ is a disjoint denumerable collection of members of $\mathcal{M}$ and Definition 1.2.2(c) implies that $\cup_{n=1}^{\infty} A_n = \cup_{n=1}^{\infty} B_n \in \mathcal{M}$.

Theorem 1.2.2. *Let $\mathcal{G}$ be a π-system and $\mathcal{D}$ a λ-system such that $\mathcal{G} \subseteq \mathcal{D}$. Then $\sigma(\mathcal{G}) \subseteq \mathcal{D}$.*

PROOF. Let $\mathcal{M}$ be the intersection of all λ-systems containing $\mathcal{G}$. The class $\mathcal{M}$ is a λ-system and $\mathcal{G} \subseteq \mathcal{M} \subseteq \mathcal{D}$. In view of Theorem 1.2.1 $\mathcal{M}$ is a σ-algebra, if $\mathcal{M}$ is a π-system and in that case $\sigma(\mathcal{G}) \subseteq \mathcal{M}$. Thus the theorem follows if we show that $\mathcal{M}$ is a π-system.

Given $C \subseteq X$, denote by $\mathcal{D}_C$ be the class of all $D \subseteq X$ such that $D \cap C \in \mathcal{M}$.

CLAIM 1. If $C \in \mathcal{M}$, then $\mathcal{D}_C$ is a λ-system.

PROOF OF CLAIM 1. First $X \in \mathcal{D}_C$ since $X \cap C = C \in \mathcal{M}$. Moreover, if $A, B \in \mathcal{D}_C$ and $A \subseteq B$, then $A \cap C, B \cap C \in \mathcal{M}$ and

$$(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C) \in \mathcal{M}.$$

Accordingly from this, $B \setminus A \in \mathcal{D}_C$. Finally, if $(A_n)_{n\in\mathbf{N}_+}$ is a disjoint denumerable collection of members of $\mathcal{D}_C$, then $(A_n \cap C)_{n\in\mathbf{N}_+}$ is disjoint denumerable collection of members of $\mathcal{M}$ and

$$(\cup_{n\in\mathbf{N}_+} A_n) \cap C = \cup_{n\in\mathbf{N}_+} (A_n \cap C) \in \mathcal{M}.$$

Thus $\cup_{n\in\mathbf{N}_+} A_n \in \mathcal{D}_C$.

CLAIM 2. If $A \in \mathcal{G}$, then $\mathcal{M} \subseteq \mathcal{D}_A$.

PROOF OF CLAIM 2. If $B \in \mathcal{G}$, $A \cap B \in \mathcal{G} \subseteq \mathcal{M}$. Thus $B \in \mathcal{D}_A$. We have proved that $\mathcal{G} \subseteq \mathcal{D}_A$ and remembering that $\mathcal{M}$ is the intersection of all λ-systems containing $\mathcal{G}$ Claim 2 follows since $\mathcal{D}_A$ is a λ-system.

To complete the proof of Theorem 1.2.2, observe that $B \in \mathcal{D}_A$ if and only if $A \in \mathcal{D}_B$. By Claim 2, if $A \in \mathcal{G}$ and $B \in \mathcal{M}$, then $B \in \mathcal{D}_A$ that is $A \in \mathcal{D}_B$. Thus $\mathcal{G} \subseteq \mathcal{D}_B$ if $B \in \mathcal{M}$. Now the definition of $\mathcal{M}$ implies that $\mathcal{M} \subseteq \mathcal{D}_B$ if $B \in \mathcal{M}$. The proof is almost finished. In fact, if $A, B \in \mathcal{M}$ then $A \in \mathcal{D}_B$ that is $A \cap B \in \mathcal{M}$. Theorem 1.2.2 now follows from Theorem 1.2.1.

Theorem 1.2.3. *Let μ and ν be positive measures on $\mathcal{M} = \sigma(\mathcal{G})$, where $\mathcal{G}$ is a π-system, and suppose $\mu(A) = \nu(A)$ for every $A \in \mathcal{G}$.*

(a) *If μ and ν are probability measures, then $\mu = \nu$.*

(b) *Suppose there exist $E_n \in \mathcal{G}$, $n \in \mathbf{N}_+$, such that $X = \cup_{n=1}^{\infty} E_n$, $E_1 \subseteq E_2 \subseteq ...$, and*

$$\mu(E_n) = \nu(E_n) < \infty, \text{ all } n \in \mathbf{N}_+.$$

Then $\mu = \nu$.

PROOF. (a) Let

$$\mathcal{D} = \{A \in \mathcal{M};\ \ \mu(A) = \ \nu(A)\}.$$

It is immediate that $\mathcal{D}$ is a λ-system and Theorem 1.2.2 implies that $\mathcal{M} = \sigma(\mathcal{G}) \subseteq \mathcal{D}$ since $\mathcal{G} \subseteq \mathcal{D}$ and $\mathcal{G}$ is a π-system.

(b) If $\mu(E_n) = \nu(E_n) = 0$ for all all $n \in \mathbf{N}_+$, then

$$\mu(X) = \lim_{n\to\infty} \mu(E_n) = 0$$

and, in a similar way, $\nu(X) = 0$. Thus $\mu = \nu$. If $\mu(E_n) = \nu(E_n) > 0$, set

$$\mu_n(A) = \frac{1}{\mu(E_n)}\mu(A \cap E_n) \text{ and } \nu_n(A) = \frac{1}{\nu(E_n)}\nu(A \cap E_n)$$

for each $A \in \mathcal{M}$. By Part (a) $\mu_n = \nu_n$ and we get

$$\mu(A \cap E_n) = \nu(A \cap E_n)$$

for each $A \in \mathcal{M}$. Theorem 1.1.2(e) now proves that $\mu = \nu$.

Theorem 1.2.3 implies that there is at most one positive measure defined on $\mathcal{R}_n$ such that the measure of any open n-cell in $\mathbf{R}^n$ equals its volume.

Next suppose $f : X \to Y$ and let $A \subseteq X$ and $B \subseteq Y$. The image of A and the inverse image of B are

$$f(A) = \{y;\ y = f(x) \text{ for some } x \in A\}$$

and

$$f^{-1}(B) = \{x;\ f(x) \in B\}$$

respectively. Note that

$$f^{-1}(Y) = X$$

and

$$f^{-1}(Y \setminus B) = X \setminus f^{-1}(B).$$

Moreover, if $(A_i)_{i \in I}$ is a collection of subsets of X and $(B_i)_{i \in I}$ is a collection of subsets of Y

$$f(\cup_{i \in I} A_i) = \cup_{i \in I} f(A_i)$$

and

$$f^{-1}(\cup_{i \in I} B_i) = \cup_{i \in I} f^{-1}(B_i).$$

Given a class $\mathcal{E}$ of subsets of Y, set

$$f^{-1}(\mathcal{E}) = \left\{ f^{-1}(B);\ B \in \mathcal{E} \right\}.$$

If $(Y, \mathcal{N})$ is a measurable space, it follows that the class $f^{-1}(\mathcal{N})$ is a σ-algebra in X. If $(X, \mathcal{M})$ is a measurable space

$$\left\{ B \in \mathcal{P}(Y);\ f^{-1}(B) \in \mathcal{M} \right\}$$

is a σ-algebra in Y. Thus, given a class $\mathcal{E}$ of subsets of Y,

$$\sigma(f^{-1}(\mathcal{E})) = f^{-1}(\sigma(\mathcal{E})).$$

Definition 1.2.3. Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces. The function $f : X \to Y$ is said to be $(\mathcal{M}, \mathcal{N})$-measurable if $f^{-1}(\mathcal{N}) \subseteq \mathcal{M}$. If we say that $f : (X, \mathcal{M}) \to (Y, \mathcal{N})$ is measurable this means that $f : X \to Y$ is an $(\mathcal{M}, \mathcal{N})$-measurable function.

Theorem 1.2.4. *Let $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ be measurable spaces and suppose $\mathcal{E}$ generates $\mathcal{N}$. The function $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable if*

$$f^{-1}(\mathcal{E}) \subseteq \mathcal{M}.$$

PROOF. The assumptions yield

$$\sigma(f^{-1}(\mathcal{E})) \subseteq \mathcal{M}.$$

Since

$$\sigma(f^{-1}(\mathcal{E})) = f^{-1}(\sigma(\mathcal{E})) = f^{-1}(\mathcal{N})$$

we are done.

Corollary 1.2.1. *A function $f : X \to \mathbf{R}$ is $(\mathcal{M}, \mathcal{R})$-measurable if and only if the set $f^{-1}(]\alpha, \infty[) \in \mathcal{M}$ for all $\alpha \in \mathbf{R}$.*

If $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable and μ is a positive measure on $\mathcal{M}$, the equation

$$\nu(B) = \mu(f^{-1}(B)),\ B \in \mathcal{N}$$

defines a positive measure ν on $\mathcal{N}$. We will write $\nu = \mu f^{-1}$, $\nu = f(\mu)$ or $\nu = \mu_f$. The measure ν is called the image measure of μ under f and f is said to transport μ to ν. Two $(\mathcal{M}, \mathcal{N})$-measurable functions $f : X \to Y$ and $g : X \to Y$ are said to be μ-equimeasurable if $f(\mu) = g(\mu)$.

As an example, let $a \in \mathbf{R}^n$ and define $f(x) = x + a$ if $x \in \mathbf{R}^n$. If $B \subseteq \mathbf{R}^n$,

$$f^{-1}(B) = \{x;\ x + a \in B\} = B - a.$$

Thus $f^{-1}(B)$ is an open n-cell if B is, and Theorem 1.2.4 proves that f is $(\mathcal{R}_n, \mathcal{R}_n)$-measurable. Now, granted the existence of volume measure v_n, for every $B \in \mathcal{R}_n$ define

$$\mu(B) = f(v_n)(B) = v_n(B - a).$$

Then $\mu(B) = v_n(B)$ if B is an open n-cell and Theorem 1.2.3 implies that $\mu = v_n$. We have thus proved the following

Theorem 1.2.5. *For any $A \in \mathcal{R}_n$ and $x \in \mathbf{R}^n$*

$$A + x \in \mathcal{R}_n$$

and

$$v_n(A + x) = v_n(A).$$

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. A measurable function ξ defined on Ω is called a random variable and the image measure P_ξ is called the probability law of ξ. We sometimes write

$$\mathcal{L}(\xi) = P_\xi.$$

Here are two simple examples.

If the range of a random variable ξ consists of n points $S = \{s_1, ..., s_n\}$ $(n \geq 1)$ and $P_\xi = \frac{1}{n} c_S$, ξ is said to have a uniform distribution in S. Note that

$$P_\xi = \frac{1}{n} \Sigma_{k=1}^n \delta_{s_k}.$$

Suppose $\lambda > 0$ is a constant. If a random variable ξ has its range in $\mathbf{N}$ and

$$P_\xi = \Sigma_{n=0}^\infty \frac{\lambda^n}{n!} e^{-\lambda} \delta_n$$

then ξ is said to have a Poisson distribution with parameter λ.

Exercises

1. Let $f : X \to Y$, $A \subseteq X$, and $B \subseteq Y$. Show that

$$f(f^{-1}(B)) \subseteq B \text{ and } f^{-1}(f(A)) \supseteq A.$$

2. Let $(X, \mathcal{M})$ be a measurable space and suppose $A \subseteq X$. Show that the function χ_A is $(\mathcal{M}, \mathcal{R})$-measurable if and only if $A \in \mathcal{M}$.

3. Suppose $(X, \mathcal{M})$ is a measurable space and $f_n : X \to \mathbf{R}$, $n \in \mathbf{N}$, a sequence of $(\mathcal{M}, \mathcal{R})$-measurable functions such that

$$\lim_{n\to\infty} f_n(x) \text{ exists and } = f(x) \in \mathbf{R}$$

for each $x \in X$. Prove that f is $(\mathcal{M}, \mathcal{R})$-measurable.

4. Suppose $f : (X, \mathcal{M}) \to (Y, \mathcal{N})$ and $g : (Y, \mathcal{N}) \to (Z, \mathcal{S})$ are measurable. Prove that $g \circ f$ is $(\mathcal{M}, \mathcal{S})$-measurable.

5. Granted the existence of volume measure v_n, show that $v_n(rA) = r^n v_n(A)$ if $r \geq 0$ and $A \in \mathcal{R}$.

6. Let μ be the counting measure on $\mathbf{Z}^2$ and $f(x, y) = x$, $(x, y) \in \mathbf{Z}^2$. The measure μ is σ-finite. Prove that the image measure $f(\mu)$ is not σ-finite.

7. Let $\mu, \nu : \mathcal{R} \to [0, \infty]$ be two positive measures such that $\mu(I) = \nu(I) < \infty$ for each open subinterval of $\mathbf{R}$. Prove that $\mu = \nu$.

8. Suppose ξ has a Poisson distribution with parameter λ. Show that $P_\xi\left[2\mathbf{N}\right] = e^{-\lambda}\cosh\lambda$.

9. Find a λ-system which is not a σ-algebra.

1.3. Lebesgue Measure

Once the problem about the existence of volume measure is solved the existence of the so called Lebesgue measure is simple to establish as will be seen in this section. We start with some concepts of general interest.

If $(X, \mathcal{M}, \mu)$ is a positive measure space, the zero set $\mathcal{Z}_\mu$ of μ is, by definition, the set at all $A \in \mathcal{M}$ such that $\mu(A) = 0$. An element of $\mathcal{Z}_\mu$ is called a null set or μ-null set. If

$$(A \in \mathcal{Z}_\mu \text{ and } B \subseteq A) \Rightarrow B \in \mathcal{M}$$

the measure space $(X, \mathcal{M}, \mu)$ is said to be complete. In this case the measure μ is also said to be complete. The positive measure space $(X, \{\phi, X\}, \mu)$, where $X = \{0, 1\}$ and $\mu = 0$, is not complete since $X \in \mathcal{Z}_\mu$ and $\{0\} \notin \{\phi, X\}$.

Theorem 1.3.1 *If* $(E_n)_{n=1}^\infty$ *is a denumerable collection of members of* $\mathcal{Z}_\mu$ *then* $\cup_{n=1}^\infty E_n \in \mathcal{Z}_\mu$.

PROOF We have

$$0 \leq \mu(\cup_{n=1}^\infty E_n) \leq \Sigma_{n=1}^\infty \mu(E_n) = 0$$

which proves the result.

Granted the existence of linear measure v_1 it follows from Theorem 1.3.1 that $\mathbf{Q} \in \mathcal{Z}_{v_1}$ since $\mathbf{Q}$ is countable and $\{a\} \in \mathcal{Z}_{v_1}$ for each real number a.

Suppose $(X, \mathcal{M}, \mu)$ is an arbitrary positive measure space. It turns out that μ is the restriction to $\mathcal{M}$ of a complete measure. To see this suppose $\mathcal{M}^-$ is the class of all $E \subseteq X$ is such that there exist sets $A, B \in \mathcal{M}$ such that

$A \subseteq E \subseteq B$ and $B \setminus A \in \mathcal{Z}_\mu$. It is obvious that $X \in \mathcal{M}^-$ since $\mathcal{M} \subseteq \mathcal{M}^-$. If $E \in \mathcal{M}^-$, choose $A, B \in \mathcal{M}$ such that $A \subseteq E \subseteq B$ and $B \setminus A \in \mathcal{Z}_\mu$. Then $B^c \subseteq E^c \subseteq A^c$ and $A^c \setminus B^c = B \setminus A \in \mathcal{Z}_\mu$ and we conclude that $E^c \in \mathcal{M}^-$. If $(E_i)_{i=1}^\infty$ is a denumerable collection of members of $\mathcal{M}^-$, for each i there exist sets $A_i, B_i \in \mathcal{M}$ such that $A_i \subseteq E \subseteq B_i$ and $B_i \setminus A_i \in \mathcal{Z}_\mu$. But then

$$\cup_{i=1}^\infty A_i \subseteq \cup_{i=1}^\infty E_i \subseteq \cup_{i=1}^\infty B_i$$

where $\cup_{i=1}^\infty A_i, \cup_{i=1}^\infty B_i \in \mathcal{M}$. Moreover, $(\cup_{i=1}^\infty B_i) \setminus (\cup_{i=1}^\infty A_i) \in \mathcal{Z}_\mu$ since

$$(\cup_{i=1}^\infty B_i) \setminus (\cup_{i=1}^\infty A_i) \subseteq \cup_{i=1}^\infty (B_i \setminus A_i).$$

Thus $\cup_{i=1}^\infty E_i \in \mathcal{M}^-$ and $\mathcal{M}^-$ is a σ-algebra.

If $E \in \mathcal{M}$, suppose $A_i, B_i \in \mathcal{M}$ are such that $A_i \subseteq E \subseteq B_i$ and $B_i \setminus A_i \in \mathcal{Z}_\mu$ for $i = 1, 2$. Then for each i, $(B_1 \cap B_2) \setminus A_i \in \mathcal{Z}_\mu$ and

$$\mu(B_1 \cap B_2) = \mu((B_1 \cap B_2) \setminus A_i) + \mu(A_i) = \mu(A_i).$$

Thus the real numbers $\mu(A_1)$ and $\mu(A_2)$ are the same and we define $\bar{\mu}(E)$ to be equal to this common number. Note also that $\mu(B_1) = \bar{\mu}(E)$. It is plain that $\bar{\mu}(\phi) = 0$. If $(E_i)_{i=1}^\infty$ is a disjoint denumerable collection of members of $\mathcal{M}$, for each i there exist sets $A_i, B_i \in \mathcal{M}$ such that $A_i \subseteq E_i \subseteq B_i$ and $B_i \setminus A_i \in \mathcal{Z}_\mu$. From the above it follows that

$$\bar{\mu}(\cup_{i=1}^\infty E_i) = \mu(\cup_{i=1}^\infty A_i) = \Sigma_{n=1}^\infty \mu(A_i) = \Sigma_{n=1}^\infty \bar{\mu}(E_i).$$

We have proved that $\bar{\mu}$ is a positive measure on $\mathcal{M}^-$. If $E \in \mathcal{Z}_{\bar{\mu}}$ the definition of $\bar{\mu}$ shows that any set $A \subseteq E$ belongs to the σ-algebra $\mathcal{M}^-$. It follows that the measure $\bar{\mu}$ is complete and its restriction to $\mathcal{M}$ equals μ.

The measure $\bar{\mu}$ is called the completion of μ and $\mathcal{M}^-$ is called the completion of $\mathcal{M}$ with respect to μ.

Definition 1.3.1 The completion of volume measure v_n on $\mathbf{R}^n$ is called Lebesgue measure on $\mathbf{R}^n$ and is denoted by m_n. The completion of $\mathcal{R}_n$ with respect to v_n is called the Lebesgue σ-algebra in $\mathbf{R}^n$ and is denoted by $\mathcal{R}_n^-$. A member of the class $\mathcal{R}_n^-$ is called a Lebesgue measurable set in $\mathbf{R}^n$ or a Lebesgue set in $\mathbf{R}^n$. A function $f : \mathbf{R}^n \to \mathbf{R}$ is said to be Lebesgue measurable if it is $(\mathcal{R}_n^-, \mathcal{R})$-measurable. Below, m_1 is written m if this notation will not lead to misunderstanding. Furthermore, $\mathcal{R}_1^-$ is written $\mathcal{R}^-$.

Theorem 1.3.2. *Suppose $E \in \mathcal{R}_n^-$ and $x \in \mathbf{R}^n$. Then $E + x \in \mathcal{R}_n^-$ and $m_n(E + x) = m_n(E)$.*

PROOF. Choose $A, B \in \mathcal{R}_n$ such that $A \subseteq E \subseteq B$ and $B \setminus A \in \mathcal{Z}_{v_n}$. Then, by Theorem 1.2.5, $A + x, B + x \in \mathcal{R}_n$, $v_n(A + x) = v_n(A) = m_n(E)$, and $(A + x) \setminus (B + x) = (A \setminus B) + x \in \mathcal{Z}_{v_n}$. Since $A + x \subseteq E + x \subseteq B + x$ the theorem is proved.

The Lebesgue σ-algebra in $\mathbf{R}^n$ is very large and contains each set of interest in analysis and probability. In fact, in most cases, the σ-algebra $\mathcal{R}_n$ is sufficiently large but there are some exceptions. For example, if $f : \mathbf{R}^n \to \mathbf{R}^n$ is continuous and $A \in \mathcal{R}_n$, the image set $f(A)$ need not belong to the class $\mathcal{R}_n$ (see e.g. the Dudley book [D]). To prove the existence of a subset of the real line, which is not Lebesgue measurable we will use the so called Axiom of Choice.

Axiom of Choice. If $(A_i)_{i \in I}$ is a non-empty collection of non-empty sets, there exists a function $f : I \to \cup_{i \in I} A_i$ such that $f(i) \in A_i$ for every $i \in I$.

Let X and Y be sets. The set of all ordered pairs (x, y), where $x \in X$ and $y \in Y$ is denoted by $X \times Y$. An arbitrary subset R of $X \times Y$ is called a relation. If $(x, y) \in R$, we write $x \sim y$. A relation is said to be an equivalence relation on X if $X = Y$ and

(i) $x \sim x$ (reflexivity)
(ii) $x \sim y \Rightarrow y \sim x$ (symmetry)
(ii) $(x \sim y$ and $y \sim z) \Rightarrow x \sim z$ (transitivity)

The equivalence class $R(x) =_{def} \{y;\ y \sim x\}$. The definition of the equivalence relation $\sim$ implies the following:

(a) $x \in R(x)$

(b) $R(x) \cap R(y) \neq \phi \Rightarrow R(x) = R(y)$
(c) $\cup_{x \in X} R(x) = X$.

An equivalence relation leads to a partition of X into a disjoint collection of subsets of X.

Let $X = \left[-\frac{1}{2}, \frac{1}{2}\right]$ and define an equivalence relation for numbers x, y in X by stating that $x \sim y$ if $x - y$ is a rational number. By the Axiom of Choice it is possible to pick exactly one element from each equivalence class. Thus there exists a subset NL of X which contains exactly one element from each equivalence class.

If we assume that $NL \in \mathcal{R}^-$ we get a contradiction as follows. Let $(r_i)_{i=1}^{\infty}$ be an enumeration of the rational numbers in $[-1, 1]$. Then

$$X \subseteq \cup_{i=1}^{\infty} (r_i + NL)$$

and it follows from Theorem 1.3.1 that $r_i + NL \notin \mathcal{Z}_m$ for some i. Thus, by Theorem 1.3.2, $NL \notin \mathcal{Z}_m$.

Now assume $(r_i + NL) \cap (r_j + NL) \neq \phi$. Then there exist $a', a'' \in NL$ such that $r_i + a' = r_j + a''$ or $a' - a'' = r_j - r_i$. Hence $a' \sim a''$ and it follows that a' and a'' belong to the same equivalence class. But then $a' = a''$. Thus $r_i = r_j$ and we conclude that $(r_i + NL)_{i \in \mathbf{N}_+}$ is a disjoint enumeration of Lebesgue sets. Now, since

$$\cup_{i=1}^{\infty} (r_i + NL) \subseteq \left[-\frac{3}{2}, \frac{3}{2}\right]$$

it follows that

$$3 \geq m(\cup_{i=1}^{\infty} (r_i + NL)) = \Sigma_{n=1}^{\infty} m(NL).$$

But then $NL \in \mathcal{Z}_m$, which is a contradiction. Thus $NL \notin \mathcal{R}^-$.

In the early 1970' Solovay [S] proved that it is consistent with the usual axioms of Set Theory, excluding the Axiom of Choice, that every subset of $\mathbf{R}$ is Lebesgue measurable.

From the above we conclude that the Axiom of Choice implies the existence of a subset of the set of real numbers which does not belong to the class $\mathcal{R}$. Interestingly enough, such an example can be given without any use of the Axiom of Choice and follows naturally from the theory of analytic sets. The interested reader may consult the Dudley book $[D]$.

Exercises

1. $(X, \mathcal{M}, \mu)$ is a positive measure space. Prove or disprove: If $A \subseteq E \subseteq B$ and $\mu(A) = \mu(B)$ then E belongs to the domain of the completion $\bar{\mu}$.

2. Prove or disprove: If A and B are not Lebesgue measurable subsets of $\mathbf{R}$, then $A \cup B$ is not Lebesgue measurable.

3. Let $(X, \mathcal{M}, \mu)$ be a complete positive measure space and suppose $A, B \in \mathcal{M}$, where $B \setminus A$ is a μ-null set. Prove that $E \in \mathcal{M}$ if $A \subseteq E \subseteq B$ (stated otherwise $\mathcal{M}^- = \mathcal{M}$).

4. Suppose $E \subseteq \mathbf{R}$ and $E \notin \mathcal{R}^-$. Show there is an $\varepsilon > 0$ such that

$$m(B \setminus A) \geq \varepsilon$$

for all $A, B \in \mathcal{R}^-$ such that $A \subseteq E \subseteq B$.

5. Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space and $(Y, \mathcal{N})$ a measurable space. Furthermore, suppose $f : X \to Y$ is $(\mathcal{M}, \mathcal{N})$-measurable and let $\nu = \mu f^{-1}$, that is $\nu(B) = \mu(f^{-1}(B))$, $B \in \mathcal{N}$. Show that f is $(\mathcal{M}^-, \mathcal{N}^-)$-measurable, where $\mathcal{M}^-$ denotes the completion of $\mathcal{M}$ with respect to μ and $\mathcal{N}^-$ the completion of $\mathcal{N}$ with respect to ν.

1.4. Carathéodory's Theorem

In these notes we exhibit two famous approaches to Lebesgue measure. One is based on the Carathéodory Theorem, which we present in this section, and the other one, due to F. Riesz, is a representation theorem of positive linear functionals on spaces of continuous functions in terms of positive measures. The latter approach, is presented in Chapter 3. Both methods depend on topological concepts such as compactness.

Definition 1.4.1. A function $\theta : \mathcal{P}(X) \to [0, \infty]$ is said to be an outer measure if the following properties are satisfied:

(i) $\theta(\phi) = 0$.
(ii) $\theta(A) \leq \theta(B)$ if $A \subseteq B$.
(iii) for any denumerable collection $(A_n)_{n=1}^{\infty}$ of subsets of X

$$\theta(\cup_{n=1}^{\infty} A_n) \leq \Sigma_{n=1}^{\infty} \theta(A_n).$$

Since

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

an outer measure θ satisfies the inequality

$$\theta(E) \leq \theta(E \cap A) + \theta(E \cap A^c).$$

If θ is an outer measure on X we define $\mathcal{M}(\theta)$ as the set of all $A \subseteq X$ such that

$$\theta(E) = \theta(E \cap A) + \theta(E \cap A^c) \text{ for all } E \subseteq X$$

or, what amounts to the same thing,

$$\theta(E) \geq \theta(E \cap A) + \theta(E \cap A^c) \text{ for all } E \subseteq X.$$

The next theorem is one of the most important in measure theory.

Theorem 1.4.1. (Carathéodory's Theorem) *Suppose θ is an outer measure. The class $\mathcal{M}(\theta)$ is a σ-algebra and the restriction of θ to $\mathcal{M}(\theta)$ is a complete measure.*

PROOF. Clearly, $\phi \in \mathcal{M}(\theta)$ and $A^c \in \mathcal{M}(\theta)$ if $A \in \mathcal{M}(\theta)$. Moreover, if $A, B \in \mathcal{M}(\theta)$ and $E \subseteq X$,

$$\theta(E) = \theta(E \cap A) + \theta(E \cap A^c)$$

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

$$+\theta(E \cap A^c \cap B) + \theta(E \cap A^c \cap B^c).$$

But

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

and

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

and we get

$$\theta(E) \geq \theta(E \cap (A \cup B)) + \theta(E \cap (A \cup B)^c).$$

It follows that $A \cup B \in \mathcal{M}(\theta)$ and we have proved that the class $\mathcal{M}(\theta)$ is an algebra. Now if $A, B \in \mathcal{M}(\theta)$ are disjoint

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

and therefore the restriction of θ to $\mathcal{M}(\theta)$ is a content.

Next we prove that $\mathcal{M}(\theta)$ is a σ-algebra. Let $(A_i)_{i=1}^{\infty}$ be a disjoint denumerable collection of members of $\mathcal{M}(\theta)$ and set for each $n \in \mathbf{N}$

$$B_n = \cup_{1 \leq i \leq n} A_i \text{ and } B = \cup_{i=1}^{\infty} A_i$$

(here $B_0 = \phi$). Then for any $E \subseteq X$,

$$\theta(E \cap B_n) = \theta(E \cap B_n \cap A_n) + \theta(E \cap B_n \cap A_n^c)$$

$$= \theta(E \cap A_n) + \theta(E \cap B_{n-1})$$

and, by induction,

$$\theta(E \cap B_n) = \Sigma_{i=1}^{n} \theta(E \cap A_i).$$

But then

$$\theta(E) = \theta(E \cap B_n) + \theta(E \cap B_n^c)$$

$$\geq \Sigma_{i=1}^{n}\theta(E\cap A_i)+\theta(E\cap B^c)$$

and letting $n\to\infty$,

$$\theta(E)\geq \Sigma_{i=1}^{\infty}\theta(E\cap A_i)+\theta(E\cap B^c)$$

$$\geq \theta(\cup_{i=1}^{\infty}(E\cap A_i))+\theta(E\cap B^c)$$

$$=\theta(E\cap B)+\theta(E\cap B^c)\geq\theta(E).$$

All the inequalities in the last calculation must be equalities and we conclude that $B\in\mathcal{M}(\theta)$ and, choosing $E=B$, results in

$$\theta(B)=\Sigma_{i=1}^{\infty}\theta(A_i).$$

Thus $\mathcal{M}(\theta)$ is a σ-algebra and the restriction of θ to $\mathcal{M}(\theta)$ is a positive measure.

Finally we prove that the the restriction of θ to $\mathcal{M}(\theta)$ is a complete measure. Suppose $B\subseteq A\in\mathcal{M}(\theta)$ and $\theta(A)=0$. If $E\subseteq X$,

$$\theta(E)\leq\theta(E\cap B)+\theta(E\cap B^c)\leq\theta(E\cap B^c)\leq\theta(E)$$

and so $B\in\mathcal{M}(\theta)$. The theorem is proved.

Exercises

1. Suppose $\theta_i:\mathcal{P}(X)\to[0,\infty[\,,\ i=1,2$, are outer measures. Prove that $\theta=\max(\theta_1,\theta_2)$ is an outer measure.

2. Suppose $a,b\in\mathbf{R}$ and $a\neq b$. Set $\theta=\max(\delta_a,\delta_b)$. Prove that

$$\{a\}\,,\{b\}\notin\mathcal{M}(\theta).$$

1.5. Existence of Linear Measure

The purpose of this section is to show the existence of linear measure on $\mathbf{R}$ using the Carathéodory Theorem and a minimum of topology.

First let us recall the definition of infimum and supremum of a non-empty subset of the extended real line. Suppose A is a non-empty subset of $[-\infty,\infty] = \mathbf{R}\cup\{-\infty,\infty\}$. We define $-\infty \leq x$ and $x \leq \infty$ for all $x \in [-\infty,\infty]$. An element $b \in [-\infty,\infty]$ is called a majorant of A if $x \leq b$ for all $x \in A$ and a minorant if $x \geq b$ for all $x \in A$. The Supremum Axiom states that A possesses a least majorant, which is denoted by $\sup A$. From this follows that if A is non-empty, then A possesses a greatest minorant, which is denoted by $\inf A$. (Actually, the Supremum Axiom is a theorem in courses where time is spent on the definition of real numbers.)

Theorem 1.5.1. (The Heine-Borel Theorem; weak form) *Let $[a,b]$ be a closed bounded interval and $(U_i)_{i\in I}$ a collection of open sets such that*

$$\cup_{i\in I} U_i \supseteq [a,b]\,.$$

Then

$$\cup_{i\in J} U_i \supseteq [a,b]$$

for some finite subset J of I.

PROOF. Let A be the set of all $x \in [a,b]$ such that

$$\cup_{i\in J} U_i \supseteq [a,x]$$

for some finite subset J of I. Clearly, $a \in A$ since $a \in U_i$ for some i. Let $c = \sup A$. There exists an i_0 such that $c \in U_{i_0}$. Let $c \in \,]a_0,b_0[\, \subseteq U_{i_0}$, where $a_0 < b_0$. Furthermore, by the very definition of least upper bound, there exists a finite set J such that

$$\cup_{i\in J} U_i \supseteq [a,(a_0+c)/2]\,.$$

Hence

$$\cup_{i\in J\cup\{i_0\}} U_k \supseteq [a,(c+b_0)/2]$$

and it follows that $c \in A$ and $c = b$. The lemma is proved.

A subset K of $\mathbf{R}$ is called compact if for every family of open subsets U_i, $i \in I$, with $\cup_{i \in I} U_i \supseteq K$ we have $\cup_{i \in J} U_i \supseteq K$ for some finite subset J of I. The Heine-Borel Theorem shows that a closed bounded interval is compact.

If $x, y \in \mathbf{R}$ and $E,\ F \subseteq \mathbf{R}$, let

$$d(x, y) = \mid x - y \mid$$

be the distance between x and y, let

$$d(x, E) = \inf_{u \in E} d(x, u)$$

be the distance from x to E, and let

$$d(E, F) = \inf_{u \in E, v \in F} d(u, v)$$

be the distance between E and F (here the infimum of the emty set equals ∞). Note that for any $u \in E$,

$$d(x, u) \leq d(x, y) + d(y, u)$$

and, hence

$$d(x, E) \leq d(x, y) + d(y, u)$$

and

$$d(x, E) \leq d(x, y) + d(y, E).$$

By interchanging the roles of x and y and assuming that $E \neq \phi$, we get

$$\mid d(x, E) - d(y, E) \mid \leq d(x, y).$$

Note that if $F \subseteq \mathbf{R}$ is closed and $x \notin F$, then $d(x, F) > 0$.

An outer measure $\theta : \mathcal{P}(\mathbf{R}) \to [0, \infty]$ is called a metric outer measure if

$$\theta(A \cup B) = \theta(A) + \theta(B)$$

for all $A, B \in \mathcal{P}(\mathbf{R})$ such that $d(A, B) > 0$.

Theorem 1.5.2. *If $\theta : \mathcal{P}(\mathbf{R}) \to [0, \infty]$ is a metric outer measure, then $\mathcal{R} \subseteq \mathcal{M}(\theta)$.*

PROOF. Let $F \in \mathcal{P}(\mathbf{R})$ be closed. It is enough to show that $F \in \mathcal{M}(\theta)$. To this end we choose $E \subseteq X$ with $\theta(E) < \infty$ and prove that

$$\theta(E) \geq \theta(E \cap F) + \theta(E \cap F^c).$$

Let $n \geq 1$ be an integer and define

$$A_n = \left\{ x \in E \cap F^c;\ d(x, F) \geq \frac{1}{n} \right\}.$$

Note that $A_n \subseteq A_{n+1}$ and

$$E \cap F^c = \cup_{n=1}^{\infty} A_n.$$

Moreover, since θ is a metric outer measure

$$\theta(E) \geq \theta((E \cap F) \cup A_n) = \theta(E \cap F) + \theta(A_n)$$

and, hence, proving

$$\theta(E \cap F^c) = \lim_{n \to \infty} \theta(A_n)$$

we are done.

Let $B_n = A_{n+1} \cap A_n^c$. It is readily seen that

$$d(B_{n+1}, A_n) \geq \frac{1}{n(n+1)}$$

since if $x \in B_{n+1}$ and

$$d(x, y) < \frac{1}{n(n+1)}$$

then

$$d(y, F) \leq d(y, x) + d(x, F) < \frac{1}{n(n+1)} + \frac{1}{n+1} = \frac{1}{n}.$$

Now

$$\theta(A_{2k+1}) \geq \theta(B_{2k} \cup A_{2k-1}) = \theta(B_{2k}) + \theta(A_{2k-1})$$
$$\geq ... \geq \Sigma_{i=1}^{k} \theta(B_{2i})$$

and in a similar way

$$\theta(A_{2k}) \geq \Sigma_{i=1}^{k} \theta(B_{2i-1}).$$

But $\theta(A_n) \leq \theta(E) < \infty$ and we conclude that

$$\Sigma_{i=1}^{\infty} \theta(B_i) < \infty.$$

We now use that

$$E \cap F^c = A_n \cup (\cup_{i=n}^{\infty} B_i)$$

to obtain

$$\theta(E \cap F^c) \leq \theta(A_n) + \Sigma_{i=n}^{\infty} \theta(B_i).$$

Now, since $\theta(E \cap F^c) \geq \theta(A_n)$,

$$\theta(E \cap F^c) = \lim_{n \to \infty} \theta(A_n)$$

and the theorem is proved.

PROOF OF THEOREM 1.1.1 IN ONE DIMENSION. Suppose $\delta > 0$. If $A \subseteq \mathbf{R}$, define

$$\theta_\delta(A) = \inf \Sigma_{k=1}^{\infty} l(I_k)$$

the infimum being taken over all open intervals I_k with $l(I_k) < \delta$ such that

$$A \subseteq \cup_{k=1}^{\infty} I_k.$$

Obviously, $\theta_\delta(\phi) = 0$ and $\theta_\delta(A) \leq \theta_\delta(B)$ if $A \subseteq B$. Suppose $(A_n)_{n=1}^{\infty}$ is a denumerable collection of subsets of $\mathbf{R}$ and let $\varepsilon > 0$. For each n there exist intervals $I_{kn}, k \in \mathbf{N}_+$, such that $l(I_{kn}) < \delta$,

$$A_n \subseteq \cup_{k=1}^{\infty} I_{kn}$$

and

$$\Sigma_{k=1}^{\infty} l(I_{kn}) \leq \theta_\delta(A_n) + \varepsilon 2^{-n}.$$

Then

$$A =_{def} \cup_{n=1}^{\infty} A_n \subseteq \cup_{k,n=1}^{\infty} I_{kn}$$

and

$$\Sigma_{k,n=1}^{\infty} l(I_{kn}) \leq \Sigma_{n=1}^{\infty} \theta_\delta(A_n) + \varepsilon.$$

Thus

$$\theta_\delta(A) \leq \Sigma_{n=1}^{\infty} \theta_\delta(A_n) + \varepsilon$$

and, since $\varepsilon > 0$ is arbitrary,

$$\theta_\delta(A) \leq \Sigma_{n=1}^{\infty}\theta_\delta(A_n).$$

It follows that θ_δ is an outer measure.

If I is an open interval it is simple to see that

$$\theta_\delta(I) \leq l(I).$$

To prove the reverse inequality, choose a closed bounded interval $J \subseteq I$. Now, if

$$I \subseteq \cup_{k=1}^{\infty} I_k$$

where each I_k is an open interval of $l(I_k) < \delta$, it follows from the Heine-Borel Theorem that

$$J \subseteq \cup_{k=1}^{n} I_k$$

for some n. Hence

$$l(J) \leq \Sigma_{k=1}^{n} l(I_k) \leq \Sigma_{k=1}^{\infty} l(I_k)$$

and it follows that

$$l(J) \leq \theta_\delta(I)$$

and, accordingly from this,

$$l(I) \leq \theta_\delta(I).$$

Thus, if I is an open interval, then

$$\theta_\delta(I) = l(I).$$

Note that $\theta_{\delta_1} \geq \theta_{\delta_2}$ if $0 < \delta_1 \leq \delta_2$. We define

$$\theta_0(A) = \lim_{\delta \to 0} \theta_\delta(A) \text{ if } A \subseteq \mathbf{R}.$$

It obvious that θ_0 is an outer measure such that $\theta_0(I) = l(I)$, if I is an open interval.

To complete the proof we show that θ_0 is a metric outer measure. To this end let $A, B \subseteq \mathbf{R}$ and $d(A, B) > 0$. Suppose $0 < \delta < d(A, B)$ and

$$A \cup B \subseteq \cup_{k=1}^{\infty} I_k$$

where each I_k is an open interval with $l(I_k) < \delta$. Let

$$\alpha = \{k;\ I_k \cap A \neq \phi\}$$

and

$$\beta = \{k;\ I_k \cap B \neq \phi\}.$$

Then $\alpha \cap \beta = \phi$,

$$A \subseteq \cup_{k\in\alpha} I_k$$

and

$$B \subseteq \cup_{k\in\beta} I_k$$

and it follows that

$$\Sigma_{k=1}^{\infty} l(I_k) \geq \Sigma_{k\in\alpha} l(I_k) + \Sigma_{k\in\beta} l(I_k)$$

$$\geq \theta_\delta(A) + \theta_\delta(B).$$

Thus

$$\theta_\delta(A \cup B) \geq \theta_\delta(A) + \theta_\delta(B)$$

and by letting $\delta \to 0$ we have

$$\theta_0(A \cup B) \geq \theta_0(A) + \theta_0(B)$$

and

$$\theta_0(A \cup B) = \theta_0(A) + \theta_0(B).$$

Finally by applying the Carathéodory Theorem and Theorem 1.5.2 it follows that the restriction of θ_0 to $\mathcal{R}$ equals v_1.

We end this section with some additional results of great interest.

Theorem 1.5.3. *For any $\delta > 0$, $\theta_\delta = \theta_0$. Moreover, if $A \subseteq \mathbf{R}$*

$$\theta_0(A) = \inf \Sigma_{k=1}^{\infty} l(I_k)$$

the infimum being taken over all open intervals I_k, $k \in \mathbf{N}_+$, such that $\cup_{k=1}^{\infty} I_k \supseteq A$.

PROOF. It follows from the definition of θ_0 that $\theta_\delta \leq \theta_0$. To prove the reverse inequality let $A \subseteq \mathbf{R}$ and choose open intervals I_k, $k \in \mathbf{N}_+$, such that $\cup_{k=1}^{\infty} I_k \supseteq A$. Then

$$\theta_0(A) \leq \theta_0(\cup_{k=1}^{\infty} I_k) \leq \Sigma_{k=1}^{\infty} \theta_0(I_k)$$

$$= \Sigma_{k=1}^{\infty} l(I_k).$$

Hence

$$\theta_0(A) \leq \inf \Sigma_{k=1}^{\infty} l(I_k)$$

the infimum being taken over all open intervals I_k, $k \in \mathbf{N}_+$, such that $\cup_{k=1}^{\infty} I_k \supseteq A$. Thus $\theta_0(A) \leq \theta_\delta(A)$, which completes the proof of Theorem 1.5.3.

Theorem 1.5.4. *If* $A \subseteq \mathbf{R}$,

$$\theta_0(A) = \inf_{\substack{U \supseteq A \\ U \text{ open}}} \theta_0(U).$$

Moreover, if $A \in \mathcal{M}(\theta_0)$,

$$\theta_0(A) = \sup_{\substack{K \subseteq A \\ K \text{ closed bounded}}} \theta_0(K).$$

PROOF. If $A \subseteq U$, $\theta_0(A) \leq \theta_0(U)$. Hence

$$\theta_0(A) \leq \inf_{\substack{U \supseteq A \\ U \text{ open}}} \theta_0(U).$$

Next let $\varepsilon > 0$ be fixed and choose open intervals I_k, $k \in \mathbf{N}_+$, such that $\cup_{k=1}^{\infty} I_k \supseteq A$ and

$$\Sigma_{k=1}^{\infty} l(I_k) \leq \theta_0(A) + \varepsilon$$

(here observe that it may happen that $\theta_0(A) = \infty$). Then the set $U =_{def} \cup_{k=1}^{\infty} I_k$ is open and

$$\theta_0(U) \leq \Sigma_{k=1}^{\infty} \theta_0(I_k) = \Sigma_{k=1}^{\infty} l(I_k) \leq \theta_0(A) + \varepsilon.$$

Thus

$$\inf_{\substack{U \supseteq A \\ U \text{ open}}} \theta_0(U) \leq \theta_0(A)$$

and we have proved that

$$\theta_0(A) = \inf_{\substack{U \supseteq A \\ U \text{ open}}} \theta_0(U).$$

If $K \subseteq A$, $\theta_0(K) \leq \theta_0(A)$ and, accordingly from this,

$$\sup_{\substack{K \subseteq A \\ K \text{ closed bounded}}} \theta_0(K) \leq \theta_0(A).$$

To prove the reverse inequality we first assume that $A \in \mathcal{M}(\theta_0)$ is bounded. Let $\varepsilon > 0$ be fixed and suppose J is a closed bounded interval containing A. Then we know from the first part of Theorem 1.5.4 already proved that there exists an open set $U \supseteq J \smallsetminus A$ such that

$$\theta_0(U) < \theta_0(J \smallsetminus A) + \varepsilon.$$

But then

$$\theta_0(J) \leq \theta_0(J \smallsetminus U) + \theta_0(U) < \theta_0(J \smallsetminus U) + \theta_0(J \smallsetminus A) + \varepsilon$$

and it follows that

$$\theta_0(A) - \varepsilon < \theta_0(J \setminus U).$$

Since $J \smallsetminus U$ is a closed bounded set contained in A we conclude that

$$\theta_0(A) \leq \sup_{\substack{K \subseteq A \\ K \text{ closed bounded}}} \theta_0(K).$$

If $A \in \mathcal{M}(\theta_0)$ let $A_n = A \cap [-n, n]$, $n \in \mathbf{N}_+$. Then given $\varepsilon > 0$ and $n \in \mathbf{N}_+$, let K_n be a closed bounded subset of A_n such that $\theta_0(K_n) > \theta_0(A_n) - \varepsilon$. Clearly, there is no loss of generality to assume that $K_1 \subseteq K_2 \subseteq K_3 \subseteq ...$ and by letting n tend to plus infinity we get

$$\lim_{n \to \infty} \theta_0(K_n) \geq \theta_0(A) - \varepsilon.$$

Hence

$$\theta_0(A) = \sup_{\substack{K \subseteq A \\ K \text{ compact}}} \theta_0(K).$$

and Theorem 1.5.4 is completely proved.

Theorem 1.5.5. *Lebesgue measure m_1 equals the restriction of θ_0 to $\mathcal{M}(\theta_0)$.*

PROOF. Recall that linear measure v_1 equals the restriction of θ_0 to $\mathcal{R}$ and $m_1 = \bar{v}_1$. First suppose $E \in \mathcal{R}^-$ and choose $A, B \in \mathcal{R}$ such that $A \subseteq E \subseteq B$ and $B \setminus A \in \mathcal{Z}_{v_1}$. But then $\theta_0(E \setminus A) = 0$ and $E = A \cup (E \setminus A) \in \mathcal{M}(\theta_0)$ since the Carathéodory Theorem gives us a complete measure. Hence $m_1(E) = v_1(A) = \theta_0(E)$.

Conversely suppose $E \in \mathcal{M}(\theta_0)$. We will prove that $E \in \mathcal{R}^-$ and $m_1(E) = \theta_0(E)$. First assume that E is bounded. Then for each positive integer n there exist open $U_n \supseteq E$ and closed bounded $K_n \subseteq E$ such that

$$\theta_0(U_n) < \theta_0(E) + 2^{-n}$$

and

$$\theta_0(K_n) > \theta_0(E) - 2^{-n}.$$

The definitions yield $A = \cup_1^\infty K_n$, $B = \cap_1^\infty U_n \in \mathcal{R}$ and

$$\theta_0(E) = \theta_0(A) = \theta_0(B) = v_1(A) = v_1(B) = m_1(E).$$

It follows that $E \in \mathcal{R}^-$ and $\theta_0(E) = m_1(E)$.

In the general case set $E_n = E \cap [-n, n]$, $n \in \mathbf{N}_+$. Then from the above $E_n \in \mathcal{R}^-$ and $\theta_0(E_n) = m_1(E_n)$ for each n and Theorem 1.5.5 follows by letting n go to infinity.

The Carathéodory Theorem can be used to show the existence of volume measure on $\mathbf{R}^n$ but we do not go into this here since its existence follows by several other means below. By passing, let us note that the Carathéodory Theorem is very efficient to prove the existence of so called Haussdorff measures (see e.g. $[F]$), which are of great interest in Geometric Measure Theory.

Exercises

1. Prove that a subset K of $\mathbf{R}$ is compact if and only if K is closed and bounded.

2. Suppose $A \in \mathcal{R}^-$ and $m(A) < \infty$. Set $f(x) = m(A \cap]-\infty, x])$, $x \in \mathbf{R}$. Prove that f is continuous.

3. Suppose $A \in \mathcal{Z}_m$ and $B = \{x^3; x \in A\}$. Prove that $B \in \mathcal{Z}_m$.

4. Let A be the set of all real numbers x such that

$$| x - \frac{p}{q} | \leq \frac{1}{q^3}$$

for infinitely many pairs of positive integers p and q. Prove that $A \in \mathcal{Z}_m$.

5. Let $I_1, ..., I_n$ be open subintervals of $\mathbf{R}$ such that

$$\mathbf{Q} \cap [0, 1] \subseteq \cup_{k=1}^{n} I_k.$$

Prove that $\Sigma_{k=1}^{n} m(I_k) \geq 1$.

6. If $E \in \mathcal{R}^-$ and $m(E) > 0$, for every $\alpha \in]0, 1[$ there is an interval I such that $m(E \cap I) > \alpha m(I)$. (Hint: $m(E) = \inf \Sigma_{k=1}^{\infty} m(I_k)$, where the infimum is taken over all intervals such that $\cup_{k=1}^{\infty} I_k \supseteq E$.)

7. If $E \in \mathcal{R}^-$ and $m(E) > 0$, then the set $E - E = \{x - y; x, y \in E\}$ contains an open non-empty interval centred at 0.(Hint: Take an interval I with $m(E \cap I) \geq \frac{3}{4} m(I)$. Set $\varepsilon = \frac{1}{2} m(I)$. If $| x | \leq \varepsilon$, then $(E \cap I) \cap (x + (E \cap I)) \neq \phi$.)

8. Let μ be the restriction of the positive measure $\Sigma_{k=1}^{\infty} \delta_{\mathbf{R}, \frac{1}{k}}$ to $\mathcal{R}$. Prove that

$$\inf_{\substack{U \supseteq A \\ U \text{ open}}} \mu(U) > \mu(A)$$

if $A = \{0\}$.

1.6. Positive Measures Induced by Increasing Right Continuous Functions

Suppose $F : \mathbf{R} \to [0, \infty[$ is a right continuous increasing function such that

$$\lim_{x \to -\infty} F(x) = 0.$$

Set

$$L = \lim_{x \to \infty} F(x).$$

We will prove that there exists a unique positive measure $\mu : \mathcal{R} \to [0, L]$ such that

$$\mu(]-\infty, x]) = F(x), \ x \in \mathbf{R}.$$

The special case $L = 0$ is trivial so let us assume $L > 0$ and introduce

$$H(y) = \inf \{x \in \mathbf{R}; F(x) \geq y\}, \ 0 < y < L.$$

The definition implies that the function H increases.

Suppose a is a fixed real number. We claim that

$$\{y \in]0, L[\,; \ H(y) \leq a\} =]0, F(a)] \cap]0, L[\,.$$

To prove this first suppose that $y \in]0, L[$ and $H(y) \leq a$. Then to each positive integer n, there is an $x_n \in [H(y), H(y) + 2^{-n}[$ such that $F(x_n) \geq y$. Then $x_n \to H(y)$ as $n \to \infty$ and we obtain that $F(H(y)) \geq y$ since F is right continuous. Thus, remembering that F increases, $F(a) \geq y$. On the other hand, if $0 < y < L$ and $0 < y \leq F(a)$, then, by the very definition of $H(y)$, $H(y) \leq a$.

We now define

$$\mu = H(v_{1|]0,L[})$$

and get

$$\mu(]-\infty, x]) = F(x), \ x \in \mathbf{R}.$$

The uniqueness follows at once from Theorem 1.2.3. Note that the measure μ is a probability measure if $L = 1$.

Exercises

1. Suppose $F : \mathbf{R} \to \mathbf{R}$ is a right continuous increasing function. Prove that there is a unique positive measure μ on $\mathcal{R}$ such that

$$\mu(]a, x]) = F(x) - F(a), \text{ if } a, x \in \mathbf{R} \text{ and } a < x.$$

2. Suppose $F : \mathbf{R} \to \mathbf{R}$ is an increasing function. Prove that the set of all discontinuity points of F is at most denumerable. (Hint: Assume first that F is bounded and prove that the set of all points $x \in \mathbf{R}$ such that $F(x+) - F(x-) > \varepsilon$ is finite for every $\varepsilon > 0$.)

3. Suppose μ is a σ-finite positive measure on $\mathcal{R}$. Prove that the set of all $x \in \mathbf{R}$ such that $\mu(\{x\}) > 0$ is at most denumerable.

4. Suppose μ is a σ-finite positive measure on $\mathcal{R}_n$. Prove that there is an at most denumerable set of hyperplanes of the type

$$x_k = c \qquad (k = 1, ..., n,\ c \in \mathbf{R})$$

with positive μ-measure.

5. Construct an increasing function $f : \mathbf{R} \to \mathbf{R}$ such that the set of discontinuity points of f equals $\mathbf{Q}$.

CHAPTER 2

INTEGRATION

Introduction

In this chapter Lebesgue integration in abstract positive measure spaces is introduced. A series of famous theorems and lemmas will be proved.

2.1. Integration of Functions with Values in $[0,\infty]$

Recall that $[0,\infty] = [0,\infty[\cup \{\infty\}$. A subinterval of $[0,\infty]$ is defined in the natural way. We denote by $\mathcal{R}_{0,\infty}$ the σ-algebra generated by all subintervals of $[0,\infty]$. The class of all intervals of the type $]\alpha,\infty]$, $0 \leq \alpha < \infty$, (or of the type $[\alpha,\infty]$, $0 \leq \alpha < \infty$) generates the σ-algebra $\mathcal{R}_{0,\infty}$ and we get the following

Theorem 2.1.1. *Let $(X,\mathcal{M})$ be a measurable space and suppose $f : X \to [0,\infty]$.*

(a) *The function f is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable if $f^{-1}(]\alpha,\infty]) \in \mathcal{M}$ for every $0 \leq \alpha < \infty$.*

(b) *The function f is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable if $f^{-1}([\alpha,\infty]) \in \mathcal{M}$ for every $0 \leq \alpha < \infty$.*

Note that the set $\{f > \alpha\} \in \mathcal{M}$ for all real α if f is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable.

If $f, g : X \to [0,\infty]$ are $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable, then $\min(f,g)$, $\max(f,g)$, and $f+g$ are $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable, since, for each $\alpha \in [0,\infty[$,

$$\min(f,g) \geq \alpha \Leftrightarrow (f \geq \alpha \text{ and } g \geq \alpha)$$

$$\max(f,g) \geq \alpha \Leftrightarrow (f \geq \alpha \text{ or } g \geq \alpha)$$

and

$$\{f+g>\alpha\}=\bigcup_{q\in\mathbf{Q}}(\{f>\alpha-q\}\cap\{g>q\}).$$

Given functions $f_n : X \to [0,\infty]$, $n = 1, 2, ...,$ $f = \sup_{n\geq 1} f_n$ is defined by the equation

$$f(x)=\sup\{f_n(x);\ n=1,2,...\}.$$

Note that

$$f^{-1}(]\alpha,\infty])=\cup_{n=1}^{\infty}f_n^{-1}(]\alpha,\infty])$$

for every real $\alpha \geq 0$ and, accordingly from this, the function $\sup_{n\geq 1} f_n$ is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable if each f_n is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable. Moreover, $f = \inf_{n\geq 1} f_n$ is given by

$$f(x)=\inf\{f_n(x);\ n=1,2,...\}.$$

Since

$$f^{-1}([0,\alpha[)=\cup_{n=1}^{\infty}f_n^{-1}([0,\alpha[)$$

for every real $\alpha \geq 0$ we conclude that the function $f = \inf_{n\geq 1} f_n$ is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable if each f_n is $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable.

Below we write

$$f_n\uparrow f$$

if f_n, $n = 1, 2, ...,$ and f are functions from X into $[0,\infty]$ such that $f_n \leq f_{n+1}$ for each n and $f_n(x) \to f(x)$ for each $x \in X$ as $n \to \infty$.

An $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable function $\varphi : X \to [0,\infty]$ is called a simple measurable function if $\varphi(X)$ is a finite subset of $[0,\infty[$. If it is neccessary to be more precise, we say that φ is a simple $\mathcal{M}$-measurable function.

Theorem 2.1.2. *Let $f : X \to [0,\infty]$ be $(\mathcal{M},\mathcal{R}_{0,\infty})$-measurable. There exist simple measurable functions φ_n, $n \in \mathbf{N}_+$, on X such that $\varphi_n \uparrow f$.*

PROOF. Given $n \in \mathbf{N}_+$, set

$$E_{in}=f^{-1}(\left[\frac{i-1}{2^n},\frac{i}{2^n}\right[),\ i\in\mathbf{N}_+$$

and

$$\rho_n = \sum_{i=1}^{\infty} \frac{i-1}{2^n} \chi_{E_{in}} + \infty \chi_{f^{-1}(\{\infty\})}.$$

It is obvious that $\rho_n \leq f$ and that $\rho_n \leq \rho_{n+1}$. Now set $\varphi_n = \min(n, \rho_n)$ and we are done.

Let $(X, \mathcal{M}, \mu)$ be a positive measure space and $\varphi : X \to [0, \infty[$ a simple measurable function. If $\alpha_1, ..., \alpha_n$ are the distinct values of the simple function φ, and if $E_i = \varphi^{-1}(\{\alpha_i\})$, $i = 1, ..., n$, then

$$\varphi = \Sigma_{i=1}^n \alpha_i \chi_{E_i}.$$

Furthermore, if $A \in \mathcal{M}$ we define

$$\nu(A) = \int_A \varphi d\mu = \Sigma_{k=1}^n \alpha_i \mu(E_i \cap A) = \Sigma_{k=1}^n \alpha_i \mu^{E_i}(A).$$

Clearly, ν is a positive measure since each term in the right side is a positive measure as a function of A. Note that

$$\int_A \alpha \varphi d\mu = \alpha \int_A \varphi d\mu \text{ if } 0 \leq \alpha < \infty$$

and

$$\int_A \varrho d\mu = a\mu(A)$$

if $a \in [0, \infty[$ and ϱ is a simple measurable function such that $\varrho = a$ on A.

If ψ is another simple measurable function and $\varphi \leq \psi$,

$$\int_A \varphi d\mu \leq \int_A \psi d\mu.$$

To see this, let $\beta_1, ..., \beta_p$ be the distinct values of ψ and $F_j = \psi^{-1}(\{\beta_j\})$, $j = 1, ..., p$. Now, putting $B_{ij} = E_i \cap F_j$,

$$\int_A \varphi d\mu = \nu(\cup_{ij}(A \cap B_{ij}))$$

$$= \Sigma_{ij} \nu(A \cap B_{ij}) = \Sigma_{ij} \int_{A \cap B_{ij}} \varphi d\mu = \Sigma_{ij} \int_{A \cap B_{ij}} \alpha_i d\mu$$

$$\leq \Sigma_{ij} \int_{A\cap B_{ij}} \beta_j d\mu = \int_A \psi d\mu.$$

In a similar way one proves that

$$\int_A (\varphi + \psi) d\mu = \int_A \varphi d\mu + \int_A \psi d\mu.$$

From the above it follows that

$$\int_A \varphi \chi_A d\mu = \int_A \Sigma_{i=1}^n \alpha_i \chi_{E_i \cap A} d\mu$$

$$= \Sigma_{i=1}^n \alpha_i \int_A \chi_{E_i \cap A} d\mu = \Sigma_{i=1}^n \alpha_i \mu(E_i \cap A)$$

and

$$\int_A \varphi \chi_A d\mu = \int_A \varphi d\mu.$$

If $f : X \to [0, \infty]$ is an $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable function and $A \in \mathcal{M}$, we define

$$\int_A f d\mu = \sup \left\{ \int_A \varphi d\mu;\ 0 \leq \varphi \leq f,\ \varphi \text{ simple measurable} \right\}$$

$$= \sup \left\{ \int_A \varphi d\mu;\ 0 \leq \varphi \leq f,\ \varphi \text{ simple measurable and } \varphi = 0 \text{ on } A^c \right\}.$$

The left member in this equation is called the Lebesgue integral of f over A with respect to the measure μ. Sometimes we also speek of the μ-integral of f over A. The two definitions of the μ-integral of a simple measurable function $\varphi : X \to [0, \infty[$ over A agree.

From now on in this section, an $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable function $f : X \to [0, \infty]$ is simply called measurable.

The following properties are immediate consequences of the definitions. The functions and sets occurring in the equations are assumed to be measurable.

(a) If f, $g \geq 0$ and $f \leq g$ on A, then $\int_A f d\mu \leq \int_A g d\mu$.

(b) $\int_A f d\mu = \int_X \chi_A f d\mu$.

(c) If $f \geq 0$ and $\alpha \in [0, \infty[$, then $\int_A \alpha f d\mu = \alpha \int_A f d\mu$.

(d) $\int_A f d\mu = 0$ if $f = 0$ and $\mu(A) = \infty$.

(e) $\int_A f d\mu = 0$ if $f = \infty$ and $\mu(A) = 0$.

If $f : X \to [0, \infty]$ is measurable and $0 < \alpha < \infty$, then $f \geq \alpha\chi_{f^{-1}([\alpha,\infty])} = \alpha\chi_{\{f\geq\alpha\}}$and

$$\int_X f d\mu \geq \int_X \alpha\chi_{\{f\geq\alpha\}} d\mu = \alpha \int_X \chi_{\{f\geq\alpha\}} d\mu.$$

This proves the so called Markov Inequality

$$\mu(f \geq \alpha) \leq \frac{1}{\alpha} \int_X f d\mu$$

where we write $\mu(f \geq \alpha)$ instead of the more precise expression $\mu(\{f \geq \alpha\})$.

Example 2.1.1. Suppose $f : X \to [0, \infty]$ is measurable and

$$\int_X f d\mu < \infty.$$

We claim that

$$\{f = \infty\} = f^{-1}(\{\infty\}) \in \mathcal{Z}_\mu.$$

To prove this we use the Markov Inequality and have

$$\mu(f = \infty) \leq \mu(f \geq \alpha) \leq \frac{1}{\alpha} \int_X f d\mu$$

for each $\alpha \in \,]0, \infty[$. Thus $\mu(f = \infty) = 0$.

Example 2.1.2. Suppose $f : X \to [0, \infty]$ is measurable and

$$\int_X f d\mu = 0.$$

We claim that

$$\{f > 0\} = f^{-1}(]0, \infty]) \in \mathcal{Z}_\mu.$$

To see this, note that

$$f^{-1}(]0, \infty]) = \cup_{n=1}^{\infty} f^{-1}(\left]\frac{1}{n}, \infty\right])$$

Furthermore, for every fixed $n \in \mathbf{N}_+$, the Markov Inequality yields

$$\mu(f > \frac{1}{n}) \leq n \int_X f d\mu = 0$$

and we get $\{f > 0\} \in \mathcal{Z}_\mu$ since a countable union of null sets is a null set.

We now come to one of the most important results in the theory.

Theorem 2.1.3. **(Monotone Convergence Theorem)** *Let $f_n : X \to [0, \infty]$, $n = 1, 2, 3, ...,$ be a sequence of measurable functions and suppose that $f_n \uparrow f$,* that is $0 \leq f_1 \leq f_2 \leq ...$ *and*

$$f_n(x) \to f(x) \text{ as } n \to \infty, \text{ for every } x \in X.$$

Then f is measurable and

$$\int_X f_n d\mu \to \int_X f d\mu \text{ as } n \to \infty.$$

PROOF. The function f is measurable since $f = \sup_{n \geq 1} f_n$.

The inequalities $f_n \leq f_{n+1} \leq f$ yield $\int_X f_n d\mu \leq \int_X f_{n+1} d\mu \leq \int_X f d\mu$ and we conclude that there exists an $\alpha \in [0, \infty]$ such that

$$\int_X f_n d\mu \to \alpha \text{ as } n \to \infty$$

and

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

To prove the reverse inequality, let φ be any simple measurable function such that $0 \leq \varphi \leq f$, let $0 < \theta < 1$ be a constant, and define, for fixed $n \in \mathbf{N}_+$,

$$A_n = \{x \in X;\ f_n(x) \geq \theta\varphi(x)\}.$$

If $\alpha_1, ..., \alpha_p$ are the distinct values of φ,

$$A_n = \cup_{k=1}^p (\{x \in X;\ f_n(x) \geq \theta\alpha_k\} \cap \{\varphi = \alpha_k\})$$

and it follows that A_n is measurable. Clearly, $A_1 \subseteq A_2 \subseteq ...$. Moreover, if $f(x) = 0$, then $x \in A_1$ and if $f(x) > 0$, then $\theta\varphi(x) < f(x)$ and $x \in A_n$ for all sufficiently large n. Thus $\cup_{n=1}^\infty A_n = X$. Now

$$\alpha \geq \int_{A_n} f_n d\mu \geq \theta \int_{A_n} \varphi d\mu$$

and we get

$$\alpha \geq \theta \int_X \varphi d\mu$$

since the map $A \to \int_A \varphi d\mu$ is a positive measure on $\mathcal{M}$. By letting $\theta \uparrow 1$,

$$\alpha \geq \int_X \varphi d\mu$$

and, hence

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

The theorem follows.

Theorem 2.1.4. (a) *Let $f, g : X \to [0, \infty]$ be measurable functions. Then*

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

(b) **(Beppo Levi's Theorem)** *If* $f_k : X \to [0, \infty]$, $k = 1, 2, ...$ *are measurable,*

$$\int_X \Sigma_{k=1}^{\infty} f_k d\mu = \Sigma_{k=1}^{\infty} \int_X f_k d\mu$$

PROOF. (a) Let $(\varphi_n)_{n=1}^{\infty}$ and $(\psi_n)_{n=1}^{\infty}$ be sequences of simple and measurable functions such that $0 \leq \varphi_n \uparrow f$ and $0 \leq \psi_n \uparrow g$. We proved above that

$$\int_X (\varphi_n + \psi_n) d\mu = \int_X \varphi_n d\mu + \int_X \psi_n d\mu$$

and, by letting $n \to \infty$, Part (a) follows from the Monotone Convergence Theorem.

(b) Part (a) and induction imply that

$$\int_X \Sigma_{k=1}^{n} f_k d\mu = \Sigma_{k=1}^{n} \int_X f_k d\mu$$

and the result follows from monotone convergence.

Theorem 2.1.5. *Suppose* $w : X \to [0, \infty]$ *is a measurable function and define*

$$\nu(A) = \int_A w d\mu, \ A \in \mathcal{M}.$$

Then ν *is a positive measure and*

$$\int_A f d\nu = \int_A f w d\mu, \ A \in \mathcal{M}$$

for every measurable function $f : X \to [0, \infty]$.

PROOF. Clearly, $\nu(\phi) = 0$. Suppose $(E_k)_{k=1}^{\infty}$ is a disjoint denumerable collection of members of $\mathcal{M}$ and set $E = \cup_{k=1}^{\infty} E_k$. Then

$$\nu(\cup_{k=1}^{\infty} E_k) = \int_E w d\mu = \int_X \chi_E w d\mu = \int_X \Sigma_{k=1}^{\infty} \chi_{E_k} w d\mu$$

where, by the Beppo Levi Theorem, the right member equals

$$\Sigma_{k=1}^{\infty} \int_X \chi_{E_k} w d\mu = \Sigma_{k=1}^{\infty} \int_{E_k} w d\mu = \Sigma_{k=1}^{\infty} \nu(E_k).$$

This proves that ν is a positive measure.

Let $A \in \mathcal{M}$. To prove the last part in Theorem 2.1.5 we introduce the class $\mathcal{C}$ of all measurable functions $f : X \to [0, \infty]$ such that

$$\int_A f d\nu = \int_A f w d\mu.$$

The indicator function of a measurable set belongs to $\mathcal{C}$ and from this we conclude that every simple measurable function belongs to $\mathcal{C}$. Furthermore, if $f_n \in \mathcal{C}$, $n \in \mathbf{N}$, and $f_n \uparrow f$, the Monotone Convergence Theorem proves that $f \in \mathcal{C}$. Thus in view of Theorem 2.1.2 the class $\mathcal{C}$ contains every measurable function $f : X \to [0, \infty]$. This completes the proof of Theorem 2.1.5.

The measure ν in Theorem 2.1.5 is written

$$\nu = w\mu$$

or

$$d\nu = w d\mu.$$

Let $(\alpha_n)_{n=1}^{\infty}$ be a sequence in $[-\infty, \infty]$. First put $\beta_k = \inf \{\alpha_k, \alpha_{k+1}, \alpha_{k+2}, ...\}$ and $\gamma = \sup \{\beta_1, \beta_2, \beta_3, ..\} = \lim_{n\to\infty} \beta_n$. We call γ the lower limit of $(\alpha_n)_{n=1}^{\infty}$ and write

$$\gamma = \liminf_{n\to\infty} \alpha_n.$$

Note that

$$\gamma = \lim_{n\to\infty} \alpha_n$$

if the limit exists. Now put $\beta_k = \sup \{\alpha_k, \alpha_{k+1}, \alpha_{k+2}, ...\}$ and $\gamma = \inf \{\beta_1, \beta_2, \beta_3, ..\} = \lim_{n\to\infty} \beta_n$. We call γ the upper limit of $(\alpha_n)_{n=1}^{\infty}$ and write

$$\gamma = \limsup_{n\to\infty} \alpha_n.$$

Note that

$$\gamma = \lim_{n\to\infty} \alpha_n$$

if the limit exists.

Given measurable functions $f_n : X \to [0, \infty]$, $n = 1, 2, ...$, the function $\liminf_{n\to\infty} f_n$ is measurable. In particular, if

$$f(x) = \lim_{n\to\infty} f_n(x)$$

exists for every $x \in X$, then f is measurable.

Theorem 2.1.6. (Fatou's Lemma) *If* $f_n : X \to [0, \infty]$, $n = 1, 2, ...,$ *are measurable*

$$\int_X \liminf_{n\to\infty} f_n d\mu \leq \liminf_{n\to\infty} \int_X f_n d\mu.$$

PROOF. Introduce

$$g_k = \inf_{n\geq k} f_n.$$

The definition gives that $g_k \uparrow \liminf_{n\to\infty} f_n$ and, moreover,

$$\int_X g_k d\mu \leq \int_X f_n d\mu, \ n \geq k$$

and

$$\int_X g_k d\mu \leq \inf_{n\geq k} \int_X f_n d\mu.$$

The Fatou Lemma now follows by monotone convergence.

Below we often write

$$\int_E f(x) d\mu(x)$$

instead of

$$\int_E f d\mu.$$

Example 2.1.3. Suppose $a \in \mathbf{R}$ and $f : (\mathbf{R}, \mathcal{R}^-) \to ([0, \infty], \mathcal{R}_{0,\infty})$ is measurable. We claim that

$$\int_{\mathbf{R}} f(x+a) dm(x) = \int_{\mathbf{R}} f(x) dm(x).$$

First if $f = \chi_A$, where $A \in \mathcal{R}^-$,

$$\int_{\mathbf{R}} f(x+a)dm(x) = \int_{\mathbf{R}} \chi_{A-a}(x)dm(x) = m(A-a) =$$

$$m(A) = \int_{\mathbf{R}} f(x)dm(x).$$

Next it is clear that the relation we want to prove is true for simple measurable functions and finally, we use the Lebesgue Dominated Convergence Theorem to deduce the general case.

Exercises

1. Suppose $f_n : X \to [0,\infty]$, $n = 1, 2, ...$, are measurable and

$$\Sigma_{k=1}^{\infty} \mu(f_n > 1) < \infty.$$

Prove that

$$\left\{ \limsup_{n\to\infty} f_n > 1 \right\} \in \mathcal{Z}_\mu \ .$$

2. Set $f_n = n^2 \chi_{\left[0,\frac{1}{n}\right]}$, $n \in \mathbf{N}_+$. Prove that

$$\int_{\mathbf{R}} \liminf_{n\to\infty} f_n dm = 0 < \infty = \liminf_{n\to\infty} \int_{\mathbf{R}} f_n dm$$

(the inequality in the Fatou Lemma may be strict).

3. Suppose $f : (\mathbf{R},\mathcal{R}^-) \to ([0,\infty], \mathcal{R}_{0,\infty})$ is measurable and set

$$g(x) = \Sigma_{k=1}^{\infty} f(x+k), \ \ x \in \mathbf{R}.$$

Show that

$$\int_{\mathbf{R}} gdm < \infty \text{ if and only if } \{f > 0\} \in \mathcal{Z}_m.$$

4. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and $f : X \to [0, \infty]$ an $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable function such that

$$f(X) \subseteq \mathbf{N}$$

and

$$\int_X f d\mu < \infty.$$

For every $t \geq 0$, set

$$F(t) = \mu(f > t) \text{ and } G(t) = \mu(f \geq t).$$

Prove that

$$\int_X f d\mu = \Sigma_{n=0}^{\infty} F(n) = \Sigma_{n=1}^{\infty} G(n).$$

2.2. Integration of Functions with Arbitrary Sign

As usual suppose $(X, \mathcal{M}, \mu)$ is a positive measure space. In this section when we speak of a measurable function $f : X \to \mathbf{R}$ it is understood that f is an $(\mathcal{M}, \mathcal{R})$-measurable function, if not otherwise stated. If $f, g : X \to \mathbf{R}$ are measurable, the sum $f + g$ is measurable since

$$\{f + g > \alpha\} = \bigcup_{q \in \mathbf{Q}} (\{f > \alpha - q\} \cap \{g > q\})$$

for each real α. Besides the function $-f$ and the difference $f - g$ are measurable. It follows that a function $f : X \to \mathbf{R}$ is measurable if and only if the functions $f^+ = \max(0, f)$ and $f^- = \max(0, -f)$ are measurable since $f = f^+ - f^-$.

We write $f \in \mathcal{L}^1(\mu)$ if $f : X \to \mathbf{R}$ is measurable and

$$\int_X \mid f \mid d\mu < \infty$$

and in this case we define

$$\int_X f d\mu = \int_X f^+ d\mu - \int_X f^- d\mu.$$

Note that

$$| \int_X f d\mu | \leq \int_X | f | d\mu$$

since $| f |= f^+ + f^-$. Moreover, if $E \in \mathcal{M}$ we define

$$\int_E f d\mu = \int_E f^+ d\mu - \int_E f^- d\mu$$

and it follows that

$$\int_E f d\mu = \int_X \chi_E f d\mu.$$

Note that

$$\int_E f d\mu = 0 \text{ if } \mu(E) = 0.$$

Sometimes we write

$$\int_E f(x) d\mu(x)$$

instead of

$$\int_E f d\mu.$$

If $f, g \in \mathcal{L}^1(\mu)$, setting $h = f + g$,

$$\int_X | h | d\mu \leq \int_X | f | d\mu + \int_X | g | d\mu < \infty$$

and it follows that $h + g \in \mathcal{L}^1(\mu)$. Moreover,

$$h^+ - h^- = f^+ - f^- + g^+ - g^-$$

and the equation

$$h^+ + f^- + g^- = f^+ + g^+ + h^-$$

gives

$$\int_X h^+ d\mu + \int_X f^- d\mu + \int_X g^- d\mu = \int_X f^+ d\mu + \int_X g^+ d\mu + \int_X h^- d\mu.$$

Thus

$$\int_X h d\mu = \int_X f d\mu + \int_X g d\mu.$$

Moreover,

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu$$

for each real α. The case $\alpha \geq 0$ follows from (c) in Section 2.1. The case $\alpha = -1$ is also simple since $(-f)^+ = f^-$ and $(-f)^- = f^+$.

Theorem 2.2.1. (Lebesgue's Dominated Convergence Theorem) S*uppose* $f_n : X \to \mathbf{R}$, $n = 1, 2, ...,$ *are measurable and*

$$f(x) = \lim_{n\to\infty} f_n(x)$$

exists for every $x \in X$. *Moreover, suppose there exists a function* $g \in \mathcal{L}^1(\mu)$ *such that*

$$\mid f_n(x) \mid \leq g(x), \text{ all } x \in X \text{ and } n \in \mathbf{N}_+.$$

Then $f \in \mathcal{L}^1(\mu)$,

$$\lim_{n\to\infty} \int_X \mid f_n - f \mid d\mu = 0$$

and

$$\lim_{n\to\infty} \int_X f_n d\mu = \int_X f d\mu$$

Proof. Since $\mid f \mid \leq g$, the function f is real-valued and measurable since f^+ and f^- are measurable. Note here that

$$f^{\pm}(x) = \lim_{n\to\infty} f_n^{\pm}(x), \text{all } x \in X.$$

We now apply the Fatous Lemma to the functions $2g- \mid f_n - f \mid, n = 1, 2, ...,$ and have

$$\int_X 2g d\mu \leq \liminf_{n\to\infty} \int_X (2g- \mid f_n - f \mid) d\mu$$

$$= \int_X 2g d\mu - \limsup_{n\to\infty} \int_X \mid f_n - f \mid d\mu.$$

But $\int_X 2g d\mu$ is finite and we get

$$\lim_{n\to\infty} \int_X \mid f_n - f \mid d\mu = 0.$$

Since

$$| \int_X f_n d\mu - \int_X f d\mu |=| \int_X (f - f_n) d\mu |\leq \int_X | f - f_n | d\mu$$

the last part in Theorem 2.2.1 follows from the first part. The theorem is proved.

Example 2.2.1. Suppose $f :]a, b[\times X \to \mathbf{R}$ is a function such that $f(t, \cdot) \in \mathcal{L}^1(\mu)$ for each $t \in]a, b[$ and, moreover, assume $\frac{\partial f}{\partial t}$ exists and

$$| \frac{\partial f}{\partial t}(t, x) |\leq g(x) \text{ for all } (t, x) \in]a, b[\times X$$

where $g \in \mathcal{L}^1(\mu)$. Set

$$F(t) = \int_X f(t, x) d\mu(x) \text{ if } t \in]a, b[\, .$$

We claim that F is differentiable and

$$F'(t) = \int_X \frac{\partial f}{\partial t}(t, x) d\mu(x).$$

To see this let $t_* \in]a, b[$ be fixed and choose a sequence $(t_n)_{n=1}^{\infty}$ in $]a, b[\setminus \{t_*\}$ which converges to t_*. Define

$$h_n(x) = \frac{f(t_n, x) - f(t_*, x)}{t_n - t_*} \text{ if } x \in X.$$

Here each h_n is measurable and

$$\lim_{n \to \infty} h_n(x) = \frac{\partial f}{\partial t}(t_*, x) \text{ for all } x \in X.$$

Furthermore, for each fixed n and x there is a $\tau_{n,x} \in]t_n, t_*[$ such that $h_n(x) = \frac{\partial f}{\partial t}(\tau_{n,x}, x)$ and we conclude that $| h_n(x) |\leq g(x)$ for every $x \in X$. Since

$$\frac{F(t_n) - F(t_*)}{t_n - t_*} = \int_X h_n(x) d\mu(x)$$

the claim above now follows from the Lebesgue Dominated Convergence Theorem.

Suppose $S(x)$ is a statement, which depends on $x \in X$. We will say that $S(x)$ holds almost (or μ-almost) everywhere if there exists an $N \in \mathcal{Z}_\mu$ such that $S(x)$ holds at every point of $X \setminus N$. In this case we write "S holds a.e. " or "S holds a.e. $[\mu]$". Sometimes we prefer to write "$S(x)$ holds a.e." or "$S(x)$ holds a.e. $[\mu]$". If the underlying measure space is a probability space, we often say "almost surely" instead of almost everywhere. The term "almost surely" is abbreviated a.s.

Suppose $f : X \to \mathbf{R}$, is an $(\mathcal{M}, \mathcal{R})$-measurable functions and $g : X \to \mathbf{R}$. If $f = g$ a.e. $[\mu]$ there exists an $N \in \mathcal{Z}_\mu$ such that $f(x) = g(x)$ for every $x \in X \setminus N$. We claim that g is $(\mathcal{M}^-, \mathcal{R})$-measurable. To see this let $\alpha \in \mathbf{R}$ and use that

$$\{g > \alpha\} = [\{f > \alpha\} \cap (X \setminus N)] \cup [\{g > \alpha\} \cap N]\,.$$

Now if we define

$$A = \{f > \alpha\} \cap (X \setminus N)$$

the set $A \in \mathcal{M}$ and

$$A \subseteq \{g > \alpha\} \subseteq A \cup N.$$

Accordingly from this $\{g > \alpha\} \in \mathcal{M}^-$ and g is $(\mathcal{M}^-, \mathcal{R})$-measurable since α is an arbitrary real number.

Next suppose $f_n : X \to \mathbf{R}$, $n \in \mathbf{N}_+$, is a sequence of $(\mathcal{M}, \mathcal{R})$-measurable functions and $f : X \to \mathbf{R}$ a function. Recall if

$$\lim_{n\to\infty} f_n(x) = f(x),\ \text{all } x \in X$$

then f is $(\mathcal{M}, \mathcal{R})$-measurable since

$$\{f > \alpha\} = \cup_{k,l \in \mathbf{N}_+} \cap_{n \geq k} \left\{f_n > \alpha + l^{-1}\right\},\ \text{all } \alpha \in \mathbf{R}.$$

If we only assume that

$$\lim_{n\to\infty} f_n(x) = f(x),\ \text{a.e. } [\mu]$$

then f need not be $(\mathcal{M}, \mathcal{R})$-measurable but f is $(\mathcal{M}^-, \mathcal{R})$-measurable. To see this suppose $N \in \mathcal{Z}_\mu$ and

$$\lim_{n\to\infty} f_n(x) = f(x),\ \text{all } x \in X \setminus N.$$

Then

$$\lim_{n\to\infty}\chi_{X\setminus N}(x)f_n(x)=\chi_{X\setminus N}(x)f(x)$$

and it follows that the function $\chi_{X\setminus N}f$ is $(\mathcal{M},\mathcal{R})$-measurable. Since $f=\chi_{X\setminus N}f$ a.e. $[\mu]$ it follows that f is $(\mathcal{M}^-,\mathcal{R})$-measurable. The next example shows that f need not be $(\mathcal{M},\mathcal{R})$-measurable.

Example 2.2.2. Let $X=\{0,1,2\}$, $\mathcal{M}=\{\phi,\{0\},\{1,2\},X\}$, and $\mu(A)=\chi_A(0)$, $A\in\mathcal{M}$. Set $f_n=\chi_{\{1,2\}}$, $n\in\mathbf{N}_+$, and $f(x)=x$, $x\in X$. Then each f_n is $(\mathcal{M},\mathcal{R})$-measurable and

$$\lim_{n\to\infty}f_n(x)=f(x)\text{ a.e. }[\mu]$$

since

$$\left\{x\in X;\ \lim_{n\to\infty}f_n(x)=f(x)\right\}=\{0,1\}$$

and $N=\{1,2\}$ is a μ-null set. The function f is not $(\mathcal{M},\mathcal{R})$-measurable.

Suppose $f,g\in\mathcal{L}^1(\mu)$. The functions f and g are equal almost everywhere with respect to μ if and only if $\{f\neq g\}\in\mathcal{Z}_\mu$. This is easily seen to be an equivalence relation and the set of all equivalence classes is denoted by $L^1(\mu)$. Moreover, if $f=g$ a.e. $[\mu]$, then

$$\int_X fd\mu=\int_X gd\mu$$

since

$$\int_X fd\mu=\int_{\{f=g\}}fd\mu+\int_{\{f\neq g\}}fd\mu=\int_{\{f=g\}}fd\mu=\int_{\{f=g\}}gd\mu$$

and, in a similar way,

$$\int_X gd\mu=\int_{\{f=g\}}gd\mu.$$

Below we consider the elements of $L^1(\mu)$ as members of $\mathcal{L}^1(\mu)$ and two members of $L^1(\mu)$ are identified if they are equal a.e. $[\mu]$. From this convention it is straight-forward to define $f+g$ and αf for all $f, g \in L^1(\mu)$ and $\alpha \in \mathbf{R}$. Moreover, we get

$$\int_X (f+g)d\mu = \int_X f d\mu + \int_X g d\mu \text{ if } f, g \in L^1(\mu)$$

and

$$\int_X \alpha f d\mu = \alpha \int_X f d\mu \text{ if } f \in L^1(\mu) \text{ and } \alpha \in \mathbf{R}.$$

Next we give two theorems where exceptional null sets enter. The first one is a mild variant of Theorem 2.2.1 and needs no proof.

Theorem 2.2.2. *Suppose $(X, \mathcal{M}, \mu)$ is a positive complete measure space and let $f_n : X \to \mathbf{R}$, $n \in \mathbf{N}_+$, be measurable functions such that*

$$\sup_{n \in \mathbf{N}_+} \mid f_n(x) \mid \leq g(x) \text{ a.e. } [\mu]$$

where $g \in L^1(\mu)$. Moreover, suppose $f : X \to \mathbf{R}$ is a function and

$$f(x) = \lim_{n \to \infty} f_n(x) \text{ a.e. } [\mu].$$

T*hen,* $f \in L^1(\mu)$,

$$\lim_{n \to \infty} \int_X \mid f_n - f \mid d\mu = 0$$

and

$$\lim_{n \to \infty} \int_X f_n d\mu = \int_X f d\mu.$$

Theorem 2.2.3. *Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space.*

(a) *If $f : (X, \mathcal{M}^-) \to ([0, \infty], \mathcal{R}_{0,\infty})$ is measurable there exists a measurable function $g : (X, \mathcal{M}) \to ([0, \infty], \mathcal{R}_{0,\infty})$ such that $f = g$* a.e. $[\mu]$.

(b) *If $f : (X, \mathcal{M}^-) \to (\mathbf{R}, \mathcal{R})$ is measurable there exists a measurable function $g : (X, \mathcal{M}) \to (\mathbf{R}, \mathcal{R})$ such that $f = g$* a.e. $[\mu]$.

PROOF. Since $f = f^+ - f^-$ it is enough to prove Part (a). There exist simple $\mathcal{M}^-$-measurable functions φ_n , $n \in \mathbf{N}_+$, such that $0 \leq \varphi_n \uparrow f$. For each fixed

n suppose $\alpha_{1n}, ..., \alpha_{k_n n}$ are the distinct values of φ_n and choose for each fixed $i = 1, ..., k_n$ a set $A_{in} \subseteq \varphi_n^{-1}(\{\alpha_{in}\})$ such that $A_{in} \in \mathcal{M}$ and $\varphi_n^{-1}(\alpha_{in}) \setminus A_{in} \in \mathcal{Z}_{\bar{\mu}}$. Set

$$\psi_n = \Sigma_{i=1}^{k_n} \alpha_{in} \chi_{A_{in}}.$$

Clearly $\psi_n(x) \uparrow f(x)$ if $x \in E =_{def} \cap_{n=1}^{\infty} (\cup_{i=1}^{k_n} A_{in})$ and $\mu(X \setminus E) = 0$. We now define $g(x) = f(x),$ if $x \in E$, and $g(x) = 0$ if $x \in X \setminus E$. The theorem is proved.

Exercises

1. Suppose f and g are real-valued measurable functions. Prove that f^2 and fg are measurable functions.

2. Suppose $f \in L^1(\mu)$. Prove that

$$\lim_{\alpha \to \infty} \int_{|f| \geq \alpha} | f | \, d\mu = 0.$$

(Here $\int_{|f| \geq \alpha}$ means $\int_{\{|f| \geq \alpha\}}$.)

3. Suppose $f \in L^1(\mu)$. Prove that to each $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\int_E | f | \, d\mu < \varepsilon$$

whenever $\mu(E) < \delta$.

4. Let $(f_n)_{n=1}^{\infty}$ be a sequence of $(\mathcal{M}, \mathcal{R})$-measurable functions. Prove that the set of all $x \in \mathbf{R}$ such that the sequence $(f_n(x))_{n=1}^{\infty}$ converges to a real limit belongs to $\mathcal{M}$.

5. Let $(X, \mathcal{M}, \mathcal{R})$ be a positive measure space such that $\mu(A) = 0$ or ∞ for every $A \in \mathcal{M}$. Show that $f \in L^1(\mu)$ if and only if $f(x) = 0$ a.e. $[\mu]$.

6. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and suppose f and g are non-negative measurable functions such that

$$\int_A f d\mu = \int_A g d\mu, \text{ all } A \in \mathcal{M}.$$

(a) Prove that $f = g$ a.e. $[\mu]$ if μ is σ-finite.
(b) Prove that the conclusion in Part (a) may fail if μ is not σ-finite.

7. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space and suppose the functions $f_n : X \to \mathbf{R}$, $n = 1, 2, ...$, are measurable. Show that there is a sequence $(\alpha_n)_{n=1}^{\infty}$ of positive real numbers such that

$$\lim_{n\to\infty} \alpha_n f_n = 0 \text{ a.e. } [\mu].$$

8. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and let $f_n : X \to \mathbf{R}$, $n = 1, 2, ...$, be a sequence in $L^1(\mu)$ which converges to f a.e. $[\mu]$ as $n \to \infty$. Suppose $f \in L^1(\mu)$ and

$$\lim_{n\to\infty} \int_X \mid f_n \mid d\mu = \int_X \mid f \mid d\mu.$$

Show that

$$\lim_{n\to\infty} \int_X \mid f_n - f \mid d\mu = 0.$$

9. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space and suppose $f \in L^1(\mu)$ is a bounded function such that

$$\int_X f^2 d\mu = \int_X f^3 d\mu = \int_X f^4 d\mu.$$

Prove that $f = \chi_A$ for an appropriate $A \in \mathcal{M}$.

10. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space and $f : X \to \mathbf{R}$ a measurable function. Prove that $f \in L^1(\mu)$ if and only if

$$\Sigma_{k=1}^{\infty} \mu(\mid f \mid \geq k) < \infty.$$

11. Suppose $f \in L^1(m)$. Prove that the series $\Sigma_{k=-\infty}^{\infty} f(x+k)$ converges for m-almost all x.

12. a) Suppose $f : \mathbf{R} \to [0, \infty[$ is Lebesgue measurable and $\int_{\mathbf{R}} f dm < \infty$. Prove that

$$\lim_{\alpha \to \infty} \alpha m(f \geq \alpha) = 0.$$

b) Find a Lebesgue measurable function $f : \mathbf{R} \to [0, \infty[$ such that $f \notin L^1(m)$, $m(f > 0) < \infty$, and

$$\lim_{\alpha \to \infty} \alpha m(f \geq \alpha) = 0.$$

2.3 Comparison of Riemann and Lebesgue Integrals

In this section we will show that the Lebesgue integral is a natural generalization of the Riemann integral. For short, the discussion is restricted to a closed and bounded interval.

Let $[a, b]$ be a closed and bounded interval and suppose $f : [a, b] \to \mathbf{R}$ is a bounded function. For any partition

$$\Delta : a = x_0 < x_1 < ... < x_n = b$$

of $[a, b]$ define

$$S_\Delta f = \Sigma_{i=1}^{n} (\sup_{]x_{i-1}, x_i]} f)(x_i - x_{i-1})$$

and

$$s_\Delta f = \Sigma_{k=1}^{n} (\inf_{]x_{i-1}, x_i]} f)(x_i - x_{i-1}).$$

The function f is Riemann integrable if

$$\inf_{\Delta} S_\Delta f = \sup_{\Delta} s_\Delta f$$

and the Riemann integral $\int_a^b f(x) dx$ is, by definition, equal to this common value.

Below an $((\mathcal{R}^-)_{[a,b]}, \mathcal{R})$-measurable function is simply called Lebesgue measurable. Furthermore, we write m instead of $m_{|[a,b]}$.

Theorem 2.3.1. *A bounded function $f : [a,b] \to \mathbf{R}$ is Riemann integrable if and only if the set of discontinuity points of f is a Lebesgue null set. Moreover, if the set of discontinuity points of f is a Lebesgue null set, then f* is *Lebesgue measurable and*

$$\int_a^b f(x)dx = \int_{[a,b]} f dm.$$

PROOF. A partition $\Delta' : a = x'_0 < x'_1 < ... < x'_{n'} = b$ is a refinement of a partition $\Delta : a = x_0 < x_1 < ... < x_n = b$ if each x_k is equal to some x'_l and in this case we write $\Delta \prec \Delta'$. The definitions give $S_\Delta f \geq S_{\Delta'} f$ and $s_\Delta f \leq s_{\Delta'} f$ if $\Delta \prec \Delta'$. We define, $\text{mesh}(\Delta) = \max_{1 \leq i \leq n}(x_i - x_{i-1})$.

First suppose f is Riemann integrable. For each partition Δ let

$$G_\Delta = f(a)\chi_{\{a\}} + \Sigma_{i=1}^n (\sup_{]x_{i-1},x_i]} f)\chi_{]x_{i-1},x_i]}$$

and

$$g_\Delta = f(a)\chi_{\{a\}} + \Sigma_{i=1}^n (\inf_{]x_{i-1},x_i]} f)\chi_{]x_{i-1},x_i]}$$

and note that

$$\int_{[a,b]} G_\Delta dm = S_\Delta f$$

and

$$\int_{[a,b]} g_\Delta dm = s_\Delta f.$$

Suppose Δ_k, $k = 1, 2, ...$, is a sequence of partitions such that $\Delta_k \prec \Delta_{k+1}$,

$$S_{\Delta_k} f \downarrow \int_a^b f(x)dx$$

and

$$s_{\Delta_k} f \uparrow \int_a^b f(x)dx$$

as $k \to \infty$. Let $G = \lim_{k\to\infty} G_{\Delta_k}$ and $g = \lim_{k\to\infty} g_{\Delta_k}$. Then G and g are $(\mathcal{R}_{[a,b]}, \mathcal{R})$-measurable, $g \leq f \leq G$, and by dominated convergence

$$\int_{[a,b]} G dm = \int_{[a,b]} g dm = \int_a^b f(x) dx.$$

But then

$$\int_{[a,b]} (G - g) dm = 0$$

so that $G = g$ a.e. $[m]$ and therefore $G = f$ a.e. $[m]$. In particular, f is Lebesgue measurable and

$$\int_a^b f(x) dx = \int_{[a,b]} f dm.$$

Set

$$N = \{x;\; g(x) < f(x) \text{ or } f(x) < G(x)\}.$$

We proved above that $m(N) = 0$. Let M be the union of all those points which belong to some partition Δ_k. Clearly, $m(M) = 0$ since M is denumerable. We claim that f is continuous off $N \cup M$. If f is not continuous at a point $c \notin N \cup M$, there is an $\varepsilon > 0$ and a sequence $(c_n)_{n=1}^{\infty}$ converging to c such that

$$\mid f(c_n) - f(c) \mid \geq \varepsilon \text{ all } n.$$

Since $c \notin M$, c is an interior point to exactly one interval of each partition Δ_k and we get

$$G_{\Delta_k}(c) - g_{\Delta_k}(c) \geq \varepsilon$$

and in the limit

$$G(c) - g(c) \geq \varepsilon.$$

But then $c \in N$ which is a contradiction.

Conversely, suppose the set of discontinuity points of f is a Lebesgue null set and let $(\Delta_k)_{k=1}^{\infty}$ is an arbitrary sequence of partitions of $[a,b]$ such that $\Delta_k \prec \Delta_{k+1}$ and $\mathrm{mesh}(\Delta_k) \to 0$ as $k \to \infty$. By assumption,

$$\lim_{k\to\infty} G_{\Delta_k}(x) = \lim_{k\to\infty} g_{\Delta_k}(x) = f(x)$$

at each point x of continuity of f. Therefore f is Lebesgue measurable and dominated convergence yields

$$\lim_{k\to\infty}\int_{[a,b]} G_{\Delta_k}dm = \int_{[a,b]} fdm$$

and

$$\lim_{k\to\infty}\int_{[a,b]} g_{\Delta_k}dm = \int_{[a,b]} fdm.$$

Thus f is Riemann integrable and

$$\int_a^b f(x)dx = \int_{[a,b]} fdm.$$

In the following we sometimes write

$$\int_A f(x)dx \qquad (A \in \mathcal{R}^-)$$

instead of

$$\int_A fdm \qquad (A \in \mathcal{R}^-).$$

In a similar way we often prefer to write

$$\int_A f(x)dx \qquad (A \in \mathcal{R}_n^-)$$

instead of

$$\int_A fdm_n \qquad (A \in \mathcal{R}_n^-).$$

Furthermore, $\int_a^b fdm$ means $\int_{[a,b]} fdm$. Here, however, a warning is motivated. It is simple to find a real-valued function f on $[0,\infty[$, which is bounded on each bounded subinterval of $[0,\infty[$, such that the generalized Riemann integral

$$\int_0^\infty f(x)dx$$

is convergent, that is

$$\lim_{b\to\infty}\int_0^b f(x)dx$$

exists and the limit is a real number, while the Riemann integral

$$\int_0^\infty \mid f(x) \mid dx$$

is divergent (take e.g. $f(x) = \frac{\sin x}{x}$). In this case the function f does not belong to $\mathcal{L}^1$ with respect to Lebesgue measure on $[0, \infty[$ since

$$\int_{[0,\infty[} \mid f \mid dm = \lim_{b \to \infty} \int_0^b \mid f(x) \mid dx = \infty.$$

Exercises

1. Let $f_n : [0, 1] \to [0, 1]$, $n \in \mathbf{N}$, be a sequence of Riemann integrable functions such that

$$\lim_{n \to \infty} f_n(x) \text{ exists } = f(x) \text{ all } x \in [0, 1]\,.$$

Show by giving an example that f need not be Riemann integrable.

2. Suppose $f_n(x) = n^2 \mid x \mid e^{-n|x|}$, $x \in \mathbf{R}$, $n \in \mathbf{N}_+$. Compute $\lim_{n\to\infty} f_n$ and $\lim_{n\to\infty} \int_{\mathbf{R}} f_n dm$.

3. Compute the following limits and justify the calculations:

a)

$$\lim_{n \to \infty} \int_0^\infty \frac{\sin(e^x)}{1 + nx^2} dx.$$

b)

$$\lim_{n \to \infty} \int_0^n (1 + \frac{x}{n})^{-n} \cos x dx.$$

c)

$$\lim_{n \to \infty} \int_0^n (1 + \frac{x}{n})^n e^{-2x} dx.$$

d)

$$\lim_{n \to \infty} \int_0^\infty (1 + \frac{x}{n})^n \exp(-(1 + \frac{x}{n})^n) dx.$$

e)

$$\lim_{n\to\infty}\int_0^n (1-\frac{x}{n})^{-n} e^{\frac{x}{2}} dx.$$

f)

$$\lim_{n\to\infty}\int_0^n (1-\frac{x}{n})^n \frac{1+nx}{n+x}\cos x dx$$

g)

$$\lim_{n\to\infty}\int_0^\infty (1+\frac{x}{n})^{n^2} e^{-nx} dx.$$

4. Let $(r_n)_{n=1}^\infty$ be an enumeration of $\mathbf{Q}$ and define

$$f(x) = \Sigma_{n=1}^\infty 2^{-n}\varphi(x-r_n)$$

where $\varphi(x) = x^{-\frac{1}{2}}$ if $0 < x < 1$ and $\varphi(x) = 0$ if $x \leq 0$ or $x \geq 1$. Show that
a)

$$\int_{-\infty}^\infty f(x)dx = 2.$$

b)

$$\int_a^b f^2(x)dx = \infty \text{ if } a < b.$$

c)

$$f < \infty \text{ a.s. } [m].$$

d)

$$\sup_{a<x<b} f(x) = +\infty \text{ if } a < b.$$

5. Suppose

$$f(t) = \int_0^\infty e^{-tx}\frac{\ln(1+x)}{1+x}dx,\ t > 0.$$

a) Show that $\int_0^\infty f(t)dt < \infty$.
b) Show that f is infinitely many times differentiable.

6. Suppose

$$f(t) = \int_0^\infty \frac{xe^{-x^2}}{x^2+t^2}dx,\ t > 0.$$

Compute

$$\lim_{t\to 0+} f(t) \text{ and } \int_0^\infty f(t)dt.$$

Finally, prove that f is differentiable.

2.4. Expectation

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and $\xi : (\Omega, \mathcal{F}) \to (S, \mathcal{S})$ a random variable. Recall that the probability law μ of ξ is given by the image measure P_ξ. By definition,

$$\int_S \chi_B d\mu = \int_\Omega \chi_B(\xi) dP$$

for every $B \in \mathcal{S}$, and, hence

$$\int_S \varphi d\mu = \int_\Omega \varphi(\xi) dP$$

for each simple $\mathcal{S}$-measurable function φ on S (we sometimes write $f \circ g = f(g)$). By monotone convergence, we get

$$\int_S f d\mu = \int_\Omega f(\xi) dP$$

for every measurable $f : S \to [0, \infty]$. Thus if $f : S \to \mathbf{R}$ is measurable, $f \in L^1(\mu)$ if and only if $f(\xi) \in L^1(P)$ and in this case

$$\int_S f d\mu = \int_\Omega f(\xi) dP.$$

In the special case when ξ is real-valued and $\xi \in L^1(P)$,

$$\int_{\mathbf{R}} x d\mu(x) = \int_\Omega \xi dP.$$

The integral in the right-hand side is called the expectation of ξ and is denoted by $E[\xi]$.

CHAPTER 3

Further Construction Methods of Measures

Introduction

In the first section of this chapter we collect some basic results on metric spaces, which every mathematician must know about. Section 3.2 gives a version of the Riesz Representation Theorem, which leads to another and perhaps simpler approach to Lebesgue measure than the Carathéodory Theorem. A reader can skip Section 3.2 without losing the continuity in this paper. The chapter also treats so called product measures and Stieltjes integrals.

3.1. Metric Spaces

The construction of our most important measures requires topological concepts. For our purpose it will be enough to restrict ourselves to so called metric spaces.

A metric d on a set X is a mapping $d : X \times X \to [0, \infty[$ such that

(a) $d(x, y) = 0$ if and only if $x = y$
(b) $d(x, y) = d(y, x)$ (symmetry)
(c) $d(x, y) \leq d(x, z) + d(z, y)$ (triangle inequality).

Here recall, if $A_1, ..., A_n$ are sets,

$$A_1 \times ... \times A_n = \{(x_1, ..., x_n);\ x_i \in A_i \text{ for all } i = 1, ..., n\}$$

A set X equipped with a metric d is called a metric space. Sometimes we write $X = (X, d)$ to emphasize the metric d. If E is a subset of the metric space (X, d), the function $d_{|E\times E}(x, y) = d(x, y)$, if $x, y \in E$, is a metric on E. Thus $(E, d_{|E\times E})$ is a metric space.

The function $\varphi(t) = \min(1, t)$, $t \geq 0$, satisfies the inequality

$$\varphi(s + t) \leq \varphi(s) + \varphi(t).$$

Therefore, if d is a metric on X, $\min(1, d)$ is a metric on X. The metric $\min(1, d)$ is a bounded metric.

The set $\mathbf{R}$ equipped with the metric $d_1(x, y) =\mid x - y \mid$ is a metric space. More generally, $\mathbf{R}^n$ equipped with the metric

$$d_n(x, y) = d_n((x_1, ..., x_n), (y_1, ..., y_n)) = \max_{1\leq k\leq n} \mid x_k - y_k \mid$$

is a metric space. If not otherwise stated, it will always be assumed that $\mathbf{R}^n$ is equipped with this metric.

Let $C\,[0, T]$ denote the vector space of all real-valued continuous functions on the interval $[0, T]$, where $T > 0$. Then

$$d_\infty(x, y) = \max_{0\leq t\leq T} \mid x(t) - y(t) \mid$$

is a metric on $C\,[0, T]$.

If (X_k, e_k), $k = 1, ..., n$, are metric spaces,

$$d(x, y) = \max_{1\leq k\leq n} e_k(x_k, y_k),\ x = (x_1, ..., x_n)\ , y = (y_1, ..., y_n)$$

is a metric on $X_1 \times ... \times X_n$. The metric d is called the product metric on $X_1 \times ... \times X_n$.

If $X = (X, d)$ is a metric space and $x \in X$ and $r > 0$, the open ball with centre at x and radius r is the set $B(x, r) = \{y \in X; d(y, x) < r\}$. If $E \subseteq X$ and E is contained in an appropriate open ball in X it is said to be bounded. The diameter of E is, by definition,

$$\text{diam } E = \sup_{x,y\in E} d(x, y)$$

and it follows that E is bounded if and only if diam $E < \infty$. A subset of X which is a union of open balls in X is called open. In particular, an open ball is an open set. The empty set is open since the union of an empty family of sets is empty. An arbitrary union of open sets is open. The class of all open subsets of X is called the topology of X. The metrics d and $\min(1, d)$ determine the same topology. A subset E of X is said to be closed if its complement E^c relative to X is open. An intersection of closed subsets of X is closed. If $E \subseteq X$, E° denotes the largest open set contained in E and E^- (or $\bar{E}$) the smallest closed set containing E. E° is the interior of E and E^- its closure. The σ-algebra generated by the open sets in X is called the Borel σ-algebra in X and is denoted by $\mathcal{B}(X)$. A positive measure on $\mathcal{B}(X)$ is called a positive Borel measure.

A sequence $(x_n)_{n=1}^\infty$ in X converges to $x \in X$ if

$$\lim_{n\to\infty} d(x_n, x) = 0.$$

If, in addition, the sequence $(x_n)_{n=1}^\infty$ converges to $y \in X$, the inequalities

$$0 \leq d(x, y) \leq d(x_n, x) + d(x_n, y)$$

imply that $y = x$ and the limit point x is unique.

If $E \subseteq X$ and $x \in X$, the following properties are equivalent:

(i) $x \in E^-$.
(ii) $B(x, r) \cap E \neq \phi$, all $r > 0$.
(iii) There is a sequence $(x_n)_{n=1}^\infty$ in E which converges to x.

If $B(x, r) \cap E = \phi$, then $B(x, r)^c$ is a closed set containing E but not x. Thus $x \notin E^-$. This proves that (i)$\Rightarrow$(ii). Conversely, if $x \notin E^-$, since $\bar{E}^c$ is open there exists an open ball $B(y, s)$ such that $x \in B(y, s) \subseteq \bar{E}^c \subseteq E^c$. Now choose $r = s - d(x, y) > 0$ so that $B(x, r) \subseteq B(y, s)$. Then $B(x, r) \cap E = \phi$. This proves (ii)$\Rightarrow$(i).

If (ii) holds choose for each $n \in \mathbf{N}_+$ a point $x_n \in E$ with $d(x_n, x) < \frac{1}{n}$ and (iii) follows. If there exists an $r > 0$ such that $B(x, r) \cap E = \phi$, then (iii) cannot hold. Thus (iii)$\Rightarrow$(ii).

If $E \subseteq X$, the set $E^- \setminus E^\circ$ is called the boundary of E and is denoted by ∂E.

A set $A \subseteq X$ is said to be dense in X if $A^- = X$. The metric space X is called separable if there is an at most denumerable dense subset of X. For example, $\mathbf{Q}^n$ is a dense subset of $\mathbf{R}^n$. The space $\mathbf{R}^n$ is separable.

Theorem 3.1.1. $\mathcal{B}(\mathbf{R}^n) = \mathcal{R}_n$.

PROOF. The σ-algebra $\mathcal{R}_n$ is generated by the open n-cells in $\mathbf{R}^n$ and an open n-cell is an open subset of $\mathbf{R}^n$. Hence $\mathcal{R}_n \subseteq \mathcal{B}(\mathbf{R}^n)$. Let U be an open subset in $\mathbf{R}^n$ and note that an open ball in $\mathbf{R}^n = (\mathbf{R}^n, d_n)$ is an open n-cell. If $x \in U$ there exist an $a \in \mathbf{Q}^n \cap U$ and a rational number $r > 0$ such that $x \in B(a, r) \subseteq U$. Thus U is an at most denumerable union of open n-cells and it follows that $U \in \mathcal{R}_n$. Thus $\mathcal{B}(\mathbf{R}^n) \subseteq \mathcal{R}_n$ and the theorem is proved.

Let $X = (X, d)$ and $Y = (Y, e)$ be two metric spaces. A mapping $f : X \to Y$ (or $f : (X, d) \to (Y, e)$ to emphasize the underlying metrics) is said to be continuous at the point $a \in X$ if for every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$x \in B(a, \delta) \Rightarrow f(x) \in B(f(a), \varepsilon).$$

Equivalently this means that for any sequence $(x_n)_{n=1}^{\infty}$ in X which converges to a in X, the sequence $(f(x_n))_{n=1}^{\infty}$ converges to $f(a)$ in Y. If f is continuous at each point of X, the mapping f is called continuous. Stated otherwise this means that

$$f^{-1}(V) \text{ is open if } V \text{ is open}$$

or

$$f^{-1}(F) \text{ is closed if } F \text{ is closed.}$$

The mapping f is said to be Borel measurable if

$$f^{-1}(B) \in \mathcal{B}(X) \text{ if } B \in \mathcal{B}(Y)$$

or, what amounts to the same thing,

$$f^{-1}(V) \in \mathcal{B}(X) \text{ if } V \text{ is open.}$$

A Borel measurable function is sometimes called a Borel function. A continuous function is a Borel function.

Example 3.1.1. Let $f : (\mathbf{R}, d_1) \to (\mathbf{R}, d_1)$ be a continuous strictly increasing function and set $\rho(x, y) = \mid f(x) - f(y) \mid$, $x, y \in \mathbf{R}$. Then ρ is a metric on $\mathbf{R}$.

Define $j(x) = x$, $x \in \mathbf{R}$. The mapping $j : (\mathbf{R},d_1)\to (\mathbf{R},\rho)$ is continuous. We claim that the map $j : (\mathbf{R},\rho) \to (\mathbf{R},d_1)$ is continuous. To see this, let $a \in \mathbf{R}$ and suppose the sequence $(x_n)_{n=1}^{\infty}$ converges to a in the metric space $(\mathbf{R},\rho)$, that is $\mid f(x_n) - f(a) \mid\to 0$ as $n \to \infty$. Let $\varepsilon > 0$. Then

$$f(x_n) - f(a) \geq f(a+\varepsilon) - f(a) > 0 \text{ if } x_n \geq a+\varepsilon$$

and

$$f(a) - f(x_n) \geq f(a) - f(a-\varepsilon) > 0 \text{ if } x_n \leq a-\varepsilon.$$

Thus $x_n \in \,]a-\varepsilon, a+\varepsilon[$ if n is sufficiently large. This proves that he map $j : (\mathbf{R},\rho) \to (\mathbf{R},d_1)$ is continuous.

The metrics d_1 and ρ determine the same topology and Borel subsets of $\mathbf{R}$.

A mapping $f : (X,d) \to (Y,e)$ is said to be uniformly continuous if for each $\varepsilon > 0$ there exists a $\delta > 0$ such that $e(f(x), f(y)) < \varepsilon$ as soon as $d(x,y) < \delta$.

If $x \in X$ and E, $F \subseteq X$, let

$$d(x,E) = \inf_{u\in E} d(x,u)$$

be the distance from x to E and let

$$d(E,F) = \inf_{u\in E, v\in F} d(u,v)$$

be the distance between E and F. Note that $d(x,E) = 0$ if and only if $x \in \bar{E}$.

If $x, y \in X$ and $u \in E$,

$$d(x,u) \leq d(x,y) + d(y,u)$$

and, hence

$$d(x,E) \leq d(x,y) + d(y,u)$$

and

$$d(x,E) \leq d(x,y) + d(y,E).$$

Next suppose $E \neq \phi$. Then by interchanging the roles of x and y, we get

$$\mid d(x,E) - d(y,E) \mid\leq d(x,y)$$

and conclude that the distance function $d(x, E)$, $x \in X$, is continuous. In fact, it is uniformly continuous. If $x \in X$ and $r > 0$, the so called closed ball $\bar{B}(x, r) = \{y \in X;\ d(y, x) \leq r\}$ is a closed set since the map $y \to d(y, x)$, $y \in X$, is continuous.

If $F \subseteq X$ is closed and $\varepsilon > 0$, the continuous function

$$\Pi^X_{F,\varepsilon} = \max(0, 1 - \frac{1}{\varepsilon} d(\cdot, F))$$

fulfils $0 \leq \Pi^X_{F,\varepsilon} \leq 1$ and $\Pi^X_{F,\varepsilon} = 1$ on F. Furthermore, $\Pi^X_{F,\varepsilon}(a) > 0$ if and only if $a \in F_\varepsilon =_{def} \{x \in X;\ d(x, F) < \varepsilon\}$. Thus

$$\chi_F \leq \Pi^X_{F,\varepsilon} \leq \chi_{F_\varepsilon}.$$

Let $X = (X, d)$ be a metric space. A sequence $(x_n)_{n=1}^\infty$ in X is called a Cauchy sequence if to each $\varepsilon > 0$ there exists a positive integer p such that $d(x_n, x_m) < \varepsilon$ for all $n, m \geq p$. If a Cauchy sequence $(x_n)_{n=1}^\infty$ contains a convergent subsequence $(x_{n_k})_{k=1}^\infty$ it must be convergent. To prove this claim, suppose the subsequence $(x_{n_k})_{k=1}^\infty$ converges to a point $x \in X$. Then

$$d(x_m, x) \leq d(x_m, x_{n_k}) + d(x_{n_k}, x)$$

can be made arbitrarily small for all sufficiently large m by choosing k sufficiently large. Thus $(x_n)_{n=1}^\infty$ converges to x.

A subset E of X is said to be complete if every Cauchy sequence in E converges to a point in E. If $E \subseteq X$ is closed and X is complete it is clear that E is complete. Conversely, if X is a metric space and a subset E of X is complete, then E is closed.

It is important to know that $\mathbf{R}$ is complete equipped with its standard metric. To see this let $(x_n)_{n=1}^\infty$ be a Cauchy sequence. There exists a positive integer such that $\mid x_n - x_m \mid < 1$ if $n, m \geq p$. Therefore

$$\mid x_n \mid \leq \mid x_n - x_p \mid + \mid x_p \mid \leq 1 + \mid x_p \mid$$

for all $n \geq p$. We have proved that the sequence $(x_n)_{n=1}^\infty$ is bounded (the reader can check that every Cauchy sequence in a metric space has this property). Now define

$$a = \sup \{x \in \mathbf{R};\ \text{there are only finitely many } n \text{ with } x_n \leq x\}.$$

The definition implies that there exists a subsequence $(x_{n_k})_{k=1}^\infty$, which converges to a (since for any $r > 0$, $x_n \in B(a, r)$ for infinitely many n). The original sequence is therefore convergent and we conclude that $\mathbf{R}$ is complete (equipped with its standard metric d_1). It is simple to prove that the product of n complete spaces is complete and we conclude that $\mathbf{R}^n$ is complete.

Let $E \subseteq X$. A family $(V_i)_{i\in I}$ of subsets of X is said to be a cover of E if $\cup_{i\in I} V_i \supseteq E$ and E is said to be covered by the $V_i's$. The cover $(V_i)_{i\in I}$ is said to be an open cover if each member V_i is open. The set E is said to be totally bounded if, for every $\varepsilon > 0$, E can be covered by finitely many open balls of radius ε. A subset of a totally bounded set is totally bounded.

The following definition is especially important.

Definition 3.1.1. A subset E of a metric space X is said to be compact if to every open cover $(V_i)_{i\in I}$ of E, there is a finite subcover of E, which means there is a finite subset J of I such that $(V_i)_{i\in J}$ is a cover of E.

If K is closed, $K \subseteq E$, and E is compact, then K is compact. To see this, let $(V_i)_{i\in I}$ be an open cover of K. This cover, augmented by the set $X \setminus K$ is an open cover of E and has a finite subcover since E is compact. Noting that $K \cap (X \setminus K) = \phi$, the assertion follows.

Theorem 3.1.2. *The following conditions are equivalent:*

(a) *E is complete and totally bounded.*

(b) *Every sequence in E contains a subsequence which converges to a point of E.*

(c) *E is compact.*

PROOF. (a)$\Rightarrow$(b). Suppose $(x_n)_{n=1}^{\infty}$ is a sequence in E. The set E can be covered by finitely many open balls of radius 2^{-1} and at least one of them must contain x_n for infinitely many $n \in \mathbf{N}_+$. Suppose $x_n \in B(a_1, 2^{-1})$ if $n \in N_1 \subseteq N_0 =_{def} \mathbf{N}_+$, where N_1 is infinite. Next $E \cap B(a_1, 2^{-1})$ can be covered by finitely many balls of radius 2^{-2} and at least one of them must contain x_n for infinitely many $n \in N_1$. Suppose $x_n \in B(a_2, 2^{-1})$ if $n \in N_2$, where $N_2 \subseteq N_1$ is infinite. By induction, we get open balls $B(a_j, 2^{-j})$ and infinite sets $N_j \subseteq N_{j-1}$ such that $x_n \in B(a_j, 2^{-j})$ for all $n \in N_j$ and $j \geq 1$.

Let $n_1 < n_2 < ...$, where $n_k \in N_k$, $k = 1, 2, ...$. The sequence $(x_{n_k})_{k=1}^{\infty}$ is a Cauchy sequence, and since E is complete it converges to a point of E .

(b)⇒(a). If E is not complete there is a Cauchy sequence in E with no limit in E. Therefore no subsequence can converge in E, which contradicts (b). On the other hand if E is not totally bounded, there is an $\varepsilon > 0$ such that E cannot be covered by finitely many balls of radius ε. Let $x_1 \in E$ be arbitrary. Having chosen $x_1, ..., x_{n-1}$, pick $x_n \in E \setminus \cup_{i=1}^{n-1} B(x_i, \varepsilon)$, and so on. The sequence $(x_n)_{n=1}^{\infty}$ cannot contain any convergent subsequence as $d(x_n, x_m) \geq \varepsilon$ if $n \neq m$, which contradicts (b).

{(a) and (b)} ⇒(c). Let $(V_i)_{i \in I}$ be an open cover of E. Since E is totally bounded it is enough to show that there is an $\varepsilon > 0$ such that any open ball of radius ε which intersects E is contained in some V_i. Suppose on the contrary that for every $n \in \mathbf{N}_+$ there is an open ball B_n of radius $\leq 2^{-n}$ which intersects E and is contained in no V_i. Choose $x_n \in B_n \cap E$ and assume without loss of generality that $(x_n)_{n=1}^{\infty}$ converges to some point x in E by eventually going to a subsequence. Suppose $x \in V_{i_0}$ and choose $r > 0$ such that $B(x, r) \subseteq V_{i_0}$. But then $B_n \subseteq B(x, r) \subseteq V_{i_0}$ for large n, which contradicts the assumption on B_n.

(c)⇒(b). If $(x_n)_{n=1}^{\infty}$ is a sequence in E with no convergent subsequence in E, then for every $x \in E$ there is an open ball $B(x, r_x)$ which contains x_n for only finitely many n. Then $(B(x, r_x))_{x \in E}$ is an open cover of E without a finite subcover.

Corollary 3.1.1. *A subset of* $\mathbf{R}^n$ *is compact if and only if it is closed and bounded.*

PROOF. Suppose K is compact. If $x_n \in K$ and $x_n \notin B(0, n)$ for every $n \in \mathbf{N}_+$, the sequence $(x_n)_{n=1}^{\infty}$ cannot contain a convergent subsequence. Thus K is bounded. Since K is complete it is closed.

Conversely, suppose K is closed and bounded. Since $\mathbf{R}^n$ is complete and K is closed, K is complete. We next prove that a bounded set is totally bounded. It is enough to prove that any n-cell in $\mathbf{R}^n$ is a union of finitely many n-cells $I_1 \times ... \times I_n$ where each interval $I_1, ..., I_n$ has a prescribed positive length. This is clear and the theorem is proved.

Corollary 3.1.2. *Suppose* $f : X \to \mathbf{R}$ *is continuous and* X *compact.*
(a) *There exists an* $a \in X$ *such that* $\max_X f = f(a)$ *and a* $b \in X$ *such that* $\min_X f = f(b)$.
(b) *The function* f *is uniformly continuous.*

PROOF. (a) For each $a \in X$, let $V_a = \{x \in X : f(x) < 1 + f(a)\}$. The open cover $(V_a)_{a \in K}$ of X has a finite subcover and it follows that f is bounded. Let $(x_n)_{n=1}^\infty$ be a sequence in X such that $f(x_n) \to \sup_K f$ as $n \to \infty$. Since X is compact there is a subsequence $(x_{n_k})_{k=1}^\infty$ which converges to a point $a \in X$. Thus, by the continuity of f, $f(x_{n_k}) \to f(a)$ as $k \to \infty$.

The existence of a minimum is proved in a similar way.

(b) If f is not uniformly continuous there exist $\varepsilon > 0$ and sequences $(x_n)_{n=1}^\infty$ and $(y_n)_{n=1}^\infty$ such that $\mid f(x_n) - f(y_n) \mid \geq \varepsilon$ and $\mid x_n - y_n \mid < 2^{-n}$ for every $n \geq 1$. Since X is compact there exists a subsequence $(x_{n_k})_{k=1}^\infty$ of $(x_n)_{n=1}^\infty$ which converges to a point $a \in X$. Clearly the sequence $(y_{n_k})_{k=1}^\infty$ converges to a and therefore

$$\mid f(x_{n_k}) - f(y_{n_k}) \mid \leq \mid f(x_{n_k}) - f(a) \mid + \mid f(a) - f(y_{n_k}) \mid \to 0$$

as $k \to \infty$ since f is continuous. But $\mid f(x_{n_k}) - f(y_{n_k}) \mid \geq \varepsilon$ and we have got a contradiction. The corollary is proved.

Example 3.1.2. Suppose $X =]0, 1]$ and define $\rho_1(x, y) = d_1(x, y)$ and $\rho_2(x, y) = \mid \frac{1}{x} - \frac{1}{y} \mid$, $x, y \in X$. As in Example 3.1.1 we conclude that the metrics ρ_1 and ρ_2 determine the same topology of subsets of X. The space (X, ρ_1)

totally bounded but not complete. However, the space (X, ρ_2) is not totally bounded but it is complete. To see this, let $(x_n)_{n=1}^{\infty}$be a Cauchy sequence in (X, ρ_2). As a Cauchy sequence it must be bounded and therefore there exists an $\varepsilon \in]0,1]$ such that $x_n \in [\varepsilon, 1]$ for all n. But then, by Corollary 3.1.1, $(x_n)_{n=1}^{\infty}$ contains a convergent subsequence in (X, ρ_1) and, accordingly from this, the same property holds in (X, ρ_2). The space (X, ρ_2) is not compact, since (X, ρ_1) is not compact, and we conclude from Theorem 3.1.2 that the space (X, ρ_2) cannot be totally bounded.

Example 3.1.3. Set $\hat{\mathbf{R}}$=$\mathbf{R}\cup\{-\infty, \infty\}$ and

$$\hat{d}(x, y) = \mid \arctan x - \arctan y \mid$$

if $x, y \in \hat{\mathbf{R}}$. Here

$$\arctan \infty = \frac{\pi}{2} \text{ and } \arctan -\infty = -\frac{\pi}{2}.$$

Example 3.1.1 shows that the standard metric d_1 and the metric $\hat{d}_{|\mathbf{R}\times\mathbf{R}}$ determine the same topology.

We next prove that the metric space $\hat{\mathbf{R}}$ is compact. To this end, consider a sequence $(x_n)_{n=1}^{\infty}$ in $\hat{\mathbf{R}}$**.** If there exists a real number M such that $\mid x_n \mid \leq M$ for infinitely many n, the sequence $(x_n)_{n=1}^{\infty}$ contains a convergent subsequence since the interval $[-M, M]$ is compact. In the opposite case, for each positive real number M, either $x_n \geq M$ for infinitely many n or $x_n \leq -M$ for infinitely many n. Suppose $x_n \geq M$ for infinitely many n for every $M \in \mathbf{N}_+$. Then $\hat{d}(x_{n_k}, \infty) = \mid \arctan x_{n_k} - \frac{\pi}{2} \mid \rightarrow 0$ as $k \rightarrow \infty$ for an appropriate subsequence $(x_{n_k})_{k=1}^{\infty}$.

The space $\hat{\mathbf{R}}$= $(\hat{\mathbf{R}},\hat{d})$ is called a two-point compactification of $\mathbf{R}$.

It is an immediate consequence of Theorem 3.1.2 that the product of finitely many compact metric spaces is compact. Thus $\hat{\mathbf{R}}^n$ equipped with the product metric is compact.

We will finish this section with several useful approximation theorems.

Theorem 3.1.3. *Suppose X is a metric space and μ positive Borel measure in X. Moreover, suppose there is a sequence $(U_n)_{n=1}^{\infty}$of open subsets of X such that*

$$X = \cup_{n=1}^{\infty} U_n$$

and

$$\mu(U_n) < \infty, \text{ all } n \in \mathbf{N}_+.$$

Then for each $A \in \mathcal{B}(X)$ and $\varepsilon > 0$, there are a closed set $F \subseteq A$ and an open set $V \supseteq A$ such that

$$\mu(V \setminus F) < \varepsilon.$$

In particular, for every $A \in \mathcal{B}(X)$,

$$\mu(A) = \inf_{\substack{V \supseteq A \\ V \text{ open}}} \mu(V)$$

and

$$\mu(A) = \sup_{\substack{F \subseteq A \\ F \text{ closed}}} \mu(F)$$

If $X = \mathbf{R}$ and $\mu(A) = \Sigma_{n=1}^{\infty} \delta_{\frac{1}{n}}(A)$, $A \in \mathcal{R}$, then $\mu(\{0\}) = 0$ and $\mu(V) = \infty$ for every open set containing $\{0\}$. The hypothesis that the sets U_n, $n \in \mathbf{N}_+$, are open (and not merely Borel sets) is very important in Theorem 3.1.3.

PROOF. First suppose that μ is a finite positive measure.

Let $\mathcal{A}$ be the class of all Borel sets A in X such that for every $\varepsilon > 0$ there exist a closed $F \subseteq A$ and an open $V \supseteq A$ such that $\mu(V \setminus F) < \varepsilon$. If F is a closed subset of X and $V_n = \left\{x; d(x, F) < \frac{1}{n}\right\}$, then V_n is open and, by Theorem 1.1.2 (f), $\mu(V_n) \downarrow v(F)$ as $n \to \infty$. Thus $F \in \mathcal{A}$ and we conclude that $\mathcal{A}$ contains all closed subsets of X.

Now suppose $A \in \mathcal{A}$. We will prove that $A^c \in \mathcal{A}$. To this end, we choose $\varepsilon > 0$ and a closed set $F \subseteq A$ and an open set $V \supseteq A$ such that $\mu(V \setminus F) < \varepsilon$. Then $V^c \subseteq A^c \subseteq F^c$ and, moreover, $\mu(F^c \setminus V^c) < \varepsilon$ since

$$V \setminus F = F^c \setminus V^c.$$

If we note that V^c is closed and F^c open it follows that $A^c \in \mathcal{A}$.

Next let $(A_i)_{i=1}^{\infty}$ be a denumerable collection of members of $\mathcal{A}$. Choose $\varepsilon > 0$. By definition, for each $i \in \mathbf{N}_+$ there exist a closed $F_i \subseteq A_i$ and an open $V_i \supseteq A_i$ such that $\mu(V_i \setminus F_i) < 2^{-i}\varepsilon$. Set

$$V = \cup_{i=1}^{\infty} V_i.$$

Then

$$\mu(V \setminus (\cup_{i=1}^{\infty} F_i)) \leq \mu(\cup_{i=1}^{\infty}(V_i \setminus F_i))$$
$$\leq \Sigma_{i=1}^{\infty} \mu(V_i \setminus F_i) < \varepsilon.$$

But

$$V \setminus (\cup_{i=1}^{\infty} F_i) = \cap_{n=1}^{\infty} \{V \setminus (\cup_{i=1}^{n} F_i)\}$$

and since μ is a finite positive measure

$$\mu(V \setminus (\cup_{i=1}^{\infty} F_i)) = \lim_{n\to\infty} \mu(V \setminus (\cup_{i=1}^{n} F_i)).$$

Accordingly, from these equations

$$\mu(V \setminus (\cup_{i=1}^{n} F_i)) < \varepsilon$$

if n is large enough. Since a union of open sets is open and a finite union of closed sets is closed, we conclude that $\cup_{i=1}^{\infty} A_i \in \mathcal{A}$. This proves that $\mathcal{A}$ is a σ-algebra. Since $\mathcal{A}$ contains each closed subset of X, $\mathcal{A} = \mathcal{B}(X)$.

We now prove the general case. Suppose $A \in \mathcal{B}(X)$. Since μ^{U_n} is a finite positive measure the previous theorem gives us an open set $V_n \supseteq A \cap U_n$ such that $\mu^{U_n}(V_n \setminus (A \cap U_n)) < \varepsilon 2^{-n}$. By eventually replacing V_n by $V_n \cap U_n$ we can assume that $V_n \subseteq U_n$. But then $\mu(V_n \setminus (A \cap U_n)) = \mu^{U_n}(V_n \setminus (A \cap U_n)) < \varepsilon 2^{-n}$.

Set $V = \cup_{n=1}^{\infty} V_n$ and note that V is open. Moreover,

$$V \setminus A \subseteq \cup_{n=1}^{\infty}(V_n \setminus (A \cap U_n))$$

and we get

$$\mu(V \setminus A) \leq \Sigma_{n=1}^{\infty} \mu(V_n \setminus (A \cap U_n)) < \varepsilon.$$

By applying the result already proved to the complement A^c we conclude there exists an open set $W \supseteq A^c$ such that

$$\mu(A \setminus W^c) = \mu(W \setminus A^c) < \varepsilon.$$

Thus if $F =_{def} W^c$ it follows that $F \subseteq A \subseteq V$ and $\mu(V \setminus F) < 2\varepsilon$. The theorem is proved.

If X is a metric space $C(X)$ denotes the vector space of all real-valued continuous functions $f : X \to \mathbf{R}$. If $f \in C(X)$, the closure of the set of all x where $f(x) \neq 0$ is called the support of f and is denoted by supp$f$. The vector space of all all real-valued continuous functions $f : X \to \mathbf{R}$ with compact support is denoted by $C_c(X)$.

Corollary 3.1.3. *Suppose μ and ν are positive Borel measures in $\mathbf{R}^n$ such that*

$$\mu(K) < \infty \text{ and } \nu(K) < \infty$$

for every compact subset K of $\mathbf{R}^n$. If

$$\int_{\mathbf{R}^n} f(x)d\mu(x) = \int_{\mathbf{R}^n} f(x)d\nu(x), \text{ all } f \in C_c(\mathbf{R}^n)$$

then $\mu = \nu$.

PROOF. Let F be closed. Clearly $\mu(B(0,i)) < \infty$ and $\nu(B(0,i)) < \infty$ for every positive integer i. Hence, by Theorem 3.1.3 it is enough to show that $\mu(F) = \nu(F)$. Now fix a positive integer i and set $K = \bar{B}(0,i) \cap F$. It is enough to show that $\mu(K) = \nu(K)$. But

$$\int_{\mathbf{R}^n} \Pi^{\mathbf{R}^n}_{K,2^{-j}}(x)d\mu(x) = \int_{\mathbf{R}^n} \Pi^{\mathbf{R}^n}_{K,2^{-j}}(x)d\nu(x)$$

for each positive integer j and letting $j \to \infty$ we are done.

A metric space X is called a standard space if it is separable and complete. Standard spaces have a series of very nice properties related to measure theory; an example is furnished by the following

Theorem 3.1.4. (Ulam's Theorem) *Let X be a standard space and suppose μ is a finite positive Borel measure on X. Then to each $A \in \mathcal{B}(X)$ and $\varepsilon > 0$ there exist a compact $K \subseteq A$ and an open $V \supseteq A$ such that $\mu(V \setminus K) < \varepsilon$.*

PROOF. Let $\varepsilon > 0$. We first prove that there is a compact subset K of X such that $\mu(K) > \mu(X) - \varepsilon$. To this end, let A be a dense denumerable subset of X and let $(a_i)_{i=1}^{\infty}$ be an enumeration of A. Now for each positive integer j, $\cup_{i=1}^{\infty} B(a_i, 2^{-j}\varepsilon) = X$, and therefore there is a positive integer n_j such that

$$\mu(\cup_{i=1}^{n_j} B(a_i, 2^{-j}\varepsilon)) > \mu(X) - 2^{-j}\varepsilon.$$

Set

$$F_j = \cup_{i=1}^{n_j} \bar{B}(a_i, 2^{-j}\varepsilon)$$

and

$$L = \cap_{j=1}^{\infty} F_j.$$

The set L is totally bounded. Since X is complete and L closed, L is complete. Therefore, the set L is compact and, moreover

$$\mu(K) = \mu(X) - \mu(L^c) = \mu(X) - \mu(\cup_{j=1}^{\infty} F_j^c)$$

$$\geq \mu(X) - \Sigma_{j=1}^{\infty} \mu(F_j^c) = \mu(X) - \Sigma_{j=1}^{\infty} (\mu(X) - \mu(F_j))$$

$$\geq \mu(X) - \Sigma_{j=1}^{\infty} 2^{-j}\varepsilon = \mu(X) - \varepsilon.$$

Depending on Theorem 3.1.3 to each $A \in \mathcal{B}(X)$ there exists a closed $F \subseteq A$ and an open $V \supseteq A$ such that $\mu(V \setminus F) < \varepsilon$. But

$$V \setminus (F \cap L) = (V \setminus F) \cup (F \setminus L)$$

and we get

$$\mu(V \setminus (F \cap L)) \leq \mu(V \setminus F) + \mu(X \setminus K) < 2\varepsilon.$$

Since the set $F \cap L$ is compact Theorem 3.1.4 is proved.

Two Borel sets in $\mathbf{R}^n$ are said to be almost disjoint if their intersection has volume measure zero.

Theorem 3.1.5. *Every open set U in $\mathbf{R}^n$ is the union of an at most denumerable collection of mutually almost disjoint cubes.*

Before the proof observe that a cube in $\mathbf{R}^n$ is the same as a closed ball in $\mathbf{R}^n$ equipped with the metric d_n.

PROOF. For each, $k \in \mathbf{N}_+$, let $\mathcal{Q}_k$be the class of all cubes of side length 2^{-k} whose vertices have coordinates of the form $i2^{-k}$, $i \in \mathbf{Z}$. Let F_1 be the union of those cubes in $\mathcal{Q}_1$ which are contained in U. Inductively, for $k \geq 1$, let F_k be the union of those cubes in $\mathcal{Q}_k$ which are contained in U and whose interiors are disjoint from $\cup_{j=1}^{k-1} F_j$. Since $d(x,\mathbf{R}^n \setminus U) > 0$ for every $x \in U$ it follows that $U = \cup_{j=1}^{\infty} F_j$.

Exercises

1. Suppose $f : (X, \mathcal{M}) \to (\mathbf{R}^d, \mathcal{R}_d)$ and $g : (X, \mathcal{M}) \to (\mathbf{R}^n, \mathcal{R}_n)$ are measurable. Set $h(x) = (f(x), g(x)) \in \mathbf{R}^{d+n}$ if $x \in X$. Prove that $h : (X, \mathcal{M}) \to (\mathbf{R}^{d+n}, \mathcal{R}_{d+n})$ is measurable.

2. Suppose $f : (X, \mathcal{M}) \to (\mathbf{R}, \mathcal{R})$ and $g : (X, \mathcal{M}) \to (\mathbf{R}, \mathcal{R})$ are measurable. Prove that fg is $(\mathcal{M}, \mathcal{R})$-measurable.

3. The function $f : \mathbf{R} \to \mathbf{R}$ is a Borel function. Set $g(x, y) = f(x)$, $(x, y) \in \mathbf{R}^2$. Prove that $g : \mathbf{R}^2 \to \mathbf{R}$ is a Borel function.

4. Suppose $f : [0, 1] \to \mathbf{R}$ is a continuous function and $g : [0, 1] \to [0, 1]$ a Borel function. Compute the limit

$$\lim_{n\to\infty} \int_0^1 f(g(x)^n)dx.$$

5. Suppose X and Y are metric spaces and $f : X \to Y$ a continuous mapping. Show that $f(E)$ is compact if E is a compact subset of X.

6. Suppose X and Y are metric spaces and $f : X \to Y$ a continuous bijection. Show that the inverse mapping f^{-1} is continuous if X is compact.

7. Construct an open bounded subset V of $\mathbf{R}$ such that $m(\partial V) > 0$.

8. The function $f : [0,1] \to \mathbf{R}$ has a continuous derivative. Prove that the set $f(K) \in \mathcal{Z}_m$ if $K = (f')^{-1}(\{0\})$.

9. Let P denote the class of all Borel probability measures on $[0,1]$ and L the class of all functions $f : [0,1] \to [-1,1]$ such that

$$\mid f(x) - f(y) \mid \leq \mid x - y \mid, \; x, y \in [0,1] \,.$$

For any $\mu, \nu \in P$, define

$$\rho(\mu, \nu) = \sup_{f \in L} \mid \int_{[0,1]} f d\mu - \int_{[0,1]} f d\nu \mid \,.$$

(a) Show that (P, ρ) is a metric space. (b) Compute $\rho(\mu, \nu)$ if μ is linear measure on $[0,1]$ and $\nu = \frac{1}{n}\Sigma_{k=0}^{n-1}\delta_{\frac{k}{n}}$, where $n \in \mathbf{N}_+$ (linear measure on $[0,1]$ is Lebesgue measure on $[0,1]$ restricted to the Borel sets in $[0,1]$).

10. Suppose μ is a finite positive Borel measure on $\mathbf{R}^n$. (a) Let $(V_i)_{i\in I}$ be a family of open subsets of $\mathbf{R}^n$ and $V = \cup_{i\in I} V_i$. Prove that

$$\mu(V) = \sup_{\substack{i_1,...,i_k \in I \\ k \in \mathbf{N}_+}} \mu(V_{i_1} \cup ... \cup V_{i_k}).$$

(b) Let $(F_i)_{i\in I}$ be a family of closed subsets of $\mathbf{R}^n$ and $F = \cap_{i\in I} F_i$. Prove that

$$\mu(F) = \inf_{\substack{i_1,...,i_k \in I \\ k \in \mathbf{N}_+}} \mu(F_{i_1} \cap ... \cap F_{i_k}).$$

$\downarrow\downarrow\downarrow$

3.2. Linear Functionals and Measures

Let X be a metric space. A mapping $T : C_c(X) \to \mathbf{R}$ is said to be a linear functional on $C_c(X)$ if

$$T(f+g) = Tf + Tg, \text{ all } f, g \in C_c(X)$$

and

$$T(\alpha f) = \alpha Tf, \text{ all } \alpha \in \mathbf{R}, \ f \in C_c(X).$$

If in addition $Tf \geq 0$ for all $f \geq 0$, T is called a positive linear functional on $C_c(X)$. In this case $Tf \leq Tg$ if $f \leq g$ since $g - f \geq 0$ and $Tg - Tf = T(g-f) \geq 0$. Note that $C_c(X) = C(X)$ if X is compact.

The main result in this section is the following

Theorem 3.2.1. (The Riesz Representation Theorem) *Suppose X is a compact metric space and let T be a positive linear functional on $C(X)$. Then there exists a unique finite positive Borel measure μ in X with the following properties:*

(a)

$$Tf = \int_X f d\mu, \ f \in C(X).$$

(b) *For every $E \in \mathcal{B}(X)$*

$$\mu(E) = \sup_{\substack{K \subseteq E \\ K \text{ compact}}} \mu(K).$$

(c) *For every $E \in \mathcal{B}(X)$*

$$\mu(E) = \inf_{\substack{V \supseteq E \\ V \text{ open}}} \mu(V).$$

The property (c) is a consequence of (b), since for each $E \in \mathcal{B}(X)$ and $\varepsilon > 0$ there is a compact $K \subseteq X \setminus E$ such that

$$\mu(X \setminus E) < \mu(K) + \varepsilon.$$

But then

$$\mu(X \setminus K) < \mu(E) + \varepsilon$$

and $X \setminus K$ is open and contains E. In a similar way, (b) follows from (c) since X is compact.

The proof of the Riesz Representation Theorem depends on properties of continuous functions of independent interest. Suppose $K \subseteq X$ is compact and $V \subseteq X$ is open. If $f : X \to [0,1]$ is a continuous function such that

$$f \leq \chi_V \text{ and } \mathrm{supp} f \subseteq V$$

we write

$$f \prec V$$

and if

$$\chi_K \leq f \leq \chi_V \text{ and } \mathrm{supp} f \subseteq V$$

we write

$$K \prec f \prec V.$$

Theorem 3.2.2. *Let K be compact subset X.*

(a) *Suppose $K \subseteq V$ where V is open. There exists a function f on X such that*

$$K \prec f \prec V.$$

(b) *Suppose X is compact and $K \subseteq V_1 \cup ... \cup V_n$, where K is compact and $V_1, ..., V_n$ are open. There exist functions $h_1, ..., h_n$ on X such that*

$$h_i \prec V_i,\ \ i = 1, ..., n$$

and

$$h_1 + ... + h_n = 1 \text{ on } K.$$

PROOF. (a) Suppose $\varepsilon = \frac{1}{2}\min_K d(\cdot, V^c)$. By Corollary 3.1.2, $\varepsilon > 0$. The continuous function $f = \Pi^X_{K,\varepsilon}$ satisfies $\chi_K \le f \le \chi_{K_\varepsilon}$, that is $K \prec f \prec K_\varepsilon$. Part (a) follows if we note that the closure $(K_\varepsilon)^-$ of K_ε is contained in V.

(b) For each $x \in K$ there exists an $r_x > 0$ such that $B(x, r_x) \subseteq V_i$ for some i. Let $U_x = B(x, \frac{1}{2}r_x)$. It is important to note that $(U_x)^- \subseteq V_i$ and $(U_x)^-$ is compact since X is compact. There exist points $x_1, ..., x_m \in K$ such that $\cup_{j=1}^m U_{x_i} \supseteq K$. If $1 \le i \le n$, let F_i denote the union of those $(U_{x_j})^-$ which are contained in V_i. By Part (a), there exist continuous functions f_i such that $F_i \prec f_i \prec V_i$, $i = 1, ..., n$. Define

$$\begin{aligned}
h_1 &= f_1 \\
h_2 &= (1 - f_1)f_2 \\
&\;\;\dots. \\
h_n &= (1 - f_1)...(1 - f_{n-1})f_n.
\end{aligned}$$

Clearly, $h_i \prec V_i$, $i = 1, ..., n$. Moreover, by induction, we get

$$h_1 + ... + h_n = 1 - (1 - f_1)...(1 - f_{n-1})(1 - f_n).$$

Since $\cup_{i=1}^n F_i \supseteq K$ we are done.

The uniqueness in Theorem 3.2.1 is simple to prove. Suppose μ_1 and μ_2 are two measures for which the theorem holds. Fix $\varepsilon > 0$ and compact $K \subseteq X$ and choose an open set V so that $\mu_2(V) \le \mu_2(K) + \varepsilon$. If $K \prec f \prec V$,

$$\mu_1(K) = \int_X \chi_K d\mu_1 \le \int_X f d\mu_1 = Tf$$

$$= \int_X f d\mu_2 \le \int_X \chi_V d\mu_2 = \mu_2(V) \le \mu_2(K) + \varepsilon.$$

Thus $\mu_1(K) \le \mu_2(K)$. If we interchange the roles of the two measures, the opposite inequality is obtained, and the uniqueness of μ follows.

To prove the existence of the measure μ in Theorem 3.2.1, define for every open V in X,

$$\mu(V) = \sup_{f \prec V} Tf.$$

Here $\mu(\phi) = 0$ since the supremum over the empty set, by convention, equals 0. Note also that $\mu(X) = T1$. Moreover, $\mu(V_1) \leq \mu(V_2)$ if V_1 and V_2 are open and $V_1 \subseteq V_2$. Now set

$$\mu(E) = \inf_{\substack{V \supseteq E \\ V \text{ open}}} \mu(V) \text{ if } E \in \mathcal{B}(X).$$

Clearly, $\mu(E_1) \leq \mu(E_2)$, if $E_1 \subseteq E_2$ and $E_{1,}E_2 \in \mathcal{B}(X)$. We therefore say that μ is increasing.

Lemma 3.2.1. (a) *If* $V_1, ..., V_n$ *are open,*

$$\mu(\cup_{i=1}^n V_i) \leq \Sigma_{i=1}^n \mu(V_i).$$

(b) *If* $E_1, E_2, ... \in \mathcal{B}(X)$,

$$\mu(\cup_{i=1}^\infty E_i) \leq \Sigma_{i=1}^\infty \mu(E_i).$$

(c) *If* $K_1, ..., K_n$ *are compact and pairwise disjoint,*

$$\mu(\cup_{i=1}^n K_i) = \Sigma_{i=1}^n \mu(K_i).$$

PROOF. (a) It is enough to prove (a) for $n = 2$. To this end first choose $g \prec V_1 \cup V_2$ and then $h_i \prec V_i$, $i = 1, 2$, such that $h_1 + h_2 = 1$ on supp g. Then

$$g = h_1 g + h_2 g$$

and it follows that

$$Tg = T(h_1 g) + T(h_2 g) \leq \mu(V_1) + \mu(V_2).$$

Thus

$$\mu(V_1 \cup V_2) \leq \mu(V_1) + \mu(V_2).$$

(b) Choose $\varepsilon > 0$ and for each $i \in \mathbf{N}_+$, choose an open $V_i \supseteq E_i$ such $\mu(V_i) < \mu(E_i) + 2^{-i}\varepsilon$. Set $V = \cup_{i=1}^{\infty} V_i$ and choose $f \prec V$. Since $\text{supp} f$ is compact, $f \prec V_1 \cup ... \cup V_n$ for some n. Thus, by Part (a),

$$Tf \leq \mu(V_1 \cup ... \cup V_n) \leq \Sigma_{i=1}^{n} \mu(V_i) \leq \Sigma_{i=1}^{\infty} \mu(E_i) + \varepsilon$$

and we get

$$\mu(V) \leq \Sigma_{i=1}^{\infty} \mu(E_i)$$

since $\varepsilon > 0$ is arbitrary. But $\cup_{i=1}^{\infty} E_i \subseteq V$ and it follows that

$$\mu(\cup_{i=1}^{\infty} E_i) \leq \Sigma_{i=1}^{\infty} \mu(E_i).$$

(c) It is enough to treat the special case $n = 2$. Choose $\varepsilon > 0$. Set $\rho = d(K_1, K_2)$ and $V_1 = (K_1)_{\rho/2}$ and $V_2 = (K_2)_{\rho/2}$. There is an open set $U \supseteq K_1 \cup K_2$ such that $\mu(U) < \mu(K_1 \cup K_2) + \varepsilon$ and there are functions $f_i \prec U \cap V_i$ such that $Tf_i > \mu(U \cap V_i) - \varepsilon$ for $i = 1, 2$. Now, using that μ increases

$$\mu(K_1) + \mu(K_2) \leq \mu(U \cap V_1) + \mu(U \cap V_2)$$

$$\leq Tf_1 + Tf_2 + 2\varepsilon = T(f_1 + f_2) + 2\varepsilon.$$

Since $f_1 + f_2 \prec U$,

$$\mu(K_1) + \mu(K_2) \leq \mu(U) + 2\varepsilon \leq \mu(K_1 \cup K_2) + 3\varepsilon$$

and, by letting $\varepsilon \to 0$,

$$\mu(K_1) + \mu(K_2) \leq \mu(K_1 \cup K_2).$$

The reverse inequality follows from Part (b). The lemma is proved.

Next we introduce the class

$$\mathcal{M} = \left\{ E \in \mathcal{B}(X);\ \mu(E) = \sup_{\substack{K \subseteq E \\ K \text{ compact}}} \mu(K) \right\}$$

Since μ is increasing $\mathcal{M}$ contains every compact set. Recall that a closed set in X is compact, since X is compact. Especially, note that ϕ and $X \in \mathcal{M}$.

COMPLETION OF THE PROOF OF THEOREM 3.2.1:

CLAIM 1. $\mathcal{M}$ contains every open set.

PROOF OF CLAIM 1. Let V be open and suppose $\alpha < \mu(V)$. There exists an $f \prec V$ such that $\alpha < Tf$. If U is open and $U \supseteq K =_{def} \text{supp} f$, then $f \prec U$, and hence $Tf \leq \mu(U)$. But then $Tf \leq \mu(K)$. Thus $\alpha < \mu(K)$ and Claim 1 follows since K is compact and $K \subseteq V$.

CLAIM 2. Let $(E_i)_{i=1}^{\infty}$ be a disjoint denumerable collection of members of $\mathcal{M}$ and put $E = \cup_{i=1}^{\infty} E_i$. Then

$$\mu(E) = \Sigma_{i=1}^{\infty} \mu(E_i)$$

and $E \in \mathcal{M}$.

PROOF OF CLAIM 2. Choose $\varepsilon > 0$ and for each $i \in \mathbf{N}_+$, choose a compact $K_i \subseteq E_i$ such that $\mu(K_i) > \mu(E_i) - 2^{-i}\varepsilon$. Set $H_n = K_1 \cup ... \cup K_n$. Then, by Lemma 3.2.1 (c),

$$\mu(E) \geq \mu(H_n) = \Sigma_{i=1}^{n} \mu(K_i) > \Sigma_{i=1}^{n} \mu(E_i) - \varepsilon$$

and we get

$$\mu(E) \geq \Sigma_{i=1}^{\infty} \mu(E_i).$$

Thus, by Lemma 3.2.1 (b), $\mu(E) = \Sigma_{i=1}^{\infty} \mu(E_i)$. To prove that $E \in \mathcal{M}$, let ε be as in the very first part of the proof and choose n such that

$$\mu(E) \leq \Sigma_{i=1}^{n} \mu(E_i) + \varepsilon.$$

Then

$$\mu(E) < \mu(H_n) + 2\varepsilon$$

and this shows that $E \in \mathcal{M}$.

CLAIM 3. Suppose $E \in \mathcal{M}$ and $\varepsilon > 0$. Then there exist a compact K and an open V such that $K \subseteq E \subseteq V$ and $\mu(V \setminus K) < \varepsilon$.

PROOF OF CLAIM 3. The definitions show that there exist a compact K and an open V such that

$$\mu(V) - \frac{\varepsilon}{2} < \mu(E) < \mu(K) + \frac{\varepsilon}{2}.$$

The set $V \setminus K$ is open and $V \setminus K \in \mathcal{M}$ by Claim 1. Thus Claim 2 implies that

$$\mu(K) + \mu(V \setminus K) = \mu(V) < \mu(K) + \varepsilon$$

and we get $\mu(V \setminus K) < \varepsilon$.

CLAIM 4. If $A \in \mathcal{M}$, then $X \setminus A \in \mathcal{M}$.

PROOF OF CLAIM 4. Choose $\varepsilon > 0$. Furthermore, choose compact $K \subseteq A$ and open $V \supseteq A$ such that $\mu(V \setminus K) < \varepsilon$. Then

$$X \setminus A \subseteq (V \setminus K) \cup (X \setminus V).$$

Now, by Lemma 3.2.1 (b),

$$\mu(X \setminus A) \leq \varepsilon + \mu(X \setminus V).$$

Since $X \setminus V$ is a compact subset of $X \setminus A$, we conclude that $X \setminus A \in \mathcal{M}$.

Claims 1, 2 and 4 prove that $\mathcal{M}$ is a σ-algebra which contains all Borel sets. Thus $\mathcal{M} = \mathcal{B}(X)$.

We finally prove (a). It is enough to show that

$$Tf \le \int_X f d\mu$$

for each $f \in C(X)$. For once this is known

$$-Tf = T(-f) \le \int_X -f d\mu \le -\int_X f d\mu$$

and (a) follows.

Choose $\varepsilon > 0$. Set $f(X) = [a, b]$ and choose $y_0 < y_1 < ... < y_n$ such that $y_1 = a$, $y_{n-1} = b$, and $y_i - y_{i-1} < \varepsilon$. The sets

$$E_i = f^{-1}([y_{i-1}, y_i[), \; i = 1, ..., n$$

constitute a disjoint collection of Borel sets with the union X. Now, for each i, pick an open set $V_i \supseteq E_i$ such that $\mu(V_i) \le \mu(E_i) + \frac{\varepsilon}{n}$ and $V_i \subseteq f^{-1}(]-\infty, y_i[)$. By Theorem 3.2.2 there are functions $h_i \prec V_i$, $i = 1, ..., n$, such that $\Sigma_{i=1}^n h_i = 1$ on supp$f$ and $h_i f \prec y_i h_i$ for all i. From this we get

$$\begin{aligned} Tf = \Sigma_{i=1}^n T(h_i f) &\le \Sigma_{i=1}^n y_i T h_i \le \Sigma_{i=1}^n y_i \mu(V_i) \\ &\le \Sigma_{i=1}^n y_i \mu(E_i) + \Sigma_{i=1}^n y_i \frac{\varepsilon}{n} \\ &\le \Sigma_{i=1}^n (y_i - \varepsilon)\mu(E_i) + \varepsilon\mu(X) + (b + \varepsilon)\varepsilon \\ &\le \Sigma_{i=1}^n \int_{E_i} f d\mu + \varepsilon\mu(X) + (b + \varepsilon)\varepsilon \\ &= \int_X f d\mu + \varepsilon\mu(X) + (b + \varepsilon)\varepsilon. \end{aligned}$$

Since $\varepsilon > 0$ is arbitrary, we get

$$Tf \le \int_X f d\mu.$$

This proves Theorem 3.2.1.

It is now simple to show the existence of volume measure in $\mathbf{R}^n$. For pedagogical reasons we first discuss the so called volume measure in the unit cube $Q = [0, 1]^n$ in $\mathbf{R}^n$.

The Riemann integral

$$\int_Q f(x)dx,$$

is a positive linear functional as a function of $f \in C(Q)$. Moreover, $T1 = 1$ and the Riesz Representation Theorem gives us a Borel probability measure μ in Q such that

$$\int_Q f(x)dx = \int_Q f d\mu.$$

Suppose $A \subseteq Q$ is a closed n-cell and $i \in \mathbf{N}_+$. Then

$$\mathrm{vol}(A) \le \int_Q \Pi^Q_{A,2^{-i}}(x)dx \le \mathrm{vol}(A_{2^{-i}})$$

and

$$\Pi^Q_{A,2^{-i}}(x) \to \chi_A(x) \text{ as } i \to \infty$$

for every $x \in \mathbf{R}^n$. Thus

$$\mu(A) = \mathrm{vol}(A).$$

The measure μ is called the volume measure in the unit cube. In the special case $n = 2$ it is called the area measure in the unit square and if $n = 1$ it is called the linear measure in the unit interval.

PROOF OF THEOREM 1.1.1. Let $\hat{\mathbf{R}}=\mathbf{R}\cup\{-\infty,\infty\}$ be the two-point compactification of $\mathbf{R}$ introduced in Example 3.1.3 and let $\hat{\mathbf{R}}^n$ denote the product of n copies of the metric space $\hat{\mathbf{R}}$. Clearly,

$$\mathcal{B}(\mathbf{R}^n) = \left\{ A \cap \mathbf{R}^n;\ A \in \mathcal{B}(\hat{\mathbf{R}}^n) \right\}.$$

Moreover, let $w : \mathbf{R}^n \to]0,\infty[$ be a continuous map such that

$$\int_{\mathbf{R}^n} w(x)dx = 1.$$

Now we define

$$Tf = \int_{\mathbf{R}^n} f(x)w(x)dx,\ f \in C(\hat{\mathbf{R}}^n).$$

Note that $T1 = 1$. The function T is a positive linear functional on $C(\hat{\mathbf{R}}^n)$ and the Riesz Representation Theorem gives us a Borel probability measure μ on $\hat{\mathbf{R}}^n$ such that

$$\int_{\mathbf{R}^n} f(x)w(x)dx = \int_{\hat{\mathbf{R}}^n} fd\mu,\ f \in C(\hat{\mathbf{R}}^n).$$

As above we get

$$\int_A w(x)dx = \mu(A)$$

for each compact n-cell in $\mathbf{R}^n$. Thus

$$\mu(\mathbf{R}^n) = \lim_{i\to\infty} \int_{[-i,i]^n} w(x)dx = 1$$

and we conclude that μ is concentrated on $\mathbf{R}^n$. Set $\mu_0(A) = \mu(A)$, $A \in \mathcal{B}(\mathbf{R}^n)$, and

$$dm_n = \frac{1}{w}d\mu_0.$$

Then, if $f \in C_c(\mathbf{R}^n)$,

$$\int_{\mathbf{R}^n} f(x)w(x)dx = \int_{\mathbf{R}^n} fd\mu_0$$

and by replacing f by f/w,

$$\int_{\mathbf{R}^n} f(x)dx = \int_{\mathbf{R}^n} fdm_n.$$

From this $m_n(A) =$vol(A) for every compact n-cell A and it follows that m_n is the volume measure on $\mathbf{R}^n$. Theorem 1.1.1 is proved.

$\uparrow\uparrow\uparrow$

3.3 q-Adic Expansions of Numbers in the Unit Interval

To begin with in this section we will discuss so called q-adic expansions of real numbers and give some interesting consequences. As an example of an application, we construct a one-to-one real-valued Borel map f defined on a proper interval such that the range of f is a Lebesgue null set. Another example exhibits an increasing continuous function G on the unit interval with the range equal to the unit interval such that the derivative of G is equal to zero almost everywhere with respect to Lebesgue measure. In the next section we will give more applications of q-adic expansions in connection with infinite product measures.

To simplify notation let $(\Omega, P, \mathcal{F}) = ([0,1[\,, v_{1|[0,1[}, \mathcal{B}([0,1[))$. Furthermore, let $q \geq 2$ be an integer and define a function $h : \mathbf{R} \to \{0, 1, 2, ..., q-1\}$ of period one such that

$$h(x) = k,\ \frac{k}{q} \leq x < \frac{k+1}{q},\ k = 0, ..., q-1.$$

Furthermore, set for each $n \in \mathbf{N}_+$,

$$\xi_n(\omega) = h(q^{n-1}\omega),\ 0 \leq \omega < 1.$$

Then

$$P[\xi_n = k] = \frac{1}{q},\ k = 0, ..., q-1.$$

Moreover, if $k_1, ..., k_n \in \{0, 1, 2, ..., q-1\}$, it becomes obvious on drawing a figure that

$$P\left[\xi_1 = k_1, ..., \xi_{n-1} = k_{n-1}\right] = \Sigma_{i=0}^{q-1} P\left[\xi_1 = k_1, ..., \xi_{n-1} = k_{n-1}, \xi_n = i\right]$$

where each term in the sum in the right-hand side has the same value. Thus

$$P\left[\xi_1 = k_1, ..., \xi_{n-1} = k_{n-1}\right] = qP\left[\xi_1 = k_1, ..., \xi_{n-1} = k_{n-1}, \xi_n = k_n\right]$$

and

$$P\left[\xi_1 = k_1, ..., \xi_{n-1} = k_{n-1}, \xi_n = k_n\right] = P\left[\xi_1 = k_1, ..., \xi_{n-1} = k_{n-1}\right] P[\xi_n = k_n].$$

By repetition,

$$P\left[\xi_1 = k_1, ...\xi_{n-1} = k_{n-1}, \xi_n = k_n\right] = \Pi_{i=1}^{n} P[\xi_i = k_i].$$

From this we get

$$P\left[\xi_1 \in A_1, ...\xi_{n-1} \in A_{n-1}, \xi_n \in A_n\right] = \Pi_{i=1}^{n} P[\xi_i \in A_i]$$

for all $A_1, ..., A_n \subseteq \{0, 1, 2, ..., q-1\}$.

Note that each $\omega \in [0, 1[$ has a so called q-adic expansion

$$\omega = \Sigma_{i=1}^{\infty} \frac{\xi_i(\omega)}{q^i}.$$

If necessary, we write $\xi_n = \xi_n^{(q)}$ to indicate q explicitly.

Let $k_0 \in \{0, 1, 2, ..., q-1\}$ be fixed and consider the event A that a number in $[0, 1[$ does not have k_0 in its q-adic expansion. The probability of A equals

$$P[A] = P[\xi_i \neq k_0, \; i = 1, 2, ...] = \lim_{n\to\infty} P[\xi_i \neq k_0, \; i = 1, 2, ..., n]$$

$$= \lim_{n\to\infty} \Pi_{i=1}^{n} P[\xi_i \neq k_0] = \lim_{n\to\infty} (\frac{q-1}{q})^n = 0.$$

In particular, if

$$D_n = \left\{\omega \in [0, 1[\, ; \; \xi_i^{(3)} \neq 1, \; i = 1, ..., n\right\}.$$

then, $D = \cap_{n=1}^{\infty} D_n$ is a P-zero set.

Set

$$f(\omega) = \Sigma_{i=1}^{\infty} \frac{2\xi_i^{(2)}(\omega)}{3^i}, \; 0 \leq \omega < 1.$$

We claim that f is one-to-one. If $0 \leq \omega, \omega' < 1$ and $\omega \neq \omega'$ let n be the least i such that $\xi_i^{(2)}(\omega) \neq \xi_i^{(2)}(\omega')$; we may assume that $\xi_n^{(2)}(\omega) = 0$ and $\xi_n^{(2)}(\omega') = 1$. Then

$$f(\omega') \geq \Sigma_{i=1}^{n} \frac{2\xi_i^{(2)}(\omega')}{3^i} = \Sigma_{i=1}^{n-1} \frac{2\xi_i^{(2)}(\omega')}{3^i} + \frac{2}{3^n}$$

$$= \Sigma_{i=1}^{n-1} \frac{2\xi_i^{(2)}(\omega)}{3^i} + \Sigma_{i=n+1}^{\infty} \frac{4}{3^i} > \Sigma_{i=1}^{\infty} \frac{2\xi_i^{(2)}(\omega)}{3^i} = f(\omega).$$

Thus f is one-to-one. We next prove that $f(\Omega) = D$. To this end choose $y \in D$. If $\xi_i^{(3)}(y) = 2$ for all $i \in \mathbf{N}_+$, then $y = 1$ which is a contradiction. If $k \geq 1$ is fixed and $\xi_k^{(3)}(y) = 0$ and $\xi_i^{(3)}(y) = 2, \; i \geq k+1$, then it is readily seen that $\xi_k^{(3)}(y) = 1$ which is a contradiction. Now define

$$\omega = \Sigma_{i=1}^{\infty} \frac{\frac{1}{2}\xi_i^{(3)}(y)}{2^i}$$

and we have $f(\omega) = y$.

Let $C_n = D_n^-$, $n \in \mathbf{N}_+$. The set $C = \cap_{n=1}^{\infty} C_n$, is called the Cantor set. The Cantor set is a compact Lebesgue zero set. The construction of the Cantor set may alternatively be described as follows. First $C_0 = [0,1]$. Then trisect C_0 and remove the middle interval $\left]\frac{1}{3}, \frac{2}{3}\right[$ to obtain $C_1 = C_0 \setminus \left]\frac{1}{3}, \frac{2}{3}\right[= \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$. At the second stage subdivide each of the closed intervals of C_1 into thirds and remove from each one the middle open thirds. Then $C_2 = C_1 \setminus (\left]\frac{1}{9}, \frac{2}{9}\right[\cup \left]\frac{7}{9}, \frac{8}{9}\right[)$. What is left from C_{n-1} is C_n defined above. The set $[0,1] \setminus C_n$ is the union of $2^n - 1$ intervals numbered I_k^n, $k = 1, ..., 2^n - 1$, where the interval I_k^n is situated to the left of the interval I_l^n if $k < l$.

Suppose n is fixed and let $G_n : [0,1] \to [0,1]$ be the unique monotone increasing continuous function, which satisfies $G_n(0) = 0, G_n(1) = 1, G_n(x) = k2^{-n}$ for $x \in I_k^n$ and which is affine on each interval of C_n. It is clear that $G_n = G_{n+1}$ on each interval I_k^n, $k = 1, ..., 2^n - 1$. Moreover, $\mid G_n - G_{n+1} \mid \leq 2^{-n-1}$ and thus

$$\mid G_n - G_{n+k} \mid \leq \Sigma_{k=n}^{n+k} \mid G_k - G_{k+1} \mid \leq 2^{-n}.$$

Let $G(x) = \lim_{n\to\infty} G_n(x)$, $0 \leq x \leq 1$. The continuous and increasing function G is constant on each removed interval and it follows that $G' = 0$ a.e. with respect to linear measure in the unit interval.The function G is called the Cantor function or Cantor-Lebesgue function.

Next we introduce the following convention, which is standard in Lebesgue integration. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and suppose $A \in \mathcal{M}$ and $\mu(A^c) = 0$. If two functions $g, h \in \mathcal{L}^1(\mu)$ agree on A,

$$\int_X g d\mu = \int_X h d\mu.$$

If a function $f : A \to \mathbf{R}$ is the restriction to A of a function $g \in \mathcal{L}^1(\mu)$ we define

$$\int_X f d\mu = \int_X g d\mu.$$

Now suppose $F : \mathbf{R} \to \mathbf{R}$ is a right continuous increasing function and let μ be the unique positive Borel such that

$$\mu(]a, x]) = F(x) - F(a) \text{ if } a, x \in \mathbf{R} \text{ and } a < x.$$

If $h \in L^1(\mu)$ and $E \in \mathcal{R}$, the so called Stieltjes integral

$$\int_E h(x) dF(x)$$

is by definition equal to

$$\int_E h d\mu.$$

If $a, b \in \mathbf{R}$, $a < b$, and F is continuous at the points a and b, we define

$$\int_a^b h(x) dF(x) = \int_I h d\mu$$

where I is any interval with boundary points a and b.

The reader should note that the integral

$$\int_{\mathbf{R}} h(x) dF(x)$$

in general is different from the integral

$$\int_{\mathbf{R}} h(x) F'(x) dx.$$

For example, if G is the Cantor function and G is extended so that $G(x) = 0$ for negative x and $G(x) = 1$ for x larger than 1, clearly

$$\int_{\mathbf{R}} h(x) G'(x) dx = 0$$

since $G'(x) = 0$ a.e. $[m]$. On the other hand, if we choose $h = \chi_{[0,1]}$,

$$\int_{\mathbf{R}} h(x) dG(x) = 1.$$

3.4. Product Measures

Suppose $(X, \mathcal{M})$ and $(Y, \mathcal{N})$ are two measurable spaces. If $A \in \mathcal{M}$ and $B \in \mathcal{N}$, the set $A \times B$ is called a measurable rectangle in $X \times Y$. The product σ-algebra $\mathcal{M} \otimes \mathcal{N}$ is, by definition, the σ-algebra generated by all measurable rectangles in $X \times Y$. If we introduce the projections

$$\pi_X(x, y) = x, \ (x, y) \in X \times Y$$

and

$$\pi_Y(x, y) = y, \ (x, y) \in X \times Y,$$

the product σ-algebra $\mathcal{M} \otimes \mathcal{N}$ is the least σ-algebra $\mathcal{S}$ of subsets of $X \times Y$, which makes the maps $\pi_X : (X \times Y, \mathcal{S}) \to (X, \mathcal{M})$ and $\pi_Y : (X \times Y, \mathcal{S}) \to (Y, \mathcal{N})$ measurable, that is $\mathcal{M} \otimes \mathcal{N} = \sigma(\pi_X^{-1}(\mathcal{M}) \cup \pi_Y^{-1}(\mathcal{N}))$..

Suppose $\mathcal{E}$ generates $\mathcal{M}$, where $X \in \mathcal{E}$, and $\mathcal{F}$ generates $\mathcal{N}$, where $Y \in \mathcal{F}$. We claim that the class

$$\mathcal{E} \boxtimes \mathcal{F} = \{E \times F; E \in \mathcal{E} \text{ and } F \in \mathcal{F}\}$$

generates the σ-algebra $\mathcal{M} \otimes \mathcal{N}$. First it is clear that

$$\sigma(\mathcal{E} \boxtimes \mathcal{F}) \subseteq \mathcal{M} \otimes \mathcal{N}.$$

Moreover, the class

$$\left\{E \in \mathcal{M}; E \times Y \in \sigma(\mathcal{E} \boxtimes \mathcal{F})\ \right\} = \mathcal{M} \cap \left\{E \subseteq X;\ \pi_X^{-1}(E) \in \sigma(\mathcal{E} \boxtimes \mathcal{F})\ \right\}$$

is a σ-algebra, which contains $\mathcal{E}$ and therefore equals $\mathcal{M}$. Thus $A \times Y \in \sigma(\mathcal{E} \boxtimes \mathcal{F})$ for all $A \in \mathcal{M}$ and, in a similar way, $X \times B \in \sigma(\mathcal{E} \boxtimes \mathcal{F})$ for all $B \in \mathcal{N}$ and we conclude that $A \times B = (A \times Y) \cap (X \times B) \in \sigma(\mathcal{E} \boxtimes \mathcal{F})$ for all $A \in \mathcal{M}$ and all $B \in \mathcal{N}$. This proves that

$$\mathcal{M} \otimes \mathcal{N} \subseteq \sigma(\mathcal{E} \boxtimes \mathcal{F})$$

and it follows that

$$\sigma(\mathcal{E} \boxtimes \mathcal{F}) = \mathcal{M} \otimes \mathcal{N}.$$

Thus

$$\sigma(\mathcal{E} \boxtimes \mathcal{F}) = \sigma(\mathcal{E}) \otimes \sigma(\mathcal{F}) \text{ if } X \in \mathcal{E} \text{ and } Y \in \mathcal{F}.$$

Since the σ-algebra $\mathcal{R}_n$ is generated by all open n-cells in $\mathbf{R}^n$, we conclude that

$$\mathcal{R}_{k+n} = \mathcal{R}_k \otimes \mathcal{R}_n.$$

Given $E \subseteq X \times Y$, define

$$E_x = \{y;\ (x, y) \in E\} \text{ if } x \in X$$

and

$$E^y = \{x;\ (x, y) \in E\} \text{ if } y \in Y.$$

If $f : X \times Y \to Z$ is a function and $x \in X$, $y \in Y$, let

$$f_x(y) = f(x, y), \text{ if } y \in Y$$

and

$$f^y(x) = f(x,y), \text{ if } x \in X.$$

Theorem 3.4.1 (a) *If* $E \in \mathcal{M} \otimes \mathcal{N}$, *then* $E_x \in \mathcal{N}$ and $E^y \in \mathcal{M}$ *for every* $x \in X$ *and* $y \in Y$.

(b) *If* $f : (X \times Y, \mathcal{M} \otimes \mathcal{N}) \to (Z, \mathcal{O})$ *is measurable, then* f_x *is* $(\mathcal{N}, \mathcal{O})$-*measurable for each* $x \in X$ *and* f^y *is* $(\mathcal{M}, \mathcal{O})$-*measurable for each* $y \in Y$.

Proof. (a) Let

$$\mathcal{S} = \{E \in \mathcal{M} \otimes \mathcal{N}; E_x \in \mathcal{N} \text{ and } E^y \in \mathcal{M} \text{ for every } x \in X \text{ and } y \in Y\}.$$

Clearly, $X \times Y \in \mathcal{S}$. Furthermore, if $E, E_1, E_2, ... \in \mathcal{S}$, $(E^c)_x = (E_x)^c \in \mathcal{N}$ and $(\cup_{i=1}^{\infty} E_i)_x = \cup_{i=1}^{\infty} (E_i)_x \in \mathcal{N}$ for every x in X and $(E^c)^y = (E^y)^c \in \mathcal{M}$ and $(\cup_{i=1}^{\infty} E_i)^y = \cup_{i=1}^{\infty} (E_i)^y \in \mathcal{M}$ for every y in Y. It follows that $\mathcal{S}$ is a σ-algebra. Furthermore, if $A \in \mathcal{M}$ and $B \in \mathcal{N}$, $(A \times B)_x = B \in \mathcal{N}$ if $x \in A$ and $(A \times B)_x = \phi \in \mathcal{N}$ if $x \notin A$ and $(A \times B)^y = A \in \mathcal{M}$ if $y \in B$ and $(A \times B)^y = \phi \in \mathcal{M}$ if $y \notin B$. Thus $A \times B \in \mathcal{S}$ and, accordingly from this, $\mathcal{S} = \mathcal{M} \otimes \mathcal{N}$.

(b) For any set $V \in \mathcal{O}$,

$$(f^{-1}(V))_x = (f_x)^{-1}(V)$$

and

$$(f^{-1}(V))^y = (f^y)^{-1}(V).$$

Part (b) now follows from (a).

Below an $(\mathcal{M}, \mathcal{R}_{0,\infty})$-measurable or $(\mathcal{M}, \mathcal{R})$-measurable function is simply called $\mathcal{M}$-measurable.

Theorem 3.4.2. *Suppose* $(X, \mathcal{M}, \mu)$ *and* $(Y, \mathcal{N}, \nu)$ *are positive* σ-*finite measurable spaces and suppose* $E \in \mathcal{M} \otimes \mathcal{N}$. *If*

$$f(x) = \nu(E_x) \text{ and } g(y) = \mu(E^y)$$

for every $x \in X$ and $y \in Y$, then f is $\mathcal{M}$-measurable, g is $\mathcal{N}$-measurable, and

$$\int_X f d\mu = \int_Y g d\nu.$$

Proof. We first assume that $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are finite positive measure spaces.

Let $\mathcal{D}$ be the class of all sets $E \in \mathcal{M} \otimes \mathcal{N}$ for which the conclusion of the theorem holds. It is clear that the class $\mathcal{G}$ of all measurable rectangles in $X \times Y$ is a subset of $\mathcal{D}$ and $\mathcal{G}$ is a π-system. Furthermore, the Beppo Levi Theorem shows that $\mathcal{D}$ is a λ-system. Therefore, using Theorem 1.2.2, $\mathcal{M} \otimes \mathcal{N} = \sigma(\mathcal{G}) \subseteq \mathcal{D}$ and it follows that $\mathcal{D} = \mathcal{M} \otimes \mathcal{N}$.

In the general case, choose a denumerable disjoint collection $(X_k)_{k=1}^{\infty}$of members of $\mathcal{M}$ and a denumerable disjoint collection $(Y_n)_{n=1}^{\infty}$of members of $\mathcal{N}$ such that

$$\cup_{k=1}^{\infty} X_k = X \text{ and } \cup_{n=1}^{\infty} Y_n = Y.$$

Set

$$\mu_k = \chi_{X_k}\mu,\ k = 1, 2, ...$$

and

$$\nu_n = \chi_{Y_n}\nu,\ n = 1, 2, ... \ .$$

Then, by the Beppo Levi Theorem, the function

$$f(x) = \int_X \Sigma_{n=1}^{\infty} \chi_E(x, y)\chi_{Y_n}(y)d\nu(y)$$

$$= \Sigma_{n=1}^{\infty} \int_X \chi_E(x, y)\chi_{Y_n}(y)d\nu(y) = \Sigma_{n=1}^{\infty} \nu_n(E_x)$$

is $\mathcal{M}$-measurable. Again, by the Beppo Levi Theorem,

$$\int_X f d\mu = \Sigma_{k=1}^{\infty} \int_X f d\mu_k$$

and

$$\int_X f d\mu = \Sigma_{k=1}^{\infty}(\Sigma_{n=1}^{\infty} \int_X \nu_n(E_x)d\mu_k(x)) = \Sigma_{k,n=1}^{\infty} \int_X \nu_n(E_x)d\mu_k(x).$$

In a similar way, the function g is $\mathcal{N}$-measurable and

$$\int_Y g d\nu = \Sigma_{n=1}^{\infty}(\Sigma_{k=1}^{\infty} \int_Y \mu_k(E^y)d\nu_n(y)) = \Sigma_{k,n=1}^{\infty} \int_Y \mu_k(E^y)d\nu_n(y).$$

Since the theorem is true for finite positive measure spaces, the general case follows.

Definition 3.4.1. If $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ are positive σ-finite measurable spaces and $E \in \mathcal{M} \otimes \mathcal{N}$, define

$$(\mu \times \nu)(E) = \int_X \nu(E_x) d\mu(x) = \int_Y \mu(E^y) d\nu(y).$$

The function $\mu \times \nu$ is called the product of the measures μ and ν.

Note that Beppo Levi's Theorem ensures that $\mu \times \nu$ is a positive measure.

Before the next theorem we recall the following convention. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and suppose $A \in \mathcal{M}$ and $\mu(A^c) = 0$. If two functions $g, h \in \mathcal{L}^1(\mu)$ agree on A,

$$\int_X g d\mu = \int_X h d\mu.$$

If a function $f : A \to \mathbf{R}$ is the restriction to A of a function $g \in \mathcal{L}^1(\mu)$ we define

$$\int_X f d\mu = \int_X g d\mu.$$

Theorem 3.4.3. *Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be positive σ-finite measurable spaces.*

(a) **(Tonelli's Theorem)** *If $h : X \times Y \to [0, \infty]$ is $(\mathcal{M} \otimes \mathcal{N})$-measurable and*

$$f(x) = \int_Y h(x, y) d\nu(y) \text{ and } g(y) = \int_X h(x, y) d\mu(x)$$

for every $x \in X$ and $y \in Y$, then f is $\mathcal{M}$-measurable, g is $\mathcal{N}$-measurable, and

$$\int_X f d\mu = \int_{X \times Y} h d(\mu \times \nu) = \int_Y g d\nu$$

(b) **(Fubini's Theorem)**

(i) If $h : X \times Y \to \mathbf{R}$ *is* $(\mathcal{M} \otimes \mathcal{N})$*-measurable and*

$$\int_X (\int_Y \mid h(x,y) \mid d\nu(y)) d\mu(x) < \infty$$

then $h \in L^1(\mu \times \nu)$*. Moreover,*

$$\int_X (\int_Y h(x,y) d\nu(y)) d\mu(x) = \int_{X \times Y} h d(\mu \times \nu) = \int_Y (\int_X h(x,y) d\mu(x)) d\nu(y)$$

(ii) If $h \in L^1((\mu \times \nu)^-)$*, then* $h_x \in L^1(\nu)$ *for* μ*-almost all* x *and*

$$\int_{X \times Y} h d(\mu \times \nu) = \int_X (\int_Y h(x,y) d\nu(y)) d\mu(x)$$

(iii) If $h \in L^1((\mu \times \nu)^-)$*, then* $h^y \in L^1(\mu)$ *for* ν*-almost all* y *and*

$$\int_{X \times Y} h d(\mu \times \nu) = \int_Y (\int_X h(x,y) d\mu(x)) d\nu(y)$$

PROOF. (a) The special case when h is a non-negative $(\mathcal{M} \otimes \mathcal{N})$-measurable simple function follows from Theorem 3.4.2. Remembering that any non-negative measurable function is the pointwise limit of an increasing sequence of simple measurable functions, the Lebesgue Monotone Convergence Theorem implies the Tonelli Theorem.

(b) PART (i) : By Part (a)

$$\infty > \int_X (\int_Y h^+(x,y) d\nu(y)) d\mu(x) = \int_{X \times Y} h^+ d(\mu \times \nu)$$

$$= \int_Y (\int_X h^+(x,y) d\mu(x)) d\nu(y)$$

and

$$\infty > \int_X (\int_Y h^-(x,y) d\nu(y)) d\mu(x) = \int_{X \times Y} h^- d(\mu \times \nu)$$

$$= \int_Y (\int_X h^-(x,y) d\mu(x)) d\nu(y).$$

Let

$$A = \left\{x \in X;\ (h^+)_x, (h^-)_x \in L^1(\nu)\right\}.$$

Then A^c is a μ-null set and we get

$$\int_A (\int_Y h^+(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} h^+ d(\mu\times\nu)$$

and

$$\int_A (\int_Y h^-(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} h^- d(\mu\times\nu).$$

Thus

$$\int_A (\int_Y h(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} hd(\mu\times\nu)$$

and, hence,

$$\int_X (\int_Y h(x,y)d\nu(y))d\mu(x) = \int_{X\times Y} hd(\mu\times\nu).$$

The other case can be treated in a similar way. The theorem is proved.

PART (ii) : We first use Theorem 2.2.3 and write $h = \varphi + \psi$ where $\varphi \in L^1(\mu\times\nu)$, ψ is $(\mathcal{M}\otimes\mathcal{N})^-$-measurable and $\psi = 0$ a.e. $[\mu\times\nu]$. Set

$$A = \left\{x \in X;\ (\varphi^+)_x, (\varphi^-)_x \in L^1(\nu)\right\}.$$

Furthermore, suppose $E \supseteq \{(x,y);\ \psi(x,y) \neq 0\}$, $E \in \mathcal{M}\otimes\mathcal{N}$ and

$$(\mu\times\nu)(E) = 0.$$

Then, by Tonelli's Theorem

$$0 = \int_X \nu(E_x)d\mu(x).$$

Let $B = \{x \in X;\ \nu(E_x) \neq 0\}$ and note that $B \in \mathcal{M}$. Moreover $\mu(B) = 0$ and if $x \notin B$, then $\psi_x = 0$ a.e. $[\nu]$ that is $h_x = \varphi_x$ a.e. $[\nu]$. Now, by Part (i)

$$\int_{X\times Y} hd(\mu\times\nu)^- = \int_{X\times Y} \varphi d(\mu\times\nu) = \int_A (\int_Y \varphi(x,y)d\nu(y))d\mu(x)$$

$$= \int_{A\cap B^c} (\int_Y \varphi(x,y) d\nu(y)) d\mu(x) = \int_{A\cap B^c} (\int_Y h(x,y) d\nu(y)) d\mu(x)$$
$$= \int_X (\int_Y h(x,y) d\nu(y)) d\mu(x).$$

Part (iii) is proved in the same manner as Part (ii). This concludes the proof of the theorem.

If $(X_i, \mathcal{M}_i)$, $i = 1, ..., n$, are measurable spaces, the product σ-algebra $\mathcal{M}_1 \otimes ... \otimes \mathcal{M}_n$ is, by definition, the σ-algebra generated by all sets of the form

$$A_1 \times ... \times A_n$$

where $A_i \in \mathcal{M}_i$, $i = 1, ..., n$. Now assume $(X_i, \mathcal{M}_i, \mu_i)$, $i = 1, ..., n$, are σ-finite positive measure spaces. By induction, we define $\nu_1 = \mu_1$ and $\nu_k = \nu_{k-1} \times \mu_k$, $k = 1, 2, ..., n$. The measure, ν_n is called the product of the measures $\mu_1, ..., \mu_n$ and is denoted by $\mu_1 \times ... \times \mu_n$. It is readily seen that

$$\mathcal{R}_n = \mathcal{R}_1 \otimes ... \otimes \mathcal{R}_1 \text{ (n factors)}$$

and

$$v_n = v_1 \times ... \times v_1 \text{ (n factors).}$$

Moreover,

$$\mathcal{R}_n^- \supseteq (\mathcal{R}_1^-)^n =_{def} \mathcal{R}_1^- \otimes ... \otimes \mathcal{R}_1^- \quad \text{(n factors).}$$

If $A \in \mathcal{P}(\mathbf{R}) \setminus \mathcal{R}_1^-$, by the Tonelli Theorem, the set $A \times \{0, ..., 0\}$ ($n - 1$ zeros) is an m_n-null set, which, in view of Theorem 3.4.1, cannot belong to the σ-algebra $(\mathcal{R}_1^-)^n$. Thus the Axiom of Choice implies that

$$\mathcal{R}_n^- \neq (\mathcal{R}_1^-)^n.$$

Clearly, the completion of the measure $m_1 \times ... \times m_1$ (n factors) equals m_n.

Sometimes we prefer to write

$$\int_{A_1 \times ... \times A_n} f(x_1, ..., x_n) dx_1 ... dx_n$$

instead of

$$\int_{A_1\times\ldots\times A_n} f(x)dm_n(x)$$

or

$$\int_{A_1\times\ldots\times A_n} f(x)dx.$$

Moreover, the integral

$$\int_{A_1}\cdots\int_{A_n} f(x_1,...,x_n)dx_1...dx_n$$

is the same as

$$\int_{A_1\times\ldots\times A_n} f(x_1,...,x_n)dx_1...dx_n.$$

Definition 3.4.2. (a) The measure

$$\gamma_1(A) = \int_A e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}},\ A\in\mathcal{R}$$

is called the canonical Gauss measure in $\mathbf{R}$.

(b) The measure

$$\gamma_n = \gamma_1\times\ldots\times\gamma_1\ (n \text{ factors})$$

is called the canonical Gauss measure in $\mathbf{R}^n$. Thus, if

$$\mid x\mid = \sqrt{x_1^2+\ldots+x_n^2},\ x=(x_1,...,x_n)\in\mathbf{R}^n$$

we have

$$\gamma_n(A) = \int_A e^{-\frac{|x|^2}{2}}\frac{dx}{\sqrt{2\pi}},\ A\in\mathcal{R}_n.$$

(c) A Borel measure μ in $\mathbf{R}$ is said to be a centred Gaussian measure if $\mu = f(\gamma_1)$ for some linear map $f:\mathbf{R}\to\mathbf{R}$.

(d) A real-valued random variable ξ is said to be a centred Gaussian random variable if its probability law is a centred Gaussian measure in $\mathbf{R}$. Stated otherwise, ξ is a real-valued centred Gaussian random variable if either

$$\mathcal{L}(\xi) = \delta_0 \text{ (abbreviated } \xi \in N(0,0))$$

or there exists a $\sigma > 0$ such that

$$\mathcal{L}(\frac{\xi}{\sigma}) = \gamma_1 \text{ (abbreviated } \xi \in N(0,\sigma)).$$

(e) A family $(\xi_t)_{t \in T}$ of real-valued random variables is said to be a centred real-valued Gaussian process if for all $t_1, ..., t_n \in T$, $\alpha_1, ..., \alpha_n \in \mathbf{R}$ and every $n \in \mathbf{N}_+$, the sum

$$\xi = \Sigma_{k=1}^{n} \alpha_k \xi_{t_k}$$

is a centred Gaussian random variable.

Exercises

1. Let $(X, \mathcal{M}, \mu)$ and $(Y, \mathcal{N}, \nu)$ be two σ-finite measure spaces. Let $f \in L^1(\mu)$ and $g \in L^1(\nu)$ and define $h(x,y) = f(x)g(y)$, $(x,y) \in X \times Y$. Prove that $h \in L^1(\mu \times \nu)$ and

$$\int_{X \times Y} h d(\mu \times \nu) = \int_X f d\mu \int_Y g d\nu.$$

2. Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and $f : X \to [0, \infty[$ a measurable function. Prove that

$$\int_X f d\mu = (\mu \times m) \{(x,y);\ 0 < y < f(x),\ x \in X\}.$$

3. Let $(X, \mathcal{M}, \mu)$ be a σ-finite measure space and $f : X \to \mathbf{R}$ a measurable function. Prove that $(\mu \times m)(\{(x, f(x));\ x \in X\ \}) = 0$.

4. Let $E \in \mathcal{R}_2^-$ and $E \subseteq [0,1] \times [0,1]$. Suppose $m(E_x) \leq \frac{1}{2}$ for m-almost all $x \in [0,1]$. Show that

$$m(\{y \in [0,1]\,; m(E^y) = 1\}) \leq \frac{1}{2}.$$

5. Let c be the counting measure on $\mathbf{R}$ restricted to $\mathcal{R}$ and

$$D = \{(x,x);\ x \in \mathbf{R}\}.$$

Define for every $A \in (\mathcal{R} \boxtimes \mathcal{R}) \cup \{D\}$,

$$\mu(A) = \int_{\mathbf{R}} (\int_{\mathbf{R}} \chi_A(x,y) dv_1(x)) dc(y)$$

and

$$\nu(A) = \int_{\mathbf{R}} (\int_{\mathbf{R}} \chi_A(x,y) dc(y)) dv_1(x).$$

(a) Prove that μ and ν agree on $\mathcal{R} \boxtimes \mathcal{R}$.
(b) Prove that $\mu(D) \neq \nu(D)$.

6. Let $I =]0,1[$ and

$$h(x,y) = \frac{x^2 - y^2}{(x^2 + y^2)^2},\ (x,y) \in I \times I.$$

Prove that

$$\int_I (\int_I h(x,y) dy) dx = \frac{\pi}{4},$$

$$\int_I (\int_I h(x,y) dx) dy = -\frac{\pi}{4}$$

and

$$\int_{I \times I} \mid h(x,y) \mid dxdy = \infty.$$

7. For $t > 0$ and $x \in \mathbf{R}$ let

$$g(t,x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}}$$

and

$$h(t,x) = \frac{\partial g}{\partial t}.$$

Given $a > 0$, prove that

$$\int_{-\infty}^{\infty} (\int_{a}^{\infty} h(t,x)dt)dx = -1$$

and

$$\int_{a}^{\infty} (\int_{-\infty}^{\infty} h(t,x)dx)dt = 0$$

and conclude that

$$\int_{[a,\infty[\times \mathbf{R}} \mid h(t,x) \mid dtdx = \infty.$$

(Hint: First prove that

$$\int_{-\infty}^{\infty} g(t,x)dx = 1$$

and

$$\frac{\partial g}{\partial t} = \frac{1}{2}\frac{\partial^2 g}{\partial x^2}.)$$

8. Given $f \in L^1(m)$, let

$$g(x) = \frac{1}{2}\int_{x-1}^{x+1} f(t)dt,\ x \in \mathbf{R}.$$

Prove that

$$\int_{\mathbf{R}} \mid g(x) \mid dx \le \int_{\mathbf{R}} \mid f(x) \mid dx.$$

9. Let $I = [0,1]$ and suppose $f : I \to \mathbf{R}$ is a Lebesgue measurable function such that

$$\int_{I\times I} \mid f(x) - f(y) \mid dxdy < \infty.$$

Prove that

$$\int_{I} \mid f(x) \mid dx < \infty.$$

10. Suppose $A \in \mathcal{R}^-$ and $f \in L^1(m)$. Set

$$g(x) = \int_{\mathbf{R}} \frac{d(y,A)f(y)}{\mid x-y \mid^2}dy,\ x \in \mathbf{R}.$$

Prove that

$$\int_A \mid g(x) \mid dx < \infty.$$

11. Suppose that the functions $f, g : \mathbf{R} \to [0, \infty[$ are Lebesgue measurable and introduce $\mu = fm$ and $\nu = gm$. Prove that the measures μ and ν are σ-finite and

$$(\mu \times \nu)(E) = \int_E f(x)g(y)dxdy \text{ if } E \in \mathcal{R}^- \otimes \mathcal{R}^-.$$

12. Suppose μ is a finite positive Borel measure on $\mathbf{R}^n$ and $f : \mathbf{R}^n \to \mathbf{R}$ a Borel measurable function. Set $g(x, y) = f(x) - f(y)$, $x, y \in \mathbf{R}^n$. Prove that $f \in L^1(\mu)$ if and only if $g \in L^1(\mu \times \mu)$.

13. A random variable ξ is non-negative and possesses the distribution function $F(x) = P[\xi \leq x]$. Prove that$E[\xi] = \int_0^\infty (1 - F(x))dx$.

14. Let (X, d) be a metric space and suppose $Y \in \mathcal{B}(X)$. Then Y equipped with the metric $d_{|Y \times Y}$ is a metric space. Prove that

$$\mathcal{B}(Y) = \{A \cap Y;\ A \in \mathcal{B}(X)\}.$$

15. The continuous bijection $f : (X, d) \to (Y, e)$ has a continuous inverse. Prove that $f(A) \in \mathcal{B}(Y)$ if $A \in \mathcal{B}(X)$

16. A real-valued function $f(x, y), x, y \in \mathbf{R}$, is a Borel function of x for every fixed y and a continuous function of y for every fixed x. Prove that f is a Borel function. Is the same conclusion true if we only assume that $f(x, y)$ is a real-valued Borel function in each variable separately?

17. Suppose $a > 0$ and

$$\mu_a = e^{-a} \sum_{n=0}^{\infty} \frac{a^n}{n!} \delta_n$$

where $\delta_n(A) = \chi_A(n)$ if $n \in \mathbf{N} = \{0, 1, 2, ...\}$ and $A \subseteq \mathbf{N}$. Prove that

$$(\mu_a \times \mu_b) s^{-1} = \mu_{a+b}$$

for all $a, b > 0$, if $s(x, y) = x + y$, $x, y \in \mathbf{N}$.

18. Suppose

$$f(t) = \int_0^{\infty} \frac{xe^{-x^2}}{x^2 + t^2} dx, \ t > 0.$$

Compute

$$\lim_{t \to 0+} f(t) \text{ and } \int_0^{\infty} f(t) dt.$$

Finally, prove that f is differentiable.

3.5 Change of Variables in Volume Integrals

If T is a non-singular n by n matrix with real entries, we claim that

$$T(v_n) = \frac{1}{\mid \det T \mid} v_n$$

(here T is viewed as a linear map of $\mathbf{R}^n$ into $\mathbf{R}^n$). Remembering Corollary 3.1.3 this means that the following linear change of variables formula holds, viz.

$$\int_{\mathbf{R}^n} f(Tx) dx = \frac{1}{\mid \det T \mid} \int_{\mathbf{R}^n} f(x) dx \text{ all } f \in C_c(\mathbf{R}^n).$$

The case $n = 1$ is obvious. Moreover, by Fubini's Theorem the linear change of variables formula is true for arbitrary n in the following cases:

(a) $Tx = (x_{\pi(1)}, ..., x_{\pi(n)})$, where π is a permutation of the numbers $1, ..., n$.

(b) $Tx = (\alpha x_1, x_2, ..., x_n)$, where α is a non-zero real number.

(c) $Tx = (x_1 + x_2, x_2, ..., x_n)$.

Recall from linear algebra that every non-singular n by n matrix T can be row-reduced to the identity matrix, that is T can by written as the product of finitely many transformations of the types in (a),(b), and (c). This proves the above linear change of variables formula.

Our main objective in this section is to prove a more general change of variable formula. To this end let Ω and Γ be open subsets of $\mathbf{R}^n$ and $G : \Omega \to \Gamma$ a C^1 diffeomorphism, that is $G = (g_1, ..., g_n)$ is a bijective continuously differentiable map such that the matrix $G'(x) = (\frac{\partial g_i}{\partial x_j}(x))_{1 \leq i,j \leq n}$ is non-singular for each $x \in \Omega$. The inverse function theorem implies that $G^{-1} : \Gamma \to \Omega$ is a C^1 diffeomorphism $[DI]$.

Theorem 3.5.1. *If f is a non-negative Borel function in Ω, then*

$$\int_\Gamma f(x)dx = \int_\Omega f(G(x)) \mid \det G'(x) \mid dx.$$

The proof of Theorem 3.5.1 is based on several lemmas.

Throughout, $\mathbf{R}^n$ is equipped with the metric

$$d_n(x, y) = \max_{1 \leq k \leq n} \mid x_k - y_k \mid .$$

Let K be a compact convex subset of Ω. Then if $x, y \in K$ and $1 \leq i \leq n$,

$$g_i(x) - g_i(y) = \int_0^1 \frac{d}{dt} g_i(y + t(x - y))dt$$

$$= \int_0^1 \Sigma_{k=1}^n \frac{\partial g_i}{\partial x_k}(y + t(x - y))(x_k - y_k)dt$$

and we get

$$d_n(G(x), G(y)) \leq M(G, K)d_n(x, y)$$

where

$$M(G, K) = \max_{1 \leq i \leq n} \Sigma_{k=1}^n \max_{z \in K} \mid \frac{\partial g_i}{\partial x_k}(z) \mid .$$

Thus if $\bar{B}(a; r)$ is a closed ball contained in K,

$$G(\bar{B}(a; r)) \subseteq \bar{B}(G(a); M(G, K)r).$$

Lemma 3.5.1. *Let* $(Q_k)_{k=1}^{\infty}$ *be a sequence of closed balls contained in* Ω *such that*

$$Q_{k+1} \subseteq Q_k$$

and

$$\text{diam } Q_k \to 0 \text{ as } k \to \infty.$$

Then, there is a unique point a *belonging to each* Q_k *and*

$$\limsup_{n\to\infty} \frac{v_n(G(Q_k))}{v_n(Q_k)} \leq \mid \det G'(a) \mid .$$

PROOF. The existence of a point a belonging to each Q_k is an immediate consequence of Theorem 3.1.2. The uniqueness is also obvious since the diameter of Q_k converges to 0 as $k \to \infty$. Set $T = G'(a)$ and $F = T^{-1}G$. Then, if $Q_k = \bar{B}(x_k; r_k)$,

$$v_n(G(Q_k)) = v_n(T(T^{-1}G(Q_k))) = \mid \det T \mid \; v_n(T^{-1}G(\bar{B}(x_k; r_k)))$$

$$\leq \mid \det T \mid v_n(\bar{B}(T^{-1}G(x_k); M(T^{-1}G; Q_k)r_k) = \mid \det T \mid \; M(T^{-1}G; Q_k)^n v_n(Q_k).$$

Since

$$\lim_{k\to\infty} M(T^{-1}G; Q_k) = 1$$

the lemma follows at once.

Lemma 3.5.2. *Let* Q *be a closed ball contained in* Ω. *Then*

$$v_n(G(Q)) \leq \int_Q \mid \det G'(x) \mid dx.$$

PROOF. Suppose there is a closed ball Q contained in Ω such that

$$v_n(G(Q)) > \int_Q \mid \det G'(x) \mid dx.$$

This will lead us to a contradiction as follows.

Choose $\varepsilon > 0$ such that

$$v_n(G(Q)) \geq (1+\varepsilon) \int_Q \mid \det G'(x) \mid dx.$$

Let $Q = \cup_1^{2^n} B_k$ where $B_1, ..., B_{2^n}$ are mutually almost disjoint closed balls with the same volume. If

$$v_n(G(B_k)) < (1+\varepsilon) \int_{B_k} \mid \det G'(x) \mid dx, \;\; k = 1, ..., 2^n$$

we get

$$v_n(G(Q)) \leq \Sigma_{k=1}^{2^n} v_n(G(B_k))$$

$$< \Sigma_{k=1}^{2^n} (1+\varepsilon) \int_{B_k} \mid \det G'(x) \mid dx = (1+\varepsilon) \int_Q \mid \det G'(x) \mid dx$$

which is a contradiction. Thus

$$v_n(G(B_k)) \geq (1+\varepsilon) \int_{B_k} \mid \det G'(x) \mid dx$$

for some k. By induction we obtain a sequence $(Q_k)_{k=1}^{\infty}$ of closed balls contained in Ω such that

$$Q_{k+1} \subseteq Q_k,$$

$$\text{diam } Q_k \to 0 \text{ as } k \to \infty$$

and

$$v_n(G(Q_k)) \geq (1+\varepsilon) \int_{Q_k} \mid \det G'(x) \mid dx.$$

But applying Lemma 3.5.1 we get a contradiction.

PROOF OF THEOREM 3.5.1. Let $U \subseteq \Omega$ be open and write $U = \cup_{i=1}^{\infty} Q_i$ where the $Q_i's$ are almost disjoint cubes as in Theorem 3.1.5. Then

$$v_n(G(U)) \leq \Sigma_{i=1}^{\infty} v_n(G(Q_i)) \leq \Sigma_{i=1}^{\infty} \int_{Q_i} \mid \det G'(x) \mid dx$$

$$= \int_U \mid \det G'(x) \mid dx.$$

Using Theorem 3.1.3 we now have that

$$v_n(G(E)) \leq \int_E \mid \det G'(x) \mid dx$$

for each Borel set $E \subseteq \Omega$. But then

$$\int_\Gamma f(x)dx \leq \int_\Omega f(G(x)) \mid \det G'(x) \mid dx$$

for each simple Borel measurable function $f \geq 0$ and, accordingly from this and monotone convergence, the same inequality holds for each non-negative Borel function f. But the same line of reasoning applies to G replaced by G^{-1} and f replaced by $f(G) \mid \det G' \mid$, so that

$$\int_\Omega f(G(x)) \mid \det G'(x) \mid dx \leq \int_\Gamma f(x) \mid \det G'(G^{-1}(x)) \mid\mid \det(G^{-1})'(x) \mid dx$$

$$= \int_\Gamma f(x)dx.$$

This proves the theorem.

Example 3.5.1. If $f : \mathbf{R}^2 \to [0, \infty]$ is $(\mathcal{R}_2, \mathcal{R}_{0,\infty})$-measurable and $0 < \varepsilon < R < \infty$, the substitution

$$G(r, \theta) = (r \cos \theta, r \sin \theta)$$

yields

$$\int_{\varepsilon < \sqrt{x_1^2 + x_2^2} < R} f(x_1, x_2)dx_1dx_2 = \int_\varepsilon^R \int_0^{2\pi} f(r \cos \theta, r \sin \theta)rdrd\theta$$

and by letting $\varepsilon \to 0$ and $R \to \infty$, we have

$$\int_{\mathbf{R}^2} f(x_1, x_2)dx_1dx_2 = \int_0^\infty \int_0^{2\pi} f(r \cos \theta, r \sin \theta)rdrd\theta.$$

The purpose of the example is to show an analogue formula for volume measure in $\mathbf{R}^n$.

Let $S^{n-1} = \{x \in \mathbf{R}^n; \mid x \mid= 1\}$ be the unit sphere in $\mathbf{R}^n$. We will define a so called surface area Borel measure σ_{n-1} on S^{n-1} such that

$$\int_{\mathbf{R}^n} f(x)dx = \int_0^\infty \int_{S^{n-1}} f(r\omega)r^{n-1}drd\sigma_{n-1}(\omega)$$

for any $(\mathcal{R}_n, \mathcal{R}_{0,\infty})$-measurable function $f : \mathbf{R}^n \to [0, \infty]$. To this end define $G : \mathbf{R}^n \setminus \{0\} \to]0, \infty[\times S^{n-1}$ by setting $G(x) = (r, \omega)$, where

$$r = \mid x \mid \text{ and } \omega = \frac{x}{\mid x \mid}.$$

Note that $G^{-1} :]0, \infty[\times S^{n-1} \to \mathbf{R}^n \setminus \{0\}$ is given by the equation

$$G^{-1}(r, \omega) = r\omega.$$

Moreover,

$$G^{-1}(]0, a] \times E) = aG^{-1}(]0, 1] \times E) \text{ if } a > 0 \text{ and } E \subseteq S^{n-1}.$$

If $E \in \mathcal{B}(S^{n-1})$ we therefore have that

$$v_n(G^{-1}(]0, a] \times E)) = a^n v_n(G^{-1}(]0, 1] \times E)).$$

We now define

$$\sigma_{n-1}(E) = nv_n(G^{-1}(]0, 1] \times E)) \text{ if } E \in \mathcal{B}(S^{n-1})$$

and

$$\rho(A) = \int_A r^{n-1} dr \text{ if } A \in \mathcal{B}(]0, \infty[).$$

Below, by abuse of language, we write $v_{n|\mathbf{R}^n \setminus \{0\}} = v_n$. Then, if $0 < a \leq b < \infty$ and $E \in \mathcal{B}(S^{n-1})$,

$$G(v_n)(]0, a] \times E) = \rho(]0, a])\sigma_{n-1}(E)$$

and

$$G(v_n)(]a, b] \times E) = \rho(]a, b])\sigma_{n-1}(E).$$

Thus, by Theorem 1.2.3,

$$G(v_n) = \rho \times \sigma_{n-1}$$

and the claim above is immediate.

To check the normalization constant in the definition of σ_{n-1}, first note that

$$v_n(\mid x \mid < R) = \int_0^R \int_{S^{n-1}} r^{n-1} dr d\sigma(\omega) = \frac{R^n}{n} \sigma_{n-1}(S^{n-1})$$

and we get

$$\frac{d}{dR}v_n(\mid x \mid< R) = R^{n-1}\sigma_{n-1}(S^{n-1}).$$

Exercises

1. Extend Theorem 3.5.1 to Lebesgue measurable functions.

2. The function $f : \mathbf{R} \rightarrow [0, \infty[$ is Lebesgue measurable and $\int_{\mathbf{R}} f dm = 1$. Determine all non-zero real numbers α such that $\int_{\mathbf{R}} hdm < \infty$, where

$$h(x) = \Sigma_{n=0}^{\infty} f(\alpha^n x + n),\ \ x \in \mathbf{R}.$$

$\downarrow\downarrow\downarrow$

3.6. Independence in Probability

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space. The random variables $\xi_k : (\Omega, P) \rightarrow (S_k, \mathcal{S}_k)$, $k = 1, ..., n$ are said to be independent if

$$P_{(\xi_1,...,\xi_n)} = \times_{k=1}^{n} P_{\xi_k}.$$

A family $(\xi_i)_{i\in I}$ of random variables is said to be independent if $\xi_{i_1}, ..., \xi_{i_n}$ are independent for any $i_1, ...i_n \in I$ with $i_k \neq i_l$ if $k \neq l$. A family of events $(A_i)_{i\in I}$ is said to be independent if $(\chi_{A_i})_{i\in I}$ is a family of independent random variables. Finally a family $(\mathcal{A}_i)_{i\in I}$ of sub-σ-algebras of $\mathcal{F}$ is said to be independent if, for any $A_i \in \mathcal{A}_i$, $i \in I$, the family $(A_i)_{i\in I}$ is a family of independent events.

Example 3.6.1. Let $q \geq 2$ be an integer. A real number $\omega \in [0, 1[$ has a q-adic expansion

$$\omega = \Sigma_{k=1}^{\infty} \frac{\xi_k^{(q)}}{q^k}.$$

The construction of the Cantor set shows that $(\xi_k^{(q)})_{k=1}^{\infty}$ is a sequence of independent random variables based on the probability space

$$([0,1[\,, v_{1|[0,1[}, \mathcal{B}([0,1[)).$$

Example 3.6.2. Let $(X, \mathcal{M}, \mu)$ be a positive measure space and let $A_i \in \mathcal{M}$, $i \in \mathbf{N}_+$, be such that

$$\Sigma_{i=1}^{\infty} \mu(A_i) < \infty.$$

The first Borel-Cantelli Lemma asserts that μ-almost all $x \in X$ lie in A_i for at most finitely many i. This result is an immediate consequence of the Beppo Levi Theorem since

$$\int_X \Sigma_{i=1}^{\infty} \chi_{A_i} d\mu = \Sigma_{i=1}^{\infty} \int_X \chi_{A_i} d\mu < \infty$$

implies that

$$\Sigma_{i=1}^{\infty} \chi_{A_i} < \infty \text{ a.e. } [\mu]\,.$$

Suppose $(\Omega, \mathcal{F}, P)$ is a probability space and let $(A_i)_{i=1}^{\infty}$ be independent events such that

$$\Sigma_{i=1}^{\infty} P\left[A_i\right] = \infty.$$

The second Borel-Cantelli Lemma asserts that almost surely A_i happens for infinitely many i.

To prove this, we use the inequality

$$1 + x \leq e^x, \; x \in \mathbf{R}$$

to obtain

$$P\left[\cap_{i=k}^{k+n} A_i^c\right] = \Pi_{i=k}^{k+n} P\left[A_i^c\right]$$

$$= \Pi_{i=k}^{k+n}(1 - P\left[A_i\right]) \leq \Pi_{i=k}^{k+n} e^{-P[A_i]} = e^{-\Sigma_{i=k}^{k+n} P[A_i]}.$$

By letting $n \to \infty$,

$$P\left[\cap_{i=k}^{\infty} A_i^c\right] = 0$$

or

$$P\left[\cup_{i=k}^{\infty} A_i\right] = 1.$$

But then

$$P\left[\cap_{k=1}^{\infty} \cup_{i=k}^{\infty} A_i\right] = 1$$

and the second Borel-Cantelli Lemma is proved.

Theorem 3.6.1. *Suppose* $\xi_1, ..., \xi_n$ *are independent random variables and* $\xi_k \in N(0,1),\ k = 1, ..., n$. *If* $\alpha_1, ..., \alpha_n \in R$, *then*

$$\Sigma_{k=1}^{n} \alpha_k \xi_k \in N(0, \Sigma_{k=1}^{n} \alpha_k^2)$$

PROOF. The case $\alpha_1, ..., \alpha_n = 0$ is trivial so assume $\alpha_k \neq 0$ for some k. We have for each open interval A,

$$P\left[\Sigma_{k=1}^{n} \alpha_k \xi_k \in A\right] = \int_{\Sigma_{k=1}^{n} \alpha_k x_k \in A} d\gamma_1(x_1)...d\gamma_1(x_n)$$

$$\int_{\Sigma_{k=1}^{n} \alpha_k x_k \in A} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(x_1^2+...+x_n^2)} dx_1...dx_n.$$

Set $\sigma = \sqrt{\alpha_1^2 + ... + \alpha_n^2}$ and let $y = Gx$ be an orthogonal transformation such that

$$y_1 = \frac{1}{\sigma}(\alpha_1 x_1 + ... + \alpha_n x_n).$$

Then, since $\det G = 1$,

$$P\left[\Sigma_{k=1}^{\infty} \alpha_k \xi_k \in A\right] = \int_{\sigma y_1 \in A} \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}(y_1^2+...+y_n^2)} dy_1...dy_n$$

$$= \int_{\sigma y_1 \in A} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y_1^2} dy_1$$

where we used Fubini's theorem in the last step. The theorem is proved.

Finally, in this section, we prove a basic result about the existence of infinite product measures. Let $\mu_k,\ k \in \mathbf{N}_+$ be Borel probability measures in $\mathbf{R}$. The space $\mathbf{R}^{\mathbf{N}_+}$ is, by definition, the set of all sequences $x = (x_k)_{k=1}^{\infty}$

of real numbers. For each $k \in \mathbf{N}_+$, set $\pi_k(x) = x_k$. The σ-algebra $\mathcal{R}^{\mathbf{N}_+}$ is the least σ-algebra $\mathcal{S}$ of subsets of $\mathbf{R}^{\mathbf{N}_+}$ which makes all the projections $\pi_k : (\mathbf{R}^{\mathbf{N}_+}, \mathcal{S}) \to (\mathbf{R}, \mathcal{R})$, $k \in \mathbf{N}_+$, measurable. Below, $(\pi_1, ..., \pi_n)$ denotes the mapping of $\mathbf{R}^{\mathbf{N}_+}$ into $\mathbf{R}^n$ defined by the equation

$$(\pi_1, ..., \pi_n)(x) = (\pi_1(x), ..., \pi_n(x)).$$

Theorem 3.6.1. *There is a unique probability measure μ on $\mathcal{R}^{\mathbf{N}_+}$ such that*

$$\mu_{(\pi_1,...,\pi_n)} = \mu_1 \times ... \times \mu_n$$

for every $n \in \mathbf{N}_+$.

The measure μ in Theorem 3.6.1 is called the product of the measures μ_k, $k \in \mathbf{N}_+$, and is often denoted by

$$\times_{k=1}^{\infty} \mu_k.$$

PROOF OF THEOREM 3.6.1. Let $(\Omega, P, \mathcal{F}) = ([0,1[\,, v_{1|[0,1[}, \mathcal{B}([0,1[)$ and set

$$\eta(\omega) = \Sigma_{k=1}^{\infty} \frac{\xi_k^{(2)}(\omega)}{2^k},\ \omega \in \Omega.$$

We already know that $P_\eta = P$. Now suppose $(k_i)_{i=1}^{\infty}$ is a strictly increasing sequence of positive integers and introduce

$$\eta' = \Sigma_{i=1}^{\infty} \frac{\xi_{k_i}^{(2)}(\omega)}{2^i},\ \omega \in \Omega.$$

Note that for each fixed positive integer n, the $\mathbf{R}^n$-valued maps $(\xi_1^{(2)}, ..., \xi_n^{(2)})$ and $(\xi_{k_1}^{(2)}, ..., \xi_{k_n}^{(2)})$ are P-equimeasurable. Thus, if $f : \Omega \to \mathbf{R}$ is continuous,

$$\int_\Omega f(\eta) dP = \lim_{n\to\infty} \int_\Omega f(\Sigma_{k=1}^{n} \frac{\xi_k^{(2)}(\omega)}{2^k}) dP(\omega)$$

$$= \lim_{n\to\infty} \int_\Omega f(\Sigma_{i=1}^{n} \frac{\xi_{k_i}^{(2)}(\omega)}{2^i}) dP(\omega) = \int_\Omega f(\eta') dP$$

and it follows that $P_{\eta'} = P_\eta = P$.

By induction, we define for each $k \in \mathbf{N}_+$ an infinite subset N_k of the set $\mathbf{N}_+ \setminus \cup_{i=1}^{k-1} N_i$ such that the set $\mathbf{N}_+ \setminus \cup_{i=1}^{k} N_i$ contains infinitely many elements and define

$$\eta_k = \Sigma_{i=1}^{\infty} \frac{\xi_{n_{ik}}^{(2)}(\omega)}{2^i}$$

where $(n_{ik})_{i=1}^{\infty}$ is an enumeration of N_k. The map

$$\Psi(\omega) = (\eta_k(\omega))_{k=1}^{\infty}$$

is a measurable map of $(\Omega, \mathcal{F})$ into $(\mathbf{R}^{\mathbf{N}_+}, \mathcal{R}^{\mathbf{N}_+})$ and

$$P_\Psi = \times_{k=1}^{\infty} \lambda_i$$

where $\lambda_i = P$ for each $i \in \mathbf{N}_+$.

For each $i \in \mathbf{N}_+$ there exists a measurable map φ_i of $(\Omega, \mathcal{F})$ into $(\mathbf{R}, \mathcal{R})$ such that $P_{\varphi_i} = \mu_i$ (see Section 1.6). The map

$$\Gamma(x) = (\varphi_i(x_i))_{i=1}^{\infty}$$

is a measurable map of $(\mathbf{R}^{\mathbf{N}_+}, \mathcal{R}^{\mathbf{N}_+})$ into itself and we get $\mu = (P_\Psi)_\Gamma$. This completes the proof of Theorem 3.6.1.

$\uparrow\uparrow\uparrow$

CHAPTER 4

MODES OF CONVERGENCE

Introduction

In this chapter we will treat a variety of different sorts of convergence notions in measure theory. So called L^2-convergence is of particular importance.

4.1. Convergence in Measure, in $L^1(\mu)$, and in $L^2(\mu)$

Let $(X, \mathcal{M}, \mu)$ be a positive measure space and denote by $\mathcal{F}(X)$ the class of measurable functions $f : (X, \mathcal{M}) \rightarrow (\mathbf{R}, \mathcal{R})$. For any $f \in \mathcal{F}(X)$, set

$$\| f \|_1 = \int_X | f(x) | \, d\mu(x)$$

and

$$\| f \|_2 = \sqrt{\int_X f^2(x) d\mu(x)}.$$

The Cauchy-Schwarz inequality states that

$$\int_X | fg | \, d\mu \leq \| f \|_2 \| g \|_2 \text{ if } f, g \in \mathcal{F}(X).$$

To prove this, without loss of generality, it can be assumed that

$$0 < \| f \|_2 < \infty \ \text{ and } 0 < \| g \|_2 < \infty.$$

We now use the inequality

$$\alpha\beta \leq \frac{1}{2}(\alpha^2 + \beta^2), \ \ \alpha, \beta \in \mathbf{R}$$

to obtain

$$\int_X \frac{|f|}{\| f \|_2} \frac{|g|}{\| g \|_2} d\mu \leq \int \frac{1}{2}\left(\frac{f^2}{\| f \|_2^2} + \frac{g^2}{\| g \|_2^2}\right) d\mu = 1$$

and the Cauchy-Schwarz inequality is immediate.

If not otherwise stated, in this section p is a number equal to 1 or 2. If it is important to emphasize the underlying measure $\| f \|_p$ is written $\| f \|_{p,\mu}$.

We now define

$$\mathcal{L}^p(\mu) = \{ f \in \mathcal{F}(X);\ \ \| f \|_p < \infty \} .$$

The special case $p = 1$ has been introduced earlier. We claim that the following so called triangle inequality holds, viz.

$$\| f + g \|_p \leq \| f \|_p + \| g \|_p \ \text{ if } f, g \in \mathcal{L}^p(\mu).$$

The case $p = 1$, follows by μ-integration of the relation

$$| f + g | \leq | f | + | g | .$$

To prove the case $p = 2$, we use the Cauchy-Schwarz inequality and have

$$\| f + g \|_2^2 \leq \| | f | + | g | \|_2^2$$

$$= \| f \|_2^2 + 2 \int_X | fg | \, d\mu + \| g \|_2^2$$

$$\leq \| f \|_2^2 + 2 \| f \|_2 \| g \|_2 + \| g \|_2^2 = (\| f \|_2 + \| g \|_2)^2$$

and the triangle inequality is immediate.

Suppose $f, g \in \mathcal{L}^p(\mu)$. The functions f and g are equal almost everywhere with respect to μ if $\{f \neq g\} \in \mathcal{Z}_\mu$. This is easily seen to be an equivalence relation and the set of all equivalence classes is denoted by $L^p(\mu)$. Below we consider the elements of $L^p(\mu)$ as members of $\mathcal{L}^p(\mu)$ and two members of $L^p(\mu)$ are identified if they are equal a.e. $[\mu]$. From this convention it is straight-forward to define $f + g$ and αf for all $f, g \in L^p(\mu)$ and $\alpha \in \mathbf{R}$ and the function $d^{(p)}(f, g) = \| f - g \|_p$ is a metric on $L^p(\mu)$. Convergence in the metric space $L^p(\mu) = (L^p(\mu), d^{(p)})$ is called convergence in $L^p(\mu)$. A sequence $(f_k)_{k=1}^\infty$ in $\mathcal{F}(X)$ converges in measure to a function $f \in \mathcal{F}(X)$ if

$$\lim_{k \to \infty} \mu(| f_k - f | > \varepsilon) = 0 \text{ all } \varepsilon > 0.$$

If the sequence $(f_k)_{k=1}^{\infty}$ in $\mathcal{F}(X)$ converges in measure to a function $f \in \mathcal{F}(X)$ as well as to a function $g \in \mathcal{F}(X)$, then $f = g$ a.e. $[\mu]$ since

$$\{\mid f - g \mid > \varepsilon\} \subseteq \left\{\mid f - f_k \mid > \frac{\varepsilon}{2}\right\} \cup \left\{\mid f_k - g \mid > \frac{\varepsilon}{2}\right\}$$

and

$$\mu(\mid f - g \mid > \varepsilon) \leq \mu(\mid f - f_k \mid > \frac{\varepsilon}{2}) + \mu(\mid f_k - g \mid > \frac{\varepsilon}{2})$$

for every $\varepsilon > 0$ and positive integer k. A sequence $(f_k)_{k=1}^{\infty}$ in $\mathcal{F}(X)$ is said to be Cauchy in measure if for every $\varepsilon > 0$,

$$\mu(\mid f_k - f_n \mid > \varepsilon) \to 0 \text{ as } k, n \to \infty.$$

By the Markov inequality, a Cauchy sequence in $L^p(\mu)$ is Cauchy in measure.

Example 4.1.1. (a) If $f_k = \sqrt{k}\chi_{[0,\frac{1}{k}]}$, $k \in \mathbf{N}_+$, then

$$\parallel f_k \parallel_{2,m} = 1 \text{ and } \parallel f_k \parallel_{1,m} = \frac{1}{\sqrt{k}}.$$

Thus $f_k \to 0$ in $L^1(m)$ as $k \to \infty$ but $f_k \nrightarrow 0$ in $L^2(m)$ as $k \to \infty$.

(b) $L^1(m) \nsubseteq L^2(m)$ since

$$\chi_{[1,\infty[}(x)\frac{1}{\mid x \mid} \in L^2(m) \setminus L^1(m)$$

and $L^2(m) \nsubseteq L^1(m)$ since

$$\chi_{]0,1]}(x)\frac{1}{\sqrt{\mid x \mid}} \in L^1(m) \setminus L^2(m).$$

Theorem 4.1.1. *Suppose $p = 1$ or 2.*

(a) *Convergence in $L^p(\mu)$ implies convergence in measure.*

(b) *If* $\mu(X) < \infty$, *then* $L^2(\mu) \subseteq L^1(\mu)$ *and convergence in* $L^2(\mu)$ *implies convergence in* $L^1(\mu)$.

Proof. (a) Suppose the sequence $(f_n)_{n=1}^{\infty}$ converges to f in $L^p(\mu)$ and let $\varepsilon > 0$. Then, by the Markov inequality,

$$\mu(\mid f_n - f \mid \geq \varepsilon) \leq \frac{1}{\varepsilon^p} \int_X \mid f_n - f \mid^p d\mu = \frac{1}{\varepsilon^p} \parallel f_n - f \parallel_p^p$$

and (a) follows at once.

(b) The Cauchy-Schwarz inequality gives for any $f \in \mathcal{F}(X)$,

$$(\int_X \mid f \mid \cdot 1 d\mu)^2 \leq \int_X f^2 d\mu \int_X 1 d\mu$$

or

$$\parallel f \parallel_1 \leq \parallel f \parallel_2 \sqrt{\mu(X)}$$

and Part (b) is immediate.

Theorem 4.1.2. *Suppose* $f_n \in \mathcal{F}(X)$, $n \in \mathbf{N}_+$.

(a) *If* $(f_n)_{n=1}^{\infty}$ *is Cauchy in measure, there is a measurable function* $f : X \to \mathbf{R}$ *such that* $f_n \to f$ *in measure as* $n \to \infty$ *and a strictly increasing sequence of positive integers* $(n_j)_{j=1}^{\infty}$ *such that* $f_{n_j} \to f$ *a.e.* $[\mu]$ *as* $j \to \infty$.

(b) *If* μ *is a finite positive measure and* $f_n \to f \in \mathcal{F}(X)$ *a.e.* $[\mu]$ *as* $n \to \infty$, *then* $f_n \to f$ *in measure.*

(c) **(Egoroff's Theorem)** *If* μ *is a finite positive measure and* $f_n \to f \in \mathcal{F}(X)$ *a.e.* $[\mu]$ *as* $n \to \infty$, *then for every* $\varepsilon > 0$ *there exists* $E \in \mathcal{M}$ *such that* $\mu(E) < \varepsilon$ and

$$\sup_{\substack{k \geq n \\ x \in E^c}} \mid f_k(x) - f(x) \mid \to 0 \text{ as } n \to \infty.$$

PROOF. (a) For each positive integer j, there is a positive integer n_j such that

$$\mu(\mid f_k - f_l \mid > 2^{-j}) < 2^{-j}, \text{ all } k, l \geq n_j.$$

There is no loss of generality to assume that $n_1 < n_2 < ...$. Set

$$E_j = \left\{ \mid f_{n_j} - f_{n_{j+1}} \mid > 2^{-j} \right\}$$

and

$$F_k = \cup_{j=k}^{\infty} E_j.$$

If $x \in F_k^c$ and $i \geq j \geq k$

$$\mid f_{n_i}(x) - f_{n_j}(x) \mid \leq \sum_{j \leq l < i} \mid f_{n_{l+1}}(x) - f_{n_l}(x) \mid$$

$$\leq \sum_{j \leq l < i} 2^{-l} < 2^{-j+1}$$

and we conclude that $(f_{n_j}(x))_{j=1}^{\infty}$ is a Cauchy sequence for every $x \in F_k^c$. Let $G = \cup_{k=1}^{\infty} F_k^c$ and note that for every fixed positive integer k,

$$\mu(G^c) \leq \mu(F_k) < \sum_{j=k}^{\infty} 2^{-j} = 2^{-k+1}.$$

Thus G^c is a μ-null set. We now define $f(x) = \lim_{j \to \infty} f_{n_j}(x)$ if $x \in G$ and $f(x) = 0$ if $x \notin G$.

We next prove that the sequence $(f_n)_{n=1}^{\infty}$ converges to f in measure. If $x \in F_k^c$ and $j \geq k$ we get

$$\mid f(x) - f_{n_j}(x) \mid \leq 2^{-j+1}.$$

Thus, if $j \geq k$

$$\mu(\mid f - f_{n_j} \mid > 2^{-j+1}) \leq \mu(F_k) < 2^{-k+1}.$$

Since

$$\mu(\mid f_n - f \mid > \varepsilon) \leq \mu(\mid f_n - f_{n_j} \mid > \frac{\varepsilon}{2}) + \mu(\mid f_{n_j} - f \mid > \frac{\varepsilon}{2})$$

if $\varepsilon > 0$, Part (a) follows at once.

(b) For each $\varepsilon > 0$,

$$\mu(\mid f_n - f \mid > \varepsilon) = \int_X \chi_{]\varepsilon,\infty[}(\mid f_n - f \mid) d\mu$$

and Part (c) follows from the Lebesgue Dominated Convergence Theorem.

(c) Set for fixed $k, n \in \mathbf{N}_+$,

$$E_{kn} = \cup_{j=n}^{\infty} \left\{ \mid f_j - f \mid > \frac{1}{k} \right\}.$$

We have

$$\cap_{n=1}^{\infty} E_{kn} \in Z_{\mu}$$

and since μ is a finite measure

$$\mu(E_{kn}) \to 0 \text{ as } n \to \infty.$$

Given $\varepsilon > 0$ pick $n_k \in \mathbf{N}_+$ such that $\mu(E_{kn_k}) < \varepsilon 2^{-k}$. Then, if $E = \cup_{k=1}^{\infty} E_{kn_k}$, $\mu(E) < \varepsilon$. Moreover, if $x \notin E$ and $j \geq n_k$

$$\mid f_j(x) - f(x) \mid \leq \frac{1}{k}.$$

The theorem is proved.

Corollary 4.1.1. *The spaces $L^1(\mu)$ and $L^2(\mu)$ are complete.*

PROOF. Suppose $p = 1$ or 2 and let $(f_n)_{n=1}^{\infty}$ be a Cauchy sequence in $L^p(\mu)$. We know from the previous theorem that there exists a subsequence $(f_{n_j})_{j=1}^{\infty}$ which converges pointwise to a function $f \in \mathcal{F}(X)$ a.e. $[\mu]$. Thus, by Fatou's Lemma,

$$\int_X \mid f - f_k \mid^p d\mu \leq \liminf_{j \to \infty} \int_X \mid f_{n_j} - f_k \mid^p d\mu$$

and it follows that $f - f_k \in L^p(\mu)$ and, hence $f = (f - f_k) + f_k \in L^p(\mu)$. Moreover, we have that $\parallel f - f_k \parallel_p \to 0$ as $k \to \infty$. This concludes the proof of the theorem.

Corollary 4.1.2. *Suppose $\xi_n \in N(0, \sigma_n^2)$, $n \in \mathbf{N}_+$, and $\xi_n \to \xi$ in $L^2(P)$ as $n \to \infty$. Then ξ is a centred Gaussian random variable.*

PROOF. We have that $\parallel \xi_n \parallel_2 = \sqrt{E\left[\xi_n^2\right]} = \sigma_n$ and $\parallel \xi_n \parallel_2 \to \parallel \xi \parallel_2 =_{def} \sigma$ as $n \to \infty$.

Suppose f is a bounded continuous function on $\mathbf{R}$. Then, by dominated convergence,

$$E\left[f(\xi_n)\right] = \int_{\mathbf{R}} f(\sigma_n x) d\gamma_1(x) \to \int_{\mathbf{R}} f(\sigma x) d\gamma_1(x)$$

as $n \to \infty$. Moreover, there exists a subsequence $(\xi_{n_k})_{k=1}^{\infty}$ which converges to ξ a.s. Hence, by dominated convergence

$$E\left[f(\xi_{n_k})\right] \to E\left[f(\xi)\right]$$

as $k \to \infty$ and it follows that

$$E\left[f(\xi)\right] = \int_{\mathbf{R}} f(\sigma x) d\gamma_1(x).$$

By using Corollary 3.1.3 the theorem follows at once.

Theorem 4.1.3. *Suppose X is a standard space and μ a positive σ-finite Borel measure on X. Then the spaces $L^1(\mu)$ and $L^2(\mu)$ are separable.*

PROOF. Let $(E_k)_{k=1}^{\infty}$ be a denumerable collection of Borel sets with finite μ-measures and such that $E_k \subseteq E_{k+1}$ and $\cup_{k=1}^{\infty} E_k = X$. Set $\mu_k = \chi_{E_k}\mu$ and first suppose that the set D_k is at most denumerable and dense in $L^p(\mu_k)$ for every $k \in \mathbf{N}_+$. Without loss of generality it can be assumed that each member of D_k vanishes off E_k. By monotone convergence

$$\int_X f d\mu = \lim_{k \to \infty} \int_X f d\mu_k, \; f \geq 0 \text{ measurable},$$

and it follows that the set $\cup_{k=1}^{\infty} D_k$ is at most denumerable and dense in $L^p(\mu)$.

From now on we can assume that μ is a finite positive measure. Let A be an at most denumerable dense subset of X and and suppose the subset $\{r_n;\ n \in \mathbf{N}_+,\}$ of $\,]0,\infty[\,$ is dense in $]0,\infty[$. Furthermore, denote by $\mathcal{U}$ the class of all open sets which are finite unions of open balls of the type $B(a, r_n)$, $a \in A$, $n \in \mathbf{N}_+$. If U is any open subset of X

$$U = \cup [V : V \subseteq U \text{ and } V \in \mathcal{U}]$$

and, hence, by the Ulam Theorem

$$\mu(U) = \sup \{\mu(V);\ V \in \mathcal{U} \text{ and } V \subseteq U\}.$$

Let $\mathcal{K}$ be the class of all functions which are finite sums of functions of the type $\kappa\chi_U$, where κ is a positive rational number and $U \in \mathcal{U}$. It follows that $\mathcal{K}$ is at most denumerable.

Suppose $\varepsilon > 0$ and that $f \in L^p(\mu)$ is non-negative. There exists a sequence of simple measurable functions $(\varphi_i)_{i=1}^{\infty}$ such that

$$0 \leq \varphi_i \uparrow f \text{ a.e. } [\mu].$$

Since $| f - \varphi_i |^p \leq f^p$, the Lebesgue Dominated Convergence Theorem shows that $\| f - \varphi_k \|_p < \frac{\varepsilon}{2}$ for an appropriate k. Let $\alpha_1, ..., \alpha_l$ be the distinct positive values of φ_k and set

$$C = 1 + \Sigma_{k=1}^{l} \alpha_k.$$

Now for each fixed $j \in \{1, ..., l\}$ we use Theorem 3.1.3 to get an open $U_j \supseteq \varphi_k^{-1}(\{\alpha_j\})$ such that $\| \chi_{U_j} - \chi_{\varphi_k^{-1}(\{\alpha_j\})} \|_p < \frac{\varepsilon}{4C}$ and from the above we get a $V_j \in \mathcal{U}$ such that $V_j \subseteq U_j$ and $\| \chi_{U_j} - \chi_{V_j} \|_p < \frac{\varepsilon}{4C}$. Thus

$$\| \chi_{V_j} - \chi_{\varphi_k^{-1}(\{\alpha_j\})} \|_p < \frac{\varepsilon}{2C}$$

and

$$\| f - \Sigma_{k=1}^{l} \alpha_j \chi_{V_j} \|_p < \varepsilon$$

Now it is simple to find a $\psi \in \mathcal{K}$ such that $\| f - \psi \|_p < \varepsilon$. From this we deduce that the set

$$\mathcal{K} - \mathcal{K} = \{g - h;\ g, h \in \mathcal{K}\}$$

is at most denumerable and dense in $L^p(\mu)$.

The set of all real-valued and infinitely many times differentiable functions defined on $\mathbf{R}^n$ is denoted by $C^{(\infty)}(\mathbf{R}^n)$ and

$$C_c^{(\infty)}(\mathbf{R}^n) = \left\{ f \in C^{(\infty)}(\mathbf{R}^n);\ \text{supp} f \text{ compact} \right\}.$$

Recall that the support supp$f$ of a real-valued continuous function f defined on $\mathbf{R}^n$ is the closure of the set of all x where $f(x) \neq 0$. If

$$f(x) = \prod_{k=1}^{n} \{\varphi(1 + x_k)\varphi(1 - x_k)\}\,,\ x = (x_1, ..., x_n) \in \mathbf{R}^n$$

where $\varphi(t) = \exp(-t^{-1})$, if $t > 0$, and $\varphi(t) = 0$, if $t \leq 0$, then $f \in C_c^{\infty}(\mathbf{R}^n)$.

The proof of the previous theorem also gives Part (a) of the following

Theorem 4.1.4. *Suppose μ is a positive Borel measure in $\mathbf{R}^n$ such that $\mu(K) < \infty$ for every compact subset K of $\mathbf{R}^n$. The following sets are dense in $L^1(\mu)$, and $L^2(\mu)$:*

(a) *the linear span of the functions*

$$\chi_I,\ I \text{ open bounded } n\text{-cell in } \mathbf{R}^n,$$

(b) $C_c^{(\infty)}(\mathbf{R}^n)$.

PROOF. a) The proof is almost the same as the proof of Theorem 4.1.3. First the E_k:s can be chosen to be open balls with their centres at the origin since each bounded set in $\mathbf{R}^n$ has finite μ-measure. Moreover, as in the proof of Theorem 4.1.3 we can assume that μ is a finite measure. Now let A be an at most denumerable dense subset of $\mathbf{R}^n$ and for each $a \in A$ let

$$R(a) = \{r > 0;\ \mu(\{x \in X;\ \mid x_k - a_k \mid = r\}) > 0 \text{ for some } k = 1, ..., n\}.$$

Then $\cup_{a \in A} R(a)$ is at most denumerable and there is a subset $\{r_n;\ n \in \mathbf{N}_+\}$ of $]0, \infty[\, \setminus \cup_{a \in A} R(a)$ which is dense in $]0, \infty[$. Finally, let $\mathcal{U}$ denote the class of all open sets which are finite unions of open balls of the type $B(a, r_n)$, $a \in A$, $n \in \mathbf{N}_+$, and proceed as in the proof of Theorem 4.1.3. The result follows by observing that the characteristic function of any member of $\mathcal{U}$ equals a finite sum of characteristic functions of open bounded n-cells a.e. $[\mu]$.

Part (b) in Theorem 4.1.4 follows from Part (a) and the following

Lemma 4.1.1. *Suppose $K \subseteq U \subseteq \mathbf{R}^n$, where K is compact and U is open. Then there exists a function $f \in C_c^{(\infty)}(\mathbf{R}^n)$* such that

$$K \prec f \prec U$$

that is

$$\chi_K \leq f \leq \chi_U \text{ and supp} f \subseteq U.$$

PROOF. Suppose $\rho \in C_c^\infty(\mathbf{R}^n)$ is non-negative, supp $\rho \subseteq B(0,1)$, and

$$\int_{\mathbf{R}^n} \rho dm_n = 1.$$

Moreover, let $\varepsilon > 0$ be fixed. For any $g \in L^1(v_n)$ we define

$$f_\varepsilon(x) = \varepsilon^{-n} \int_{\mathbf{R}^n} g(y)\rho(\varepsilon^{-1}(x-y))dy.$$

Since

$$\mid g \mid \max_{\mathbf{R}^n} \mid \frac{\partial^{k_1+...+k_n}\rho}{\partial x_1^{k_1}...\partial x_n^{k_n}} \mid \in L^1(v_n), \text{ all } k_1, ..., k_n \in \mathbf{N}$$

the Lebesgue Dominated Convergent Theorem shows that $f_\varepsilon \in C^\infty(\mathbf{R}^n)$. Here $f_\varepsilon \in C_c^\infty(\mathbf{R}^n)$ if g vanishes off a bounded subset of $\mathbf{R}^n$. In fact,

$$\text{supp } f_\varepsilon \subseteq (\text{supp } g)_\varepsilon.$$

Now choose a positive number $\varepsilon \leq \frac{1}{2}d(K, U^c)$ and define $g = \chi_{K_\varepsilon}$. Since

$$f_\varepsilon(x) = \int_{\mathbf{R}^n} g(x - \varepsilon y)\rho(y)dy$$

we also have that $f_\varepsilon(x) = 1$ if $x \in K$. The lemma is proved.

Example 4.1.2. Suppose $f \in L^1(m_n)$ and let $g : \mathbf{R}^n \to \mathbf{R}$ be a bounded Lebesgue measurable function. Set

$$h(x) = \int_{\mathbf{R}^n} f(x-y)g(y)dy, \ x \in \mathbf{R}^n.$$

We claim that h is continuous.

To see this first note that

$$h(x+\Delta x) - h(x) = \int_{\mathbf{R}^n} (f(x+\Delta x - y) - f(x-y))g(y)dy$$

and

$$\mid h(x+\Delta x) - h(x) \mid \leq K \int_{\mathbf{R}^n} \mid f(x+\Delta x - y) - f(x-y) \mid dy$$

$$= K \int_{\mathbf{R}^n} \mid f(\Delta x + y) - f(y) \mid dy$$

if $\mid g(x) \mid \leq K$ for every $x \in \mathbf{R}^n$. Now first choose $\varepsilon > 0$ and then $\varphi \in C_c(\mathbf{R}^n)$ such that

$$\parallel f - \varphi \parallel_1 < \varepsilon.$$

Using the triangle inequality, we get

$$\mid h(x+\Delta x) - h(x) \mid \leq K(2 \parallel f - \varphi \parallel_1 + \int_{\mathbf{R}^n} \mid \varphi(\Delta x + y) - \varphi(y) \mid dy)$$

$$\leq K(2\varepsilon + \int_{\mathbf{R}^n} \mid \varphi(\Delta x + y) - \varphi(y) \mid dy)$$

where the right hand side is smaller than $3K\varepsilon$ if $\mid \Delta x \mid$ is sufficiently small. This proves that h is continuous.

Example 4.1.3. Suppose $A \in \mathcal{R}_n^-$ and $m_n(A) > 0$. We claim that the set

$$A - A = \{x - x; x \in A\}$$

contains a neighbourhood of the origin.

To show this there is no loss of generality to assume that $m_n(A) < \infty$. Set

$$f(x) = m_n(A \cap (A + x)), \ x \in \mathbf{R}^n.$$

Note that

$$f(x) = \int_{\mathbf{R}^n} \chi_A(y)\chi_A(y-x)dy$$

and Example 4.1.2 proves that f is continuous. Since $f(0) > 0$ there exists a $\delta > 0$ such that $f(x) > 0$ if $\mid x \mid< \delta$. In particular, $A \cap (A+x) \neq \phi$ if $\mid x \mid< \delta$, which proves that

$$B(0,\delta) \subseteq A - A.$$

The following three examples are based on the Axiom of Choice.

Example 4.1.4. Let NL be the non-Lebesgue measurable set constructed in Section 1.3. Furthermore, assume $A \subseteq \mathbf{R}$ is Lebesgue measurable and $A \subseteq NL$. We claim that $m(A) = 0$. If not, there exists a $\delta > 0$ such that $B(0,\delta) \subseteq A - A \subseteq NL - NL$. If $0 < r < \delta$ and $r \in \mathbf{Q}$, there exist $a, b \in NL$ such that

$$a = b + r.$$

But then $a \neq b$ and at the same time a and b belong to the same equivalence class, which is a contradiction. Accordingly from this, $m(A) = 0$.

Example 4.1.5. Suppose $A \subseteq \left[-\frac{1}{2}, \frac{1}{2}\right]$ is Lebesgue measurable and $m(A) > 0$. We claim there exists a non-Lebesgue measurable subset of A. To see this note that

$$A = \cup_{i=1}^{\infty}((r_i + NL) \cap A)$$

where $(r_i)_{i=1}^{\infty}$ is an enumeration of the rational numbers in the interval $[-1, 1]$. If each set $(r_i + NL) \cap A$, is Lebesgue measurable

$$m(A) = \Sigma_{i=1}^{\infty} m((r_i + NL) \cap A)$$

and we conclude that $m((r_i + NL) \cap A) > 0$ for an appropriate i. But then $m(NL \cap (A - r_i)) > 0$ and $NL \cap (A - r_i) \subseteq NL$, which contradicts Example 4.1.4. Hence $(r_i + NL) \cap A$ is non-Lebesgue measurable for an appropriate i.

If A is a Lebesgue measurable subset of the real line of positive Lebesgue measure, we conclude that A contains a non-Lebesgue measurable subset.

Example 4.1.6. Set $I = [0,1]$. We claim there exist a continuous function $f : I \to I$ and a Lebesgue measurable set $L \subseteq I$ such that $f^{-1}(L)$ is not Lebesgue measurable.

First recall from Section 3.3 the construction of the Cantor set C and the Cantor function G. First $C_0 = [0,1]$. Then trisect C_0 and remove the middle interval $\left]\frac{1}{3}, \frac{2}{3}\right[$ to obtain $C_1 = C_0 \setminus \left]\frac{1}{3}, \frac{2}{3}\right[= \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]$. At the second stage subdivide each of the closed intervals of C_1 into thirds and remove from each one the middle open thirds. Then $C_2 = C_1 \setminus (\left]\frac{1}{9}, \frac{2}{9}\right[\cup \left]\frac{7}{9}, \frac{8}{9}\right[)$. We repeat the process and what is left from C_{n-1} is C_n. The set $[0,1] \setminus C_n$ is the union of $2^n - 1$ intervals numbered I_k^n, $k = 1, ..., 2^n - 1$, where the interval I_k^n is situated to the left of the interval I_l^n if $k < l$. The Cantor set $C = \cap_{n=1}^{\infty} C_n$.

Suppose n is fixed and let $G_n : [0,1] \to [0,1]$ be the unique the monotone increasing continuous function, which satisfies $G_n(0) = 0, G_n(1) = 1, G_n(x) = k2^{-n}$ for $x \in I_k^n$ and which is linear on each interval of $C_{n.}$ It is clear that $G_n = G_{n+1}$ on each interval I_k^n, $k = 1, ..., 2^n - 1$. The Cantor function is defined by the limit $G(x) = \lim_{n \to \infty} G_n(x)$, $0 \le x \le 1$.

Now define

$$h(x) = \frac{1}{2}(x + G(x)), \ x \in I$$

where G is the Cantor function. Since $h : I \to I$ is a strictly increasing and continuous bijection, the inverse function $f = h^{-1}$ is a continuous bijection from I onto I. Set

$$A = h(I \setminus C)$$

and

$$B = h(C).$$

Recall from the definition of G that G is constant on each removed interval I_k^n and that h takes each removed interval onto an interval of half its length. Thus $m(A) = \frac{1}{2}$ and $m(B) = 1 - m(A) = \frac{1}{2}$.

By the previous example there exists a non-Lebesgue measurable subset M of B. Put $L = h^{-1}(M)$. The set L is Lebesgue measurable since $L \subseteq C$ and C is a Lebesgue null set. However, the set $M = f^{-1}(L)$ is not Lebesgue measurable.

Exercises

1. Let $(X, \mathcal{M}, \mu)$ be a finite positive measure space and suppose $\varphi(t) = \min(t, 1)$, $t \geq 0$. Prove that $f_n \to f$ in measure if and only if $\varphi(\mid f_n - f \mid) \to 0$ in $L^1(\mu)$.

2. Let $\mu = m_{|[0,1]}$. Find measurable functions $f_n : [0,1] \to [0,1]$, $n \in \mathbf{N}_+$, such that $f_n \to 0$ in $L^2(\mu)$ as $n \to \infty$,

$$\liminf_{n\to\infty} f_n(x) = 0 \text{ all } x \in [0,1]$$

and

$$\limsup_{n\to\infty} f_n(x) = 1 \text{ all } x \in [0,1] \, .$$

3. If $f \in \mathcal{F}(X)$ set

$$\| f \|_0 = \inf \{\alpha \in [0,\infty] \, ; \; \mu(\mid f \mid > \alpha) \leq \alpha\} \, .$$

Let

$$L^0(\mu) = \{f \in \mathcal{F}(X); \;\; \| f \|_0 < \infty\}$$

and identify functions in $L^0(\mu)$ which agree a.e. $[\mu]$.

(a) Prove that $d^{(0)} = \| f - g \|_0$ is a metric on $L^0(\mu)$ and that the corresponding metric space is complete.

(b) Show that $\mathcal{F}(X) = L^0(\mu)$ if μ is a finite positive measure.

4. Suppose $L^p(X, \mathcal{M}, \mu)$ is separable, where $p = 1$ or 2. Show that $L^p(X, \mathcal{M}^-, \bar{\mu})$ is separable.

5. Suppose g is a real-valued, Lebesgue measurable, and bounded function of period one. Prove that

$$\lim_{n\to\infty} \int_{-\infty}^{\infty} f(x)g(nx)dx = \int_{-\infty}^{\infty} f(x)dx \int_0^1 g(x)dx$$

for every $f \in L^1(m)$.

6. Let $h_n(t) = 2 + \sin nt$, $0 \leq t \leq 1$, and $n \in \mathbf{N}_+$. Find real constants α and β such that

$$\lim_{n\to\infty} \int_0^1 f(t)h_n(t)dt = \alpha \int_0^1 f(t)dt$$

and

$$\lim_{n\to\infty} \int_0^1 \frac{f(t)}{h_n(t)}dt = \beta \int_0^1 f(t)dt$$

for every real-valued Lebesgue integrable function f on $[0,1]$.

7. If $k = (k_1, ..., k_n) \in \mathbf{N}_+^n$, set $e_k(x) = \Pi_{i=1}^n \sin k_i x_i$, $x = (x_1, ..., x_n) \in \mathbf{R}^n$, and $\mid k \mid = (\Sigma_{i=1}^n k_i^2)^{\frac{1}{2}}$. Prove that

$$\lim_{|k|\to\infty} \int_{\mathbf{R}^n} f e_k dm_n = 0$$

for every $f \in L^1(m_n)$.

8. Suppose $f \in L^1(m_n)$, where m_n denotes Lebesgue measure on $\mathbf{R}^n$. Compute the following limit and justify the calculations:

$$\lim_{|h|\to\infty} \int_{\mathbf{R}^n} \mid f(x+h) - f(x) \mid dx.$$

4.2 Orthogonality

Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space. If $f, g \in L^2(\mu)$, let

$$\langle f, g \rangle =_{def} \int_X fg d\mu$$

be the so called scalar product of f and g. The Cauchy-Schwarz inequality

$$\mid \langle f, g \rangle \mid \leq \parallel f \parallel_2 \parallel g \parallel_2$$

shows that the map $f \to \langle f, g \rangle$ of $L^2(\mu)$ into $\mathbf{R}$ is continuous. Observe that

$$\| f+g \|_2^2 = \| f \|_2^2 + 2\langle f, g \rangle + \| g \|_2^2$$

and from this we get the so called Parallelogram Law

$$\| f+g \|_2^2 + \| f-g \|_2^2 = 2(\| f \|_2^2 + \| g \|_2^2).$$

We will say that f and g are orthogonal (abbr. $f \perp g$) if $\langle f, g \rangle = 0$. Note that

$$\| f+g \|_2^2 = \| f \|_2^2 + \| g \|_2^2 \text{ if and only if } f \perp g.$$

Since $f \perp g$ implies $g \perp f$, the relation $\perp$ is symmetric. Moreover, if $f \perp h$ and $g \perp h$ then $(\alpha f + \beta g) \perp h$ for all $\alpha, \beta \in \mathbf{R}$. Thus $h^\perp =_{def} \{ f \in L^2(\mu);\ f \perp h \}$ is a subspace of $L^2(\mu)$, which is closed since the map $f \to \langle f, h \rangle$, $f \in L^2(\mu)$ is continuous. If M is a subspace of $L^2(\mu)$, the set

$$M^\perp =_{def} \cap_{h \in M} h^\perp$$

is a closed subspace of $L^2(\mu)$. The function $f = 0$ if and only if $f \perp f$.

If M is a subspace of $L^2(\mu)$ and $f \in L^2(\mu)$ there exists at most one point $g \in M$ such that $f - g \in M^\perp$. To see this, let $g_0, g_1 \in M$ be such that $f - g_k \in M^\perp$, $k = 0, 1$. Then $g_1 - g_0 = (f - g_0) - (f - g_1) \in M^\perp$ and hence $g_1 - g_0 \perp g_1 - g_0$ that is $g_0 = g_1$.

Theorem 4.2.1. *Let M be a closed subspace in $L^2(\mu)$ and suppose $f \in L^2(\mu)$. Then there exists a unique point $g \in M$ such that*

$$\| f - g \|_2 \leq \| f - h \|_2 \text{ all } h \in M.$$

Moreover,

$$f - g \in M^\perp.$$

The function g in Theorem 4.2.1 is called the projection of f on M and is denoted by $\mathrm{Proj}_M\ f$.

PROOF OF THEOREM 4.2.1. Set

$$d =_{def} d^{(2)}(f, M) = \inf_{g \in M} \| f - g \|_2 .$$

and let $(g_n)_{n=1}^{\infty}$ be a sequence in M such that

$$d = \lim_{n\to\infty} \| f - g_n \|_2 .$$

Then, by the Parallelogram Law

$$\| (f-g_k)+(f-g_n) \|_2^2 + \| (f-g_k)-(f-g_n) \|_2^2 = 2(\| f-g_k \|_2^2 + \| f-g_n \|_2^2)$$

that is

$$4 \| f - \frac{1}{2}(g_k + g_n) \|_2^2 + \| g_n - g_k \|_2^2 = 2(\| f - g_k \|_2^2 + \| f - g_n \|_2^2)$$

and, since $\frac{1}{2}(g_k + g_n) \in M$, we get

$$4d^2 + \| g_n - g_k \|_2^2 \leq 2(\| f - g_k \|_2^2 + \| f - g_n \|_2^2).$$

Here the right hand converges to $4d^2$ as k and n go to infinity and we conclude that $(g_n)_{n=1}^{\infty}$ is a Cauchy sequence. Since $L^2(\mu)$ is complete and M closed there exists a $g \in M$ such that $g_n \to g$ as $n \to \infty$. Moreover,

$$d = \| f - g \|_2 .$$

We claim that $f - g \in M^{\perp}$. To prove this choose $h \in M$ and $\alpha > 0$ arbitrarily and use the inequality

$$\| (f - g) + \alpha h \|_2^2 \geq \| f - g \|_2^2$$

to obtain

$$\| f - g \|_2^2 + 2\alpha\langle f - g, h\rangle + \alpha^2 \| h \|_2^2 \geq \| f - g \|_2^2$$

and

$$2\langle f - g, h\rangle + \alpha \| h \|_2^2 \geq 0.$$

By letting $\alpha \to 0$, $\langle f-g, h\rangle \geq 0$ and replacing h by $-h$, $\langle f-g, h\rangle \leq 0$. Thus $f - g \in h^{\perp}$ and it follows that $f - g \in M^{\perp}$.

The uniqueness in Theorem 4.2.1 follows from the remark just before the formulation of Theorem 4.2.1. The theorem is proved.

A linear mapping $T : L^2(\mu) \to \mathbf{R}$ is called a linear functional on $L^2(\mu)$. If $h \in L^2(\mu)$, the map $h \to \langle f, h\rangle$ of $L^2(\mu)$ into $\mathbf{R}$ is a continuous linear functional on $L^2(\mu)$. It is a very important fact that every continuous linear functional on $L^2(\mu)$ is of this type.

Theorem 4.2.2. *Suppose T is a continuous linear functional on $L^2(\mu)$. Then there exists a unique $w \in L^2(\mu)$ such that*

$$Tf = \langle f, w \rangle \text{ all } f \in L^2(\mu).$$

PROOF. Uniqueness: If $w, w' \in L^2(\mu)$ and $\langle f, w \rangle = \langle f, w' \rangle$ for all $f \in L^2(\mu)$, then $\langle f, w - w' \rangle = 0$ for all $f \in L^2(\mu)$. By choosing $f = w - w'$ we get $f \perp f$ that is $w = w'$.

Existence: The set $M =_{def} T^{-1}(\{0\})$ is closed since T is continuous and M is a linear subspace of $L^2(\mu)$ since T is linear. If $M = L^2(\mu)$ we choose $w = 0$. Otherwise, pick a $g \in L^2(\mu) \setminus M$. Without loss of generality it can be assumed that $Tg = 1$ by eventually multiplying g by a scalar. The previous theorem gives us a vector $h \in M$ such that $u =_{def} g - h \in M^{\perp}$. Note that $0 <\| u \|_2^2 = \langle u, g - h \rangle = \langle u, g \rangle$.

To conclude the proof, let fixed $f \in L^2(\mu)$ be fixed, and use that $(Tf)g - f \in M$ to obtain

$$\langle (Tf)g - f, u \rangle = 0$$

or

$$(Tf)\langle g, u \rangle = \langle f, u \rangle.$$

By setting

$$w = \frac{1}{\| u \|_2^2} u$$

we are done.

$$\downarrow\downarrow\downarrow$$

4.3. The Haar Basis and Wiener Measure

In this section we will show the existence of Brownian motion with continuous paths as a consequence of the existence of linear measure λ in the unit interval. The so called Wiener measure is the probability law on $C[0,1]$ of real-valued Brownian motion in the time interval $[0,1]$. The Brownian motion process is named after the British botanist Robert Brown (1773-1858). It was suggested by Lous Bachelier in 1900 as a model of stock price fluctuations and later by Albert Einstein in 1905 as a model of the physical phenomenon Brownian motion. The existence of the mathematical Brownian motion process was first established by Norbert Wiener in the twenties. Wiener also proved that the model can be chosen such that the path $t \to W(t)$, $0 \leq t \leq 1$, is continuous a.s. Today Brownian motion is a very important concept in probability, financial mathematics, partial differential equations and in many other fields in pure and applied mathematics.

Suppose n is a non-negative integer and set $I_n = \{0, ..., n\}$. A sequence $(e_i)_{i \in I_n}$ in $L^2(\mu)$ is said to be orthonormal if $e_i \perp e_j$ for all $i \neq j$, $i, j \in I_n$ and $\| e_i \| = 1$ for each $i \in I_n$. If $(e_i)_{i \in I_n}$ is orthonormal and $f \in L^2(\mu)$,

$$f - \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \perp e_j \text{ all } j \in I$$

and Theorem 4.2.1 shows that

$$\| f - \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \|_2 \leq \| f - \Sigma_{i \in I_n} \alpha_i e_i \|_2 \text{ all real } \alpha_1, ..., \alpha_n.$$

Moreover

$$\| f \|_2^2 = \| f - \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \|_2^2 + \| \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \|_2^2$$

and we get

$$\Sigma_{i \in I_n} \langle f, e_i \rangle^2 \leq \| f \|_2^2 .$$

We say that $(e_n)_{n \in I_n}$ is an orthonormal basis in $L^2(\mu)$ if it is orthonormal and

$$f = \Sigma_{i \in I_n} \langle f, e_i \rangle e_i \text{ all } f \in L^2(\mu).$$

A sequence $(e_i)_{i=0}^{\infty}$ in $L^2(\mu)$ is said to be orthonormal if $(e_i)_{i=0}^{n}$ is orthonormal for each non-negative integer n. In this case, for each $f \in L^2(\mu)$,

$$\Sigma_{i=0}^{\infty} \langle f, e_i \rangle^2 \leq \| f \|_2^2$$

and the series

$$\Sigma_{i=0}^{\infty} \langle f, e_i \rangle e_i$$

converges since the sequence

$$(\Sigma_{i=0}^{n}\langle f, e_i\rangle e_i)_{n=0}^{\infty}$$

of partial sums is a Cauchy sequence in $L^2(\mu)$. We say that $(e_i)_{i=0}^{\infty}$ is an orthonormal basis in $L^2(\mu)$ if it is orthonormal and

$$f = \Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i \text{ for all } f \in L^2(\mu).$$

Theorem 4.3.1. *An orthonormal sequence $(e_i)_{i=0}^{\infty}$ in $L^2(\mu)$ is a basis of $L^2(\mu)$ if*

$$(\langle f, e_i\rangle = 0 \text{ all } i \in \mathbf{N}) \Rightarrow f = 0$$

Proof. Let $f \in L^2(\mu)$ and set

$$g = f - \Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i.$$

Then, for any $j \in \mathbf{N}$,

$$\langle g, e_j\rangle = \langle f - \Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i, e_j\rangle$$

$$= \langle f, e_j\rangle - \Sigma_{i=0}^{\infty}\langle f, e_i\rangle\langle e_i, e_j\rangle = \langle f, e_j\rangle - \langle f, e_j\rangle = 0.$$

Thus $g = 0$ or

$$f = \Sigma_{i=0}^{\infty}\langle f, e_i\rangle e_i.$$

The theorem is proved.

As an example of an application of Theorem 4.3.1 we next construct an orthonormal basis of $L^2(\lambda)$, where λ is linear measure in the unit interval. Set

$$H(t) = \chi_{[0,\frac{1}{2}[}(t) - \chi_{[\frac{1}{2},1]}(t),\ t \in \mathbf{R}$$

Moreover, define $h_{00}(t) = 1,\ 0 \leq t \leq 1$, and for each $n \geq 1$ and $j = 1, ..., 2^{n-1}$,

$$h_{jn}(t) = 2^{\frac{n-1}{2}} H(2^{n-1}t - j + 1),\ 0 \leq t \leq 1.$$

Stated otherwise, we have for each $n \geq 1$ and $j = 1, ..., 2^{n-1}$

$$h_{jn}(t) = \begin{cases} 2^{\frac{n-1}{2}}, \frac{j-1}{2^{n-1}} \leq t < \frac{j-\frac{1}{2}}{2^{n-1}}, \\ \\ -2^{\frac{n-1}{2}}, \frac{j-\frac{1}{2}}{2^{n-1}} \leq t \leq \frac{j}{2^{n-1}}, \\ \\ 0, \text{ elsewhere in } [0,1]. \end{cases}$$

It is simple to show that the sequence $h_{00,}h_{jn}, j = 1, ..., 2^{n-1}, n \geq 1$, is orthonormal in $L^2(\lambda)$. We will prove that the same sequence constitute an orthonormal basis of $L^2(\lambda)$. Therefore, suppose $f \in L^2(\lambda)$ is orthogonal to each of the functions $h_{00,}h_{jn}, j = 1, ..., 2^{n-1}, n \geq 1$. Then for each $n \geq 1$ and $j = 1, ..., 2^{n-1}$

$$\int_{\frac{j-1}{2^{n-1}}}^{\frac{j-\frac{1}{2}}{2^{n-1}}} f d\lambda = \int_{\frac{j-\frac{1}{2}}{2^{n-1}}}^{\frac{j}{2^{n-1}}} f d\lambda$$

and, hence,

$$\int_{\frac{j-1}{2^{n-1}}}^{\frac{j}{2^{n-1}}} f d\lambda = \frac{1}{2^{n-1}} \int_0^1 f d\lambda = 0$$

since

$$\int_0^1 f d\lambda = \int_0^1 f h_{00} d\lambda = 0.$$

Thus

$$\int_{\frac{j}{2^{n-1}}}^{\frac{k}{2^{n-1}}} f d\lambda = 0, \ 1 \leq j \leq k \leq 2^{n-1}$$

and we conclude that

$$\int_0^1 1_{[a,b]} f d\lambda = \int_a^b f d\lambda = 0, \ 0 \leq a \leq b \leq 1.$$

Accordingly from this, $f = 0$ and we are done.

The above basis $(h_k)_{k=0}^{\infty} = (h_{00,}h_{11}, h_{12}, h_{22}, h_{13}, h_{23}, h_{33}, h_{43}, ...)$ of $L^2(\lambda)$ is called the Haar basis.

Let $0 \leq t \leq 1$ and define for fixed $k \in \mathbf{N}$

$$a_k(t) = \int_0^1 \chi_{[0,t]}(x) h_k(x) dx = \int_0^t h_k d\lambda$$

so that

$$\chi_{[0,t]} = \Sigma_{k=0}^{\infty} a_k(t) h_k \text{ in } L^2(\lambda).$$

Then, if $0 \leq s, t \leq 1$,

$$\min(s,t) = \int_0^1 \chi_{[0,s]}(x)\chi_{[0,t]}(x)dx = \langle \Sigma_{k=1}^{\infty} a_k(s) h_k, \chi_{[0,t]} \rangle$$

$$= \Sigma_{k=0}^{\infty} a_k(s) \langle h_k, \chi_{[0,t]} \rangle = \Sigma_{k=0}^{\infty} a_k(s) a_k(t).$$

Note that

$$t = \Sigma_{k=0}^{\infty} a_k^2(t).$$

If $(G_k)_{k=0}^{\infty}$ is a sequence of $N(0,1)$ distributed random variables based on a probability space $(\Omega, \mathcal{F}, P)$ the series

$$\Sigma_{k=0}^{\infty} a_k(t) G_k$$

converges in $L^2(P)$ and defines a Gaussian random variable which we denote by $W(t)$. From the above it follows that $(W(t))_{0 \leq t \leq 1}$ is a real-valued centred Gaussian stochastic process with the covariance

$$E\left[W(s)W(t)\right] = \min(s,t).$$

Such a process is called a real-valued Brownian motion in the time interval $[0,1]$.

Recall that

$$(h_{00,} h_{11}, h_{12}, h_{22}, h_{13}, h_{23}, h_{33}, h_{43}, ...) = (h_k)_{k=0}^{\infty}.$$

We define

$$(a_{00,} a_{11}, a_{12}, a_{22}, a_{13}, a_{23}, a_{33}, a_{43}, ...) = (a_k)_{k=0}^{\infty}$$

and

$$(G_{00,} G_{11}, G_{12}, G_{22}, G_{13}, G_{23}, G_{33}, G_{43}, ...) = (G_k)_{k=0}^{\infty}.$$

It is important to note that for fixed n,

$$a_{jn}(t) = \int_0^t \chi_{[0,t]}(x) h_{jn}(x) dx \neq 0 \text{ for at most one } j.$$

Set

$$U_0(t) = a_{00}(t) G_{00}$$

and

$$U_n(t) = \Sigma_{j=1}^{2^{n-1}} a_{jn}(t) G_{jn},\ n \in \mathbf{N}_+.$$

We know that

$$W(t) = \Sigma_{n=0}^{\infty} U_n(t) \text{ in } L^2(P)$$

for fixed t.

The space $C\,[0,1]$ will from now on be equipped with the metric

$$d(x,y) = \parallel x - y \parallel_\infty$$

where $\parallel x \parallel_\infty = \max_{0\leq t\leq 1} \mid x(t) \mid$. Recall that every $x \in C\,[0,1]$ is uniformly continuous. From this, remembering that $\mathbf{R}$ is separable, it follows that the space $C\,[0,1]$ is separable. Since $\mathbf{R}$ is complete it is also simple to show that the metric space $C\,[0,1]$ is complete. Finally, if $x_n \in C\,[0,1]$, $n \in \mathbf{N}$, and

$$\Sigma_{n=0}^{\infty} \parallel x_n \parallel_\infty < \infty$$

the series

$$\Sigma_{n=0}^{\infty} x_n$$

converges since the partial sums

$$s_n = \Sigma_{k=0}^{n} x_k,\ k \in \mathbf{N}$$

forms a Cauchy sequence.

We now define

$$\Theta = \left\{\omega \in \Omega; \Sigma_{n=0}^{\infty} \parallel U_n \parallel_\infty < \infty\right\}.$$

Here $\Theta \in \mathcal{F}$ since

$$\parallel U_n \parallel_\infty = \sup_{\substack{0\leq t\leq 1\\ t\in\mathbf{Q}}} \mid U_n(t) \mid$$

for each n. Next we prove that $\Omega \setminus \Theta$ is a null set.

To this end let $n \geq 1$ and note that

$$P\left[\parallel U_n \parallel_\infty > 2^{-\frac{n}{4}}\right] \leq P\left[\max_{1\leq j\leq 2^{n-1}} (\parallel a_{jn} \parallel_\infty \mid G_{jn} \mid) > 2^{-\frac{n}{4}}\right].$$

But

$$\parallel a_{jn} \parallel_\infty = \frac{1}{2^{\frac{n+1}{2}}}$$

and, hence,

$$P\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right] \leq 2^{n-1} P\left[\mid G_{00} \mid > 2^{\frac{n}{4}+\frac{1}{2}}\right].$$

Since

$$x \geq 1 \Rightarrow P\left[\mid G_{00} \mid \geq x\right] \leq 2\int_x^\infty ye^{-y^2/2}\frac{dy}{x\sqrt{2\pi}} \leq e^{-x^2/2}$$

we get

$$P\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right] \leq 2^n e^{-2^{n/2}}$$

and conclude that

$$E\left[\sum_{n=0}^\infty 1_{\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right]}\right] = \sum_{n=0}^\infty P\left[\| U_n \|_\infty > 2^{-\frac{n}{4}}\right] < \infty.$$

From this and the Beppo Levi Theorem (or the first Borel-Cantelli Lemma) $P[\Theta] = 1$.

The trajectory $t \to W(t,\omega)$, $0 \leq t \leq 1$, is continuous for every $\omega \in \Theta$. Without loss of generality, from now on we can therefore assume that all trajectories of Brownian motion are continuous (by eventually replacing Ω by Θ).

Suppose

$$0 \leq t_1 < ... < t_n \leq 1$$

and let $I_1, ..., I_n$ be open subintervals of the real line. The set

$$S(t_1,...,t_n; I_1,...,I_n) = \{x \in C[0,1]\,;\; x(t_k) \in I_k,\; k = 1,...,n\}$$

is called an open n-cell in $C[0,1]$. A set in $C[0,T]$ is called an open cell if there exists an $n \in \mathbf{N}_+$ such that it is an open n-cell. The σ-algebra generated by all open cells in $C[0,1]$ is denoted by $\mathcal{C}$. The construction above shows that the map

$$W : \Omega \to C[0,1]$$

which maps ω to the trajectory

$$t \to W(t,\omega),\; 0 \leq t \leq 1$$

is $(\mathcal{F},\mathcal{C})$-measurable. The image measure P_W is called Wiener measure in $C[0,1]$.

The Wiener measure is a Borel measure on the metric space $C\,[0,1]$. We leave it as an excersice to prove that

$$\mathcal{C} = \mathcal{B}(C\,[0,1]).$$

$\uparrow\uparrow\uparrow$

CHAPTER 5

DECOMPOSITION OF MEASURES

Introduction

In this section a version of the fundamental theorem of calculus for Lebesgue integrals will be proved. Moreover, the concept of differentiating a measure with respect to another measure will be developped. A very important result in this chapter is the so called Radon-Nikodym Theorem.

5.1. Complex Measures

Let $(X, \mathcal{M})$ be a measurable space. Recall that if $A_n \subseteq X$, $n \in \mathbf{N}_+$, and $A_i \cap A_j = \phi$ if $i \neq j$, the sequence $(A_n)_{n \in \mathbf{N}_+}$ is called a disjoint denumerable collection. The collection is called a measurable partition of A if $A = \cup_{n=1}^{\infty} A_n$ and $A_n \in \mathcal{M}$ for every $n \in \mathbf{N}_+$.

A complex function μ on $\mathcal{M}$ is called a complex measure if

$$\mu(A) = \Sigma_{n=1}^{\infty} \mu(A_n)$$

for every $A \in \mathcal{M}$ and measurable partition $(A_n)_{n=1}^{\infty}$ of A. Note that $\mu(\phi) = 0$ if μ is a complex measure. A complex measure is said to be a real measure if it is a real function. The reader should note that a positive measure need not be a real measure since infinity is not a real number. If μ is a complex measure $\mu = \mu_{\mathrm{Re}} + i\mu_{\mathrm{Im}}$, where μ_{Re} =Re μ and μ_{Im} =Im μ are real measures.

If $(X, \mathcal{M}, \mu)$ is a positive measure and $f \in L^1(\mu)$ it follows that

$$\lambda(A) = \int_A f d\mu, \ A \in \mathcal{M}$$

is a real measure and we write $d\lambda = f d\mu$.

A function $\mu : \mathcal{M} \to [-\infty, \infty]$ is called a signed measure measure if

(a) $\mu : \mathcal{M} \to]-\infty, \infty]$ or $\mu : \mathcal{M} \to [-\infty, \infty[$
(b) $\mu(\phi) = 0$
and
(c) for every $A \in \mathcal{M}$ and measurable partition $(A_n)_{n=1}^{\infty}$ of A,

$$\mu(A) = \Sigma_{n=1}^{\infty} \mu(A_n)$$

where the latter sum converges absolutely if $\mu(A) \in \mathbf{R}$.

Here $-\infty - \infty = -\infty$ and $-\infty + x = -\infty$ if $x \in \mathbf{R}$. The sum of a positive measure and a real measure and the difference of a real measure and a positive measure are examples of signed measures and it can be proved that there are no other signed measures (see Folland [F]). Below we concentrate on positive, real, and complex measures and will not say more about signed measures here.

Suppose μ is a complex measure on $\mathcal{M}$ and define for every $A \in \mathcal{M}$

$$\mid \mu \mid (A) = \sup \Sigma_{n=1}^{\infty} \mid \mu(A_n) \mid,$$

where the supremum is taken over all measurable partitions $(A_n)_{n=1}^{\infty}$ of A. Note that $\mid \mu \mid (\phi) = 0$ and

$$\mid \mu \mid (A) \geq \mid \mu(B) \mid \text{ if } A, B \in \mathcal{M} \text{ and } A \supseteq B.$$

The set function $\mid \mu \mid$ is called the total variation of μ or the total variation measure of μ. It turns out that $\mid \mu \mid$ is a positive measure. In fact, as will shortly be seen, $\mid \mu \mid$ is a finite positive measure.

Theorem 5.1.1. *The total variation $\mid \mu \mid$ of a complex measure is a positive measure.*

PROOF. Let $(A_n)_{n=1}^{\infty}$ be a measurable partition of A.

For each n, suppose $a_n <| \mu | (A_n)$ and let $(E_{kn})_{k=1}^{\infty}$ be a measurable partition of A_n such that

$$a_n < \Sigma_{k=1}^{\infty} | \mu(E_{kn}) | .$$

Since $(E_{kn})_{k,n=1}^{\infty}$ is a partition of A it follows that

$$\Sigma_{n=1}^{\infty} a_n < \Sigma_{k,n=1}^{\infty} | \mu(E_{kn}) | \leq | \mu | (A).$$

Thus

$$\Sigma_{n=1}^{\infty} | \mu | (A_n) \leq | \mu | (A).$$

To prove the opposite inequality, let $(E_k)_{k=1}^{\infty}$ be a measurable partition of A. Then, since $(A_n \cap E_k)_{n=1}^{\infty}$ is a measurable partition of E_k and $(A_n \cap E_k)_{k=1}^{\infty}$ a measurable partition of A_n,

$$\Sigma_{k=1}^{\infty} | \mu(E_k) | = \Sigma_{k=1}^{\infty} | \Sigma_{n=1}^{\infty} \mu(A_n \cap E_k) |$$

$$\leq \Sigma_{k,n=1}^{\infty} | \mu(A_n \cap E_k) | \leq \Sigma_{n=1}^{\infty} | \mu | (A_n)$$

and we get

$$| \mu | (A) \leq \Sigma_{n=1}^{\infty} | \mu | (A_n).$$

Thus

$$| \mu | (A) = \Sigma_{n=1}^{\infty} | \mu | (A_n).$$

Since $| \mu | (\phi) = 0$, the theorem is proved.

Theorem 5.1.2. *The total variation $| \mu |$ of a complex measure μ is a finite positive measure.*

PROOF. Since

$$| \mu | \leq | \mu_{\mathrm{Re}} | + | \mu_{\mathrm{Im}} |$$

there is no loss of generality to assume that μ is a real measure.

Suppose $| \mu | (E) = \infty$ for some $E \in \mathcal{M}$. We first prove that there exist disjoint sets $A, B \in \mathcal{M}$ such that

$$A \cup B = E$$

and

$$| \mu(A) |> 1 \text{ and } | \mu | (B) = \infty.$$

To this end let $c = 2(1+ | \mu(E) |)$ and let $(E_k)_{k=1}^{\infty}$ be a measurable partition of E such that

$$\Sigma_{k=1}^{n} | \mu(E_k) |> c$$

for some sufficiently large n. There exists a subset N of $\{1, ..., n\}$ such that

$$| \Sigma_{k\in N}\mu(E_k) |> \frac{c}{2}.$$

Set $A = \cup_{k\in N} E_k$ and $B = E \setminus A$. Then $| \mu(A) |> \frac{c}{2} \geq 1$ and

$$| \mu(B) |=| \mu(E) - \mu(A) |$$

$$\geq| \mu(A) | - | \mu(E) |> \frac{c}{2} - | \mu(E) |= 1.$$

Since $\infty =| \mu | (E) =| \mu | (A)+ | \mu | (B)$ we have $| \mu | (A) = \infty$ or $| \mu | (B) = \infty$. If $| \mu | (B) < \infty$ we interchange A and B and have $| \mu(A) |> 1$ and $| \mu | (B) = \infty$.

Suppose $| \mu | (X) = \infty$. Set $E_0 = X$ and choose disjoint sets $A_0, B_0 \in \mathcal{M}$ such that

$$A_0 \cup B_0 = E_0$$

and

$$| \mu(A_0) |> 1 \text{ and } | \mu | (B_0) = \infty.$$

Set $E_1 = B_0$ and choose disjoint sets $A_1, B_1 \in \mathcal{M}$ such that

$$A_1 \cup B_1 = E_1$$

and

$$| \mu(A_1) |> 1 \text{ and } | \mu | (B_1) = \infty.$$

By induction, we find a measurable partition $(A_n)_{n=0}^{\infty}$ of the set $A =_{def} \cup_{n=0}^{\infty} A_n$ such that $| \mu(A_n) |> 1$ for every n. Now, since μ is a complex measure,

$$\mu(A) = \Sigma_{n=0}^{\infty} \mu(A_n).$$

But this series cannot converge, since the general term does not tend to zero as $n \to \infty$. This contradiction shows that $| \mu |$ is a finite positive measure.

If μ is a real measure we define

$$\mu^+ = \frac{1}{2}(| \mu | + \mu)$$

and

$$\mu^- = \frac{1}{2}(| \mu | - \mu).$$

The measures μ^+ and μ^- are finite positive measures and are called the positive and negative variations of μ, respectively . The representation

$$\mu = \mu^+ - \mu^-$$

is called the Jordan decomposition of μ.

Exercises

1. Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space and $d\lambda = f d\mu$, where $f \in L^1(\mu)$. Prove that $d | \lambda | = | f | d\mu$.

2. Suppose λ, μ, and ν are real measures defined on the same σ-algebra and $\lambda \leq \mu$ and $\lambda \leq \nu$. Prove that

$$\lambda \leq \min(\mu, \nu)$$

where

$$\min(\mu, \nu) = \frac{1}{2}(\mu + \nu - | \mu - \nu |).$$

3. Suppose $\mu : \mathcal{M} \to \mathbf{C}$ is a complex measure and $f, g : X \to \mathbf{R}$ measurable functions. Show that

$$| \mu(f \in A) - \mu(g \in A) | \leq | \mu | (f \neq g)$$

for every $A \in \mathcal{R}$.

5.2. The Lebesque Decomposition and the Radon-Nikodym Theorem

Let μ be a positive measure on $\mathcal{M}$ and λ a positive or complex measure on $\mathcal{M}$. The measure λ is said to be absolutely continuous with respect to μ (abbreviated $\lambda << \mu$) if $\lambda(A) = 0$ for every $A \in \mathcal{M}$ for which $\mu(A) = 0$. If we define

$$\mathcal{Z}_\lambda = \{A \in \mathcal{M};\ \lambda(A) = 0\}$$

it follows that $\lambda << \mu$ if and only if

$$\mathcal{Z}_\mu \subseteq \mathcal{Z}_\lambda.$$

For example, $\gamma_n << v_n$ and $v_n << \gamma_n$.

The measure λ is said to be concentrated on $E \in \mathcal{M}$ if $\lambda = \lambda^E$, where $\lambda^E(A) =_{def} \lambda(E \cap A)$ for every $A \in \mathcal{M}$. This is equivalent to the hypothesis that $A \in \mathcal{Z}_\lambda$ if $A \in \mathcal{M}$ and $A \cap E = \phi$. Thus if $E_1, E_2 \in \mathcal{M}$, where $E_1 \subseteq E_2$, and λ is concentrated on E_1, then λ is concentrated on E_2. Moreover, if $E_1, E_2 \in \mathcal{M}$ and λ is concentrated on both E_1 and E_2, then λ is concentrated on $E_1 \cap E_2$. Two measures λ_1 and λ_2 are said to be mutually singular (abbreviated $\lambda_1 \perp \lambda_2$) if there exist disjoint measurable sets E_1 and E_2 such that λ_1 is concentrated on E_1 and λ_2 is concentrated on E_2.

Theorem 5.2.1. *Let μ be a positive measure and λ, λ_1, and λ_2 complex measures.*

(i) *If $\lambda_1 << \mu$ and $\lambda_2 << \mu$, then $(\alpha_1\lambda_1 + \alpha_2\lambda_2) << \mu$ for all complex numbers α_1 and α_2.*

(ii) *If $\lambda_1 \perp \mu$ and $\lambda_2 \perp \mu$, then $(\alpha_1\lambda_1 + \alpha_2\lambda_2) \perp \mu$ for all complex numbers α_1 and α_2.*

(iii) *If $\lambda << \mu$ and $\lambda \perp \mu$, then $\lambda = 0$.*

(iv) *If $\lambda << \mu$, then $\mid \lambda \mid << \mu$.*

PROOF. The properties (i) and (ii) are simple to prove and are left as exercises.

To prove (iii) suppose $E \in \mathcal{M}$ is a μ-null set and $\lambda = \lambda^E$. If $A \in \mathcal{M}$, then $\lambda(A) = \lambda(A \cap E)$ and $A \cap E$ is a μ-null set. Since $\lambda << \mu$ it follows that $A \cap E \in Z_\lambda$ and, hence, $\lambda(A) = \lambda(A \cap E) = 0$. This proves (iii)

To prove (iv) suppose $A \in \mathcal{M}$ and $\mu(A) = 0$. If $(A_n)_{n=1}^{\infty}$ is measurable partition of A, then $\mu(A_n) = 0$ for every n. Since $\lambda << \mu$, $\lambda(A_n) = 0$ for every n and we conclude that $\mid \lambda \mid (A) = 0$. This proves (vi).

Theorem 5.2.2. *Let μ be a positive measure on $\mathcal{M}$ and λ a complex measure on $\mathcal{M}$. Then the following conditions are equivalent:*

(a) $\lambda << \mu$.

(b) *To every $\varepsilon > 0$ there corresponds a $\delta > 0$ such that $\mid \lambda(E) \mid < \varepsilon$ for all $E \in \mathcal{M}$ with $\mu(E) < \delta$.*

If λ is a positive measure, the implication $(a) \Rightarrow (b)$ in Theorem 5.2.2 is, in general, wrong. To see this take $\mu = \gamma_1$ and $\lambda = v_1$. Then $\lambda << \mu$ and if we choose $A_n = [n, \infty[$, $n \in \mathbf{N}_+$, then $\mu(A_n) \to 0$ as $n \to \infty$ but $\lambda(A_n) = \infty$ for each n.

PROOF. (a)$\Rightarrow$(b). If (b) is wrong there exist an $\varepsilon > 0$ and sets $E_n \in \mathcal{M}$, $n \in \mathbf{N}_+$, such that $\mid \lambda(E_n) \mid \geq \varepsilon$ and $\mu(E_n) < 2^{-n}$. Set

$$A_n = \cup_{k=n}^{\infty} E_k \text{ and } A = \cap_{n=1}^{\infty} A_n.$$

Since $A_n \supseteq A_{n+1} \supseteq A$ and $\mu(A_n) < 2^{-n+1}$, it follows that $\mu(A) = 0$ and using that $\mid \lambda \mid (A_n) \geq \mid \lambda(E_n) \mid$, Theorem 1.1.2 (f) implies that

$$\mid \lambda \mid (A) = \lim_{n \to \infty} \mid \lambda \mid (A_n) \geq \varepsilon.$$

This contradicts that $\mid \lambda \mid << \mu$.

(b)$\Rightarrow$(a). If $E \in \mathcal{M}$ and $\mu(E) = 0$ then to each $\varepsilon > 0$, $\mid \lambda(E) \mid < \varepsilon$, and we conclude that $\lambda(E) = 0$. The theorem is proved.

Theorem 5.2.3. *Let μ be a σ-finite positive measure and λ a real measure on $\mathcal{M}$.*

(a) **(The Lebesgue Decomposition of λ)** *There exists a unique pair of real measures λ_a and λ_s on $\mathcal{M}$ such that*

$$\lambda = \lambda_a + \lambda_s, \ \lambda_a << \mu, \text{ and } \lambda_s \perp \mu.$$

If λ is a finite positive measure, λ_a and λ_s are finite positive measures.

(b) **(The Radon-Nikodym Theorem)** *There exits a unique $g \in L^1(\mu)$ such that*

$$d\lambda_a = g d\mu.$$

If λ is a finite positive measure, $g \geq 0$ a.e. $[\mu]$.

The proof of Theorem 5.2.3 is based on the following

Lemma 5.2.1. *Let (X, M, μ) be a finite positive measure space and suppose $f \in L^1(\mu)$.*

(a) *If $a \in \mathbf{R}$ and*

$$\int_E f d\mu \leq a\mu(E), \text{ all } E \in \mathcal{M}$$

then $f \leq a$ a.e. $[\mu]$.

(b) *If $b \in \mathbf{R}$ and*

$$\int_E f d\mu \geq b\mu(E), \text{ all } E \in \mathcal{M}$$

then $f \geq b$ a.e. $[\mu]$.

PROOF. (a) Set $g = f - a$ so that

$$\int_E g d\mu \leq 0, \ \textit{all } E \in \mathcal{M}.$$

Now choose $E = \{g > 0\}$ to obtain

$$0 \geq \int_E g d\mu = \int_X \chi_E g d\mu \geq 0$$

as $\chi_E g \geq 0$ a.e. $[\mu]$. But then Example 2.1.2 yields $\chi_E g = 0$ a.e. $[\mu]$ and we get $E \in Z_\mu$. Thus $g \leq 0$ a.e. $[\mu]$ or $f \leq a$ a.e. $[\mu]$.

Part (b) follows in a similar way as Part (a) and the proof is omitted here.

PROOF. Uniqueness: (a) Suppose $\lambda_a^{(k)}$ and $\lambda_s^{(k)}$ are real measures on $\mathcal{M}$ such that

$$\lambda = \lambda_a^{(k)} + \lambda_s^{(k)},\ \lambda_a^{(k)} << \mu,\ \text{and } \lambda_s^{(k)} \perp \mu$$

for $k = 1, 2$. Then

$$\lambda_a^{(1)} - \lambda_a^{(2)} = \lambda_s^{(2)} - \lambda_s^{(1)}$$

and

$$\lambda_a^{(1)} - \lambda_a^{(2)} << \mu \text{ and } \lambda_a^{(1)} - \lambda_a^{(2)} \perp \mu.$$

Thus by applying Theorem 5.2.1, $\lambda_a^{(1)} - \lambda_a^{(2)} = 0$ and $\lambda_a^{(1)} = \lambda_a^{(2)}$. From this we conclude that $\lambda_s^{(1)} = \lambda_s^{(2)}$.

(b) Suppose $g_k \in L^1(\mu), k = 1, 2,$ and

$$d\lambda_a = g_1 d\mu = g_2 d\mu.$$

Then $hd\mu = 0$ where $h = g_1 - g_2$. But then

$$\int_{\{h>0\}} hd\mu = 0$$

and it follows that $h \leq 0$ a.e. $[\mu]$. In a similar way we prove that $h \geq 0$ a.e. $[\mu]$. Thus $h = 0$ in $L^1(\mu)$, that is $g_1 = g_2$ in $L^1(\mu)$.

Existence: The beautiful proof that follows is due to von Neumann.

First suppose that μ and λ are finite positive measures and set $\nu = \lambda + \mu$. Clearly, $L^1(\lambda) \supseteq L^1(\nu) \supseteq L^2(\nu)$. Moreover, if $f : X \to \mathbf{R}$ is measurable

$$\int_X \mid f \mid d\lambda \leq \int_X \mid f \mid d\nu \leq \sqrt{\int_X f^2 d\nu} \sqrt{\nu(X)}$$

and from this we conclude that the map

$$f \to \int_X f d\lambda$$

is a continuous linear functional on $L^2(\nu)$. Therefore, in view of Theorem 4.2.2, there exists a $g \in L^2(\nu)$ such that

$$\int_X f d\lambda = \int_X f g d\nu \text{ all } f \in L^2(\nu).$$

Suppose $E \in \mathcal{M}$ and put $f = \chi_E$ to obtain

$$0 \leq \lambda(E) = \int_E g d\nu$$

and, since $\nu \geq \lambda$,

$$0 \leq \int_E g d\nu \leq \nu(E).$$

But then Lemma 5.2.1 implies that $0 \leq g \leq 1$ a.e. $[\nu]$. Therefore, without loss of generality we can assume that $0 \leq g(x) \leq 1$ for all $x \in X$ and, in addition, as above

$$\int_X f d\lambda = \int_X f g d\nu \text{ all } f \in L^2(\nu)$$

that is

$$\int_X f(1-g) d\lambda = \int_X f g d\mu \text{ all } f \in L^2(\nu).$$

Put $A = \{0 \leq g < 1\}$, $S = \{g = 1\}$, $\lambda_a = \lambda^A$, and $\lambda_s = \lambda^S$. Note that $\lambda = \lambda^A + \lambda^S$. The choice $f = \chi_S$ gives $\mu(S) = 0$ and hence $\lambda_s \perp \mu$. Moreover, the choice

$$f = (1 + ... + g^n)\chi_E$$

where $E \in \mathcal{M}$, gives

$$\int_E (1 - g^{n+1}) d\lambda = \int_E (1 + ... + g^n) g d\mu.$$

By letting $n \to \infty$ and using monotone convergence

$$\lambda(E \cap A) = \int_E h d\mu.$$

where

$$h = \lim_{n \to \infty} (1 + ... + g^n) g.$$

Since h is non-negative and

$$\lambda(A) = \int_X h d\mu$$

it follows that $h \in L^1(\mu)$. Moreover, the construction above shows that $\lambda = \lambda_a + \lambda_s$.

In the next step we assume that μ is a σ-finite positive measure and λ a finite positive measure. Let $(X_n)_{n=1}^{\infty}$ be a measurable partition of X such that $\mu(X_n) < \infty$ for every n. Let n be fixed and apply Part (a) to the pair μ^{X_n} and λ^{X_n} to obtain finite positive measures $(\lambda^{X_n})_a$ and $(\lambda^{X_n})_s$ such that

$$\lambda^{X_n} = (\lambda^{X_n})_a + (\lambda^{X_n})_s,\ (\lambda^{X_n})_a << \mu^{X_n},\ \text{and } (\lambda^{X_n})_s \perp \mu^{X_n}$$

and

$$d(\lambda^{X_n})_a = h_n d\mu^{X_n} \text{ (or } (\lambda^{X_n})_a = h_n \mu^{X_n})$$

where $0 \leq h_n \in L^1(\mu^{X_n})$. Without loss of generality we can assume that $h_n = 0$ off X_n and that $(\lambda^{X_n})_s$ is concentrated on $A_n \subseteq X_n$ where $A_n \in \mathcal{Z}_\mu$. In particular, $(\lambda^{X_n})_a = h_n \mu$. Now

$$\lambda = h\mu + \Sigma_{n=1}^{\infty} (\lambda^{X_n})_s$$

where

$$h = \Sigma_{n=1}^{\infty} h_n$$

and

$$\int_X h d\mu \leq \lambda(X) < \infty.$$

Thus $h \in L^1(\mu)$. Moreover, $\lambda_s =_{def} \Sigma_{n=1}^{\infty} (\lambda^{X_n})_s$ is concentrated on $\cup_{n=1}^{\infty} A_n \in \mathcal{Z}_\mu$. Hence $\lambda_s \perp \mu$.

Finally if λ is a real measure we apply what we have already proved to the positive and negative variations of λ and we are done.

Example 5.2.1. Let λ be Lebesgue measure in the unit interval and μ the counting measure in the unit interval restricted to the class of all Lebesgue measurable subsets of the unit interval. Clearly, $\lambda << \mu$. Suppose there is an

$h \in L^1(\mu)$ such that $d\lambda = hd\mu$. We can assume that $h \geq 0$ and the Markov inequality implies that the set $\{h \geq \varepsilon\}$ is finite for every $\varepsilon > 0$. But then

$$\lambda(h \in]0,1]) = \lim_{n\to\infty} \lambda(h \geq 2^{-n}) = 0$$

and it follows that $1 = \lambda(h = 0) = \int_{\{h=0\}} hd\mu = 0$, which is a contradiction.

Corollary 5.2.1. *Suppose μ is a real measure. Then there exists*

$$h \in L^1(\mid \mu \mid)$$

such that $\mid h(x) \mid = 1$ for all $x \in X$ and

$$d\mu = hd \mid \mu \mid .$$

PROOF. Since $\mid \mu(A) \mid \leq \mid \mu \mid (A)$ for every $A \in \mathcal{M}$, the Radon-Nikodym Theorem implies that $d\mu = hd \mid \mu \mid$ for an appropriate $h \in L^1(\mid \mu \mid)$. But then $d \mid \mu \mid = \mid h \mid d \mid \mu \mid$ (see Exercise 1 in Chapter 5.1). Thus

$$\mid \mu \mid (E) = \int_E \mid h \mid d \mid \mu \mid, \text{ all } E \in \mathcal{M}$$

and Lemma 5.2.1 yields $h = 1$ a.e. $[\mid \mu \mid]$. From this the theorem follows at once.

Theorem 5.2.4. (Hahn's Decomposition Theorem) *Suppose μ is a real measure. There exists an $A \in \mathcal{M}$ such that*

$$\mu^+ = \mu^A \text{ and } \mu^- = -\mu^{A^c}.$$

PROOF. Let $d\mu = hd \mid \mu \mid$ where $\mid h \mid = 1$. Note that $hd\mu = d \mid \mu \mid$. Set $A = \{h = 1\}$. Then

$$d\mu^+ = \frac{1}{2}(d \mid \mu \mid + d\mu) = \frac{1}{2}(h+1)d\mu = \chi_A d\mu$$

and

$$d\mu^{-} = d\mu^{+} - d\mu = (\chi_A - 1)d\mu = -\chi_{A^c} d\mu.$$

The theorem is proved.

If a real measure λ is absolutely continuous with respect to a σ-finite positive measure μ, the Radon-Nikodym Theorem says that $d\lambda = f d\mu$ for an approprite $f \in L^1(\mu)$. We sometimes write

$$f = \frac{d\lambda}{d\mu}$$

and call f the Radon-Nikodym derivate of λ with respect to μ.

Exercises

1. Suppose μ and $\nu_n, n \in\mathbf{N}$, are positive measures defined on the same σ-algebra and set $\theta = \Sigma_{n=0}^{\infty} \nu_n$. Prove that
   a) $\theta \perp \mu$ if $\nu_n \perp \mu$, all $n \in \mathbf{N}$.
   b) $\theta << \mu$ if $\nu_n << \mu$, all $n \in \mathbf{N}$.

2. Suppose μ is a real measure and $\mu = \lambda_1 - \lambda_2$, where λ_1 and λ_2 are finite positive measures. Prove that $\lambda_1 \geq \mu^+$ and $\lambda_2 \geq \mu^-$.

3. Let λ_1 and λ_2 be mutually singular complex measures on the same σ-algebra. Show that $\mid \lambda_1 \mid \perp \mid \lambda_2 \mid$.

4. Let $(X, \mathcal{M}, \mu)$ be a σ-finite positive measure space and suppose λ and τ are two probability measures defined on the σ-algebra $\mathcal{M}$ such that $\lambda << \mu$ and $\tau << \mu$. Prove that

$$\sup_{A \in \mathcal{M}} \mid \lambda(A) - \tau(A) \mid = \frac{1}{2} \int_X \mid \frac{d\lambda}{d\mu} - \frac{d\tau}{d\mu} \mid d\mu.$$

5.3. The Wiener Maximal Theorem and the Lebesgue Differentiation Theorem

We say that a Lebesgue measurable function f in $\mathbf{R}^n$ is locally Lebesgue integrable and belongs to the class $L^1_{loc}(m_n)$ if $f\chi_K \in L^1(m_n)$ for each compact subset K of $\mathbf{R}^n$. In a similar way $f \in L^1_{loc}(v_n)$ if f is a Borel function such that $f\chi_K \in L^1(v_n)$ for each compact subset K of $\mathbf{R}^n$. If $f \in L^1_{loc}(m_n)$, we define the average $A_r f(x)$ of f on the open ball $B(x,r)$ as

$$A_r f(x) = \frac{1}{m_n(B(x,r))} \int_{B(x,r)} f(y)dy.$$

It follows from dominated convergence that the map $(x,r) \to A_r f(x)$ of $\mathbf{R}^n \times]0,\infty[$ into $\mathbf{R}$ is continuous. The Hardy-Littlewood maximal function f^* is, by definition, $f^* = \sup_{r>0} A_r \mid f \mid$ or, stated more explicitly,

$$f^*(x) = \sup_{r>0} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} \mid f(y) \mid dy,\ x \in \mathbf{R}^n.$$

The function $f^* : (\mathbf{R}^n, \mathcal{B}(\mathbf{R}^n)) \to ([0,\infty], \mathcal{R}_{0,\infty})$ is measurable since

$$f^* = \sup_{\substack{r>0 \\ r\in\mathbf{Q}}} A_r \mid f \mid .$$

Theorem 5.3.1. (Wiener's Maximal Theorem) *There exists a positive constant $C = C_n < \infty$ such that for all $f \in L^1(m_n)$,*

$$m_n(f^* > \alpha) \leq \frac{C}{\alpha} \parallel f \parallel_1 \text{ if } \alpha > 0.$$

The proof of the Wiener Maximal Theorem is based on the following remarkable result.

Lemma 5.3.1. *Let $\mathcal{C}$ be a collection of open balls in $\mathbf{R}^n$ and set $V = \cup_{B\in\mathcal{C}} B$. If $c < m_n(V)$ there exist pairwise disjoint $B_1, ..., B_k \in \mathcal{C}$ such that*

$$\Sigma_{i=1}^k m_n(B_i) > 3^{-n} c.$$

PROOF. Let $K \subseteq V$ be compact with $m_n(K) > c$, and suppose $A_1, ..., A_p \in \mathcal{C}$ cover K. Let B_1 be the largest of the $A_i's$ (that is, B_1 has maximal radius), let B_2 be the largest of the $A_i's$ which are disjoint from B_1, let B_3 be the largest of the $A_i's$ which are disjoint from $B_1 \cup B_2$, and so on until the process stops after k steps. If $B_i = B(x_i, r_i)$ put $B_i^* = B(x_i, 3r_i)$. Then $\cup_{i=1}^k B_i^* \supseteq K$ and

$$c < \Sigma_{i=1}^k m_n(B_i^*) = 3^n \Sigma_{i=1}^k m_n(B_i).$$

The lemma is proved.

PROOF OF THEOREM 5.3.1. Set

$$E_\alpha = \{f^* > \alpha\}.$$

For each $x \in E_\alpha$ choose an $r_x > 0$ such that $A_{r_x} \mid f \mid (x) > \alpha$. If $c < m_n(E_\alpha)$, by Lemma 5.3.1 there exist $x_1, ..., x_k \in E_\alpha$ such that the balls $B_i = B(x_i, r_{x_i})$, $i = 1, ..., k$, are mutually disjoint and

$$\Sigma_{i=1}^k m_n(B_i) > 3^{-n} c.$$

But then

$$c < 3^n \Sigma_{i=1}^k m_n(B_i) < \frac{3^n}{\alpha} \Sigma_{i=1}^k \int_{B_i} \mid f(y) \mid dy \leq \frac{3^n}{\alpha} \int_{\mathbf{R}^n} \mid f(y) \mid dy.$$

The theorem is proved.

Theorem 5.3.2. *If* $f \in L^1_{loc}(m_n)$,

$$\lim_{r \to 0} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} f(y) dy = f(x) \text{ a.e. } [m_n].$$

PROOF. Clearly, there is no loss of generality to assume that $f \in L^1(m_n)$. Suppose $g \in C_c(\mathbf{R}^n) =_{def} \{f \in C(\mathbf{R}^n);\ f(x) = 0 \text{ if } \mid x \mid \text{ large enough}\}$. Then

$$\lim_{r \to 0} A_r g(x) = g(x) \text{ all } x \in \mathbf{R}^n.$$

Since $A_r f - f = A_r(f-g) - (f-g) + A_r g - g$,

$$\overline{\lim_{r\to 0}} \mid A_r f - f \mid \leq (f-g)^* + \mid f-g \mid .$$

Now, for fixed $\alpha > 0$,

$$m_n(\overline{\lim_{r\to 0}} \mid A_r f - f \mid > \alpha)$$

$$\leq m_n((f-g)^* > \frac{\alpha}{2}) + m_n(\mid f-g \mid > \frac{\alpha}{2})$$

and the Wiener Maximal Theorem and the Markov Inequality give

$$m_n(\overline{\lim_{r\to 0}} \mid A_r f - f \mid > \alpha)$$

$$\leq (\frac{2C}{\alpha} + \frac{2}{\alpha}) \parallel f - g \parallel_1 .$$

Remembering that $C_c(\mathbf{R}^n)$ is dense in $L^1(m_n)$, the theorem follows at once.

If $f \in L^1_{loc}(m_n)$ we define the so called Lebesgue set L_f to be

$$L_f = \left\{ x;\ \lim_{r\to 0} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} \mid f(y) - f(x) \mid dy = 0 \right\}.$$

Note that if q is real and

$$E_q = \left\{ x;\ \lim_{r\to 0} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} \mid f(y) - q \mid dy = \mid f(x) - q \mid \right\}$$

then $m_n(\cup_{q\in\mathbf{Q}} E^c_q) = 0$. If $x \in \cap_{q\in\mathbf{Q}} E_q$,

$$\overline{\lim_{r\to 0}} \frac{1}{m_n(B(x,r))} \int_{B(x,r)} \mid f(y) - f(x) \mid dy \leq 2 \mid f(x) - q \mid$$

for all rational numbers q and it follows that $m_n(L^c_f) = 0$.

A family $\mathcal{E}_{x,\alpha} = (E_{x,r})_{r>0}$ of Borel sets in $\mathbf{R}^n$ is said to shrink nicely to a point x in $\mathbf{R}^n$ if $E_{x,r} \subseteq B(x,r)$ for each r and there is a positive constant α, independent of r, such that $m_n(E_{x,r}) \geq \alpha m_n(B(x,r))$.

Theorem 5.3.3. (The Lebesgue Differentiation Theorem) *Suppose $f \in L^1_{loc}(m_n)$ and $x \in L_f$. Then*

$$\lim_{r \to 0} \frac{1}{m_n(E_{x,r})} \int_{E_{x,r}} \mid f(y) - f(x) \mid dy = 0$$

and

$$\lim_{r \to 0} \frac{1}{m_n(E_{x,r})} \int_{E_{x,r}} f(y) dy = f(x).$$

PROOF. The result follows from the inequality

$$\frac{1}{m_n(E_{x,r})} \int_{E_{x,r}} \mid f(y) - f(x) \mid dy \leq \frac{1}{\alpha m_n(B(x,r))} \int_{B(x,r)} \mid f(y) - f(x) \mid dy.$$

Theorem 5.3.4. *Suppose λ is a real or positive measure on $\mathcal{R}_n$ and suppose $\lambda \perp v_n$. If λ is a positive measure it is assumed that $\lambda(K) < \infty$ for every compact subset of $\mathbf{R}^n$. Then*

$$\lim_{r \to 0} \frac{\lambda(E_{x,r})}{v_n(E_{x,r})} = 0 \text{ a.e. } [v_n]$$

If $E_{x,r} = B(x,r)$ and λ is the counting measure $c_{\mathbf{Q}^n}$ restricted to $\mathcal{R}_n$ then $\lambda \perp v_n$ but the limit in Theorem 5.3.4 equals plus infinity for all $x \in \mathbf{R}^n$. The hypothesis "$\lambda(K) < \infty$ for every compact subset of $\mathbf{R}^n$" in Theorem 5.3.4 is not superflous.

PROOF. Since $\mid \lambda(E) \mid \leq \mid \lambda \mid (E)$ if $E \in \mathcal{R}_n$, there is no restriction to assume that λ is a positive measure (cf. Theorem 3.1.4). Moreover, since

$$\frac{\lambda(E_{x,r})}{v_n(E_{x,r})} \leq \frac{\lambda(B(x,r))}{\alpha v_n(B(x,r))}$$

it can be assumed that $E_{x,r} = B(x,r)$. Note that the function $\lambda(B(\cdot, r))$ is Borel measurable for fixed $r > 0$ and $\lambda(B(x, \cdot))$ left continuous for fixed $x \in \mathbf{R}^n$.

Suppose $A \in \mathcal{Z}_\lambda$ and $v_n = (v_n)^A$. Given $\delta > 0$, it is enough to prove that $F \in \mathcal{Z}_{v_n}$ where

$$F = \left\{ x \in A;\ \overline{\lim_{r \to 0}} \frac{\lambda(B(x,r))}{m_n(B(x,r))} > \delta \right\}$$

To this end let $\varepsilon > 0$ and use Theorem 3.1.3 to get an open $U \supseteq A$ such that $\lambda(U) < \varepsilon$. For each $x \in F$ there is an open ball $B_x \subseteq U$ such that

$$\lambda(B_x) > \delta v_n(B_x).$$

If $V = \cup_{x \in F} B_x$ and $c < v_n(V)$ we use Lemma 5.3.1 to obtain $x_1, ..., x_k$ such that $B_{x_1}, ..., B_{x_k}$ are pairwise disjoint and

$$c < 3^n \Sigma_{i=1}^k v_n(B_{x_i}) < 3^n \delta^{-1} \Sigma_{i=1}^k \lambda(B_{x_i})$$

$$\leq 3^n \delta^{-1} \lambda(U) < 3^n \delta^{-1} \varepsilon.$$

Thus $v_n(V) \leq 3^n \delta^{-1} \varepsilon$. Since $V \supseteq F \in \mathcal{R}_n$ and $\varepsilon > 0$ is arbitrary, $v_n(F) = 0$ and the theorem is proved.

Corollary 5.3.1. *Suppose $F : \mathbf{R} \to \mathbf{R}$ is an increasing function. Then $F'(x)$ exists for almost all x with respect to linear measure.*

PROOF. Let D be the set of all points of discontinuity of F. Suppose $-\infty < a < b < \infty$ and $\varepsilon > 0$. If $a < x_1 < ... < x_n < b$, where $x_1, ..., x_n \in D$ and

$$F(x_k+) - F(x_k-) \geq \varepsilon,\ \ k = 1, ..., n$$

then

$$n\varepsilon \leq \Sigma_{k=1}^n (F(x_k+) - F(x_k-)) \leq F(b) - F(a).$$

Thus $D \cap [a, b]$ is at most denumerable and it follows that D is at most denumerable. Set $H(x) = F(x+) - F(x)$, $x \in \mathbf{R}$, and let $(x_j)_{j=0}^N$ be an enumeration of the members of the set $\{H > 0\}$. Moreover, for any $a > 0$,

$$\sum_{|x_j| < a} H(x_j) \leq \sum_{|x_j| < a} (F(x_j+) - F(x_j-))$$

$$\leq F(a) - F(-a) < \infty.$$

Now, if we introduce

$$\nu(A) = \Sigma_1^N H(x_j)\delta_{x_j}(A),\ A \in \mathcal{R}$$

then ν is a positive measure such that $\nu(K) < \infty$ for each compact subset K of $\mathbf{R}$. Furthermore, if h is a non-zero real number,

$$\mid \frac{1}{h}(H(x+h) - H(x) \mid \leq \frac{1}{h}(H(x+h) + H(x)) \leq 4\frac{1}{4 \mid h \mid}\nu(B(x, 2 \mid h \mid)$$

and Theorem 5.3.4 implies that $H'(x) = 0$ a.e. $[v_1]$. Therefore, without loss of generality it may be assumed that F is right continuous and, in addition, there is no restriction to assume that $F(+\infty) - F(-\infty) < \infty$.

By Section 1.6 F induces a finite positive Borel measure μ such that

$$\mu(]x, y]) = F(y) - F(x) \text{ if } x < y.$$

Now consider the Lebesgue decomposition

$$d\mu = f dv_1 + d\lambda$$

where $f \in L^1(v_1)$ and $\lambda \perp v_1$. If $x < y$,

$$F(y) - F(x) = \int_x^y f(t)dt + \lambda(]x, y])$$

and the previous two theorems imply that

$$\lim_{y \downarrow x} \frac{F(y) - F(x)}{y - x} = f(x) \text{ a.e. } [v_1]$$

If $y < x$,

$$F(x) - F(y) = \int_y^x f(t)dt + \lambda(]y, x])$$

and we get

$$\lim_{y \uparrow x} \frac{F(y) - F(x)}{y - x} = f(x) \text{ a.e. } [v_1]\,.$$

The theorem is proved.

Exercises

1. Suppose $F : \mathbf{R} \to \mathbf{R}$ is increasing and let $f \in L^1_{loc}(v_1)$ be such that $F'(x) = f(x)$ a.e. $[v_1]$. Prove that

$$\int_x^y f(t)dt \leq F(y) - F(x) \text{ if } -\infty < x \leq y < \infty.$$

5.4. Absolutely Continuous Functions and Functions of Bounded Variation

Throughout this section a and b are reals with $a < b$ and to simplify notation we set $m_{a,b} = m_{|[a,b]}$. If $f \in L^1(m_{a,b})$ we know from the previous section that the function

$$(If)(x) =_{def} \int_a^x f(t)dt, \ \ a \leq x \leq b$$

has the derivative $f(x)$ a.e. $[m_{a,b}]$, that is

$$\frac{d}{dx} \int_a^x f(t)dt = f(x) \text{ a.e. } [m_{a,b}].$$

Our next main task will be to describe the range of the linear map I.

A function $F : [a, b] \to \mathbf{R}$ is said to be absolutely continuous if to every $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\Sigma_{i=1}^n \mid b_i - a_i \mid < \delta \text{ implies } \Sigma_{i=1}^n \mid F(b_i) - F(a_i) \mid < \varepsilon$$

whenever $]a_1, b_1[, ...,]a_n, b_n[$ are disjoint open subintervals of $[a, b]$. It is obvious that an absolutely continuous function is continuous. It can be proved that the Cantor function is not absolutely continuous.

Theorem 5.4.1. *If $f \in L^1(m_{a,b})$, then If is absolutely continuous.*

PROOF. There is no restriction to assume $f \geq 0$. Set

$$d\lambda = f dm_{a,b}.$$

By Theorem 5.2.2, to every $\varepsilon > 0$ there exists a $\delta > 0$ such that $\lambda(A) < \varepsilon$ for each Lebesgue set A in $[a, b]$ such that $m_{a,b}(A) < \delta$. Now restricting A to be a finite disjoint union of open intervals, the theorem follows.

Suppose $-\infty \leq \alpha < \beta \leq \infty$ and $F :]\alpha, \beta[\rightarrow \mathbf{R}$. For every $x \in]\alpha, \beta[$ we define

$$T_F(x) = \sup \Sigma_{i=1}^n \mid F(x_i) - F(x_{i-1}) \mid$$

where the supremum is taken over all positive integers n and all choices $(x_i)_{i=0}^n$ such that

$$\alpha < x_0 < x_1 < ... < x_n = x.$$

The function $T_F :]\alpha, \beta[\rightarrow [0, \infty]$ is called the total variation of F. Note that T_F is increasing. If T_F is a bounded function, F is said to be of bounded variation. A bounded increasing function on $\mathbf{R}$ is of bounded variation. Therefore the difference of two bounded increasing functions on $\mathbf{R}$ is of bounded variation. Interestingly enough, the converse is true. In the special case $]\alpha, \beta[= \mathbf{R}$ we write $F \in BV$ if F is of bounded variation.

Theorem 5.4.2. *Suppose $F \in BV$.*

(a) *The functions $T_F + F$ and $T_F - F$ are increasing and*

$$F = \frac{1}{2}(T_F + F) - \frac{1}{2}(T_F - F).$$

In particular, F is differentiable almost everywhere with respect to linear measure.

(b) *If F is right continuous, then so is T_F.*

PROOF. (a) Let $x < y$ and $\varepsilon > 0$. Choose $x_0 < x_1 < ... < x_n = x$ such that

$$\Sigma_{i=1}^n \mid F(x_i) - F(x_{i-1}) \mid \geq T_f(x) - \varepsilon.$$

Then

$$T_F(y) + F(y)$$

$$\geq \Sigma_{i=1}^{n} \mid F(x_i) - F(x_{i-1}) \mid + \mid F(y) - F(x) \mid + (F(y) - F(x)) + F(x)$$
$$\geq T_F(x) - \varepsilon + F(x)$$

and, since $\varepsilon > 0$ is arbitrary, $T_F(y) + F(y) \geq T_F(x) + F(x)$. Hence $T_F + F$ is increasing. Finally, replacing F by $-F$ it follows that the function $T_F - F$ is increasing.

(b) If $c \in \mathbf{R}$ and $x > c$,

$$T_f(x) = T_F(c) + \sup \Sigma_{i=1}^{n} \mid F(x_i) - F(x_{i-1}) \mid$$

where the supremum is taken over all positive integers n and all choices $(x_i)_{i=0}^{n}$ such that

$$c = x_0 < x_1 < ... < x_n = x.$$

Suppose $T_F(c+) > T_F(c)$ where $c \in \mathbf{R}$. Then there is an $\varepsilon > 0$ such that

$$T_F(x) - T_F(c) > \varepsilon$$

for all $x > c$. Now, since F is right continuous at the point c, for fixed $x > c$ there exists a partition

$$c < x_{11} < ... < x_{1n_1} = x$$

such that

$$\Sigma_{i=2}^{n_1} \mid F(x_{1i}) - F(x_{1i-1}) \mid > \varepsilon.$$

But

$$T_F(x_{11}) - T_F(c) > \varepsilon$$

and we get a partition

$$c < x_{21} < ... < x_{2n_2} = x_{11}$$

such that

$$\Sigma_{i=2}^{n_2} \mid F(x_{2i}) - F(x_{2i-1}) \mid > \varepsilon.$$

Summing up we have got a partition of the interval $[x_{21}, x]$ with

$$\Sigma_{i=2}^{n_2} \mid F(x_{2i}) - F(x_{2i-1}) \mid + \Sigma_{i=2}^{n_1} \mid F(x_{1i}) - F(x_{1i-1}) \mid > 2\varepsilon.$$

By repeating the process the total variation of F becomes infinite, which is a contradiction. The theorem is proved.

Theorem 5.4.3. *Suppose* $F : [a,b] \to \mathbf{R}$ *is absolutely continuous. Then there exists a unique* $f \in L^1(m_{a,b})$ *such that*

$$F(x) = F(a) + \int_a^x f(t)dt, \ a \le x \le b.$$

In particular, the range of the map I *equals the set of all real-valued absolutely continuous maps on* $[a,b]$.

PROOF. Set $F(x) = F(a)$ if $x \le a$ and $F(x) = F(b)$ if $x \ge b$. There exists a $\delta > 0$ such that

$$\Sigma_{i=1}^n \mid b_i - a_i \mid < \delta \text{ implies } \Sigma_{i=1}^n \mid F(b_i) - F(a_i) \mid < 1$$

whenever $]a_1, b_1[, ...,]a_n, b_n[$ are disjoint subintervals of $[a,b]$. Let p be the least positive integer such that $a + p\delta \ge b$. Then $T_F \le p$ and $F \in BV$. Let $F = G - H$, where $G = \frac{1}{2}(T_F + F)$ and $H = \frac{1}{2}(T_F - F)$. There exist finite positive Borel measures λ_G and λ_H such that

$$\lambda_G(]x,y]) = G(y) - G(x), \ x \le y$$

and

$$\lambda_H(]x,y]) = H(y) - H(x), \ x \le y.$$

If we define $\lambda = \lambda_G - \lambda_H$,

$$\lambda(]x,y]) = F(y) - F(x), \ x \le y.$$

Clearly,

$$\lambda(]x,y[) = F(y) - F(x), \ x \le y$$

since F is continuous.

Our next task will be to prove that $\lambda << v_1$. To this end, suppose $A \in \mathcal{R}$ and $v_1(A) = 0$. Now choose $\varepsilon > 0$ and let $\delta > 0$ be as in the definition of the absolute continuity of F on $[a,b]$. For each $k \in \mathbf{N}_+$ there exists an open set

$V_k \supseteq A$ such that $v_1(V_k) < \delta$ and $\lim_{k\to\infty} \lambda(V_k) = \lambda(A)$. But each fixed V_k is a disjoint union of open intervals $(]a_i, b_i[)_{i=1}^{\infty}$ and hence

$$\Sigma_{i=1}^{n} \mid b_i - a_i \mid < \delta$$

for every n and, accordingly from this,

$$\Sigma_{i=1}^{\infty} \mid F(b_i) - F(a_i) \mid \leq \varepsilon$$

and

$$\mid \lambda(V_k) \mid \leq \Sigma_{i=1}^{\infty} \mid \lambda(]a_i, b_i[) \mid \leq \varepsilon.$$

Thus $\mid \lambda(A) \mid \leq \varepsilon$ and since $\varepsilon > 0$ is arbitrary, $\lambda(A) = 0$. From this $\lambda << v_1$ and the theorem follows at once.

Suppose $(X, \mathcal{M}, \mu)$ is a positive measure space. From now on we write $f \in L^1(\mu)$ if there exist a $g \in \mathcal{L}^1(\mu)$ and an $A \in \mathcal{M}$ such that $A^c \in \mathcal{Z}_\mu$ and $f(x) = g(x)$ for all $x \in A$. Furthermore, we define

$$\int_X f d\mu = \int_X g d\mu$$

(cf the discussion in Section 2). Note that $f(x)$ need not be defined for every $x \in X$.

Corollary 5.4.1. *A function $f : [a, b] \to \mathbf{R}$ is absolutely continuous if and only if the following conditions are true:*
(i) $f'(x)$ exists for $m_{a,b}$-almost all $x \in [a, b]$
(ii) $f' \in L^1(m_{a,b})$
(iii) $f(x) = f(a) + \int_a^x f'(t)dt$, all $x \in [a, b]$.

Exercises

1. Suppose $f : [0, 1] \to \mathbf{R}$ satisfies $f(0) = 0$ and

$$f(x) = x^2 \sin \frac{1}{x^2} \text{ if } 0 < x \leq 1.$$

Prove that f is differentiable everywhere but f is not absolutely continuous.

2. Suppose α is a positive real number and f a function on $[0,1]$ such that $f(0) = 0$ and $f(x) = x^{\alpha} \sin \frac{1}{x}$, $0 < x \leq 1$. Prove that f is absolutely continuous if and only if $\alpha > 1$.

3. Suppose $f(x) = x \cos(\pi/x)$ if $0 < x < 2$ and $f(x) = 0$ if $x \in \mathbf{R} \backslash \,]0,2[$. Prove that f is not of bounded variation on $\mathbf{R}$.

4 A function $f : [a,b] \to \mathbf{R}$ is a Lipschitz function, that is there exists a positive real number C such that

$$\mid f(x) - f(y) \mid \leq C \mid x - y \mid$$

for all $x, y \in [a,b]$. Show that f is absolutely continuous and $\mid f'(x) \mid \leq C$ a.e. $[m_{a,b}]$.

5. Suppose $f : [a,b] \to \mathbf{R}$ is absolutely continuous. Prove that

$$T_g(x) = \int_a^x \mid f'(t) \mid dt, \; a < x < b$$

if g is the restriction of f to the open interval $]a,b[$.

6. Suppose f and g are real-valued absolutely continuous functions on the compact interval $[a,b]$. Show that the function $h = \max(f,g)$ is absolutely continuous and $h' \leq \max(f', g')$ a.e. $[m_{a,b}]$.

$\downarrow\downarrow\downarrow$

5.5. Conditional Expectation

Let $(\Omega, \mathcal{F}, P)$ be a probability space and suppose $\xi \in L^1(P)$. Moreover, suppose $\mathcal{G} \subseteq \mathcal{F}$ is a σ-algebra and set

$$\mu(A) = P[A], \ A \in \mathcal{G}$$

and

$$\lambda(A) = \int_A \xi dP, \ A \in \mathcal{G}.$$

It is trivial that $\mathcal{Z}_\mu = \mathcal{Z}_P \cap \mathcal{G} \subseteq \mathcal{Z}_\lambda$ and the Radon-Nikodym Theorem shows there exists a unique $\eta \in L^1(\mu)$ such that

$$\lambda(A) = \int_A \eta d\mu \text{ all } A \in \mathcal{G}$$

or, what amounts to the same thing,

$$\int_A \xi dP = \int_A \eta dP \text{ all } A \in \mathcal{G}.$$

Note that η is $(\mathcal{G}, \mathcal{R})$-measurable. The random variable η is called the conditional expectation of ξ given $\mathcal{G}$ and it is standard to write $\eta = E[\xi \mid \mathcal{G}]$.

A sequence of σ-algebras $(\mathcal{F}_n)_{n=1}^\infty$ is called a filtration if

$$\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subseteq \mathcal{F}.$$

If $(\mathcal{F}_n)_{n=1}^\infty$ is a filtration and $(\xi_n)_{n=1}^\infty$ is a sequence of real valued random variables such that for each n,

(a) $\xi_n \in L^1(P)$
(b) ξ_n is $(\mathcal{F}_n, \mathcal{R})$-measurable
(c) $E\left[\xi_{n+1} \mid \mathcal{F}_n\right] = \xi_n$

then $(\xi_n, \mathcal{F}_n)_{n=1}^\infty$ is called a martingale. There are very nice connections between martingales and the theory of differentiation (see e.g Billingsley $[B]$ and Malliavin $[M]$).

↑↑↑

CHAPTER 6

COMPLEX INTEGRATION

Introduction

In this section, in order to illustrate the power of Lebesgue integration, we collect a few results, which often appear with uncomplete proofs at the undergraduate level.

6.1. Complex Integrand

So far we have only treated integration of functions with their values in $\mathbf{R}$ or $[0, \infty]$ and it is the purpose of this section to discuss integration of complex valued functions.

Suppose $(X, \mathcal{M}, \mu)$ is a positive measure. Let $f, g \in L^1(\mu)$. We define

$$\int_X (f + ig) d\mu = \int_X f d\mu + i \int_X g d\mu.$$

If α and β are real numbers,

$$\int_X (\alpha + i\beta)(f + ig) d\mu = \int_X ((\alpha f - \beta g) + i(\alpha g + \beta f)) d\mu$$

$$= \int_X (\alpha f - \beta g) d\mu + i \int_X (\alpha g + \beta f) d\mu$$

$$= \alpha \int_X f d\mu - \beta \int_X g d\mu + i\alpha \int_X g d\mu + i\beta \int_X f d\mu$$

$$= (\alpha + i\beta)\left(\int_X f d\mu + i \int_X g d\mu\right)$$

$$= (\alpha + i\beta) \int_X (f + ig) d\mu.$$

We write $f \in L^1(\mu;\mathbf{C})$ if Re f, Im $f \in L^1(\mu)$ and have, for every $f \in L^1(\mu;\mathbf{C})$ and complex α,

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

Clearly, if $f, g \in L^1(\mu;\mathbf{C})$, then

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

Now suppose μ is a complex measure on $\mathcal{M}$. If

$$f \in L^1(\mu;\mathbf{C}) =_{def} L^1(\mu_{\mathrm{Re}};\mathbf{C}) \cap L^1(\mu_{\mathrm{Im}};\mathbf{C})$$

we define

$$\int_X f d\mu = \int_X f d\mu_{\mathrm{Re}} + i \int_X f d\mu_{\mathrm{Im}}.$$

It follows for every $f, g \in L^1(\mu;\mathbf{C})$ and $\alpha \in \mathbf{C}$ that

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

and

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

$\downarrow\downarrow\downarrow$

6.2. The Fourier Transform

Below, if $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n) \in \mathbf{R}^n$, we let

$$\langle x, y \rangle = \Sigma_{k=1}^n x_k y_k.$$

and

$$\mid x \mid = \sqrt{\langle x, y \rangle}.$$

If μ is a complex measure on $\mathcal{R}_n$ (or $\mathcal{R}_n^-$) the Fourier transform $\hat{\mu}$ of μ is defined by

$$\hat{\mu}(y) = \int_{\mathbf{R}^n} e^{-i\langle x,y \rangle} d\mu(x), \ y \in \mathbf{R}^n.$$

Note that

$$\hat{\mu}(0) = \mu(\mathbf{R}^n).$$

The Fourier transform of a function $f \in L^1(m_n; \mathbf{C})$ is defined by

$$\hat{f}(y) = \hat{\mu}(y) \text{ where } d\mu = f dm_n.$$

Theorem 6.2.1. *The canonical Gaussian measure γ_n in $\mathbf{R}^n$ has the Fourier transform*

$$\hat{\gamma}_n(y) = e^{-\frac{|y|^2}{2}}.$$

PROOF. Since

$$\gamma_n = \gamma_1 \otimes ... \otimes \gamma_1 \ (n \text{ factors})$$

it is enough to consider the special case $n = 1$. Set

$$g(y) = \hat{\gamma}_1(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} \cos xy dx.$$

Note that $g(0) = 1$. Since

$$\mid \frac{\cos x(y+h) - \cos xy}{h} \mid \leq \mid x \mid$$

the Lebesgue Dominated Convergence Theorem yields

$$g'(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} -xe^{-\frac{x^2}{2}} \sin xy dx$$

(Exercise: Prove this by using Example 2.2.1). Now, by partial integration,

$$g'(y) = \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{x^2}{2}} \sin xy \right]_{x=-\infty}^{x=\infty} - \frac{y}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} \cos xy dx$$

that is

$$g'(y) + yg(y) = 0$$

and we get

$$g(y) = e^{-\frac{y^2}{2}}.$$

If $\xi = (\xi_1, ..., \xi_n)$ is an $\mathbf{R}^n$-valued random variable with $\xi_k \in L^1(P)$, $k = 1, ..., n$, the characteristic function c_ξ of ξ is defined by

$$c_\xi(y) = E\left[e^{i\langle \xi, y\rangle}\right] = \hat{P}_\xi(-y),\ y \in \mathbf{R}^n.$$

For example, if $\xi \in N(0, \sigma)$, then $\xi = \sigma G$, where $G \in N(0, 1)$, and we get

$$c_\xi(y) = E\left[e^{i\langle G, \sigma y\rangle}\right] = \hat{\gamma}_1(-\sigma y)$$

$$= e^{-\frac{\sigma^2 y^2}{2}}.$$

Choosing $y = 1$ results in

$$E\left[e^{i\xi}\right] = e^{-\frac{1}{2}E\left[\xi^2\right]} \text{ if } \xi \in N(0, \sigma).$$

Thus if $(\xi_k)_{k=1}^n$ is a centred real-valued Gaussian process

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \exp(-\frac{1}{2}E\left[(\Sigma_{k=1}^n y_k \xi_k)^2\right]$$

$$= \exp(-\frac{1}{2}\Sigma_{k=1}^n y_k^2 E\left[\xi_k^2\right] - \Sigma_{1\leq j<k\leq n} y_j y_k E\left[\xi_j \xi_k\right]).$$

In particular, if

$$E\left[\xi_j \xi_k\right] = 0,\ j \neq k$$

we see that

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \Pi_{k=1}^n e^{-\frac{y_k^2}{2}E\left[\xi_k^2\right]}$$

or

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \Pi_{k=1}^n E\left[e^{i y_k \xi_k}\right].$$

Stated otherwise, the Fourier tranforms of the measures $P_{(\xi_1,...,\xi_n)}$ and $\times_{k=1}^n P_{\xi_k}$ agree. Below we will show that complex measures in $\mathbf{R}^n$ with the same Fourier transforms are equal and we get the following

Theorem 6.2.2. *Let $(\xi_k)_{k=1}^n$ be a centred real-valued Gaussian process with uncorrelated components, that is*

$$E\left[\xi_j \xi_k\right] = 0,\ j \neq k.$$

Then the random variables $\xi_1, ..., \xi_n$ are independent.

6.3 Fourier Inversion

Theorem 6.3.1. *Suppose $f \in L^1(m_n)$. If $\hat{f} \in L^1(m_n)$ and f is bounded and continuous*

$$f(x) = \int_{\mathbf{R}^d} e^{i\langle y,x\rangle} \hat{f}(y) \frac{dy}{(2\pi)^n},\ x \in \mathbf{R}^n.$$

PROOF. Choose $\varepsilon > 0$. We have

$$\int_{\mathbf{R}^n} e^{i\langle y,x\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \hat{f}(y) \frac{dy}{(2\pi)^n} = \int_{\mathbf{R}^n} f(u) \left\{ \int_{\mathbf{R}^n} e^{i\langle y,x-u\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \frac{dy}{(2\pi)^n} \right\} du$$

where the right side equals

$$\int_{\mathbf{R}^n} f(u) \left\{ \int_{\mathbf{R}^n} e^{i\langle v,\frac{x-u}{\varepsilon}\rangle} e^{-\frac{1}{2}|v|^2} \frac{dv}{\sqrt{2\pi}^n} \right\} \frac{du}{\sqrt{2\pi}^n \varepsilon^n} = \int_{\mathbf{R}^n} f(u) e^{-\frac{1}{2\varepsilon^2}|u-x|^2} \frac{du}{\sqrt{2\pi}^n \varepsilon^n}$$

$$= \int_{\mathbf{R}^n} f(x+\varepsilon z) e^{-\frac{1}{2}|z|^2} \frac{dz}{\sqrt{2\pi}^n}.$$

Thus

$$\int_{\mathbf{R}^n} e^{i\langle y,x\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \hat{f}(y) \frac{dy}{(2\pi)^n} = \int_{\mathbf{R}^n} f(x+\varepsilon z) e^{-\frac{1}{2}|z|^2} \frac{dz}{\sqrt{2\pi}^n}.$$

By letting $\varepsilon \to 0$ and using the Lebesgue Dominated Convergence Theorem, Theorem 6.3.1 follows at once.

Recall that $C_c^\infty(\mathbf{R}^n)$ denotes the class of all functions $f : \mathbf{R}^n \to \mathbf{R}$ with compact support which are infinitely many times differentiable. If $f \in C_c^\infty(\mathbf{R}^n)$ then $\hat{f} \in L^1(m_n)$. To see this, suppose $y_k \neq 0$ and use partial integration to obtain

$$\hat{f}(y) = \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f(x)dx = \frac{1}{iy_k} \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f'_{x_k}(x)dx$$

and

$$\hat{f}(y) = \frac{1}{(iy_k)^l} \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f^{(l)}_{x_k}(x)dx,\ l \in \mathbf{N}.$$

Thus

$$| y_k |^l | \hat{f}(y) | \leq \int_{\mathbf{R}^n} | f^{(l)}_{x_k}(x) | \, dx, \; l \in \mathbf{N}$$

and we conclude that

$$\sup_{y \in \mathbf{R}^n} (1+ | y |)^{n+1} | \hat{f}(y) | < \infty.$$

and, hence, $\hat{f} \in L^1(m_n)$.

Corollary 6.3.1. *If* $f \in C_c^\infty(\mathbf{R}^n)$, *then* $\hat{f} \in L^1(m_n)$ *and*

$$f(x) = \int_{\mathbf{R}^n} e^{i\langle y,x\rangle} \hat{f}(y) \frac{dy}{(2\pi)^n}, \; x \in \mathbf{R}^n.$$

Corollary 6.3.2 *If* μ *is a complex Borel measure in* $\mathbf{R}^n$ *and* $\hat{\mu} = 0$, *then* $\mu = 0$.

PROOF. Choose $f \in C_c^\infty(\mathbf{R}^n)$. We multiply the equation $\hat{\mu}(-y) = 0$ by $\frac{\hat{f}(y)}{(2\pi)^n}$ and integrate over $\mathbf{R}^n$ with respect to Lebesgue measure to obtain

$$\int_{\mathbf{R}^n} f(x) d\mu(x) = 0.$$

Since $f \in C_c^\infty(\mathbf{R}^n)$ is arbitrary it follows that $\mu = 0$. The theorem is proved.

6.4. Non-Differentiability of Brownian Paths

Let ND denote the set of all real-valued continuous function defined on the unit interval which are not differentiable at any point. It is well known that ND is non-empty. In fact, if ν is Wiener measure on $C\,[0,1]$, $x \in ND$ a.e. $[\nu]$. The purpose of this section is to prove this important property of Brownian motion.

Let $W = (W(t))_{0\leq t\leq 1}$ be a real-valued Brownian motion in the time interval $[0,1]$ such that every path $t \to W(t)$, $0 \leq t \leq 1$ is continuous. Recall that

$$E\left[W(t)\right] = 0$$

and

$$E\left[W(s)W(t)\right] = \min(s,t).$$

If

$$0 \leq t_0 \leq ... \leq t_n \leq 1$$

and $1 \leq j < k \leq n$

$$E\left[(W(t_k) - W(t_{k-1}))(W(t_j) - W(t_{j-1})\right]$$

$$= E\left[(W(t_k)W(t_j)\right] - E\left[W(t_k)W(t_{j-1})\right] - E\left[W(t_{k-1})W(t_j)\right] + E\left[W(t_{k-1})W(t_{j-1})\right]$$

$$= t_j - t_{j-1} - t_j + t_{j-1} = 0.$$

From the previous section we now infer that the random variables

$$W(t_1) - W(t_0), ..., W(t_n) - W(t_{n-1})$$

are independent.

Theorem 7. *The function* $t \to W(t)$, $0 \leq t \leq 1$ *is not differentiable at any point* $t \in [0,1]$ a.s. $[P]$.

PROOF. Without loss of generality we assume the underlying probability space is complete. Let $c, \varepsilon > 0$ and denote by $B(c,\varepsilon)$ the set of all $\omega \in \Omega$ such that

$$\mid W(t) - W(s) \mid < c \mid t - s \mid \text{ if } t \in [s-\varepsilon, s+\varepsilon] \cap [0,1]$$

for some $s \in [0,1]$. It is enough to prove that the set

$$\bigcup_{j=1}^{\infty}\bigcup_{k=1}^{\infty} B(j, \frac{1}{k}).$$

is of probability zero. From now on let $c, \varepsilon > 0$ be fixed. It is enough to prove $P\left[B(c,\varepsilon)\right] = 0$.

Set

$$X_{n,k} = \max_{k \le j < k+3} \mid W(\frac{j+1}{n}) - W(\frac{j}{n}) \mid$$

for each integer $n > 3$ and $k \in \{0, ..., n-3\}$.

Let $n > 3$ be so large that

$$\frac{3}{n} \le \varepsilon.$$

We claim that

$$B(c, \varepsilon) \subseteq \left[\min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right].$$

If $\omega \in B(c, \varepsilon)$ there exists an $s \in [0, 1]$ such that

$$\mid W(t) - W(s) \mid \le c \mid t - s \mid \text{ if } t \in [s - \varepsilon, s + \varepsilon] \cap [0, 1].$$

Now choose $k \in \{0, ..., n-3\}$ such that

$$s \in \left[\frac{k}{n}, \frac{k}{n} + \frac{3}{n} \right].$$

If $k \le j < k + 3$,

$$\mid W(\frac{j+1}{n}) - W(\frac{j}{n}) \mid \le \mid W(\frac{j+1}{n}) - W(s) \mid + \mid W(s) - W(\frac{j}{n}) \mid$$

$$\le \frac{6c}{n}$$

and, hence, $X_{n,k} \le \frac{6c}{n}$. Now

$$B(c, \varepsilon) \subseteq \left[\min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right]$$

and it is enough to prove that

$$\lim_{n \to \infty} P \left[\min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right] = 0.$$

But

$$P \left[\min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right] \le \sum_{k=0}^{n-3} P \left[X_{n,k} \le \frac{6c}{n} \right]$$

$$= (n-2)P\left[X_{n,0} \leq \frac{6c}{n}\right] \leq nP\left[X_{n,0} \leq \frac{6c}{n}\right]$$

$$= n(P\left[\mid W(\frac{1}{n}) \mid \leq \frac{6c}{n}\right])^3 = n(P(\mid W(1) \mid \leq \frac{6c}{\sqrt{n}})^3$$

$$\leq n(\frac{12c}{\sqrt{2\pi n}})^3.$$

where the right side converges to zero as $n \to \infty$. The theorem is proved.

Recall that a function of bounded variation possesses a derivative a.e. with respect to Lebesgue measure. Therefore, with probability one, a Brownian path is not of bounded variation. In view of this an integral of the type

$$\int_0^1 f(t)dW(t)$$

cannot be interpreted as an ordinary Stieltjes integral. Nevertheless, such an integral can be defined by completely different means and is basic in, for example, financial mathematics.

↑↑↑

CHAPTER 6

COMPLEX INTEGRATION

Introduction

In this section, in order to illustrate the power of Lebesgue integration, we collect a few results, which often appear with uncomplete proofs at the undergraduate level.

6.1. Complex Integrand

So far we have only treated integration of functions with their values in **R** or $[0, \infty]$ and it is the purpose of this section to discuss integration of complex valued functions.

Suppose $(X, \mathcal{M}, \mu)$ is a positive measure. Let $f, g \in L^1(\mu)$. We define

$$\int_X (f + ig) d\mu = \int_X f d\mu + i \int_X g d\mu.$$

If α and β are real numbers,

$$\int_X (\alpha + i\beta)(f + ig) d\mu = \int_X ((\alpha f - \beta g) + i(\alpha g + \beta f)) d\mu$$

$$= \int_X (\alpha f - \beta g) d\mu + i \int_X (\alpha g + \beta f) d\mu$$

$$= \alpha \int_X f d\mu - \beta \int_X g d\mu + i\alpha \int_X g d\mu + i\beta \int_X f d\mu$$

$$= (\alpha + i\beta)(\int_X f d\mu + i \int_X g d\mu)$$

$$= (\alpha + i\beta) \int_X (f + ig) d\mu.$$

We write $f \in L^1(\mu;\mathbf{C})$ if $\mathrm{Re}\, f$, $\mathrm{Im}\, f \in L^1(\mu)$ and have, for every $f \in L^1(\mu;\mathbf{C})$ and complex α,

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

Clearly, if $f, g \in L^1(\mu;\mathbf{C})$, then

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

Now suppose μ is a complex measure on $\mathcal{M}$. If

$$f \in L^1(\mu;\mathbf{C}) =_{def} L^1(\mu_{\mathrm{Re}};\mathbf{C}) \cap L^1(\mu_{\mathrm{Im}};\mathbf{C})$$

we define

$$\int_X f d\mu = \int_X f d\mu_{\mathrm{Re}} + i \int_X f d\mu_{\mathrm{Im}}.$$

It follows for every $f, g \in L^1(\mu;\mathbf{C})$ and $\alpha \in \mathbf{C}$ that

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

and

$$\int_X (f+g) d\mu = \int_X f d\mu + \int_X g d\mu.$$

$\downarrow\downarrow\downarrow$

6.2. The Fourier Transform

Below, if $x = (x_1, ..., x_n)$ and $y = (y_1, ..., y_n) \in \mathbf{R}^n$, we let

$$\langle x, y \rangle = \Sigma_{k=1}^n x_k y_k.$$

and

$$\mid x \mid = \sqrt{\langle x, y \rangle}.$$

If μ is a complex measure on $\mathcal{R}_n$ (or $\mathcal{R}_n^-$) the Fourier transform $\hat{\mu}$ of μ is defined by

$$\hat{\mu}(y) = \int_{\mathbf{R}^n} e^{-i\langle x,y \rangle} d\mu(x), \ y \in \mathbf{R}^n.$$

Note that

$$\hat{\mu}(0) = \mu(\mathbf{R}^n).$$

The Fourier transform of a function $f \in L^1(m_n; \mathbf{C})$ is defined by

$$\hat{f}(y) = \hat{\mu}(y) \text{ where } d\mu = f dm_n.$$

Theorem 6.2.1. *The canonical Gaussian measure γ_n in $\mathbf{R}^n$ has the Fourier transform*

$$\hat{\gamma}_n(y) = e^{-\frac{|y|^2}{2}}.$$

PROOF. Since

$$\gamma_n = \gamma_1 \otimes ... \otimes \gamma_1 \ (n \text{ factors})$$

it is enough to consider the special case $n = 1$. Set

$$g(y) = \hat{\gamma}_1(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} \cos xy dx.$$

Note that $g(0) = 1$. Since

$$\mid \frac{\cos x(y+h) - \cos xy}{h} \mid \leq \mid x \mid$$

the Lebesgue Dominated Convergence Theorem yields

$$g'(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbf{R}} -xe^{-\frac{x^2}{2}} \sin xy dx$$

(Exercise: Prove this by using Example 2.2.1). Now, by partial integration,

$$g'(y) = \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{x^2}{2}} \sin xy \right]_{x=-\infty}^{x=\infty} - \frac{y}{\sqrt{2\pi}} \int_{\mathbf{R}} e^{-\frac{x^2}{2}} \cos xy dx$$

that is

$$g'(y) + yg(y) = 0$$

and we get

$$g(y) = e^{-\frac{y^2}{2}}.$$

If $\xi = (\xi_1, ..., \xi_n)$ is an $\mathbf{R}^n$-valued random variable with $\xi_k \in L^1(P)$, $k = 1, ..., n$, the characteristic function c_ξ of ξ is defined by

$$c_\xi(y) = E\left[e^{i\langle \xi, y\rangle}\right] = \hat{P}_\xi(-y),\ y \in \mathbf{R}^n.$$

For example, if $\xi \in N(0, \sigma)$, then $\xi = \sigma G$, where $G \in N(0, 1)$, and we get

$$c_\xi(y) = E\left[e^{i\langle G, \sigma y\rangle}\right] = \hat{\gamma}_1(-\sigma y)$$

$$= e^{-\frac{\sigma^2 y^2}{2}}.$$

Choosing $y = 1$ results in

$$E\left[e^{i\xi}\right] = e^{-\frac{1}{2}E\left[\xi^2\right]} \text{ if } \xi \in N(0, \sigma).$$

Thus if $(\xi_k)_{k=1}^n$ is a centred real-valued Gaussian process

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \exp(-\frac{1}{2}E\left[(\Sigma_{k=1}^n y_k \xi_k)^2\right]$$

$$= \exp(-\frac{1}{2}\Sigma_{k=1}^n y_k^2 E\left[\xi_k^2\right] - \Sigma_{1\leq j<k\leq n} y_j y_k E\left[\xi_j \xi_k\right]).$$

In particular, if

$$E\left[\xi_j \xi_k\right] = 0,\ j \neq k$$

we see that

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \Pi_{k=1}^n e^{-\frac{y_k^2}{2}E\left[\xi_k^2\right]}$$

or

$$E\left[e^{i\Sigma_{k=1}^n y_k \xi_k}\right] = \Pi_{k=1}^n E\left[e^{i y_k \xi_k}\right].$$

Stated otherwise, the Fourier tranforms of the measures $P_{(\xi_1,...,\xi_n)}$ and $\times_{k=1}^n P_{\xi_k}$ agree. Below we will show that complex measures in $\mathbf{R}^n$ with the same Fourier transforms are equal and we get the following

Theorem 6.2.2. *Let $(\xi_k)_{k=1}^n$ be a centred real-valued Gaussian process with uncorrelated components, that is*

$$E\left[\xi_j \xi_k\right] = 0,\ j \neq k.$$

Then the random variables $\xi_1, ..., \xi_n$ are independent.

6.3 Fourier Inversion

Theorem 6.3.1. *Suppose $f \in L^1(m_n)$. If $\hat{f} \in L^1(m_n)$ and f is bounded and continuous*

$$f(x) = \int_{\mathbf{R}^d} e^{i\langle y,x\rangle} \hat{f}(y) \frac{dy}{(2\pi)^n},\ x \in \mathbf{R}^n.$$

PROOF. Choose $\varepsilon > 0$. We have

$$\int_{\mathbf{R}^n} e^{i\langle y,x\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \hat{f}(y) \frac{dy}{(2\pi)^n} = \int_{\mathbf{R}^n} f(u) \left\{ \int_{\mathbf{R}^n} e^{i\langle y,x-u\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \frac{dy}{(2\pi)^n} \right\} du$$

where the right side equals

$$\int_{\mathbf{R}^n} f(u) \left\{ \int_{\mathbf{R}^n} e^{i\langle v,\frac{x-u}{\varepsilon}\rangle} e^{-\frac{1}{2}|v|^2} \frac{dv}{\sqrt{2\pi}^n} \right\} \frac{du}{\sqrt{2\pi}^n \varepsilon^n} = \int_{\mathbf{R}^n} f(u) e^{-\frac{1}{2\varepsilon^2}|u-x|^2} \frac{du}{\sqrt{2\pi}^n \varepsilon^n}$$

$$= \int_{\mathbf{R}^n} f(x+\varepsilon z) e^{-\frac{1}{2}|z|^2} \frac{dz}{\sqrt{2\pi}^n}.$$

Thus

$$\int_{\mathbf{R}^n} e^{i\langle y,x\rangle} e^{-\frac{\varepsilon^2}{2}|y|^2} \hat{f}(y) \frac{dy}{(2\pi)^n} = \int_{\mathbf{R}^n} f(x+\varepsilon z) e^{-\frac{1}{2}|z|^2} \frac{dz}{\sqrt{2\pi}^n}.$$

By letting $\varepsilon \to 0$ and using the Lebesgue Dominated Convergence Theorem, Theorem 6.3.1 follows at once.

Recall that $C_c^\infty(\mathbf{R}^n)$ denotes the class of all functions $f : \mathbf{R}^n \to \mathbf{R}$ with compact support which are infinitely many times differentiable. If $f \in C_c^\infty(\mathbf{R}^n)$ then $\hat{f} \in L^1(m_n)$. To see this, suppose $y_k \neq 0$ and use partial integration to obtain

$$\hat{f}(y) = \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f(x) dx = \frac{1}{iy_k} \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f'_{x_k}(x) dx$$

and

$$\hat{f}(y) = \frac{1}{(iy_k)^l} \int_{\mathbf{R}^d} e^{-i\langle x,y\rangle} f^{(l)}_{x_k}(x) dx,\ l \in \mathbf{N}.$$

Thus

$$| y_k |^l | \hat{f}(y) | \leq \int_{\mathbf{R}^n} | f^{(l)}_{x_k}(x) | \, dx, \, l \in \mathbf{N}$$

and we conclude that

$$\sup_{y \in \mathbf{R}^n} (1+ | y |)^{n+1} | \hat{f}(y) | < \infty.$$

and, hence, $\hat{f} \in L^1(m_n)$.

Corollary 6.3.1. *If* $f \in C_c^\infty(\mathbf{R}^n)$, *then* $\hat{f} \in L^1(m_n)$ *and*

$$f(x) = \int_{\mathbf{R}^n} e^{i\langle y,x \rangle} \hat{f}(y) \frac{dy}{(2\pi)^n}, \, x \in \mathbf{R}^n.$$

Corollary 6.3.2 *If* μ *is a complex Borel measure in* $\mathbf{R}^n$ *and* $\hat{\mu} = 0$, *then* $\mu = 0$.

PROOF. Choose $f \in C_c^\infty(\mathbf{R}^n)$. We multiply the equation $\hat{\mu}(-y) = 0$ by $\frac{\hat{f}(y)}{(2\pi)^n}$ and integrate over $\mathbf{R}^n$ with respect to Lebesgue measure to obtain

$$\int_{\mathbf{R}^n} f(x) d\mu(x) = 0.$$

Since $f \in C_c^\infty(\mathbf{R}^n)$ is arbitrary it follows that $\mu = 0$. The theorem is proved.

6.4. Non-Differentiability of Brownian Paths

Let ND denote the set of all real-valued continuous function defined on the unit interval which are not differentiable at any point. It is well known that ND is non-empty. In fact, if ν is Wiener measure on $C\,[0,1]$, $x \in ND$ a.e. $[\nu]$. The purpose of this section is to prove this important property of Brownian motion.

Let $W = (W(t))_{0 \leq t \leq 1}$ be a real-valued Brownian motion in the time interval $[0, 1]$ such that every path $t \to W(t)$, $0 \leq t \leq 1$ is continuous. Recall that

$$E\left[W(t)\right] = 0$$

and

$$E\left[W(s)W(t)\right] = \min(s, t).$$

If

$$0 \leq t_0 \leq ... \leq t_n \leq 1$$

and $1 \leq j < k \leq n$

$$E\left[(W(t_k) - W(t_{k-1}))(W(t_j) - W(t_{j-1})\right]$$

$$= E\left[(W(t_k)W(t_j)\right] - E\left[W(t_k)W(t_{j-1})\right] - E\left[W(t_{k-1})W(t_j)\right] + E\left[W(t_{k-1})W(t_{j-1})\right]$$

$$= t_j - t_{j-1} - t_j + t_{j-1} = 0.$$

From the previous section we now infer that the random variables

$$W(t_1) - W(t_0), ..., W(t_n) - W(t_{n-1})$$

are independent.

Theorem 7. *The function* $t \to W(t)$, $0 \leq t \leq 1$ *is not differentiable at any point* $t \in [0, 1]$ a.s. $[P]$.

PROOF. Without loss of generality we assume the underlying probability space is complete. Let $c, \varepsilon > 0$ and denote by $B(c, \varepsilon)$ the set of all $\omega \in \Omega$ such that

$$\mid W(t) - W(s) \mid < c \mid t - s \mid \text{ if } t \in [s - \varepsilon, s + \varepsilon] \cap [0, 1]$$

for some $s \in [0, 1]$. It is enough to prove that the set

$$\bigcup_{j=1}^{\infty} \bigcup_{k=1}^{\infty} B(j, \frac{1}{k}).$$

is of probability zero. From now on let $c, \varepsilon > 0$ be fixed. It is enough to prove $P\left[B(c, \varepsilon)\right] = 0$.

Set

$$X_{n,k} = \max_{k \le j < k+3} \mid W(\frac{j+1}{n}) - W(\frac{j}{n}) \mid$$

for each integer $n > 3$ and $k \in \{0, ..., n-3\}$.

Let $n > 3$ be so large that

$$\frac{3}{n} \le \varepsilon.$$

We claim that

$$B(c, \varepsilon) \subseteq \left[\min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right].$$

If $\omega \in B(c, \varepsilon)$ there exists an $s \in [0, 1]$ such that

$$\mid W(t) - W(s) \mid \le c \mid t - s \mid \text{ if } t \in [s - \varepsilon, s + \varepsilon] \cap [0, 1].$$

Now choose $k \in \{0, ..., n-3\}$ such that

$$s \in \left[\frac{k}{n}, \frac{k}{n} + \frac{3}{n} \right].$$

If $k \le j < k + 3$,

$$\mid W(\frac{j+1}{n}) - W(\frac{j}{n}) \mid \le \mid W(\frac{j+1}{n}) - W(s) \mid + \mid W(s) - W(\frac{j}{n}) \mid$$

$$\le \frac{6c}{n}$$

and, hence, $X_{n,k} \le \frac{6c}{n}$. Now

$$B(c, \varepsilon) \subseteq \left[\min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right]$$

and it is enough to prove that

$$\lim_{n \to \infty} P \left[\min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right] = 0.$$

But

$$P \left[\min_{0 \le k \le n-3} X_{n,k} \le \frac{6c}{n} \right] \le \sum_{k=0}^{n-3} P \left[X_{n,k} \le \frac{6c}{n} \right]$$

$$= (n-2)P\left[X_{n,0} \leq \frac{6c}{n}\right] \leq nP\left[X_{n,0} \leq \frac{6c}{n}\right]$$

$$= n(P\left[\mid W(\frac{1}{n}) \mid \leq \frac{6c}{n}\right])^3 = n(P(\mid W(1) \mid \leq \frac{6c}{\sqrt{n}})^3$$

$$\leq n(\frac{12c}{\sqrt{2\pi n}})^3.$$

where the right side converges to zero as $n \to \infty$. The theorem is proved.

Recall that a function of bounded variation possesses a derivative a.e. with respect to Lebesgue measure. Therefore, with probability one, a Brownian path is not of bounded variation. In view of this an integral of the type

$$\int_0^1 f(t)dW(t)$$

cannot be interpreted as an ordinary Stieltjes integral. Nevertheless, such an integral can be defined by completely different means and is basic in, for example, financial mathematics.

↑↑↑

REFERENCES

[B] Billingsley, P. (1995) Probability and Measure. Wiley&Sons.

[D] Dudley, R. M. (1989) Real Analysis and Probability. Wadsworth&Brooks.

[DI] Diedonné, J. (1960) Foundations of Modern Analysis. Academic Press.

[F] Folland, G. B. (1999) Real Analysis; Modern Techniques and Their Applications. Second Edition. Wiley&Sons.

[M] Malliavin, P. (1995) Integration and Probability. Springer.

[R] Rudin, W. (1966) Real and Complex Analysis. McGraw-Hill.

[S] Solovay R. M. (1970) A model of set theory in which every set of reals is Lebesgue measurable. Annals of Mathematics **92**, 1-56.