Measure and Integration

Xue-Mei Li
with assistance from Henri Elad Altman
Imperial College London

February 21, 2023

Contents

Chapter 1

Introduction

1.1 Module Information

Prerequisite Real Analysis (M2PM1), Metric Spaces and Topology (M2PM5), and basic Probability. Measure and Integration is a foundational course, underlies analysis modules. It brings together many concepts previously taught separately, for example integration and taking expectation, reconciling discrete random variables with continuous random variables.

Contents. Measurable sets, sigma-algebras, Measurable functions, Measures. Integration with respect to measures, L_p spaces, Modes of convergences, basic important convergence theorems (Dominated convergence theorem, Fatou's lemma etc), useful inequalities, properties of integrals.

This module is related to: Probability theory, Applied probability, Markov processes, Fourier analysis and theory of distributions, Functional analysis, ODE's, Ergodic theory, Probability theory, Stochastic Filtering, Applied Stochastic Processes.

1.2 Course Structure

Assessment

- CW 1. (5%) Thursday 5 November, Deadline: Thursday 19 November
- CW 2. (5%) Thursday 3 December, Deadline: Thursday 17 December
- May Exam (90%): 2 hour exam (year 3), 2.5-hour exam (year 4/Msc).

Problem sheets: not assessed, they are the best material for preparing for the exam.

Questions and Answers sessions: Wednesdays 11-12 am

Problem classes: Tuesdays Weeks 2-4-6-8-10

References:

- Real Analysis by H. L. Royden, Third edition
- Real Analysis, Modern Techniques and their applications, by G. B. Folland
- Measure Theory by P. R. Halmos
- Measures, Integrals and Martingales, by René L. Schilling, Cambridge UP, 2005. 176-89. Web.

How to use these notes? A 'remark' is material which is helpful with our understanding, which may or may not be covered in the lectures. An 'exercise' may appear in the weekly problem sheet, and is in the notes for one or more of the following reasons: (1) the conclusion is interesting, (2) the proof is representative, (3) proving it improves our understanding.

If a section /paragraph is given a $*$ sign, it is not covered in class (or not 'officially' covered in class). They are in the notes for the background or considered to be useful to further our understanding.

1.3 Prologue

Measure theory is not only a fundamental tool to measure/compute, it also provides the basis for several fundamental concepts in mathematics.

1.4 The mass transport problem

Monge was concerned with the cost of moving material from a mine to a construction site, *Monge's problem* (1781) is: What is the optimal way to transport a pile of sand?

Let us think of μ as the original sand mass distribution, a measure μ. We also think of the target mass as a measure ν. A map T is a transport map if it pushes μ to ν. The optimal transport problem from μ to ν is then formulated as follows: find a map $x \mapsto T(x)$ such that the cost (distance)

$$\int_{\mathbf{R}} |x - T(x)|\mu(dx)$$

is minimised among all transport maps. We may also use other cost functions, e.g. kinetic energy or take into consideration the geometry of the space.

Example 1.1 Shift of a block of mass of height one of uniform density on $[0, n]$ to $[1, n+1]$. There are more than one transport maps: $T_1(x) = x + 1$ for any $x \in [0, n]$; and

$$T_2(x) = \begin{cases} n + 1 - x, & x \in [0, 1) \\ x, & x \in [1, n] \end{cases},$$

are two different examples.

Example 1.2 Can one find a map shifting 1 unit of mass at location 0 to location 1 and -1 of equal mass?

In this case, the measures involved are $\mu = \delta_0$, $\nu = \frac{1}{2}\delta_1 + \frac{1}{2}\delta_{-1}$ and no transport map exists.

Kantorovich's formulation (1940's). Find a measure on $\Omega \times \Omega$ that minimises

$$\int |x - y|\tilde{\mu}(dx, dy)$$

and has marginals μ, ν respectively, i.e. $\tilde{\mu}(\Omega \times B) = \nu(B)$, $\tilde{\mu}(A \times \Omega) = \mu(A)$. This is called a transport plan.

Find a transport plan for Example 2. (Hint: this is a measure on $\{(0,0), (1,0), (0,1), (1,1)\}$.

Optimize the transport cost involved in distribution of pastries from bakeries to coffee shops. E.g. Bakery 1 delivers 50 pastries to Cafe 1, 30 each to Cafe shops 2 and 3, Bakery 2 delivers 20 to Cafe 2, 30 to cafe 3.

1.5 Coin tossing

We will need to understand what is meant by taking limits of functions and measures.

Example 1.3 Toss a fair coin 1000 times. Denote by X_i the result of the ith toss: $X_i = 1$ if we get head and $X_i = -1$ if we get tail. Mathematically, X_i can be represented by a *random variable*. Assuming that the coin is fair and that all tosses are independent, then the random variables $X_i, i \geq 1$ are independent and identically distributed (i.i.d.), with probability distribution $\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$, i.e.

$$\mathbf{P}(\{X_i = 1\}) = \mathbf{P}(\{X_i = -1\}) = \frac{1}{2}.$$

The **Law of Large Numbers (LLN)** is a theorem ensuring that the empirical mean of a large number of tosses converges to the expectation of a single toss

$$\frac{1}{N}\sum_{k=1}^{n} X_i \underset{N\to\infty}{\longrightarrow} \mathbf{E}(X_1) = 0$$

almost-surely. Thus, if you toss a coin 1000 times, you expect the empirical mean to be very close to 0. This is not suprising in practice. However the formulation and proof of the LLN (not covered in this module) is challenging, and requires sophisticated tools of measure theory and probability.

The **Central Limit Theorem (CLT)** describes how the empirical mean fluctuates. Namely, we have the convergence *in law*

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N} X_i \underset{N\to\infty}{\longrightarrow} \mathbf{N}(0,1)$$

where $\mathbf{N}(0,1)$ denotes a standard Gaussian distribution distributions. Thus, the deviation of the empirical mean away from 0 will be distributed , for large N, like a Gaussian curve.

The (LLN) and (CLT) are two fundamental results of probability theory, with far-reaching consequences. These results will not be proved in this module, but the key notions behind will be introduced.

1.6 Extending Riemann's theory of integration

1.6.1 Inadequacy of Riemann Integrals

Let $f : [a,b] \to \mathbf{R}$ be a bounded function. Riemann's theory of integration allows to define $\int_a^b f(x)\,dx$ provided that f is Riemann-integrable. Recall the following definition:

Definition 1.6.1 For any partition $P = \{a = t_0 < t_1 < \ldots < t_n\}$ of $[a,b]$, we set

$$L(f,P) = \sum_{i=1}^{n}(t_i - t_{i-1}) \inf_{t\in[t_{i-1},t_i]} f(t), \qquad U(f,P) = \sum_{i=1}^{n}(t_i - t_{i-1}) \sup_{t\in[t_{i-1},t_i]} f(t),$$

and define

$$\underline{\int} f = \sup_P L(f,P), \qquad \overline{\int} f = \inf_P U(f,P),$$

where the infimum and supremum are taken over all partitions of $[0,1]$. Then f is Riemann-integrable if $\underline{\int} f$ and $\overline{\int} f$ coincide. In this case we define the Riemann integral $\int_a^b f(x)\,dx$ to be this common value.

Heuristically, f being Riemann-integrable means that one can approximate the area under the graph of f by rectangles $[t_{i-1}, t_i] \times [0, f(s_i)]$, where $P = \{a = t_0 < t_1 < \ldots < t_n\}$ is a partition of $[a, b]$, and s_i is any point in $[t_{i-1}, t_i]$. Intervals are thus the "building blocks" of Riemann's theory.

Unfortunately, in many interesting cases, f is not Riemann-integrable.

Example 1.4 An example of non-Riemann-integrable function is the Dirichlet function on $[0, 1]$, denoted by $\mathbf{1_Q}$: $\mathbf{1_Q}(x) = 1$ for $x \in Q$ and $\mathbf{1_Q}(x) = 0$, for $x \in [0, 1] \setminus Q$. Indeed one has $U(\mathbf{1_Q}, P) = 1$ and $L(\mathbf{1_Q}, P) = 0$ along any partition P of $[0, 1]$, so that the lower and upper integrals are distinct:

$$\underline{\int} \mathbf{1_Q} := \sup_P L(\mathbf{1_Q}, P) = 0, \qquad \overline{\int} \mathbf{1_Q} := \inf_P U(\mathbf{1_Q}, P) = 1.$$

Same with the function $\mathbf{1}_{[0,1]\setminus \mathbf{Q}}$ defined on $[0, 1]$ by $\mathbf{1}_{[0,1]\setminus \mathbf{Q}}(0) = 1$ for $x \in Q$ and $\mathbf{1}_{[0,1]\setminus \mathbf{Q}}(x) = 1$ for $x \in [0, 1] \setminus Q$.

There are more irrational points than rational points, perhaps $\int_0^1 f_1(x)dx$ can be defined to be 1? Can we quantify how big are the sets of rational and irrational number, to see that the latter is larger than the former?

One inconvenient feature of the Riemann integral is its lack of stability under taking limits. Indeed, one can construct a sequence of bounded Riemann integrable functions f_n such that $f_n(x) \to f(x)$ for every $x \in [0, 1]$, but f is not Riemann integrable, as the following exmple shows:

Example 1.5 Let $Q = \{q_1, q_2, \ldots\}$ be an enumeration of rational numbers. Define $f_n(x) = 1$ if $x \in \{q_1, q_2, \ldots, q_n\}$ and set $f_n(x) = 0$ otherwsie. Then f_n is Riemann interagble and $\int f_n(x)dx = 1$, but f is the Dirichlet function which is not Riemann integrable.

1.6.2 Idea of Lebesgue's integral: allow for more sets as "building blocks"

Although the Dirichlet is not Riemann integrable, seen from another perspective, it is a simple function, simply the indicator of the set **Q**. The idea behind Lebesgue's integral is to extend the family of "building blocks" beyond intervals, allowing for more general sets such as **Q**.

The question then arises: what kind of subsets of the plane or the real line can be measured? Classically, the basic sets are: intervals on **R**; triangles, squares, polygons, squares on $\mathbf{R}^2$. For instance, relating the area of a circle to its length is an old mathematical problem, accomplished by 'exhaustions' (by the ancient Greeks 5th BC, the Achimedes of Syracus : 2-3 BC), which corresponds to taking limit in the modern language.

To extend our theory, we therefore begin with simple sets we can measure, and gradually build up more measurable sets by taking complements, unions, and approximations. This translates into the definition of σ-algebras. The new building blocks of our theory will now be measurable sets, i.e. elements of the Lebesgue σ-algebra. The way in which we assign a measure/weight to these sets will correspond to the notion of Lebesgue measure.

1.7 Observables of Random Variables

Besides allowing to extend Riemann's theory of integration, measure theory allows to define integrals in a much more general setting: on any given measure space (to be defined below), one can build up the associated measure theory. One important example corresponds to probability spaces.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, let X be a real-valued random variable, and let $f : \Omega \to \mathbf{R}$ be a Borel measurable function. If $\mathbf{P}(\{X = i\}) = p_i$ for all $i \geq 1$, with $\sum_i p_i = 1$, the expectation of X is given by

$$\mathbf{E}f(X) = \sum_i p_i f(i)$$

On the other hand, if X is a random variable with standard Gaussian normal distribution, then

$$\mathbf{E}X = \int \frac{1}{\sqrt{2\pi}} e^{-x^2/2} f(x) dx.$$

Althought they look different, both expectations are integrals $\int f(y)\, d\mu(y)$ where μ is probability distribution of X. Both are also given by $\int_\Omega X(\omega) d\mathbf{P}(\omega)$. These concepts use integration theory with respect to a measure.

Chapter 2

Measurable sets

Given a set $\mathcal{X}$, the collection of its subsets is its power set. We denote it by $2^{\mathcal{X}}$. What kind of subsets can be measured? What axioms should they satisfy? If a subset A can be measured, this ought to lead to some understanding of its complement $A^c = X \setminus A$. Moreover, if A, B can be measured, it is natural to postulate that $A \cup B$ can be measured. Actually, we further require limits of measurable sets to be measurable, so that if $(A_n)_{n\geq 1}$ is a sequence of measurable functions, $\bigcup_{n\geq 1} A_n$ should be measurable as well.

2.1 Set Algebra

Denote by A^c the complement of a subset A.

de Morgan's identities:

$$(\cap_{A\in\Lambda} A_\alpha)^c = (\cup A \in \Lambda_\alpha)^c$$
$$(\cup_{A\in\Lambda} A_\alpha)^c = (\cap A \in \Lambda_\alpha)^c$$

This holds for any number of sets, not necessarily countable number of.

Distribution Law

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C), \qquad A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$

Distribution law holds for also for infinite number of sets. For example

$$(\bigcup_i A_i) \cap (\bigcup_i B_j) = \bigcup_{i,j} (A_i \cap B_j).$$

Limits If A_n is an increasing sequence of subsets, we set

$$\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n.$$

Similarly if A_n is a decreasing sequence of subsets, we define

$$\lim_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n.$$

Let $f : X \to Y$ be a map, define

$$f^{-1}(A) = \{x \in X : f(x) \in A\}.$$

Taking pre-image respects set operations:

Lemma 2.1.1

$$f^{-1}\Big(\bigcup B_i\Big) = \bigcup f^{-1}(B_i),$$
$$f^{-1}\Big(\bigcap B_i\Big) = \bigcap f^{-1}(B_i),$$
$$f^{-1}(A \setminus B) = f^{-1}(A) \setminus f^{-1}(B).$$

2.2 Measurable Space

Let $\mathcal{X}$ be a non-empty set. We study collections of subsets of $\mathcal{X}$. Let ϕ denote the empty set.

Definition 2.2.1 A collection of subsets of X, denoted by $\mathcal{A}$, is a (Boolean) algebra if the following property holds.

- (**Non-emptiness**) $\phi \in \mathcal{A}$.
- (**Closedness under complements**) If $B \in \mathcal{A}$ then $B^c \in \mathcal{A}$;
- (**Closedness under finite unions**) If $B \in \mathcal{A}$, $C \in \mathcal{A}$, then $B \cup C \in \mathcal{A}$.

The assumption $\phi \in \mathcal{A}$ can be equally replaced by the condition that $\mathcal{A}$ is not empty.

We would like to be able to take limits.

Definition 2.2.2 A collection $\mathcal{F}$ of subsets of X is a σ**-algebra** if:

- **non-emptiness:** $\phi \in \mathcal{F}$.
- **closedness under complements:** if $B \in \mathcal{F}$ then $B^c \in \mathcal{F}$.
- **closedness under countable unions:** if we have a sequence $(B_n)_{n\geq 1}$ with $B_n \in \mathcal{F}$ for all $n \geq 1$, then $\bigcup_{n=1}^{\infty} B_n \in \mathcal{F}$.

Elements of $\mathcal{F}$ are called **measurable sets** and $(\mathcal{X}, \mathcal{F})$ is said to be a **measurable space**.

Remark 2.2.1 Let $\mathcal{F}$ be a σ-algebra over $\mathcal{X}$. Then it follows from the definition that:

- $\mathcal{X} \in \mathcal{F}$
- If $A_1, \dots, A_n$ is a finite collection of elements of $\mathcal{F}$, then $\cup_{i=1}^n A_i \in \mathcal{F}$
- by de Morgan's identities, if $B_n \in \mathcal{F}$ for all $n \geq 1$, then $\cap_{n=1}^\infty B_n \in \mathcal{F}$.
- If $A, B \in \mathcal{F}$ then so does $A \setminus B$.

Examples:

1. The smallest σ-algebra over $\mathcal{X}$ is $\{\phi, \mathcal{X}\}$, the *coarse* σ-algebra
2. The largest σ-algebra over $\mathcal{X}$ is $2^{\mathcal{X}}$, the *discrete* σ-algebra
3. If $A \subset \mathcal{X}$, then $\mathcal{F} = \{\phi, \mathcal{X}, A, A^c\}$ is a σ-algebra.
4. The collection $\mathcal{F}$ of subsets A of $\mathcal{X}$ such that either A or A^c is countable, is a σ-algebra.

Exercise 2.2.1 Assume that $\mathcal{X}$ is infinite, and let $\mathcal{F} := \{A \in 2^{\mathcal{X}} : A \text{ is finite or } A^c \text{ is finite}\}$. Show that $\mathcal{F}$ is an algebra but not a σ-algebra.

Proposition 2.2.2 *The intersection of any number of σ-algebras is a σ-algebra. If $\mathcal{C}$ is any collection of subsets of a set $\mathcal{X}$, then there always exists a smallest σ-algebra containing $\mathcal{C}$.*

Proof The first point is straightforward to prove and left as an exercise. For the second point, it suffices to consider

$$\mathcal{A} = \bigcap_{\substack{\mathcal{B}\,\sigma-\text{algebra} \\ \mathcal{C}\subset\mathcal{B}}} \mathcal{B}.$$

In virtue of the first point, $\mathcal{A}$ is a σ-algebra. Moreover, $\mathcal{A}$ contains $\mathcal{C}$, and by contsruction $\mathcal{A} \subset \mathcal{B}$ for any other σ-algebra $\mathcal{B}$ containing $\mathcal{C}$, as claimed. $\square$

The intersection of σ-algebras is often denoted by $\wedge$. For example, $\mathcal{F}_1 \wedge \mathcal{F}_2$ denotes the intersection of $\mathcal{F}_1$ and $\mathcal{F}_2$.

Definition 2.2.3 If $\mathcal{C}$ is a collection of subsets of X, we denote by $\sigma(C)$ the smallest σ-algebra containing C. We call $\sigma(\mathcal{C})$ the σ-algebra generated by C.

Example 2.1 Let $\mathcal{X}$ with a finite partition $A_1, \ldots, A_n$, by which we mean that $A_1, \ldots, A_n$ are disjoint and $\mathcal{X} = \cup_{k=1}^n A_k$. Let $\mathcal{C} = \{A_1, \ldots, A_n\}$. Then $\sigma(\mathcal{C})$ consists of all possible unions of the sets A_i:

$$\sigma(\mathcal{C}) = \left\{ \bigcup_{i \in I} A_i, \quad I \subset \{1, \ldots n\} \right\}.$$

There are 2^n such choices: we can include or not include a particular set and every element of $\mathcal{F}$ will come from such a union. In this particular case, the σ-algebra consists of a finite number of elements.

Exercise 2.2.2 Assume that $\mathcal{X}$ is infinite, and let $\mathcal{F}$ be as in Exercise 2.2.1. Find the σ-algebra generated by $\mathcal{F}$.

2.2.1 Countability of σ-algebras

We will see below that a σ-algebra with an infinite number of elements is uncountable. So a σ-algebra is either generated by a partition as in Example 2.1 above, or has an uncountable number of elements.

Proposition 2.2.3 *A σ-algebra $\mathcal{F}$ is either finite or uncountable.*

Proof will be released later in November.

2.2.2 Fundamental Examples and Exercises

Proposition 2.2.4 *If $f : \Omega \to \mathcal{X}$ is a map, and $(\mathcal{X}, \mathcal{B})$ a measurable space. Then f pulls back the σ-algebra $\mathcal{G}$ to a σ-algebra on Ω. This is the collection of pre-images*

$$\sigma(f) := \{f^{-1}(A) : A \in \mathcal{B}\}$$

is a σ-algebra. This is called the σ-algebra generated by f, and also called the pre-image σ-algebra.

Proof. Exercise.

Example 2.2 Let $\mathbf{R}$ be endowed with the Borel σ-algebra $\mathcal{B}(\mathbf{R})$. If $f : X \to \mathbf{R}$ is of the form

$$f(x) = \sum_{j=1}^n a_j \mathbf{1}_{A_j}(x), \qquad x \in X$$

where $a_j \in \mathbf{R}$ are distinct and the A_j form a partition of X (i.e. $\cup_{j=1}^n A_j = X$ and the A_j are pairwise disjoint), then $\sigma(f)$ is generated by the finite collection of sets $\{A_1, \ldots, A_n\}$.

Example 2.3 To see the role played by the assumption that A_j are disjoint, we give an example where A_j are not disjoint. Set for example

$$f(x) = \mathbf{1}_{[1,2]} + 3\mathbf{1}_{[1,3]} + 4\mathbf{1}_{(-2,1]}.$$

Then

$$f(x) = \begin{cases} 4, & x \in (-2,1) \cup (1,2] \\ 3, & x \in (2,3] \\ 8, & x = 1 \\ 0, & \text{otherwise .} \end{cases}$$

The σ-algebra generated by f is given by the partition:

$$\{\{1\}, (-2,1) \cup (1,2], (2,3], (-\infty,-2] \cup (3,\infty)\}.$$

If $f(x) = \mathbf{1}_{[1,2]} + 3\mathbf{1}_{[1,3]}$, then $\sigma(f)$ is generated by the collection of subsets:

$$\{[1,2], (2,3], (-\infty,1) \cup (3,\infty), \phi, \mathbf{R}\}.$$

Example 2.4 If $\mathcal{F}$ is a σ-algebra and $A \subset X$, then $\{A \cap B : B \in \mathcal{F}\}$ defines a σ-algebra on A, called the trace of the σ-algebra $\mathcal{F}$.

Exercise 2.2.3 Given two σ-algebras $\mathcal{F}_1$ and $\mathcal{F}_2$, we denote by $\mathcal{F}_1 \vee \mathcal{F}_2$ the smallest σ-algebra containing both $\mathcal{F}_1$ and $\mathcal{F}_2$. Show that $\mathcal{F}_1 \vee \mathcal{F}_2$ can equivalently be characterised by the expressions:

- $\mathcal{F}_1 \vee \mathcal{F}_2 = \sigma\{A \cup B \,|\, A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2\}$,
- $\mathcal{F}_1 \vee \mathcal{F}_2 = \sigma\{A \cap B \,|\, A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2\}$.

Definition 2.2.4 A collection of subsets $\mathcal{E}$ is an elementary family if

- $\phi \in \mathcal{E}$;
- If $A, B \in \mathcal{E}$, then $A \cap B \in \mathcal{E}$;
- if $A \in \mathcal{E}$ then A^c is the finite union of disjoint sets from $\mathcal{E}$.

Exercise 2.2.4 Show that if $\mathcal{E}$ is an elementary family, then the collection $\mathcal{A}$ of finite disjoint unions of members of $\mathcal{E}$ is an algebra.

Proof If $A, B \in \mathcal{E}$, then $B^c = \cup_{l=1}^n B_l$ where $B_l \in \mathcal{E}$ are disjoint. So $A \setminus B = A \cap B^c = \cup_{l=1}^n (A \cap B_l) \in \mathcal{A}$. Also, $A \cup B = (A \setminus B) \cup B \in \mathcal{A}$. By induction if $A_i \in \mathcal{E}$ then $\cup_{i=1}^n A_i \in \mathcal{A}$. This means that the finite union of sets from the elementary family is in $\mathcal{A}$ and $\mathcal{A}$ is closed under taking unions.

We show $\mathcal{A}$ is closed under complement. Let $A = \cup_{j=1}^{n} A_j \in \mathcal{A}$ where $A_j \in \mathcal{E}$ are disjoint. Then $A_j^c = \sum_{m=1}^{n_j} A_j^m$ where $A_j^m \in \mathcal{E}$. Since $A^c = \cap_{j=1}^{n}(A_j^c)$, by associativity, so A^c is a finite union of intersections of sets from $\mathcal{E}$, therefore in $\mathcal{A}$. □

We call the following sets of the following form half open intervals: $(a, b]$, ϕ or (a, ∞), $(-\infty, a)$ and ϕ, where a, b are real numbers.

Example 2.5 The collection of unions of a finite number of disjoint half open intervals is an algebra.

This follows from taking $\mathcal{E}$ to be the collection of half open intervals and use the earlier exercise.

2.3 Borel σ-algebras

2.3.1 Definition

Definition 2.3.1 If $\mathcal{X}$ is a complete separable metric space , the smallest σ-algebra generated by the open sets is called Borel σ-algebra, it is denoted by $\mathcal{B}(\mathcal{X})$.

If X is a separable metric space, the metric topology satisfies the second axiom of countability, i.e. there exists a countable base. This countable base generates $\mathcal{B}(X)$.

2.3.2 The case of R

Example 2.6 We know the following about $\mathcal{B}(\mathbf{R})$, the Borel σ-algebra over $\mathbf{R}$.

1. If $a \in \mathbf{R}$, $\{a\} = \cap_{n=1}^{\infty}(a - \frac{1}{n}, a + \frac{1}{n}) \in \mathcal{B}(\mathbf{R})$.
2. Any countable subset of $\mathbf{R}$ is a Borel set.
3. Any intervals of any form is a Borel set.
4. $\mathcal{B}(\mathbf{R})$ is generated by the set of open intervals. This follows from the proposition below.

Proposition 2.3.1 *Every open set of* $\mathbf{R}$ *is a countable union of disjoint open sets. This decomposition into disjoint open sets is unique up to ordering.*

Proof Denote by $\mathcal{A}$ the collection of open intervals. Let O be an open set. Let $x \in O$. Then x is contained in an open interval, the interval is a subset of O. The set of the left end points of such interval has an infimum. Let us call it a_x. Let b_x denote the supremum of the right end of the intervals. Then $I_x = (a_x, b_x)$ is the maximal interval such that $x \in I_x \subset O$ Given $x, y \in O$ either they are disjoint or they overlap in which case they are the same. Pick up a rational number q_x from I_x. Then the rational numbers are all distinct. So $\{I_x : x \in O\}$ contains only a countable number of distinct open intervals. □

2.3.3 Borel σ-algbera on a subset of a metric space

Let $(\mathcal{X}, d)$ be a metric space, and $\mathcal{Y} \subset \mathcal{X}$ be a subset of $\mathcal{X}$. Then $\mathcal{Y}$ (endowed with the restriction of the metric d to $\mathcal{Y}$) is still a metric space, its open sets are the subsets of the form $U \cap \mathcal{Y}$, where U is an open subset of $\mathcal{X}$. One may consider the corresponding Borel σ-algebra $\mathcal{B}(\mathcal{Y})$. It turns out that the following holds

Proposition 2.3.2 *$\mathcal{B}(\mathcal{Y})$ coincides with the trace on $\mathcal{Y}$ of $\mathcal{B}(\mathcal{X})$, that is*

$$\mathcal{B}(\mathcal{Y}) = \{A \cap \mathcal{Y},\ A \in \mathcal{B}(\mathcal{X})\},$$

(see Example 2.4 above).

Proof By Example 2.4, we know that $\mathcal{A} := \{A \cap \mathcal{Y},\ A \in \mathcal{B}(\mathcal{X})\}$ is a σ-algebra on $\mathcal{Y}$. Any open subset of $\mathcal{Y}$ is of the form $U \cap \mathcal{Y}$, where U is an open (hence Borel) subset of $\mathcal{X}$, hence $\mathcal{B}(\mathcal{Y}) \subset \mathcal{A}$. Conversely, the collection

$$\{A \in \mathcal{B}(\mathcal{X}) :\ A \cap \mathcal{Y}\}$$

is shown at once to be a σ-algebra on $\mathcal{X}$, and it contains open subsets of $\mathcal{X}$. Therefore it contains, hence coincides with, $\mathcal{B}(\mathcal{X})$. So $\mathcal{A} \subset \mathcal{B}(\mathcal{Y})$, and the claim follows. □

Remark 2.3.3 In the special case where $\mathcal{Y}$ is itself a Borel measurable subset of $\mathcal{X}$, then for any $A \in \mathcal{B}(\mathcal{X})$, we also have $A \cap \mathcal{Y} \in \mathcal{B}(\mathcal{X})$, hence it follows from the above proposition that

$$\mathcal{B}(\mathcal{Y}) = \{B \in \mathcal{B}(\mathcal{X}) :\ B \subset \mathcal{Y}\}.$$

In particular, in this case, $\mathcal{B}(\mathcal{Y}) \subset \mathcal{B}(\mathcal{X})$.

2.4 Product σ-algebras

Let $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ be two measurable spaces. In general, subsets of $\mathcal{X} \times \mathcal{Y}$ which are product sets, i.e. of the form $A \times B$ with $A \in \mathcal{F}$ and $B \in \mathcal{G}$, do not constitute a σ-algebra. Rather, we introduce the following definition.

Definition 2.4.1 We define the product σ-algebra (also called tensor σ-algebra) on the product space $\mathcal{X} \times \mathcal{Y}$ to be that generated by product sets

$$\mathcal{F} \otimes \mathcal{G} = \sigma(\{A \times B : A \in \mathcal{F}, B \in \mathcal{G}\}).$$

2.4.1 Product of a countable family of σ-algebras *

The above definition of product extends to any number of measurable spaces. We give the definition for the product of a countable number of measurable spaces.

Definition 2.4.2 Let $(E_n, \mathcal{F}_n)$ be measurable spaces. The tensor or product σ-algebra on the product space $\Pi_{n=1}^{\infty} \mathbf{E}_n$ is

$$\sigma\big(\{\Pi_{n\geq 1} A_n, : A_n \in \mathcal{F}_n\}\big).$$

Remark 2.4.1 By stability of a σ-algebra under countable intersections, the above σ-algebra is also generated by measurable cylinders, i.e. subsets of the form $\Pi_{k=1}^{n} A_k \times \Pi_{k>n} E_k$ for some $n \geq 1$ and with $A_1 \in \mathcal{F}_1, \ldots, A_n \in \mathcal{F}_n$.

Let $(E_n, d_n), n \geq 1$, be metric spaces. The product topology on $E = \Pi_{n\geq 1} E_n$ is the coarsest topology such that the projections $p_n : E \to E_n$ are continuous, it is generated by sets of the form $p_n^{-1}(U_n)$ where $n \geq 1$ and U_n is an open set of E_n. The product topology is induced by the distance d on E defined by

$$d(x, y) = \sum_{n=1}^{\infty} \frac{d_n(x_n, y_n) \wedge 1}{2^n},$$

for any two elements $x = (x_n)_{n\geq 1}$ and $y = (y_n)_{n\geq 1}$ of E. We may consider the Borel σ-algebra $\mathcal{B}(E)$ of the product metric space E, and investigate how it relates with the product σ-algebra $\otimes_{n\geq 1}\mathcal{B}(E_n)$.

Lemma 2.4.2 $\otimes_{n\geq 1}\mathcal{B}(E_n) \subset \mathcal{B}(E)$.

Proof By Remark 2.4.1 above, it suffices to prove that, for all $n \geq 1$ and for all $A_1 \in \mathcal{B}(E_1), \ldots, A_n \in \mathcal{B}(E_n)$, we have $A := A_1 \times \ldots \times A_n \times \Pi_{k>n} E_k \in \mathcal{B}(E)$. Note that the claim is true if $A_1, \ldots, A_n$ are all open subsets, as then A is an open, hence Borel subset of E. Thus, for all $A_2, \ldots A_n$ open subsets of $E_2, \ldots, E_n$, respectively, the collection

$$\mathcal{A}_1 := \{A_1 \in \mathcal{B}(E_1) : A_1 \times \ldots \times A_n \times \Pi_{k>n} E_k \in \mathcal{B}(E)\}$$

contains all open subsets of E_1. It is also a σ-algebra. Hence $\mathcal{A}_1 = \mathcal{B}(E_1)$, proving that

$$A_1 \times \ldots \times A_n \times \Pi_{k>n} E_k \in \mathcal{B}(E)$$

whenever $A_1 \in \mathcal{B}(E_1)$, and $A_2, \ldots A_n$ are open subsets of $E_2, \ldots, E_n$, respectively. Porceeding by induction over $i = 1, \ldots, n$, we similarly show that the claim still holds if $A_1, \ldots, A_i$ are Borel measurable and $A_{i+1}, \ldots, A_n$ are open. Taking $i = n$ yields the claim. □

Remark 2.4.3 An alternative reasoning is as follows. The coordinate mappings $p_n : E \to E_n$ being continuous, by Proposition 4.3.3 (see below), they are therefore measurable with respect to the Borel σ-algebra on the product space. Hence, for all $n \geq 1$, and all $A_1 \in \mathcal{B}(E_1), \ldots, A_n \in \mathcal{B}(E_n)$,

$$A_1 \times \ldots \times A_n \times \Pi_{k>n} E_k = \cap_{i=1}^{n} P_i^{-1}(A_i) \in \mathcal{B}(E)$$

which yields the claim.

We now make the additional assumption that the E_n are all separable.

Remark 2.4.4 If X is a separable metric space, the metric topology satisfies the second axiom of countability, i.e. there exists a countable base. It suffices e.g. to consider the (countable) collection of open balls

$$B(x,r) = \{y \in X : d(x,y) < r\},$$

for $r > 0$ rational, and for $x \in D$, where D is a countable dense subset of X. This countable base generates $\mathcal{B}(X)$.

Theorem 2.4.5 *Let (E_n) be separable metric spaces and set $E = \Pi_{n=1}^{\infty} E_n$. Then*

$$\mathcal{B}(E) = \otimes_{n=1}^{\infty} \mathcal{B}(E_n).$$

Proof In view of the above lemma, there only remains to prove that $\mathcal{B}(E) \subset \otimes_{n=1}^{\infty} \mathcal{B}(E_n)$. For all $n \geq 1$, let $\mathcal{D}_n$ be a countable base for the topology of E_n. Then subsets of E of the form $U_1 \times \ldots \times U_n \times \Pi_{k>n} E_k$, for $n \geq 1$, and $U_1 \in \mathcal{D}_1, \ldots, U_n \in \mathcal{D}_n$ form a countable base for the product topology on E. That is, any open subset W of E can be written in the form $W = \cup_{j\geq 1} W_j$, with $W_j = U_1 \times \ldots \times U_{n(j)} \times \Pi_{k>n(j)} E_k$, where, for all $k = 1, \ldots, n(j)$, U_k is an element of $\mathcal{D}_k$ (in particular it is open, hence Borel). In particular, for all $j \geq 1$, $W_j \in \otimes_{n=1}^{\infty} \mathcal{B}(E_n)$. Therefore $W \in \otimes_{n=1}^{\infty} \mathcal{B}(E_n)$, and this being true for any open subset W of E, the claim follows. □

2.5 A warning

We end this chapter with a warning:

Remark 2.5.1 if $\mathcal{C}$ is a sollection of subsets of a set $\mathcal{X}$ and $\mathcal{F} = \sigma(\mathcal{C})$, this does *not* mean in general that every measurable set is of the form $\cup_{i=1}^{\infty} C_i$ with $C_i \in \mathbf{C}$. Taking for example $\mathcal{C}$ to be the collection $\{(-\infty, a],\ \text{where } a \in \mathbf{R}\}$, then $\sigma(\mathcal{C}) = \mathcal{B}(\mathbf{R})$, however if $C_i = (-\infty, a_i]$ are elements of $\mathcal{C}$, then $\cup_{i=1}^{\infty} C_i$ is an interval of the form $(-\infty, a)$ or $(-\infty, a]$ with $a = \sup\{a_i, i \geq 0\}$, and the Borel set $\{-1, 1\}$, for instance, is not of this form. More generally, as soon as $\mathcal{C}$ is an infinite collection of sets, one cannot hope to represent a generic element of $\sigma(\mathcal{C})$ as some countable combination of elements of $\mathcal{C}$.

2.6 Background material*

A **topological space** $(\mathcal{X}, \mathcal{T})$ consists of a set $\mathcal{X}$ equipped with a topology $\mathcal{T}$, *i.e.* a subset $\mathcal{T} \subset 2^{\mathcal{X}}$ such that:

- $\phi \in \mathcal{T}$ and $\mathcal{X} \in \mathcal{T}$.
- If $\{A_0, A_1, \ldots, A_N\} \subset \mathcal{T}$, then $\bigcap_{n=0}^{N} A_n \in \mathcal{T}$.
- If $\mathcal{A} \subset \mathcal{T}$, then $\bigcup_{A \in \mathcal{A}} A \in \mathcal{T}$.

In other words, $\mathcal{T}$ is closed under arbitrary unions and finite intersections. Elements of $\mathcal{T}$ are called **open sets**. A function between topological spaces is **continuous** if the pre-images of open sets are open sets.

Given a topological space $(\mathcal{X}, \mathcal{T})$, we define $\mathcal{B}(\mathcal{X})$ to be the smallest σ-algebra on $\mathcal{X}$ containing $\mathcal{T}$. This particular σ-algebra is called the **Borel σ-algebra** of $\mathcal{X}$. In other words, the Borel σ-algebra is the smallest σ-algebra such that all open sets are measurable. We denote by $\mathcal{B}_b(\mathcal{X})$ the (Banach) space of all Borel-measurable and bounded functions from $\mathcal{X}$ to **R** equipped with the norm

$$\|\varphi\|_\infty = \sup_{x \in \mathcal{X}} |\varphi(x)| \; . \tag{2.1}$$

We denote by $\mathcal{C}_b(\mathcal{X})$ the (Banach) space of all continuous and bounded functions from $\mathcal{X}$ to **R** equipped with the same norm as in (2.1).

The open sets of a metric space gives a topology on the metric space. Borel σ-algebras on a complete separable metric space is specially nice.

When discussing Borel σ-algebras, we will always assume that $\mathcal{X}$ is a complete separable metric space.

A metric space is compact if any covering of it by open sets has a sub-covering of finite open sets. A discrete metric space (whose subsets are all open sets) is compact if and only if it is finite. (e.g. **Z** and **N** with the usual distance $d(x, y) = |x - y|$ is not compact.) A subset of a metric space is compact if it is compact as a metric space with the induced metric. It is relatively compact if its closure is compact. A metric space is sequentially compact if every sequence of its elements has a convergent sub-sequence (with limit in the metric space of course). It is totally bounded if for any $\epsilon > 0$ it has a finite covering by open balls of side ϵ. A metric space is complete if every Cauchy sequence converges.

It is a theorem that a metric space is compact if and only if it is complete and totally bounded. A metric space is compact if and only if it is sequentially compact.

A subset of a metric space is relatively compact if it is sequentially compact (the limit may not need to belong to the subset).

Chapter 3

Measures

3.1 Basics of Measures

A measure assigns a number to every measurable set.

Definition 3.1.1 A measure μ on the measurable space $(\mathcal{X}, \mathcal{F})$ is a map $\mu\colon \mathcal{F} \to [0, \infty]$ with the following properties;

- $\mu(\phi) = 0$.
- (σ-additive Propertices) If $\{A_n\}_{n>0}$ is a countable collection of elements of $\mathcal{F}$ that are all disjoint, then one has
$$\mu(\bigcup_{n>0} A_n) = \sum_{n>0} \mu(A_n).$$

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

Definition 3.1.2 We say that:

- μ is a finite measure if $\mu(\mathcal{X}) < \infty$.
- μ is a probability measure if $\mu(\mathcal{X}) = 1$.
- μ is **σ-finite** if there exists $A_1 \subset A_2 \subset \dots$ such that $\mu(A_i) < \infty$ and $\mathcal{X} = \cup_{i=1}^{\infty} A_j$. This is equivalent to the statement that there exists an increasing sequence of measurable sets $\mathcal{X}_n$ with $\mu(\mathcal{X}_n) < \infty$ such that $\mathcal{X} = \cup_{n=1}^{\infty} \mathcal{X}_n$. The sequence is called an exhausting sequence.

Convention. From now on we study only σ-finite measures.

Example 3.1 If $\mathcal{X} = \{1, 2, \ldots, n\}$, let $\mathcal{F} = 2^{\mathcal{X}}$. Then $(\mathcal{X}, \mathcal{F})$ is a σ-algebra. If we set $\mu(i) = a_i$, this defines a measure.

Proposition 3.1.1 *Let $(\Omega, \mathcal{F})$ be a measurable space. Then*

1. *If $A, B \in \mathcal{F}$ and $A \subset B$ then $\mu(B \setminus A) = \mu(B) - \mu(A)$.*
2. **Monotonicity.** *If $A \subset B$ then $\mu(A) \le \mu(B)$.*
3. **(Finite) additivity.** *If $A_1, A_2 \in \mathcal{F}$ and $A_1 \cap A_2 = \phi$ then $\mu(A_1 \cup A_2) = \mu(A_1) + \mu(A_2)$.*
4. **Continuity from below.** *If $A_n \in \mathcal{F}$, A_n increases, then $\mu(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n)$.*
5. **Continuity from above.** *If $A_n \in \mathcal{F}$, A_n decreases and $\mu(A_1) < \infty$, then*
$$\mu(\cap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n).$$

Proof Claim 1 follows from that $B = A \cup (B \setminus A)$ is a disjoint union whenever $A \subset B$. Hence $\mu(B) = \mu(A) + \mu(B \setminus A)$.

Claim 2 follows by claim 1, claim 3 follows from taking $A_i = \phi$ for $i \ge 3$.

We prove claim 4. If $A_1 \subset A_2 \subset A_3 \subset$ ..., set $A_0 = \phi$, $A = \cup_{i=1}^{\infty} A_i$ Also set

$$B_1 = A_1, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_3 \setminus A_2, \ldots,$$

and so on. Then $\{B_i\}$ are pairwise disjoint and $\cup_{i=1}^{\infty} B_i = \cup_{i=1}^{\infty} A_i = A$. Then,

$$\mu(A) = \sum_{i=1}^{\infty} \mu(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mu(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} [\mu(A_i) - \mu(A_{i-1})] = \lim_{n\to\infty} \mu(A_n).$$

Prove claim 5. We assume that $\mu(A_1) < \infty$ and proceed as for claim 3. $\square$

Example 3.2
1. Let $\mathcal{X}$ be a set endowed with the discrete σ-algebra $2^{\mathcal{X}}$. Set
$$\mu(A) = |A|, \qquad A \subset \mathcal{X},$$
where $|A|$ denotes the cardinality of A. Then μ is a measure, called the counting measure on $(\mathcal{X}, 2^{\mathcal{X}})$.
2. Let $(\mathcal{X}, \mathcal{G})$ be a measurable space and $x \in \mathcal{X}$. The δ measure at x, defined by
$$\delta_x(A) = \begin{cases} 1, & \text{if } x \in A, \\ 0, & \text{if } x \notin A, \end{cases}$$
is a probability measure.

3. A discrete measure P on a countable space $\Omega = \{\omega_1, \omega_2, \dots\}$, is uniquely determined by the sequence of numbers $(p_i)_{i \geq 1}$, where

$$p_i = P(\{\omega_i\}), \qquad i \geq 1.$$

Then, if $A \subset \Omega$, we have

$$P(A) = \sum_{i=1}^{\infty} p_i \delta_{\omega_i}(A) = \sum_{i=1}^{\infty} p_i \mathbf{1}_A(\omega_i).$$

In particular P is a probability measure if and only if $\sum_{i=1}^{\infty} p_i = 1$.

The last example above contains many examples of measures on countable sets you already know:

- Let Y be a random variable on a two state space $\mathcal{X} = \{0, 1\}$ with Bernoulli distribution of parameter p: $\mathbf{P}(Y = 0) = p$ and $\mathbf{P}(Y = 1) = 1 - p$. Then the probability distribution of Y is the measure on $\mathcal{X}$ given by:
$$\mathbf{P}_Y := p\, \delta_0 + (1 - p)\, \delta_1.$$
- Let X be a binomial random variable of parameter p on $\{0, \dots, n\}$. The probability distribution of X is the measure P_X on $\{0, \dots, n\}$ given by
$$P_X = \sum_{k=0}^{n}$$
- Let X be a Poisson random variable of parameter λ on $\{0, 1, 2, \dots\}$. The probability distribution of X is the measure P_X on $\{0, 1, 2 \dots, \}$ given by
$$P_X(A) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} \delta_k(A).$$

Not all measures can be written as linear combinations of Dirac measures as above: here are two examples of more complicated measures.

Example 3.3 The interval $[0, 1]$ equipped with its Borel σ-algebra admits a unique probability measure λ such that $\lambda([a, b]) = b - a$, for all $a, b \in [0, 1]$, $a \leq b$. The existence and uniqueness of such a measure, called the Lebesgue measure on $[0, 1]$, will be shown below.

Example 3.4 The measure

$$\mathbf{P}(A) = \int_A e^{-x}\, dx,$$

to be defined below, is a probability measrue on the half-line $\mathbf{R}_+$ equipped with $\mathcal{B}(\mathbf{R}_+)$. In such a situation, where the measure has a density with respect to Lebesgue measure, we will also use the shorthand notation $\mathbf{P}(dx) = e^{-x}\, dx$.

Exercise 3.1.1 If $\mu : \mathcal{F} \to [0, \infty]$ satisfies $\mu(\phi) = 0$, is finitely additive, and continuous from below, then it is a measure.

Exercise 3.1.2 Let **N** be endowed with the discrete σ-algebra $2^{\mathbf{N}}$. Let $\mu : 2^{\mathbf{N}} \to [0, \infty]$ be defined by $\mu(A) = \sum_{n \in A} 2^{-n}$ if A is finite, and $\mu(A) = +\infty$ otherwise.

1. Is μ additive?

2. Is μ a measure on $(\mathbf{N}, 2^{\mathbf{N}})$?

3.2 Uniqueness statements

We often know that a property holds for a sub-collection of a σ-algebra, and want to show it holds for every set in the σ-algebra. The difficulty with this is that σ-algebras are in generally complicated. If we have a finite partition of $\mathcal{X}$, the σ-algebra contains a finite number of sets. Otherwise it has uncountably many elements. See §2.2.1.

There are a number of powerful techniques, one of them is the $\pi - \lambda$ theorem.

3.2.1 $\pi - \lambda$ theorem

Definition 3.2.1 A non-empty collection of sub-sets $\mathcal{C}$ is a π-system if $A, B \in C$ implies that $A \cap B \in C$.

Definition 3.2.2 A collection $\mathcal{C}$ of sub-sets is called a λ-system if

1. $\Omega \in \mathcal{C}$

2. If $A, B \in \mathcal{C}$ and $A \subset B$ then $B \setminus A \in \mathcal{C}$.

3. If $(A_n)_{n \geq 1}$ is a non-decreasing sequence of elements of $\mathcal{C}$, i.e. $A_n \in \mathcal{C}$ and $A_n \subset A_{n+1}$ for all $n \geq 1$, then $\cup_{n \geq 1} A_n \in \mathcal{C}$.

Note that a σ-algebra is both a π-system and a λ system.

Proposition 3.2.1 (not given in class) *1. Show that the intersection of any number of λ-systems is a λ-systems. Therefore given any collections of subsets there exists a smallest λ-system containing it.*

2. If a λ system $\mathcal{A}$ is also a π-system, then it is a σ-algebra.

Proof Part (1) is trivial. To see a set which is simultaneously a π-system and a λ-system is a σ-algebra, remark that taking unions is obtained from taking intersections and complements. In turn, taking countable unions is obtained from taking finite unions and monotone limits, whence the claim. $\square$

Theorem 3.2.2 *If $\mathcal{C}$ is a π-system, then the smallest λ-system generated by $\mathcal{C}$, $\lambda(\mathcal{C})$, is a σ-algebra, and $\lambda(\mathcal{C}) = \sigma(\mathcal{C})$.*

Proof Firstly

$$\begin{aligned}\mathcal{F}_1 :&= \{A : A \cap \mathcal{C} \subset \lambda(\mathcal{C})\} \\ &= \{A \in \lambda(\mathcal{C}) : A \cap F \in \lambda(\mathcal{C}), \forall F \in \mathcal{C}\}\end{aligned}$$

is a λ-system containing $\mathcal{C}$, which implies $\mathcal{F}_1 = \lambda(\mathcal{C})$. Indeed, let $F \in \mathcal{C}$, then (1) $\Omega \cap F = F \in \mathcal{C}$, and so $\Omega \in \mathcal{F}_1$; (2) If $A \subset B$, with $A, B \in \mathcal{F}_1$,

$$(B \setminus A) \cap F = B \cap F \setminus (A \cap F) \in \mathcal{F}_1$$

(3).If $A_n \in \mathcal{F}_1$, $A_n \cap F \in \lambda(C)$, so $(\bigcup A_n) \cap F = \bigcup_n (A_n \cap F) \in \lambda(\mathcal{C})$.

Similarly,

$$\begin{aligned}\mathcal{F}_2 :&= \{A : A \cap \lambda(\mathcal{C}) \subset \lambda(\mathcal{C})\} \\ &= \{A \in \lambda(\mathcal{C}) : A \cap E \in \lambda(\mathcal{C}), \forall E \in \lambda(\mathcal{C})\}\end{aligned}$$

is a λ-system containing $\mathcal{C}$, and therefore $\mathcal{F}_2 = \lambda(\mathcal{C})$. Thus $\lambda(\mathcal{C})$ is both a λ-system and a π-system, and is therefore a σ-algebra. Since it contains $\mathcal{C}$, we obtain $\sigma(\mathcal{C}) \subset \lambda(\mathcal{C})$. But, since any σ-algebra is a λ-system, we also have $\lambda(\mathcal{C}) \subset \sigma(\mathcal{C})$, and the equality follows. $\square$

We present now a very useful result, which follows at once from Theorem 3.2.2.

Corollary 3.2.3 *[$\pi - \lambda$ Theorem.] If a λ-system contains a π-system $\mathcal{C}$, then it contains the σ-algebra generated by $\mathcal{C}$.*

3.3 Measure determining sets

Theorem 3.3.1 *Suppose that $(\Omega, \mathcal{F})$ is a measurable space, and $\mathcal{C}$ is a π-system generating $\mathcal{F}$. Let μ and ν be two measures which agree on $\mathcal{C}$.*

1. *If $\mu(\Omega) = \nu(\Omega) < \infty$, then $\mu = \nu$*
2. *More generally, if there exists an increasing sequence of subsets $\Omega_k \in \mathcal{C}$, such that $\Omega = \cup_{k\geq 1}\Omega_k$ and $\mu(\Omega_k) = \nu(\Omega_k) < \infty$ for all $k \geq 1$, then $\mu = \nu$.*

Proof (1) First assume that $\mu(\Omega) = \nu(\Omega) < \infty$. Let

$$\mathcal{G} = \{A \in \mathcal{F} : \mu(A) = \nu(A)\}.$$

Then $\Omega \in \mathcal{G}$ by assumption. Moreover, if A_n is a non-decreasing sequence of measurable sets in $\mathcal{G}$,

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(A_n) = \lim_{n\to\infty} \nu(A_n) = \nu(\bigcup_{n=1}^{\infty} A_n)$$

Thus, $\mathcal{G}$ is closed under taking lower limit. Let $A \subset B$, $A, B \in \mathcal{G}$, then by additive property,

$$\mu(B \setminus A) = \mu(B) - \mu(A) = \nu(B) - \nu(A) = \nu(B \setminus A).$$

This means $B \setminus A \in \mathcal{G}$, and therefore $\mathcal{G}$ is a λ-system containing a π-system generating $\mathcal{F}$. By the $\pi - \lambda$-Theorem, $\mathcal{G} \supset \mathcal{F}$ and $\mu = \nu$ on $\mathcal{F}$, which proves the first point.

(2) For the second point, let $\mathcal{F}_k = \{A \cap \Omega_k : A \in \mathcal{F}\}$ denote the trace σ-algebras on Ω_k, and denote by μ_k and ν_k the restrictions to Ω_k of the measures μ and ν:

$$\forall A \in \mathcal{F},\ \mu_k(A) = \mu(A \cap \Omega_k),\ \nu_k(A) = \nu(A \cap \Omega_k).$$

Applying the first point to μ_k and ν_k, we deduce that $\mu_k = \nu_k$. Therefore, by lower contiunity of measures, we obtain, for all $A \in \mathcal{F}$,

$$\mu(A) = \lim_{k\to\infty} \mu(A \cap \Omega_k) = \lim_{k\to\infty} \nu(A \cap \Omega_k) = \nu(A),$$

completing the proof. $\square$

Example 3.5 Let $\Omega = \{1, 2, 3, 4\}$. Let $\mathcal{G} = \{\{1, 2\}, \{1, 3\}, \{3, 4\}\}$, which generates the discrete σ-algebra $\mathcal{F}$ (the power set), but not a π-system. The measure μ and ν agree on $\mathcal{G}$, not on $\mathcal{F}$.

$$\mu(1) = 1/6, \quad \mu(2) = 2/6, \quad \mu(3) = 1/6, \quad \mu(4) = 2/6,$$

$$\nu(1) = 2/6, \quad \nu(2) = 1/6, \quad \nu(3) = 0, \quad \nu(4) = 3/6.$$

3.4 Construction of measures from pre-measures

If $(\mathcal{X}, \mathcal{F})$ is a measurable space, constructing a non-trivial measure μ on $(\mathcal{X}, \mathcal{F})$, is in general, a very complicated task. One of the main goals we are aiming at is the construction of the Lebesgue measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. This will require introducing further set-theoretical notions. We start with some definitons.

3.4.1 Negligible sets and complete σ-algebras

Definition 3.4.1 A measurable set A on a measure space $(\mathcal{M}, \mathcal{F}, \mu)$ is said to be a 'null set' if $\mu(A) = 0$. A subset N of $\mathcal{M}$ is negligible if there exists a null set $A \in \mathcal{F}$ such that $N \subset A$.

Remark 3.4.1 A negligible set N is not necessarily measurable. Therefore we may not write that $\mu(N) = 0$, simply because $\mu(N)$ is not defined.

Definition 3.4.2 Given a measure space $(\mathcal{M}, \mathcal{F}, \mu)$, $\mathcal{F}$ is said to be complete with respect to μ if it contains every negligible null set. The **completion** of $\mathcal{F}$ with respect to μ, denoted by $\bar{\mathcal{F}}$, is the σ-algebra given by $\bar{\mathcal{F}} = \mathcal{A} \vee \mathcal{N}$, where $\mathcal{N}$ is the collection of negligible sets.

3.4.2 A completion theorem*

Let $\mathcal{A}_i$ be two collections of subsets, $\varrho_i : \mathcal{A}_i \to [0, \infty]$, $i = 1, 2$. In the sequel we say ϱ_2 is an extension of ϱ_1 if $\mathcal{A}_2$ contains $\mathcal{A}_1$ and $\varrho_1(A) = \varrho_2(A)$ whenever $A \in \mathcal{A}_1$.

Let $(\mathcal{M}, \mathcal{F}, \mu)$ be a measure space.

Proposition 3.4.2 *The completion of $\mathcal{F}$ with respect to μ can be expressed as*

$$\begin{aligned}\bar{\mathcal{F}} &= \{A \cup N, A \in \mathcal{F}, N \in \mathcal{N}\} \\ &= \{A \subset \mathcal{M} : \exists B_1, B_2 \in \mathcal{F},\ \mu(B_2 \setminus B_1) = 0,\ B_1 \subset A \subset B_2\}.\end{aligned}$$

Moreover, we can extend uniquely μ into a measure $\bar{\mu}$ on $\bar{\mathcal{F}}$ by setting $\bar{\mu}(A \cup N) = \mu(A)$ for $A \in \mathcal{F}$ and $N \in \mathcal{N}$, and $\bar{\mathcal{F}}$ is complete w.r.t. $\bar{\mu}$. Furthermore, $\bar{\mathcal{F}}$ is the smallest σ-algebra containing $\mathcal{F}$ and to which μ can be extended in a way that $\bar{\mathcal{F}}$ is complete with respect to μ.

Proof The proof of the first two assumptions is similar to the proof of Theorem B, Section 13 (Chapter III) in Halmos's book. We show the last claim: assume that $\mathcal{G}$ is a sigma-algebra containing $\mathcal{F}$ and on which μ can be extended into a measure ν in a way that $\mathcal{G}$ is complete with respect to ν. By definition of $\mathcal{N}$, any $N \in \mathcal{N}$ satisfies $N \subset A$ for some $A \in \mathcal{F}$, with $\mu(A) = 0$. Since $\mathcal{F} \subset \mathcal{G}$ and ν is an extension of μ to $\mathcal{G}$, we also have $\nu(A) = 0$. Hence A is negligible with repect to ν. Since $\mathcal{G}$ is complete w.r.t. ν, we deduce that $N \in \mathcal{G}$. Thus $\mathcal{F} \cup \mathcal{N} \subset \mathcal{G}$, so since $\mathcal{G}$ is a σ-algebra, we have $\bar{\mathcal{F}} = \sigma(\mathcal{F} \cup \mathcal{N}) \subset \mathcal{G}$. Moreover, for all $A \in \mathcal{F}$ and $N \in \mathcal{N}$, we have $\nu(A) \leq \nu(A \cup N) \leq \nu(A) + \nu(N) = \nu(A)$, so $\nu(A \cup N) = \mu(A) = \bar{\mu}(A \cup N)$. Thus ν is an extension of $\bar{\mu}$ to $\mathcal{G}$, which yields the claim. $\square$

Remark 3.4.3 Ususally the extension $\bar{\mu}$ of μ to $\bar{\mathcal{F}}$ will still be denoted by μ.

3.4.3 Outer Measure

Definition 3.4.3 A function $\mu^* : 2^{\mathcal{X}} \to [0, \infty]$ is said to be an outer measure over $\mathcal{X}$ if we have:

(1) $\mu^*(\phi) = 0$,

(2) **Monotonicity.** If $A \subset B$ then $\mu^*(A) \le \mu^*(B)$.

(3) **countable sub-additivity.**

$$\mu^*(\cup_{n=1}^\infty A_n) \le \sum_{n=1}^\infty \mu^*(A_n).$$

Exercise 3.4.1

- Let $\mu^*(\phi) = 0$ and $\mu^*(A) = \infty$ if $A \subset \mathcal{X}$ is not empty. Show that μ^* is an outer-measure.
- If an outer measure μ^* satisfies that $\mu^*(A \cup B) = \mu^*(A) + \mu^*(B)$ for any $A, B \in \mathcal{F}$ where $\mathcal{F}$ is a σ-algebra, what can you say about μ^*?

Definition 3.4.4 Let $\mathcal{A}$ be a collection of subsets containing $\phi, \mathcal{X}$. A map $\mathcal{A} \to [0, \infty)$ is a **pre-measure** if

- $\varrho(\phi) = 0$, and
- for any $A_n \in \mathcal{A}$ pairwise disjoint **with** $\cup_{n=1}^\infty A_n \in \mathcal{A}$,

$$\varrho(\cup_{n=1}^\infty A_n) = \sum_{n=1}^\infty \varrho(A_n).$$

Outer measures can be obtained from any suitable function on a given collection of subsets, by covering a set with sets from this collection. A collection of set $\{A_\alpha, \alpha \in \Lambda\}$ is said to be a cover of a set E if $E \subset \cup_{\alpha \in \Lambda} A_\alpha$. We are interested in covers consisting of a countable number of elements.

Proposition 3.4.4 *Let $\mathcal{A}$ be a non-empty collection of subsets with $\phi \in \mathcal{A}, \mathcal{X} \in \mathcal{A}$. Let $\varrho : \mathcal{A} \to [0, \infty]$ be a function such that $\varrho(\phi) = 0$. We define a function on $2^{\mathcal{X}}$ as below. For any $E \subset \mathcal{X}$,*

$$\mu^*(E) = \inf\left\{ \sum_{j=1}^\infty \varrho(A_j) : E \subset \cup_{j=1}^\infty A_j, \quad A_j \in \mathcal{A} \right\}. \tag{3.1}$$

Then μ^* is an outer measure.

Proof (1) Since every set is covered by $\{\mathcal{X}\}$, the definition makes sense, and $\mu^*(\phi) = 0$. (2) If $A \subset B$, then every cover of B is a cover A and so $\mu^*(B) \ge \mu^*(A)$.

(3) To show countable sub-additivity, take any sequence of subsets A_j. For any $\epsilon > 0$ there exist $B_j^k \in \mathcal{A}$ with $A_j \subset \cup_k B_j^k$ such that

$$\mu^*(A_j) \geq \sum_{k=1}^{\infty} \varrho(B_j^k) - \frac{\epsilon}{2^j}.$$

Summing over j,

$$\sum_{j=1}^{\infty} \mu^*(A_j) \geq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \varrho(B_j^k) - \epsilon.$$

Since $\{B_j^k\}$ is a cover of $\cup_{j=1}^{\infty} A_j$,

$$\mu^*(\cup_{j=1}^{\infty} A_j) \leq \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \varrho(B_j^k).$$

Hence,

$$\sum_j \mu^*(A_j) \geq \mu^*(\cup_{j=1}^{\infty} A_j) - \epsilon$$

for any $\epsilon > 0$. Take $\epsilon \to 0$ to see $\mu^*(\cup_{j=1}^{\infty} A_j) \leq \sum_j \mu^*(A_j)$, thus completing the proof. □

How do we obtain additive from sub-additive? Let us single out sets that are strongly additive. Take a set A, we then have two collections of subsets: subsets of A on the one hand, and subsets of A^c on the other hand. If we take one from each collection and take their union, we want to prove that μ^* is additive along this decomposition.

Definition 3.4.5 A subset A of $\mathcal{X}$ is said to be μ^*-measurable (in the sense of Caratheodory) if for any set $B \subset \mathcal{X}$, we have

$$\mu^*(B) = \mu^*(B \cap A) + \mu^*(B \cap A^c). \tag{3.2}$$

We have already sub-additivity, the non-trivial part of (3.2) is

$$\mu^*(B) \geq \mu^*(B \cap A) + \mu^*(B \cap A^c). \tag{3.3}$$

Theorem 3.4.5 *[Caratheodory Theorem] If μ^* is an outer measure on $\mathcal{X}$ and $\mathcal{G}$ the collection of μ^*-measurable sets, then*

(1) $\mathcal{G}$ is a σ-algebra.

(2) The restriction of μ^* to $\mathcal{G}$ is a measure.

(3) $\mathcal{G}$ is complete w.r.t. μ^*.

Proof The proof is deferred to §3.4.5. □

3.4.4 Extension of pre-measures

Proposition 3.4.6 *If $\mathcal{A}$ is an algebra, and ϱ is a pre-measure on $\mathcal{A}$, we define μ^* by (3.1). Then the following statements hold.*

1. *Every set in $\mathcal{A}$ is μ^*-measurable.*
2. *μ^* and ϱ agree on $\mathcal{A}$.*

In particular, μ^* is a measure on $\sigma(\mathcal{A})$.

Proof The proof is deferred to §3.4.5. □

Proposition 3.4.7 ** If ϱ is a pre-measure on an algebra $\mathcal{A}$, we define μ^* by (3.1) and ϱ is σ-finite, then μ^* is its unique extension to $\mathcal{G}$ and therefore to $\sigma(\mathcal{A})$.*

Proof The proof is deferred to §3.4.5. □

3.4.5 Proof for Caratheodory's Theorem

We return to prove some of the theorems and propositions given earlier.,

Theorem 3.4.5 (Caratheodory's Theorem) If μ^* is an outer measure on $\mathcal{X}$, then the collection $\mathcal{G}$ of μ^*-measurable sets is a σ-algebra, the restriction of μ^* to $\mathcal{G}$ is a measure, and $\mathcal{G}$ is complete w.r.t. μ^*.

Proof Throughout the proof let us write $\mu = \mu^*$ for simplicity. (1) $\phi \in \mathcal{G}$ and $\mu^*(\phi) = 0$. (2) Since the roles played by A and A^c are symmetric, $A \in \mathcal{G}$ implies that $A^c \in \mathcal{G}$.

(3) Suppose $A, F \in \mathcal{G}$, we show $A \cup F \in \mathcal{G}$.

Take any $B \in \mathcal{X}$. We apply (3.2) first with $A \in \mathcal{G}$ and then with $F \in \mathcal{G}$:

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

We use $A \cup F = (A \cap F) \cup (A \cap F^c) \cup (A^c \cap F)$, as well as $(A \cup F)^c = A^c \cap F^c$, and apply sub-additivity

$$\begin{aligned}\mu(B) &= \mu(B \cap A \cap F) + \mu(B \cap A \cap F)^c + \mu(B \cap A^c \cap F) + \mu(B \cap A^c \cap F^c)\\ &\geq \mu(B \cap (A \cup F)) + \mu(B \cap (A \cup F)^c).\end{aligned}$$

This shows that $A \cup F \in \mathcal{G}$. Also, if $A, F \in \mathcal{G}$ are disjoint,

$$\mu(A \cup F) = \mu((A \cup F) \cap F) + \mu((A \cup F) \cap F^c) = \mu(F) + \mu(A).$$

(4) Let $A_n \in \mathcal{G}$, we show $A = \cup_{n=1}^{\infty} A_n \in \mathcal{G}$. Set $F_n = \cup_{j=1}^{n} A_j$. Since $\mathcal{G}$ is stable under finite unions and under taking the complement, we may assume that the A_n are disjoint. Then for any $B \in \mathcal{X}$, using $A_j \in \mathcal{G}$,

$$\begin{aligned}
\mu(B \cap F_n) &= \mu(B \cap F_n \cap A_n) + \mu(B \cap F_n \cap (A_n)^c) \\
&= \mu(B \cap A_n) + \mu(B \cap F_{n-1}) \\
&= \mu(B \cap A_n) + \mu(B \cap F_{n-1} \cap A_{n-1}) + \mu(B \cap F_{n-1} \cap A_{n-1}^c) \\
&= \mu(B \cap A_n) + \mu(B \cap A_{n-1}) + \mu(B \cap F_{n-2}) = \dots \\
&= \sum_{j=1}^{n} \mu(B \cap A_j).
\end{aligned}$$

Thus using $F_n \in \mathcal{G}$,

$$\mu(B) = \mu(B \cap F_n) + \mu(B \cap (F_n)^c) = \sum_{j=1}^{n} \mu(B \cap A_j) + \mu(B \cap (F_n)^c)$$

$$\geq \sum_{j=1}^{n} \mu(B \cap A_j) + \mu(B \cap A^c).$$

Taking $n \to \infty$, then using $\cup(B \cap A_j) = B \cap A$ and sub-additivity,

$$\begin{aligned}
\mu(B) &\geq \sum_{j=1}^{\infty} \mu(B \cap A_j) + \mu(B \cap A^c) \\
&\geq \mu(B \cap A) + \mu(B \cap A^c) \geq \mu(B).
\end{aligned}$$

Hence $A \in \mathcal{G}$ and $\sum_{j=1}^{\infty} \mu(B \cap A_j) = \mu(B \cap A)$ for any $B \in \mathcal{G}$. Take $B = A$ to conclude the σ-additivity. We proved that $\mathcal{G}$ is a σ-algebra and μ is a measure on it.

(5) Finally if $\mu(A) = 0$ and $F \subset A$, then for any $B \subset \mathcal{X}$,

$$\mu(B) \leq \mu(B \cap F) + \mu(B \cap (F)^c) \leq 0 + \mu(B).$$

Hence all the inequalities above are equalities,

$$\mu(B) = \mu(B \cap F) + \mu(B \cap (F)^c)$$

and $F \in \mathcal{G}$. Therefore $\mathcal{G}$ contains all negligible subsets F for μ and $\mu(F) = 0$. $\square$

In the sequel, we shall assume $\mathcal{A}$ to be an algebra. In that case, we can use the following, convenient lemma:

Lemma 3.4.8 *Let $\mathcal{A}$ be an algebra and $\varrho : \mathcal{A} \to [0, \infty]$ an additive function. In words, for all $A, B \in \mathcal{A}$ disjoint, we have:*

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

Then ϱ is

1. *monotone: for any $A, B \in \mathcal{A}$ with $A \subset B$, $\varrho(A) \leq \varrho(B)$*

2. *sub-additive: for any $A_1, \ldots, A_n \in \mathcal{A}$, $\varrho(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n \varrho(A_i)$*

Proof By additivity for all $A, B \in \mathcal{A}$ with $A \subset B$, we have

$$\varrho(B) = \varrho(A) + \varrho(B \setminus A) \geq \varrho(A),$$

which yields monotonicity. Now, for $A_1, \ldots, A_n \in \mathcal{A}$, we set

$$B_i = A_i \setminus \left(\bigcup_{j=1}^{i-1} A_j \right) \in \mathcal{A}, \qquad i = 1, \ldots, n,$$

so that the B_i are disjoint. Hence, using successively the additivity and monotonicity of ϱ,

$$\varrho\left(\bigcup_{i=1}^n A_i \right) = \varrho\left(\bigcup_{i=1}^n B_i \right) = \sum_{i=1}^n \varrho(B_i) \leq \sum_{i=1}^n \varrho(A_i),$$

which yields the requested sub-additivity. $\square$

Proposition3.4.6. If ϱ is a pre-measure on an algebra $\mathcal{A}$, we define μ^* by (3.1). Then:

1. $\mu^* = \varrho$ on $\mathcal{A}$.

2. every set in $\mathcal{A}$ is μ^*-measurable.

In particular, μ^* is a measure on $\overline{\sigma(\mathcal{A})}$, the completion of $\sigma(\mathcal{A})$ w.r.t μ^*.

Proof (1) Let $B \in \mathcal{A}$, then B is a cover for itself and $\mu^*(B) \leq \varrho(B)$. As before write $\mu = \mu^*$ for simplicity. We now prove the reverse inequality. Suppose that $B \subset \cup_n A_n$ where $A_n \in \mathcal{A}$. Then $B = \cup_n \tilde{A}_n$, where $\tilde{A}_n = (B \cap A) \setminus (\cup_{i<n} A_i)$. Now $\tilde{A}_n \in \mathcal{A}$ are disjoint, and $B \in \mathcal{A}$. Applying σ-additivity for pre-measures, and monotonicity, we get

$$\varrho(B) = \sum_{n=1}^\infty \varrho(\tilde{A}_n) \leq \sum_{n=1}^\infty \varrho(A_n).$$

Take infimum over all covers,

$$\varrho(B) \leq \inf \sum_{n=1}^\infty \varrho(A_n) = \mu(B).$$

We have showed that ϱ and μ agree on $\mathcal{A}$.

(2) We now show that any $A \in \mathcal{A}$ is μ^*-measurable. Take any $B \subset X$, we want to show

$$\mu^*(B) \geq \mu^*(A \cap B) + \mu^*(A^c \cap B), \tag{3.4}$$

For any $\epsilon > 0$ choose a cover $\{B_j\}$ of B from $\mathcal{A}$ with

$$\mu^*(B) \geq \sum_{j=1}^{\infty} \varrho(B_j) - \epsilon.$$

We first use $A \in \mathcal{A} \subset \mathcal{G}$, $\varrho = \mu^*$ on $\mathcal{A}$ and σ- sub-additivity of μ, and

$$\begin{aligned}\sum_{j=1}^{\infty} \varrho(B_j) &= \sum_{j=1}^{\infty} \varrho(A \cap B_j) + \sum_{j=1}^{\infty} \varrho(A^c \cap B_j) \\ &\geq \mu^*(A \cap \cup_j B_j) + \mu^*(A^c \cap \cup_j B_j) \\ &\geq \mu^*(A \cap B) + \mu^*(A^c \cap B).\end{aligned}$$

In the last step we use $\cup_j B_j \supset B$ and the monotonicity of the outer measure. Therefore

$$\mu^*(B) \geq \mu^*(A \cap B) + \mu^*(A^c \cap B) - \epsilon.$$

Taking $\epsilon \to 0$, we obtain

$$\mu^*(B) \geq \mu^*(A \cap B) + \mu^*(A^c \cap B),$$

so that A is μ^*-measurable (the reverse inequality follows by sub-additivity), which proves the second claim. For the final statement, note that by 3.4.5, μ^* defines a measure on $\mathcal{G}$, and that $\mathcal{G}$ is complete w.r.t. μ^*. Since $\mathcal{A} \subset \mathcal{G}$, therefore $\overline{\sigma(\mathcal{A})} \subset \mathcal{G}$, and the claim follows. □

Proposition 3.4.9 ** Let ϱ be a σ-finite pre-measure on an algebra $\mathcal{A}$, and define μ^* by (3.1). Then μ^* is the unique measure extending ϱ to $\mathcal{G}$, and therefore the unique measure on $\sigma(\mathcal{A})$, extending ϱ.*

Proof Step 1. We first show that if ν is another measure on the μ^*-measurable set $\mathcal{G}$ extending ϱ, then $\nu \leq \mu^*$. Indeed if $E \in \mathcal{G}$, $E \subset \cup A_j$ where $A_j \in \mathcal{A}$, then by sub-additivity,

$$\nu(E) \leq \nu(\cup A_j) \leq \sum_j \nu(A_j) = \sum_j \varrho(A_j).$$

This holds for any covering of E by sets from $\mathcal{A}$, hence

$$\nu(E) \leq \mu^*(E).$$

Step 2. Suppose that $E \in \mathcal{G}$ is such that $\mu^*(E) < \infty$, we show that $\mu^*(E) = \nu(E)$. We can find a cover A_j of E from $\mathcal{A}$ with $\sum_j \mu^*(A_j) \leq \mu^*(E) + \epsilon$. We may and will assume that the cover is disjoint. Set $A = \cup_{j=1}^{\infty} A_j$. Since $E \subset A$,

$$\mu^*(A \setminus E) = \mu^*(A) - \mu^*(E) \leq \sum_j \mu^*(A_j) - \mu^*(E) < \epsilon. \tag{3.5}$$

Also, since $E \subset A$,

$$\begin{aligned}\mu^*(E) \le \mu^*(A) &\le \sum_j \mu^*(A_j)\\ &\overset{\text{agree on } \mathcal{A}}{=} \sum_j \nu(A_j) = \nu(A)\\ &= \nu(E) + \nu(A \setminus E)\\ &\overset{\nu \le \mu^*}{\le} \nu(E) + \mu^*(A \setminus E) \le \nu(E) + \epsilon,\end{aligned}$$

where we have used (3.5) in the last inequality. Taking $\epsilon \to 0$, we see that $\mu^*(E) = \nu(E)$.

Step 3. Let ϱ be σ-finite. By the assumption there exists an increasing sequence of sets $\mathcal{X}_n$ from $\mathcal{A}$ such that $\varrho(\mathcal{X}_n) < \infty$ and $\bigcup_{n \ge 1} \mathcal{X}_n = \mathcal{X}$. Then using step 2 we can complete the proof (exercise). □

3.4.6 Construction of Lebesgue-Stieltjes / Lebesgue measure

We want to construct a measure λ on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, such that the measure of a half open interval is its length, and such that it is translation invariant, i.e. $\lambda(A + t) = \lambda(A)$ for any $t \in \mathbf{R}$ and any Borel set A, where

$$A + t = \{a + t : a \in A\}$$

is the translate of A by the number t. We will show that there exists a unique such measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ called the Lebesgue measure. This will be one specific example of Lebesgue-Stieltjes measure.

Definition 3.4.6 A collection of subsets $\mathcal{E}$ is an elementary family if

- $\phi \in \mathcal{E}$;
- If $A, B \in \mathcal{E}$, then $A \cap B \in \mathcal{E}$;
- if $A \in \mathcal{E}$ then A^c is a finite union of disjoint sets from $\mathcal{E}$.

Exercise 3.4.2 Show that if $\mathcal{E}$ is an elementary family, then the collection $\mathcal{A}$ of finite disjoint unions of members of $\mathcal{E}$ is an algebra.

In this section, by half open interval we mean sets of the form $(a, b]$, (c, ∞), $(-\infty, b]$, or ϕ, where a, b, c are finite numbers. For simplicity we write a generic half interval as $(c, d]$, where $c \in \mathbf{R} \cup \{-\infty\}$, and d is a real number **or ∞. In the latter case by $(c, \infty]$, in this chapter, we really mean $(c, \infty] \cap \mathbf{R}$.**

Definition 3.4.7 Henceforth, in this section, let $\mathcal{E}$ denote the collection of half open intervals and let $\mathcal{A}$ be the collection of finite unions of disjoint half open intervals.

We have seen that

Proposition 3.4.10 *$\mathcal{A}$ is an algebra.*

Remark 3.4.11 Every element of $\mathcal{A}$ can be written as a finite disjoint union of maximal intervals (this means the interval cannot be extended within A):

$$A = \bigcup_{k=1}^{n} (a_k, b_k], \tag{3.6}$$

this is unique if we order them by $a_k < b_k < a_{k+1}$. Let us call this the canonical representation. If A is not bounded from above, $b_n = \infty$, by $(a_n, b_n]$ we mean $(a_n, b_n] \cap \mathbf{R}$.

Let $F : \mathbf{R} \to \mathbf{R}$ be an increasing and right continuous function. Then it has at most a countable number of discontinuities, all of which are of jump type (i.e. left and right limits exist). We define

$$F(\infty) = \lim_{x\to\infty} F(x), \qquad F(-\infty) = \lim_{x\to-\infty} F(x). \tag{3.7}$$

Definition 3.4.8 For $A = \bigcup_{j=1}^{n} (a_j, b_j] \in \mathcal{A}$, written in its canonical representation , we set

$$\varrho(A) = \sum_{j=1}^{n} (F(b_j) - F(a_j)). \tag{3.8}$$

We also set $\varrho(\phi) = 0$.

Remark 3.4.12 This definition is independent of the expression for A. Indeed, suppose that $A = (c, d]$ is written also as $A = \bigcup_{j=1}^{n} (a_j, b_j]$, then up to a re-ordering,

$$c = a_1 \le b_1 = a_2 \le b_2 = a_3 \le \cdots \le b_n = d.$$

This gives a telescopic sum,

$$\sum_{j=1}^{n} (F(b_j) - F(a_j)) = F(b_n) - F(a_1) = F(d) - F(c).$$

So when A is a single interval, there is no ambiguity in the definition of ϱ.

If $A = \sum_{k=1}^{p} (c_k, d_k]$ is written in its canonical representation, so each interval is a maximal interval, then there is partition $\{\Lambda_k, k = 1, \dots, p\}$ of $\{1, \dots, n\}$ such that

$$\bigcup_{j\in\Lambda_k} (a_j, b_j] = (c_k, d_k].$$

(recall that a partition of a set I is a collection of disjoint subsets of I whose union is I). In fact, it suffices to consider

$$\Lambda_k := \{j = 1, \dots, n : (a_j, b_j] \subset (c_k, d_k]\}.$$

Hence

$$\sum_{j=1}^{n} \varrho((a_j, b_j]) = \sum_{k=1}^{p} \sum_{j\in\Lambda_k} \varrho((a_j, b_j]) = \sum_{k=1}^{p} \varrho((c_k, b_k]).$$

Proposition 3.4.13 *If $F : \mathbf{R} \to \mathbf{R}$ is a non-decreasing and right continuous function, then ϱ defined by (3.8) is a pre-measure on $\mathcal{A}$.*

Proof We have $\varrho(\phi) = 0$, so it remains to prove σ-additivity. We start by showing that ϱ is additive. Let $A, B \in \mathcal{A}$ be disjoint. Then there exists $E_j \in \mathcal{E}$ such that

$$A = \bigcup_{j=1}^{n} E_j, \qquad B = \bigcup_{j=n+1}^{m} E_j.$$

Then

$$\varrho(A \cup B) = \sum_{j=1}^{m} \varrho(E_j) = \sum_{j=1}^{n} \varrho(E_j) + \sum_{j=n+1}^{m} \varrho(E_j) = \varrho(A) + \varrho(B).$$

Hence ϱ is indeed additive. By Lemma 3.4.8, we in particular deduce that ϱ is monotone and sub-additive.

Step 2. Let now $(A_j)_{j\geq 1}$ be a sequence of disjoint elements of $\mathcal{A}$ such that

$$A := \bigcup_{j\geq 1} A_j$$

is in $\mathcal{A}$. Using monotonicity followed by finite additivity,

$$\varrho(A) \geq \varrho(\bigcup_{j=1}^{n} A_j) = \sum_{j=1}^{n} \varrho(A_j).$$

Since this holds for every n,

$$\varrho(A) \geq \sum_{j=1}^{\infty} \varrho(A_j).$$

Step 3. We now prove the reverse inequality. Since $A \in \mathcal{A}$,

$$A = \bigcup_{k=1}^{n} (a_k, b_k]. \tag{3.9}$$

We show that $\varrho(A) \leq \sum_{j=1} \varrho(A_j)$. As argued before, in Remark 3.4.12, by additivity of ϱ we reduce this to the case A is one single half open interval.

Step 4. We first assume $A = (c, d]$ is bounded. We may also assumme that $c < d$, otherwise $\varrho((c, d]) = \varrho(\phi) = 0$ and the requested inequality is trivially satisfied. Since, for all $j \geq 1$, A_j is a finite union of bounded disjoint half open intervals, the whole collection of these intervals is countable and we write it $\{(a_k, b_k], k \geq 1\}$. In particular, we have

$$\bigcup_{j=1}^{\infty} A_j = \bigcup_{k=1}^{\infty} (a_k, b_k], \qquad \text{and} \qquad \sum_{j=1}^{\infty} \varrho(A_j) = \sum_{k=1}^{\infty} \varrho((a_k, b_k]). \tag{3.10}$$

The idea is to approximate $(c, d]$ by a compact interval $[c+\delta, d]$ and $(a_k, b_k]$ by a larger open interval $(a_k, b_k+\delta_k)$, we can then choose a finite cover and use finite additivity.

Let $\epsilon > 0$. By right-continuity of F, we may choose $\delta > 0$ such that $\delta < d - c$ and

$$F(c+\delta) - F(c) < \epsilon.$$

Similarly, for all $k \geq 1$, we choose a δ_k so that

$$F(b_k + \delta_k) - F(b_k) \leq 2^{-k}\epsilon.$$

The compact set $[c+\delta, d]$ is covered by the sets $(a_k, b_k]$ and therefore by the larger open sets $(a_k, b_k+\delta_k)$. We can choose a finite covering, so that for some N,

$$[c+\delta, d] \subset \cup_{k=1}^N (a_k, b_k + \delta_k).$$

In particular, we have the inclusion

$$(c+\delta, d] \subset \cup_{k=1}^N (a_k, b_k + \delta_k],$$

between subsets that are in $\mathcal{A}$. By monotonicity and sub-additivity of ϱ,

$$\begin{aligned}
\varrho((c,d]) = F(d) - F(c) = F(d) - F(c+\delta) + F(c+\delta) - F(c) \quad &\leq \varrho((c+\delta, d]) + \epsilon \\
&\leq \varrho\big(\cup_{k=1}^N (a_k, b_k+\delta_k]\big) + \epsilon \\
&\leq \sum_{k=1}^N \varrho((a_k, b_k]) + \sum_{k=1}^N \frac{\epsilon}{2^k} + \epsilon \\
&\leq \sum_{j=1}^\infty \varrho(A_j) + 2\epsilon,
\end{aligned}$$

where the last inequality follows from (3.10). Taking $\epsilon \to 0$, we conclude that

$$\varrho(A) = \varrho((c,d]) \leq \sum_j \varrho(A_j)$$

as requested. Together with step 2, we have proved

$$\varrho(A) = \sum_j \varrho(A_j)$$

for A a bounded half open interval.

Step 5. Now assume $A = (-\infty, b]$. Then, for every n with $-n < b$,

$$\varrho(A) = F(b) - F(-\infty) = \varrho((-n, b]) + F(-n) - F(-\infty).$$

Note that

$$(-n, b] = (-n, b] \cap \bigcup_{j=1}^{\infty} A_j = \bigcup_{j=1}^{\infty} (-n, b] \cap A_j.$$

Since $(-n, b] \cap A_j \in \mathcal{A}$ for all $j \geq 1$, we use step 4 to conclude that

$$\varrho((-n, b]) = \sum_{j=1}^{\infty} \varrho((-n, b] \cap A_j).$$

Hence, invoking monotonicity of ϱ, we obtain

$$\varrho(A) = \sum_{j=1}^{\infty} \varrho((-n, b] \cap A_j) + F(-n) - F(-\infty) \leq \sum_{j=1}^{\infty} \varrho(A_j) + F(-n) - F(-\infty).$$

Taking $n \to \infty$, since $F(-n) \to F(-\infty)$ we have proved the requested inequality. The same can be shown for $A = (a, \infty)$, $a \in \mathbf{R}$, or for $A = \mathbf{R} = (-\infty, +\infty]$. We have completed the proof. $\square$

As in (3.1), we define

$$\mu^*(E) = \inf \left\{ \sum_{j=1}^{\infty} \varrho(A_j) : E \subset \bigcup_{j=1}^{\infty} A_j, \quad A_j \in \mathcal{A} \right\}.$$

Since each $A \in \mathcal{A}$ is a finite disjoint union of members in $\mathcal{E}$, by re-arranging we have

$$\begin{aligned} \mu^*(E) &= \inf \left\{ \sum_{j=1}^{\infty} \varrho(E_j) : E \subset \bigcup_{j=1}^{\infty} E_j, \quad E_j \in \mathcal{E} \right\} \\ &= \inf \left\{ \sum_{j=1}^{\infty} \varrho((a_j, b_j]) : E \subset \bigcup_{j=1}^{\infty} (a_j, b_j] \right\}. \end{aligned}$$

Theorem 3.4.14 *There exists a measure μ_F on the Borel σ-algebra $\mathcal{B}(\mathbf{R})$ such that $\mu_F = \varrho$ on $\mathcal{A}$. It is the unique measure on $\mathcal{B}(\mathbf{R})$ that agrees with ϱ on $\mathcal{A}$.*

Proof Existence of μ_F follows from Caratheodory's Theorem. The uniqueness statement follows from Theorem 3.3.1. $\square$

Proposition 3.4.15 *μ_F extends to a measure (still denoted by μ_F) on the completion $\bar{\mathcal{B}}(\mathbf{R})$ of $\mathcal{B}(\mathbf{R})$.*

Proof Let $\mathcal{G}$ be the collection of $\mu*$ measurable subsets of $\mathbf{R}$, and $\mathcal{N}$ the collection of μ_F-negligible subsets. Since $\mathcal{G}$ is complete w.r.t. μ^*, and since μ^* coincides with μ_F on $\mathcal{B}(\mathbf{R})$, we deduce that $\mathcal{N} \subset \mathcal{G}$. But we also have $\mathcal{B}(\mathbf{R}) \subset \mathcal{G}$, and hence $\bar{\mathcal{B}}(\mathbf{R}) \subset \mathcal{G}$. So the restriction of μ^* to $\bar{\mathcal{B}}(\mathbf{R})$ provides the requested extension. $\square$

Remark 3.4.16 * Note that, by Proposition 3.4.2, such an extension is unique.

Remark 3.4.17 If $A = \cup_{j=1}^{\infty}(a_j, b_j]$ is a disjoint union, then $\mu_F(A) = \sum_{j=1}^{\infty} \varrho((a_j, b_j])$. This follows from that the union is a Borel set, and σ-additive property fo the measure. The subscript F will be omitted if there is no risk of confusion.

Definition 3.4.9

- The measure μ_F constructed in the theorem above is called Lebesgue-Stieltjes measure associated to F.
- If F is the identity map, this is called the Lebesgue measure on **R** and will be denoted by λ.
- The completion $\bar{\mathcal{B}}(\mathbf{R})$ of the Borel σ-algebra with respect to λ is called the Lebesgue σ-algebra, its elements are called Lebesgue-measurable sets.

Theorem 3.4.18 *If B is a μ^*-measurable set, then*

$$\begin{aligned}\mu_F(B) &= \inf\{\mu_F(O) : B \subset O,\ O \text{ is open}\}\\ &= \sup\{\mu_F(D) : D \subset B,\ D \text{ is closed}\}.\end{aligned}$$

This holds in particular if B is a Borel measurable set.

Proof This is left as exercise. □

Recall that $B\Delta A = (B \setminus A) \bigcup (A \setminus B)$.

Remark 3.4.19 If B is μ^*-measurable set with finite measure, then for every $\epsilon > 0$ there exists a set A which is a finite union of open intervals, such that $\mu_F(B\Delta A) < \epsilon$.

3.4.7 Example of a non-Lebesgue measurable set*

Proposition 3.4.20 $\bar{\mathcal{B}}(\mathbf{R}) \neq 2^{\mathbf{R}}$*, i.e. there exists a subset of* **R** *that is not Lebesgue measurable.*

Proof Let us define an equivalent relation: $x \sim y$ if and only if $x - y \in Q$. Using Axiom of choice we can choose the representative of the equivalent relations to be in $(0, 1]$. This quotient set $V = \mathbf{R}/\sim$ as a subset of $(0, 1]$ is called the Vitali set. If $q_1 \neq q_2$ are two rational numbers, the translates $V + q_1$ and $V + q_2$ are disjoint. Let $A = Q \cap (-1, 1]$, then

$$(0, 1] \subset \cup_{q\in A}(V + q) \subset [-1, 2]. \tag{3.11}$$

The second inequality follows from the inclusions $V \subset (0, 1]$ and $A \subset (-1, 1]$. To prove the first one, take $y \in (0, 1]$, then there exists $q \in (-1, 1] \cap Q = A$ such that $y = q + [y]$, where $[y]$ is the representative of y in $(0, 1]$. Thus $y \in \cup_{q\in A}(V + q)$ and the left hand side inclusion is proved. Let us assume by contradiction that V is Lebesgue measurable. What can its Lebesgue measure be? By

translation invariance, $\lambda(V+q)=\lambda(V)$ for all $q\in A$. Moreover, the sets $V+q$, $q\in A$, are pairwise disjoint. Therefore
$$\lambda(\cup_{q\in A}(V+Q))=\sum_{q\in A}\lambda(V),$$
which equals either 0 (if $\lambda(V)=0$) or $+\infty$ (if $\lambda(V)>0$). But in view of (3.11), we also have $1\le\lambda(\cup_{q\in A}(V+Q))\le 3$, which is a contradiction. Therefore V is not Lebesgue measurable. $\square$

Exercise 3.4.3 Show that there does not exist a translation invariant countably additive measure on $(\mathbf{R},2^{\mathbf{R}})$ which assigns an interval its length.

3.4.8 Examples

Example 3.6 Let $a,b\in\mathbf{R}$ with $a<b$. then, since $\mu_F(\{b\})\le\mu_F((b-1,b])=F(b)-F(b-1)<\infty$,
$$\mu_F((a,b))=\mu_F((a,b])-\mu_F(\{b\}).$$
Observe that $\mu_F(\{b\})\ne 0$ precisely when b is a discontinuity of F. Indeed, since $\mu_F((b-1,b])<\infty$, by continuity of μ_F from above,
$$\mu_F(\{b\})=\mu_F(\cap_{n=1}^{\infty}\{(b-\frac{1}{n},b]\})=\lim_{n\to\infty}\mu_F((b-\frac{1}{n},b])=F(b)-\lim_{n\to\infty}F(b-\frac{1}{n})=F(b)-F(b-).$$

Example 3.7 Since $F(x)=x$ is continuous, any countable subset of $\mathbf{R}$ has Lebesgue measure zero.

Example 3.8 Let $\alpha\in\mathbf{R}$ be fixed, and let
$$F(x)=\begin{cases}\alpha, & x<0,\\ \alpha+3, & x\ge 0.\end{cases}$$
Then $\mu_F=3\delta_0$. Indeed, for any interval $(a,b]$,
$$\mu_F((a,b])=\begin{cases}3, & \text{if } 0\in(a,b],\\ 0, & \text{if } 0\notin(a,b],\end{cases}$$
so that μ_F and $3\delta_0$ coincide on $\mathcal{E}$. To see that they are equal, note that $\mathcal{E}$ is a π-system generating $\mathcal{B}(\mathbf{R})$ and that $\mu_F(\mathbf{R})=3=3\delta_0(\mathbf{R})<\infty$.

Example 3.9 If Y is a random variable on a probability space $(\Omega,\mathcal{F},\mathbf{P})$, set $F(x)=\mathbf{P}(Y\le x)$. Then F is right continuous and increasing, it determines a Lebesgue-Stieltjes measure μ_F. Since $F(-\infty)=0$, Also,
$$\mu_F((a,b])=F(b)-F(a)=\mathbf{P}(a<Y\le b).$$

3.5 Lebesgue Measure on $\mathbf{R}^n$

By a similar construction, we obtain a Lebesgue measure on $\mathbf{R}^n$. We do not get into details.

Definition 3.5.1 There is a unique measure λ on $\mathcal{B}(\mathbf{R}^n)$ such that, for all $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbf{R}$ with $a_i \leq b_i$,

$$\lambda((a_1, b_1] \times \cdots \times (a_n, b_n]) = \Pi_{i=1}^n (b_i - a_i).$$

This is called the Lebesgue measure on $\mathbf{R}^n$, it extends to the completion of $\mathcal{B}(\mathbf{R}^n)$.

Remark 3.5.1 The Lebesgue measure is invariant under translations and rotations.

For $A \subset \mathbf{R}$ and $t \in \mathbf{R}$, let us denote by $t + A$ the shifted set

$$t + A = \{t + a,\, a \in A\}.$$

Theorem 3.5.2 *Any shift of a Borel set is a Borel set, i.e. if $A \in \mathcal{B}(\mathbf{R})$, then for all $t \in \mathbf{R}$, $t + A \in \mathcal{B}(\mathbf{R})$. Moreover, the Lebesgue measure is translation invariant on the Borel σ-algebra, i.e.*

$$\forall A \in \mathcal{B}(\mathbf{R}), \qquad \lambda(A) = \lambda(t + A).$$

Proof Let

$$\mathcal{C} = \{A \in \mathcal{B}(\mathbf{R}) : \forall t \in \mathbf{R},\, t + A \in \mathcal{B}(\mathbf{R})\}.$$

$\mathcal{C}$ is a σ-algebra. Moreover, if U is an open set, then for all $t \in \mathbf{R}$, $t + U$ is an open set as well, hence $\mathcal{C}$ contains all open sets. Therefore $\mathcal{C}$ contains the σ-algebra genereated by open sets, i.e. $\mathcal{C} = \mathcal{B}(\mathbf{R})$. Now we show the second point. Let

$$\mathcal{A} = \{A \in \mathcal{B}(\mathbf{R}) : \forall t \in \mathbf{R},\, \lambda(t + A) = \lambda(A)\}.$$

Then $\mathcal{A}$ contains the π-system of finite unions of disjoint half intervals. Now $\Omega \in \mathcal{A}$. If $A \subset B$ with $A, B \in \mathcal{A}$,

$$t + B \setminus A = (t + B) \setminus (t + A),$$

$t + A \subset t + B$, hence

$$\mu(t + B \setminus A) = \mu(t + B) - \mu(t + A) = \mu(B) - \mu(A) = \mu(B \setminus A).$$

Thus $B \setminus A \in \mathcal{A}$.

If $A_n \in \mathcal{A}$ is an increasing sequence, so is $t + A_n$, and $\bigcup_n (t + A_n) = t + \bigcup_n A_n$,

$$\mu(t + \bigcup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mu(t + A_n) = \lim_{n\to\infty} \mu(A_n) = \mu(\bigcup_n A_n).$$

Thus $\mathcal{A}$ is a λ system, concluding the proof. $\square$

Proposition 3.5.3 *Two non-trivial translation-invariant Borel measures on* **R** *which assign a finite measure to bounded intervals are constant multiples of each other.*

Proof Let μ and ν be translation invariant measures. Then, for all integers $p, q \geq 1$,

$$\mu((0,1]) = q\mu((0,\frac{1}{q}]), \qquad \mu([0,p/q]) = p\mu((0,\frac{1}{q}]) = (p/q)\mu((0,1]).$$

Similarly, $\nu([0,\frac{p}{q})) = (p/q)\,\nu((0,1])$. The quanitities $\nu((0,1])$ and $\nu((0,1])$ are non-zero, for otherwise the measures would be trivial. Set $c = \mu([0,1)]/\nu([0,1)$. Then $\mu([0,p/q]) = (p/q)\mu([0,1)] = c\nu([0,p/q))$. Hence, $\mu(I) = c\nu(I)$ for all $I \in \mathcal{C}$, where $\mathcal{C}$ denotes the collection of all half-open intervals with finite rational length. Now, $\mathcal{C}$ is a π-system and $\sigma(\mathcal{C}) = \mathcal{B}(\mathbf{R})$. Moreover, we have $\cup_{k\geq 1}(-k,k] = \mathbf{R}$, and μ and $c\nu$ assign the same finite mass to $(-k,k] \in \mathcal{C}$ for all $k \geq 1$. By Theorem 3.3.1, we deduce that $\mu = c\nu$. $\square$

Chapter 4

Measurable maps

4.1 Definition

Let $(\Omega, \mathcal{F})$ and $(\mathcal{X}, \mathcal{G})$ be two measurable spaces.

Definition 4.1.1 A map $f : \Omega \to \mathcal{X}$ is said to be measurable from $(\Omega, \mathcal{F})$ to $(\mathcal{X}, \mathcal{G})$ if, for all $A \in \mathcal{G}$, $f^{-1}(A) \in \mathcal{F}$.

When it is clear which σ-algebras are being considered, in the situation above, we will often just say that the map f is measurable.

Definition 4.1.2 Assume Ω and $\mathcal{X}$ are topological spaces endowed with their respective Borel σ-algebras $\mathcal{B}(\Omega)$ and $\mathcal{B}(\mathcal{C})$. We call a function $f : \Omega \to \mathcal{X}$ **(Borel) measurable** if it is measurable from $(\Omega, \mathcal{B}(\Omega))$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.

Definition 4.1.3 Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and $(\mathcal{X}, \mathcal{G})$ be a measurable space. If X is a measurable map from $(\Omega, \mathcal{F})$ to $(\mathcal{X}, \mathcal{G})$, we call X a random variable.

Exercise 4.1.1 Let $(\mathcal{X}, \mathcal{B})$ be a measurable space and Ω a set. A function $f : \Omega \to \mathcal{X}$ is always measurable when Ω is endowed with the pre-image σ-algebra $\sigma(f)$. Moreover, $\sigma(f)$ is the smallest σ-algebra on Ω such that $f : \Omega \to \mathcal{X}$ is measurable.

4.2 Examples

Example 4.1 If $\mathcal{X}$ is a discrete space, let the power set be its σ-algebra. Let $\mathcal{Y}$ be another space with a σ-algebra. Then any function $f : \mathcal{X} \to \mathcal{Y}$ is measurable.

If f is identically 1 on a set A and zero everywhere else, it is called the *indicator function* of A. We denote it by $\mathbf{1}_A$:

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

The indicator function $\mathbf{1}_A$ of a set is a measurable function if and only if A is a measurable set.

Example 4.2 If $f : E \to \mathbf{R}$ takes only a finite number of values $\{a_1, \ldots, a_k\}$. Then, f is measurable if and only if, for all i, $\{f = a_i\}$ is measurable. Also, f can be written in the following form: $f = \sum_{i=1}^k a_i \mathbf{1}_{A_j}$. (Just take $A_j = \{x : f(x) = a_j\}$).

4.3 Properties of measurable functions

Proposition 4.3.1 *If $f : (\mathcal{X}_1, \mathcal{F}_1) \to (\mathcal{X}_2, \mathcal{F}_2)$ and $g : (\mathcal{X}_2, \mathcal{F}_2) \to (\mathcal{X}_3, \mathcal{F}_3)$ are measurable functions then so $f \circ g : \mathcal{X}_1 \to \mathcal{X}_3$ is measurable.*

Proof Let $A \in \mathcal{F}_3$ then $f^{-1}(A) \in \mathcal{F}_2$. Furthermore,

$$(f \circ g)^{-1}(A) = \{x \in \mathcal{X}_1 : f \circ g \in A\} = \{x \in \mathcal{X}_1 : g \in f^{-1}(A)\} \in \mathcal{F}_1$$

using g is measurable. □

Proposition 4.3.2 *Let $(\Omega, \mathcal{F})$ and $(\mathcal{X}, \mathcal{B})$ be measurable spaces. Suppose that $\mathcal{B}$ is generated by $\mathcal{C}$ and $f : \Omega \to \mathcal{X}$. If $f^{-1}(A) \in \mathcal{F}$ whenever $A \in \mathcal{C}$, then f is measurable.*

Proof $\mathcal{B}' = \{A \in \mathcal{B} : f^{-1}(A) \in \mathcal{F}\} \supset \mathcal{C}$ is a σ-algebra, *check with Lemma 2.1.1*. It must therefore agree with $\mathcal{B} = \sigma(\mathcal{C})$. □

Here are some important consequences of the above proposition.

Proposition 4.3.3 *If X, Y are metric spaces and $f : X \to Y$ is a continuous map, then f is Borel measurable.*

Proof Continuity means that the pre-image of an open set is an open set. Hence, $f^{-1}(U) \in \mathcal{B}(X)$ for all open subset U of Y. Since $\mathcal{B}(Y)$ is generated by open sets, we deduce that f is Borel measurable. □

Proposition 4.3.4 *Let$f_1 : (\mathcal{X}, \mathcal{F}) \to (\mathcal{X}_1, \mathcal{F}_1)$ and $f_2 : (\mathcal{X}, \mathcal{F}) \to (\mathcal{X}_2, \mathcal{F}_2)$ be measurable functions. Then*

$$\psi = (f_1, f_2) : \mathcal{X} \to \mathcal{X}_1 \times \mathcal{X}_2$$

is measurable, when the target space is endowed by the product σ-algebra $\mathcal{F}_1 \otimes \mathcal{F}_2$.

Proof Apply Proposition 4.3.2, it is sufficient to know that $h^{-1}(A \times B)$, where $A \in \mathcal{F}_1$ and $\mathcal{B} \in \mathcal{F}_2$, is measurable. But,

$$\psi^{-1}(A \times B) = f_1^{-1}(A) \cap f_2^{-1}(B) \in \mathcal{F},$$

completing the proof. □

4.4 Measurability with respect to the σ-algebra generated by a map

An enhanced version of the composition rule for $f \circ g$ can be proved in the case where $\mathcal{A}_2$, the σ-algebra on the second set $\mathcal{X}_2$, corresponds to the σ-algebra $\sigma(f)$ generated by f.

Let $(\mathcal{X}_1, \mathcal{A}_1)$ and $(\mathcal{X}_3, \mathcal{A}_3)$ be measurable spaces, and $\mathcal{X}_2$ be a set. Let $f : \mathcal{X}_2 \to \mathcal{X}_3$ be a map, and let $\sigma(f)$ be the σ-algebra on $\mathcal{X}_2$ generated by f.

Proposition 4.4.1 *A map $g : \mathcal{X}_1 \to \mathcal{X}_2$ is measurable from $(\mathcal{X}_1, \mathcal{A}_1)$ to $(\mathcal{X}_2, \sigma(f))$ if and only if $f \circ g$ is measurable from $(\mathcal{X}_1, \mathcal{A}_1)$ to $(\mathcal{X}_3, \mathcal{A}_3)$.*

Proof Since f is measurable from $(\mathcal{X}_2, \sigma(f))$ to $(\mathcal{X}, \mathcal{A}_3)$, by the composition rule, if g is measurable from $(\mathcal{X}_1, \mathcal{A}_1)$ to $(\mathcal{X}_2, \sigma(f))$, then $f \circ g$ is measurable from $(\mathcal{X}_1, \mathcal{A}_1)$ to $(\mathcal{X}_3, \mathcal{A}_3)$. Conversely, assume that $f \circ g$ is measurable from $(\mathcal{X}_1, \mathcal{A}_1)$ to $(\mathcal{X}_3, \mathcal{A}_3)$, then for any $C \in \mathcal{A}_3$, we have

$$g^{-1}(f^{-1}(C)) = (f \circ g)^{-1}(C) \in \mathcal{A}_1.$$

Since $\sigma(f)$ consists of all subsets of the form $f^{-1}(C)$, for $C \in \mathcal{X}_3$, we deduce that g is measurable from $(\mathcal{X}_1, \mathcal{A}_1)$ to $(\mathcal{X}_2, \sigma(f))$. □

4.5 Measurable maps with values in a subset

Let $(\Omega, \mathcal{F})$ and $(\mathcal{X}, \mathcal{G})$ be measurable spaces, and let $\mathcal{Y} \subset \mathcal{X}$. Let us denote by $i : \mathcal{Y} \to \mathcal{X}$ the inclusion map. Recall that $\mathcal{Y}$ can be endowed with the trace $\mathcal{H}$ of the σ-algebra $\mathcal{G}$, see Example2.4 above. Let $f : \Omega \to \mathcal{Y}$ be a map, one can then also consider $i \circ f$, which is a map from Ω to $\mathcal{X}$, therefore there are two possible definitions for the measurability of f. The following Proposition states that these coincide.

Proposition 4.5.1 *Let $f : \Omega \to \mathcal{Y}$. Then f is measurable from $(\Omega, \mathcal{F})$ to $(\mathcal{Y}, \mathcal{H})$ if and only if $i \circ f$ is measurable from $(\Omega, \mathcal{F})$ to $(\mathcal{X}, \mathcal{G})$.*

Proof Note that

$$\sigma(i) = \{i^{-1}(A), A \in \mathcal{G}\} = \{A \cap \mathcal{Y}, A \in \mathcal{G}\} = \mathcal{H}.$$

Hence the result follows from Proposition 4.4.1 above. □

In particular, one obtains the following consequence for the case when $\mathcal{X}$ is a metric space.

Corollary 4.5.2 *Let $\mathcal{X}$ be a metric space and let $\mathcal{Y}$ be a subset of $\mathcal{X}$. Then $f : \Omega \to \mathcal{Y}$ is measurable from $(\Omega, \mathcal{F})$ to $(\mathcal{Y}, \mathcal{B}(\mathcal{Y}))$ if and only if it is measurable from $(\Omega, \mathcal{F})$ to $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$.*

Proof This follows from the above Proposition, noting that $\mathcal{B}(\mathcal{Y})$ is the trace of $\mathcal{B}(\mathcal{X})$ on $\mathcal{Y}$, see Proposition 2.3.2. □

4.6 Real-valued measurable functions

The Borel σ-algebra of **R** is generated by the set of open intervals, equally they are generated by the set of all closed intervals. We state some of these below, they follow straightforwardly from Proposition 4.3.2.

Proposition 4.6.1 *Let $(X, \mathcal{B})$ be a measurable space. A function $f : \mathcal{X} \to$ **R** is Borel measurable if one of the following conditions hold:*

1. *for each real number a, $\{x : f(x) > a\}$ is a measurable set;*
2. *for each real number a, $\{x : f(x) < a\}$ is a measurable set;*
3. *for each real number a, $\{x : f(x) \geq a\}$ is a measurable set;*
4. *for each real number a, $\{x : f(x) \leq a\}$ is a measurable set;*

For the first just note that sets of the form (a, ∞) generates the Borel σ-algebra, apply Proposition 4.3.2 to. conclude. Similarly the other statements chan be shown.

Proposition 4.6.2 *Let $(X, \mathcal{F})$ be a measurable space. If $f, g : X \to$ **R** are measurable functions, then so are the following functions.*

1. $f + g$
2. fg,
3. $\max(f, g)$
4. $f^+ = \max(f, 0)$, $f^- = \max(-f, 0)$
5. $|f| = f^+ + f^-$.

Proof Firstly $(f, g) : X \to \mathbf{R}^2$ is Borel measurable. Set $\psi : X \to \mathbf{R}^2$ by

$$\psi(x) = (f(x), g(x))$$

and define $\varphi : \mathbf{R}^2 \to \mathbf{R}$ by $\varphi(x, y) = x + y$, a continuous function. Then $f + g = \varphi \circ \psi$ is a composition of measurable function, so measurable. Choosing $\varphi(x, y) = xy$, for the analogous proof for fg. Now for any $a > 0$, $\{\max(f, g) < a\} = \{f < a\} \cap \{g < a\} \in \mathcal{F}$, so $\max(f, g)$ is measurable. Part 4 follows from part 3, and 5 follows part 1 and part 4. $\square$

Remark 4.6.3 Note the relation $\max(f, g) = \frac{(f+g)+|f-g|}{2}$.

Definition 4.6.1 Let $\overline{\mathbf{R}} = \mathbf{R} \cup \{-\infty, +\infty\}$. One can endow $\overline{\mathbf{R}}$ with the distance

$$d(x, y) = |F(x) - F(y)|, \qquad x, y \in \overline{\mathbf{R}},$$

with $F(x) = \arctan(x), x \in \mathbf{R}$, and $F(\pm\infty) = \lim_{u \to \pm\infty} \arctan(u) = \pm\pi/2$. We then denote by $\mathcal{B}(\overline{\mathbf{R}})$ the associated Borel σ-algebra.

Exercise 4.6.1

1. Show that open subsets of $\overline{\mathbf{R}}$ are of the form i) U, where U is an open subset of $\mathbf{R}$, ii) $[-\infty, a)$ where $a \in \mathbf{R}$, iii) $(b, +\infty]$, where $b \in \mathbf{R}$, or any union of i), ii) and iii).

2. Using Remark 2.3.3, deduce that $\mathcal{B}(\mathbf{R}) \subset \mathcal{B}(\overline{\mathbf{R}})$ and that $\mathcal{B}([0, +\infty]) \subset \mathcal{B}(\overline{\mathbf{R}})$. Show also that $\mathcal{B}(\overline{\mathbf{R}})$ is generated by the collection of intervals $(t, +\infty] \subset \overline{\mathbf{R}}$, for $t \in \mathbf{R}$.

3. Show that f is measurable from $(\mathcal{X}, \mathcal{F})$ to $(\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}}))$ iff $\{f = \omega\} \in \mathcal{F}$ for $\omega \in \{-\infty, +\infty\}$ and $f^{-1}(A) \in \mathcal{F}$ for any $A \in \mathcal{B}(\mathbf{R})$.

Remark 4.6.4 By Proposition 4.5.1, if $(\mathcal{X}, \mathcal{A})$ is a measurable space, then a function $f : \mathcal{X} \to \mathbf{R}$ is measurable from $(\mathcal{X}, \mathcal{A})$ to $(\mathbb{R}, \mathcal{B}(\mathbf{R}))$ if and only if $i \circ f$ is measurable from $(\mathcal{X}, \mathcal{A})$ to $(\overline{\mathbf{R}}, \mathcal{B}(\overline{\mathbf{R}})$, where $i : \mathbf{R} \to \overline{\mathbf{R}}$ is the inclusion map.

Proposition 4.6.5 *Let $(X, \mathcal{B})$ be a measurable space. If $f_n : X \to \overline{\mathbf{R}}$ are measurable functions, so are the following:*

1. $\sup_{\{n \geq 1\}} f_n$,

2. $\inf_{\{n \geq 1\}} f_n$;

3. $\limsup f_n$,

4. $\liminf f_n$.

Proof Firstly

$$\{\sup_n f_n > a\} = \cup_n \{f_n \geq a\},$$

belongs to $\mathcal{F}$, as each $\{f_n \geq a\}$ does (the above is $\{x : \inf_{n\geq 1} f_n(x) < a\} = \cup_n \{x : f_n(x) < a\}$), so $\sup_{\{n\geq 1\}} f_n$ is measurable. Also,

$$\{\inf_{n\geq 1} f_n < a\} = \cap_n \{f_n < a\} \in \mathcal{F},$$

showing that $\inf_{\{n\geq 1\}} f_n$ is measurable. Since

$$\limsup_{n\to\infty} f_n = \inf_n \sup_{k\geq n} f_k, \qquad \liminf f_n. = \sup_n \sup_{k\geq n} f_k,$$

they are both measurable. □

Proposition 4.6.6 *If f_n is a sequence of Borel measurable functions with $f = \lim_{n\to\infty} f_n(x)$ exists at every x, then f is a Borel measurable function.*

Example 4.3 If $(\mathcal{X}, \mathcal{F})$ and $(\mathcal{Y}, \mathcal{G})$ are measurable spaces. Suppose that μ is a measure. Let $f : \mathcal{X} \to \mathcal{Y}$ and $g : \mathcal{X} \to \mathcal{Y}$. Let $A = \{f \neq g\}$.

1. Suppose both f and g are measurable, then $f - g$ is a measurable function, consequently, $A = \{f - g \neq 0\}$ is a measurable set.
2. Suppose that $\mathcal{F}$ is complete with respect to μ. Suppose A is contained in a null measurable set. Since $\mathcal{F}$ is complete, then A is measurable and $\mu(A) = 0$. Then f is measurable implies that g is measurable. Indeed, for any $B \in \mathcal{G}$,

$$\{g \in B\} = (\{f \in B\} \cap A^c) \bigcup (\{g \in B\} \cap A) \in \mathcal{F}.$$

   We use that $(\{g \in B\} \cap A)$ is contained in a null set and therefore measurable, A^c and $\{f \in B\}$ are measurable.

Definition 4.6.2 Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space and $(\mathcal{Y}, \mathcal{G})$ be a measurable space. Let f, g two measurable maps from $(\mathcal{X}, \mathcal{F})$ to $(\mathcal{Y}, \mathcal{G})$. If $f = g$ on a set of full measure, i.e. $\mu(\{f \neq g\}) = 0$, we say that $f = g$ almost-everywhere. This is abbreviated by $f = g$ a.e.

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $(\mathcal{Y}, \mathcal{G})$ a measurable space, and X, Y two random variables from $(\Omega, \mathcal{F})$ to $(\mathcal{Y}, \mathcal{G})$. We say that $X = Y$ almost-surely if $X = Y$ on an event of full probaility, i.e. $\mathbf{P}(\{X \neq Y\}) = 0$. This is abbreviated by $X = Y$ a.s.

Example 4.4 (This is not intended to be delivered in class) If f_n is a sequence of Borel measurable functions. Set

$$f(x) = \begin{cases} \lim_{n\to\infty} f_n(x), & \text{if the limit exists at } x, \\ 0, & \text{otherwise} \end{cases}$$

Then f is measurable.

Proof Set

$$U := \{x : \lim_{n\to\infty} f_n(x) \text{ exists}\} = \{x : \limsup_{n\to\infty} f_n(x) = \liminf_{n\to\infty} f_n(x)\}$$

Since $\limsup_{n\to\infty} f_n(x)$ and $\liminf_{n\to\infty} f_n(x)$ are both measurable, U is a measurable set. Define

$$Z := (\liminf f_n, \quad \limsup f_n).$$

(Let $\Delta = \{(b,b) : b \in \mathbf{R}\}$ denote the the diagonal subset of $\mathbf{R}^2$, then $U = Z^{-1}(\Delta)$.)

Note that $f(x) = 0$ on U^c. Then for any $a < 0$,

$$\{x : \lim_{n\to\infty} f_n(x) < a\} = \{x : \limsup_{n\to\infty} f_n(x) < a\},$$

which is measurable since $\limsup_{n\to\infty} f_n$ is measurable. Let $a \geq 0$, then

$$\{x : \lim_{n\to\infty} f_n(x) < a\} = \{x : \limsup_{n\to\infty} f_n(x) < a\} \cup U,$$

which is also a measurable set. □

In analysis we frequently encounter Lebesgue measurable sets and functions, this refers to the σ-algebra

4.7 Simple functions

A simple functions is a real-valued measurable function taking on a finite number of distinct values. It has representation as linear combinations of indicator functions of measurable sets. Representations are not necessarily unique. For example take the domain space to be $\mathbf{R}$, let $A_1 = [1,2]$, $A_2 = (2,6]$, and $A_3 = (-\infty,1) \cup (4,\infty)$. Then, $f = 2\mathbf{1}_{A_1} + 3\mathbf{1}_{A_2} + 0\mathbf{1}_{A_3}$ is a simple function. We often do not include the last term with zero value, but sometimes it is included for the convenience that $A_1 \cup A_2 \cup A_3 = \mathbf{R}$. The representation is not unique,

$$f = 2\mathbf{1}_{A_1\cap Q} + 3\mathbf{1}_{A_2\cap Q} + 2\mathbf{1}_{A_1\cap Q^c} + 3\mathbf{1}_{A_2\cap Q^c},$$

by we also have,

$$f = 2\mathbf{1}_{([1,3]} + \mathbf{1}_{(2,3]} + 3\mathbf{1}_{(3,6]}.$$

Let us now give one popular definition for simple functions.

Definition 4.7.1 Let $(\mathcal{X}, \mathcal{F})$ be a measurable space. A function of the form $f(x) = \sum_{i=1}^N a_i \mathbf{1}_{A_i}(x)$ where $a_i \in \mathbf{R}$ and $A_i \in \mathcal{F}_i$ is said to be a simple function.

Proposition 4.7.1 *If A_i are measurable sets, then $f = \sum_{i=1}^\infty a_i \mathbf{1}_{A_i}$ is measurable.*

Proof If B is a Borel set,

$$f^{-1}(A) = \cup_{\{i: a_i \in B\}} A_i.$$

This is a countable union and belongs to $\mathcal{F}$ if every $A_i \in \mathcal{F}$. □

Remark 4.7.2 Suppose that $f = \sum_{i=1}^{\infty} a_i \mathbf{1}_{A_i}$ is measurable. If the numbers a_i are **distinct**, and $\{A_i\}$ are **disjoint sets**, then A_i is necessarily measurable. Indeed, $f^{-1}(\{a_i\}) = A_i$. So for f to be measurable, A_i must belong to $\mathcal{F}$.

If $a_1 = a_2$, we can conclude $f^{-1}(\{a_1\}) = A_1 \cup A_2$ is measurable, but we may not have any further information to conclude the measurability of the sets A_1 and A_2.

Definition 4.7.2 Suppose that a simple function f takes the following values $\{a_i, i = 1, \dots, N\}$. Then

$$f(x) = \sum_{i=1}^{n} a_i \mathbf{1}_{f^{-1}(a_i)}$$

is said to be the canonical representation for f. Set $A_i = f^{-1}(a_i)$, then A_i are pairwise **disjoint measurable sets** and $\bigcup_{i=1}^{N} A_i = \mathcal{X}$.

Exercise 4.7.1 Write the following function in the form $\sum_{i=1}^{\infty} a_i \mathbf{1}_{A_i}$ where $\{A_i\}$ are pairwise disjoint.

$$f(x) = \begin{cases} 2, & x \in A_1 = (-2\pi, -\pi] \\ 3, & x \in A_2 = [1, 4) \cup (7, 8) \\ \frac{1}{2}, & x \in A_3 = (4, 6] \\ 5, & x \in A_4 = (9, \infty) \cap Q \\ 0, & \text{otherwise} \end{cases}$$

Solution: The canonical representation is

$$f(x) = 2\mathbf{1}_{A_1} + 3\mathbf{1}_{A_2} + \frac{1}{2}\mathbf{1}_{A_3} + 5\mathbf{1}_{A_4}.$$

We see that $\{A_1, A_2, A_3, A_4, (\cup_{i=1}^{4} A_i)^c\}$ is a partition of $\mathcal{X}$.

Example 4.5 A step function $f : \mathbf{R} \to \mathbf{R}$ is a piecewise constant function. Steps functions with a finite number of steps are simple functions.

Theorem 4.7.3 *Let $(\mathcal{X}, \mathcal{G})$ be a measurable space. Then,*

- *Every measurable function $f : \mathcal{X} \to \mathbf{R}$ is the pointwise limit of simple functions:*

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

  where $f_n \in \mathcal{E}(\mathcal{G})$. Furthermore we can arrange so that $|f_n| \le |f|$.

  (If f is bounded, the convergence is uniform.)

- *If $f \geq 0$ we may choose f_n to be non-negative and the sequence $(f_n)_{n\geq 1}$ non- decreasing so $f = \sup_{\{n\geq 1\}} f_n$.*

Proof *Step 1.* We prove the second claim first. Let us consider dyadic partitions of the interval $[0, n]$. Let us cut each length 1 interval into 2^n pieces of equal length (this is called a dyadic partition). The dyadic partitions have an advantage: any partition points from current level contains all partition points from the previous levels.

Let us define the measurable sets:

$$A_j^n = \left\{ x : \frac{j}{2^n} < f(x) \leq \frac{j+1}{2^n} \right\}, \qquad 0 \leq j \leq n(2^n - 1).$$

We approximate f with the value $\frac{j}{2^n}$ on A_j^n and define

$$g_n = \sum_{j=0}^{n(2^n-1)} \frac{j}{2^n} \mathbf{1}_{A_j^n}.$$

Then on $\{x : 0 \leq f(x) < n\}$, $|f_n(x) - f(x)| \leq \frac{1}{2^n}$, so

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

Also f_n increases with n and $f_n(x) \leq f(x)$ for every x.

Step 2. For the first claim, we write $f = f^+ - f^-$ where

$$f^+ = \max(f, 0) \qquad f^- = \max(-f, 0).$$

Let $(f^+)_n$ be an non-decreasing sequence of simple functions with limit f^+. Similarly, Let $(f^-)_n$ be an non-decreasing sequence of simple functions with limit f^-. Set

$$g_n = (f^+)_n - (f^-)_n.$$

Then

$$|g_n(x)| = (f^+)_n(x) + (f^-)_n(x) \leq f^+(x) + f^-(x) = |f(x)|.$$

For every $x \in f^{-1}([-n, n])$,

$$\begin{aligned} |f(x) - g_n(x)| &= |(f(x) - (f^+)_n + (f^-)_n)(x)| \\ &\leq |f^+(x) - (f^+)_n(x)| + |f^-(x) - (f^-)_n(x)| \\ &\leq \frac{2}{2^n}. \end{aligned}$$

This means that for every x, $\lim_{n\to\infty} g_n(x) \to f(x)$. If $|f| \leq M \leq n_0$ is bounded, then for $n \geq n_0$,

$$|f(x) - g_n(x)| \leq \frac{2}{2^n}$$

for all $x \in \mathcal{X}$, showing the uniform convergence. $\square$

4.8 Factorisation Lemma

Lemma 4.8.1 (The factorisation Lemma) *Let $(\mathcal{X}, \mathcal{G})$ be a measurable space and Ω a set. Let $Y : \Omega \to \mathcal{X}$ be a measurable function and consider the measurable space $(\Omega, \sigma(Y))$. Then $X : \Omega \to \mathbf{R}$ is $\sigma(Y)$-measurable if and only if there exists a measurable function $f : \mathcal{X} \to \mathbf{R}$ such that $X = f \circ Y$. Moreover, if X is non-negative, f may be chosen to be non-negative.*

Proof Recall that $\sigma(Y) = \{Y^{-1}(B) : B \in \mathcal{G})$. Suppose that $f : \mathcal{X} \to \mathbf{R}$ is measurable, since Y is measurable with respect to $\sigma(Y)/\mathcal{G}$, then the composition of the measurable functions $f \circ Y$ is measurable.

For the converse, suppose that X is $\sigma(Y)$-measurable. We first assume $X = \mathbf{1}_A$ where $A \in \sigma(Y)$. Then $A = \{Y^{-1}(B)\}$ for some $B \in \mathcal{G}$ and

$$\mathbf{1}_B \circ Y = \mathbf{1}_{\{\omega : Y(\omega) \in B\}} = X.$$

So X is the composition of the measurable function $\mathbf{1}_B$ with Y.

Suppose that X takes only a countable number of values (a_n) and write $A_n = X^{-1}(\{a_n\})$. Since X is $\sigma(Y)$-measurable, there exist sets $B_n \in \mathcal{G}$ such that $Y^{-1}(B_n) = A_n$. Define now the sets

$$C_n = B_n \setminus \bigcup_{p<n} B_p.$$

These sets are disjoint and one has again $Y^{-1}(C_n) = A_n$. Setting $f(x) = a_n$ for $x \in C_n$ and $f(x) = 0$ for $x \in \mathcal{X} \setminus \bigcup_n C_n$, we see that f has the required property.

Suppose that $X \geq 0$, we may approximate it with an increasing sequence of simple functions X_n. By the paragraph above, there exists $g_n : \mathcal{X} \to \mathbf{R}_+$ simple functions such that $X_n = g_n(Y)$. In fact, g_n is an increasing sequence. Let $g = \sup_n g_n \geq 0$. Since g is a pointwise limit of measurable functions, g is also measurable. We note $g(Y(\omega)) = \sup_n g_n(Y(\omega)) = \sup_n X_n(\omega)) = X(\omega)$. Finally, in the general case where X takes values in $\mathbf{R}$ we complete the proof by exploiting the decomposition $X = X^+ - X^-$.

□

4.9 Pushed-forward measure

Definition 4.9.1 Let $(\mathcal{X}_1, \mathcal{B}_1)$ and $(\mathcal{X}_2, \mathcal{B}_2)$ be measurable spaces, and μ_1 a measure on $(\mathcal{X}_1, \mathcal{B}_1)$. Let $f : \mathcal{X}_1 \to \mathcal{X}_2$ be a measurable function. We define a measure μ_2 on $\mathcal{B}_2$ by setting

$$\mu_2(A) = \mu_1(f^{-1}(A)) = \mu_1(\{x : f(x) \in A\}).$$

This is called the **pushed-forward measure** and denoted by $f_*(\mu_1)$ or sometimes as $\mu \circ^{-1}$.

Exercise 4.9.1 Show that $f_*(\mu_1)$ is a measure.

Definition 4.9.2 If $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space and $X : \Omega \to \mathbf{R}$ is a random variable, $X_*\mathbf{P}$ is called the probability distribution of the random variable.

Example 4.6 If $X : \Omega \to \{0, 1\}$, then $X_*\mathbf{P}(\{0\}) = \mathbf{P}(X = 0)$ and $X_*\mathbf{P} = \mathbf{P}(X = 0)\delta_0 + \mathbf{P}(X = 1)\delta_1$.

4.10 Appendix*

4.10.1 Littlewood's three principles

Every Borel measurable set on $\mathbf{R}$ is nearly a finite union of intervals. Every (measurable) function is nearly continuous, every convergent sequence of measurable functions is nearly uniform convergent.

4.10.2 Product σ-algebras

Let I be an arbitrary index and let $(E_\alpha, \mathcal{F}_\alpha), \alpha \in I$ be a family of measurable spaces. The tensor σ-algebra, also called the product σ-algebra, on $E = \Pi_{\alpha \in I} E_\alpha$ is defined to be

$$\otimes_{\alpha \in I} \mathcal{F}_\alpha = \sigma\{\pi_\alpha^{-1}(A_\alpha) : A_\alpha \in \mathcal{F}_\alpha, \alpha \in I\}.$$

Here $\pi_\alpha : E = \Pi_{\alpha \in I} E_\alpha \to E_\alpha$ is the projection (also called the coordinate map). If $x = (x_\alpha, \alpha \in I)$ is an element of E, $\pi_\alpha(x) = x_\alpha$. The tensor σ-algebra is the smallest one such that for all $\alpha \in I$, the mapping

$$\pi_\alpha : (E, \otimes_{\alpha \in I} \mathcal{F}_\alpha) \to (E_\alpha, \mathcal{F}_\alpha)$$

is measurable.

Let $\pi_i : \mathbf{R}^n \to \mathbf{R}$ be the projection to the ith component. The tensor σ-algebra $\otimes_n \mathcal{B}(\mathbf{R})$ is

$$\otimes_n \mathcal{B}(\mathbf{R}) = \sigma\{\pi_i^{-1}(A) : A \in \mathcal{B}(\mathbf{R})\}.$$

For example $\pi_1^{-1}(A) = A \times \mathbf{R} \times \cdots \times \mathbf{R}$.

4.10.3 The Borel σ-algebra on the Wiener Space

The set of all maps from $[0, 1]$ to $\mathbf{R}^d$ can be denoted as the product space $(\mathbf{R}^d)^{[0,1]}$. The tensor σ-algebra $\otimes_{[0,1]} \mathcal{B}(\mathbf{R}^d)$ is the smallest space such that the projections π_t where $\pi_t(\sigma) = \sigma(t)$ is measurable. The product topology is the smallest topology such that the projections are continuous. Hence the Borel σ-algebra of the product topology is larger than the tensor σ-algebra.

Let $W^d = C([0,T], \mathbf{R}^d)$ be the space of continuous functions from $[0,1]$ to $\mathbf{R}^n$. It is a Banach space with the supremum norm:

$$\|g\|_W := \sup_{t\in[0,1]} \|g(t)\|.$$

Any measurable set in the tensor σ-algebra is determined by the a countable number of projections. The action to determine whether a function is continuous or not cannot be determined by a countable number of operations. Hence W^d is not a measurable set in the tensor σ-algebra. (once we know a function is continuous, the function can be determined by its values on a countable dense set.)

Example 4.7 A sample continuous real valued stochastic process $(X_t, 0 \le t \le 1)$ on a probability space $(\Omega, \mathcal{F}, P)$ can be considered as a function, denoted by X, from $(\Omega, \mathcal{F})$ to W:

$$X(\omega)(t) = X_t(\omega).$$

For each time t, and $a \in \mathbf{R}^n$, $\{\omega : |X_t(\omega) - a| < \epsilon\}$ belongs to $\mathcal{F}$ as X_t is measurable.

We show that the function $X : \Omega \to W$ is Borel measurable. Take $f \in W$ and the open ball in the Banach space W centred at f with radius $\varepsilon > 0$. Its pre-image by X is:

$$\begin{aligned}\{\omega : \sup_{0\le t\le 1} |X_t(\omega) - f_{t_i}| < \epsilon\} &= \{\omega : \sup_{0\le t_i\le 1, t_i\in Q} |X_{t_i}(\omega) - f_{t_i}| < \epsilon\} \\ &= \cap_{0\le t_i\le 1, t_i\in Q}\{\omega : |X_{t_i}(\omega) - f_{t_i}| < \epsilon\}.\end{aligned}$$

Since for each i, $\{\omega : |X_{t_i}(\omega) - f_{t_i}| < \epsilon\} \in \mathcal{F}$, the intersection of the countable number of such sets belongs to $\mathcal{F}$.

Chapter 5

Integration

5.1 Integration

Throughout this chapter $(\mathcal{X}, \mathcal{F}, \mu)$ is a measure space with a σ-finite measure μ. (Previously I assume the measure is complete, this assumption is not needed for defining integration.)

By the integral of a Borel measurable function $f : \mathcal{X} \to \mathbf{R}$, we mean a value assigned to this function. Let us call this value $I(f)$. So $I : \mathcal{H} \to \mathbf{R}$ is a function on a set of functions $\mathcal{H}$. For I to qualify to be an integral it should have some properties which we describe below.

Property 1 (IP) Let $\mathcal{H}$ be a class of Borel measurable functions $f : \mathcal{X} \to \mathbf{R}$. Suppose that to every $f \in \mathcal{H}$, we assign a number or ∞ which we denote by $I(f)$. We say I satisfies the IP if the following holds:

1. (Linearity) If $f, g \in \mathcal{H}$ and $a, b \in \mathbf{R}$, we have

$$I(af + bg) = aI(f) + bI(g).$$

2. (Monotonicity) If $f \leq g$ then $I(f) \leq I(g)$

Notation. We use $\int f$ for the integral of f. To emphasize the measure used we also use $\int f d\mu$ for the same integral. Sometime we emphasize the domain of integration as follows $\int_{\mathcal{X}} f d\mu$. Sometimes we explicitly express the argument of integration by a dummy variable: $\int f(x)\mu(dx)$, this is also written as $\int f(x)d\mu(x)$. Also if A is a measurable set, we set $\int_A f = \int f\mathbf{1}_A$ etc.

5.2 Outline of the construction

The procedure for defining the integral of a function $f : \mathcal{X} \to \mathbf{R}$ w.r.t. a measure μ is as follows.

1. If $f = \sum_{i=1}^N a_i \mathbf{1}_{A_i}$ is a simple function (the assumption that A_i are measurable is included in the definition), we define
$$\int_{\mathcal{X}} f \, d\mu = \sum_{i=1}^N a_i \, \mu(A_i).$$
2. If f is a positive function we define
$$\int_{\mathcal{X}} f \, d\mu = \sup_g \int g \, d\mu$$
where the supremum is taken over simple functions g with $0 \le g \le f$. We say f is integrable if $\int f \, d\mu < \infty$.
3. Let $f : \mathcal{X} \to \mathbf{R}$ be a measurable function. Let $f^+ = \max(f(x), 0)$ and $f^-(x) = \max(-f(x), 0)$ be the positive and negative parts of f, then $f = f^+ - f^-$. If both f^+ and f^- have finite integrals we say f is integrable and define
$$\int_{\mathcal{X}} f \, d\mu = \int_{\mathcal{X}} f^+ \, d\mu - \int_{\mathcal{X}} f^- \, d\mu.$$
The set of integrable functions is denoted by L_1. Observe that in this case
$$\int |f| d\mu = \int f^+ d\mu + \int f^- d\mu.$$

5.3 Integral of simple functions

Let us denote by $\mathcal{S}$ the set of measurable functions defined as follows:
$$\mathcal{S} = \left\{ \sum_{i=1}^n c_i \mathbf{1}_{A_i} : \quad A_i \in \mathcal{F}, \; c_i \in \mathbf{R}, \qquad n = 1, 2, \ldots \right\}.$$
Elements of $\mathcal{S}$ are called simple functions. We also denote by $\mathcal{S}^+$ the set of positive simple functions:
$$\mathcal{S}^+ = \left\{ \sum_{i=1}^n c_i \mathbf{1}_{A_i} : \quad A_i \in \mathcal{F}, \; c_i > 0, \quad n = 1, 2, \ldots \right\}.$$

A simple function can have more than one representations. Take for example
$$f(x) = \begin{cases} 1, & x \in [0, 1] \\ 1, & x \in (2, 3] \end{cases}$$

Then

$$f(x) = 1_{[0,1]}(x) + 1_{[2,3]}(x) = 1_{[0,1)}(x) + \mathbf{1}_{\{1\}}(x) + 1_{[2,3]}(x).$$

There exists a unique canonical representation, up to re-arranging the orders of the sets. By the canonical form we mean

$$f = \sum_{i=1}^{n} a_i \mathbf{1}_{\{f^{-1}(a_i)\}},$$

where a_i are the set of non-zero values of f (they are of course distinct numbers), this is unique up to re-ordering.

Exercise 5.3.1 A simple functions is a measurable function taking only a finite number of values, a measurable function taking only a finite number of values is a simple function.

Proof If f is a measurable function taking only a finite number of non-zero values $\{c_1, \dots, c_n\}$, then it can be expressed in the form of $\sum_{i=1}^n c_i \mathbf{1}_{f^{-1}(c_i)}$ and $f^{-1}(c_i)$ is measurable.

We show the converse: $f = \sum_{i=1}^n a_i \mathbf{1}_{A_i}$ only takes a finite number of values.

Remark 5.3.1 An easier way to show that f takes only finitely many values is as below: f is a linear combination of n indicator functions. Since an indicator function takes only value 0 or 1, such a function thus can take at most 2^n values.

We next show this in at hard way, the proof itself has some value. First suppose that

$$f = a_1 \mathbf{1}_{A_1} + a_2 \mathbf{1}_{A_2}.$$

Set

$$B_1 = A_1 \setminus A_2, \quad B_2 = A_2 \setminus A_1, \quad B_3 = A_1 \cap A_2.$$

Then

$$f = a_1 \mathbf{1}_{B_1} + a_2 \mathbf{1}_{B_2} + (a_1 + a_2) \mathbf{1}_{B_3},$$

so f takes at most the following values: $0, a_1, a_2, a_1 + a_2$.

We now prove by induction to show that every f of the form $\sum_{i=1}^n a_i \mathbf{1}_{A_i}$ can be written in the form

$$f = \sum_{i=j}^{m} b_j \mathbf{1}_{B_j},$$

where B_j are disjoint. We have shown this holds for $n \leq 2$. Suppose this holds for $k \leq n$. Then,

$$f = \sum_{i=1}^{n+1} a_i \mathbf{1}_{A_i} = \sum_{i=j}^{m} b_j \mathbf{1}_{B_j} + a_{n+1} \mathbf{1}_{A_{n+1}},$$

where B_i are disjoint. Set

$$C_j = B_j \setminus A_{n+1}, \quad j = 1, \dots, m,$$

$$C_{m+j} = B_j \cap A_{n+1}, \qquad C_{2m+1} = A_{n+1} \setminus (\bigcup_{j=1}^{n} B_m).$$

Then C_j are disjojnt sets and

$$f = \sum_{j=1}^{m} b_j \mathbf{1}_{C_j} + a_{n+1} \mathbf{1}_{C_{2m+1}} + \sum_{k=1}^{m} (b_k + a_{n+1}) \mathbf{1}_{B_k \cap A_{n+1}}.$$

Thus f can be written as a finite linear combination of indicator functions of disjoint sets, among other things this also shows that f takes a finite number of values. □

Exercise 5.3.2 1. If f, g are simple functions, so are $f + g$ and fg. In particular $\mathcal{S}$ is a vector space.

2. If f is a simple function, f^+, f^- are both simple functions.

Proof Let us denote by a_i the distinct values of f and b_j the distinct values for g. We write

$$f = \sum_{i=1}^{n} a_i \mathbf{1}_{\{f^{-1}(a_i)\}}, \qquad g = \sum_{j=1}^{m} b_j \mathbf{1}_{\{f^{-1}(b_j)\}},$$

Set

$$E_i = f^{-1}(a_i), \qquad F_j = g^{-1}(b_j).$$

Then $(E_i \cap F_j) \cap (E_k \cap F_l) = \phi$ if $(i, j) \neq (k, l)$. It is clear that

$$f + g = \sum_{1 \le i \le n, 1 \le j \le m} (a_i + b_j) \mathbf{1}_{\{E_i \cap F_j\}}.$$

showing $f + g$ is a simple function. The proof for fg is similar. It is clear that $f^+ = \max(f, 0)$, $f^- = \min(f, 0)$ are also simple functions. They are obtained by flipping a_i to $\max(a_i, 0)$ or $\max(-a_i, 0)$. □

Definition 5.3.1 Let $f = \sum_{i=1}^{n} a_i \mathbf{1}_{A_i}$, in the canonical representation. Suppose that $\mu(A_i) < \infty$ for every i, we then set

$$I(f) = \sum_{i=1}^{n} a_i \mu(A_i).$$

If in addition $f \in \mathcal{S}^+$, then $a_i > 0$, we may remove the assumption $\mu(A_i) = \infty$, $I(f)$ is still defined non-ambiguously.

Of course $I(f)$ takes the value ∞ if and only if one of the A_i's has infinite measure. Observe that if $f = 0$ almost-everywhere then $I(f) = 0$.

Lemma 5.3.2 *If $f \in \mathcal{S}$ is represented as*

$$f = \sum_{j=1}^{M} d_j \mathbf{1}_{B_j}$$

where B_i are pairwise disjoint, then

$$I(f) = \sum_{i=1}^{M} d_i \mu(B_i).$$

Proof Observe that

$$\bigcup_{\{j: d_j = a_i\}} B_j = A_i.$$

Also $\mu(A_i) = \sum_{\{j: d_j = a_i\}} \mu(B_j)$ by additive property of measures. Hence

$$\begin{aligned} I(f) = \sum_{i=1}^{n} a_i \mu(A_i) &= \sum_{i=1}^{n} a_i \sum_{\{j: d_j = a_i\}} \mu(B_j) \\ &= \sum_{j=1}^{M} d_j \mu(B_j), \end{aligned}$$

completing the proof. (It is clear if $\mu(A_i) = \infty$ then one of the $\mu(B_j) = \infty$.) $\square$

We show next that I has the IP.

Proposition 5.3.3 *Let $f, g \in \mathcal{S}$, vanishing outside of a set of finite measure. Then*

1. *for any $a, b \in \mathbf{R}$,*
$$I(af + bg) = aI(f) + bI(g).$$
2. *If $f \le g$ a.e., then $I(f) \le I(g)$.*

Proof Let $\{A_i\}$ (respectively $\{B_i\}$) be the sets in the canonical representation for f (respectively for g). Let $A_0 = f^{-1}(0)$ and $B_0 = f^{-1}(0)$. Then, $A_i \cap B_j$ forms a finite collection of disjoint measurable sets, denote this collection by $\{E_i : i \le N\}$. Then

$$f = \sum_{i=1}^{N} a_i \mathbf{1}_{E_i}, \qquad g = \sum_{i=1}^{N} b_i \mathbf{1}_{E_i}$$

and

$$af + bg = \sum_{i=1}^{N} (a\, a_i + b\, b_i) \mathbf{1}_{E_i}.$$

Use Lemma 5.3.2, we see that

$$I(af+bg)=\sum_{i=1}^{N}(a\,a_i+b\,b_i)\mu(E_i)=aI(f)+bI(g).$$

Next if $f \le g$ a.e. then $\{f>g\}$ is a measurable set (by comparing their values on E_i as above), and

$$I(f)-I(g)=I(f-g)\ge 0.$$

This proved the monotonicity. □

Example 5.1 The Dirichlet function f can be written as $f=\mathbf{1}_{Q^c}$. Its integral w.r.t. the Lebesgue measure on $[0,1]$ is $\lambda(Q^c\cap[0,1])=1$.

Example 5.2 Let a,b be real numbers. Consider the measure space $([a,b],\mathcal{B}([a,b])$. Let μ be the the Lebesgue measure on $[a,b]$. It is determined by its value on sets of the form $A=\cup_{i=1}^n(c_i,d_i)$ (union of disjoint intervals),

$$\mu(A)=\sum_i^n(d_i-c_i).$$

Consider $a=t_0<t_1<\cdots<t_n=b$. Let

$$f(x)=f_0\mathbf{1}_{\{0\}}(x)+\sum_{j=0}^{n-1}a_j\mathbf{1}_{(t_j,t_{j+1}]}(x).$$

Since $\mu(\{0\})=0$,

$$\int_a^b f(x)dx=\sum_{j=0}^{n-1}a_j(t_{j+1}-t_j).$$

5.4 Integration of non-negative functions

We will allow a non-negative function to take values in the extended real line $\overline{\mathbf{R}}$. In the sequel, a non-negative (potentially infinite) function $f:\mathcal{X}\to[0,+\infty]$ will be said to be measurable if it is measurable from $(\mathcal{X},\mathcal{F})$ to $([0,\infty],\mathcal{B}([0,\infty]))$. Note that, by Proposition 4.5.1 above, since $\mathcal{B}([0,\infty])\subset\mathcal{B}(\overline{\mathbf{R}})$ (see Exercise 4.6.1) this is equivalent to saying that f is measurable from $(\mathcal{X},\mathcal{F})$ to $(\overline{\mathbf{R}},\mathcal{B}(\overline{\mathbf{R}}))$. Recall that f is measurable from $(\mathcal{X},\mathcal{F})$ to $([0,\infty],\mathcal{B}([0,\infty]))$ iff $\{f<\infty\}$ is measurable and $f^{-1}(A)\in\mathcal{F}$ for any $A\in\mathcal{B}(\mathbf{R})$.

The principle for a measurable function $f:\mathcal{X}\to[0,\infty]$ is as follows: if $f=\infty$ on a set of non-zero measure, we should assign ∞ to its integral; if $f<\infty$ almost everywhere, we ignore those values that are ∞. This fact is encapsulated in the following, comprehensive, definition.

Definition 5.4.1 If $f : \mathcal{X} \to [0, \infty]$ is a measurable function, we define the integral of f to be

$$\int f d\mu = \sup\{I(g) : 0 \le g \le f, g \in \mathcal{S}\}.$$

We say f is integrable if $\int f d\mu < \infty$.

Remark 5.4.1 If μ is a finite measure, then a bounded measurable function is integrable. A special case is when μ is the Lebesgue measure restricted to any bounded interval.

Definition 5.4.2 If A is a measurable set and $f\mathbf{1}_A$ is integrable, we define

$$\int_A f\, d\mu = \int f\, 1_A\, d\mu.$$

Proposition 5.4.2 *1. If $f \in \mathcal{S}^+$ then $I(f) = \int f d\mu$.*

2. the integral satisfies monotonicity.

Proof Firstly, let $f \in \mathcal{S}^+$. We may take $g = f \in \mathcal{S}^+$, so

$$\sup\{I(g) : g \le f, g \in \mathcal{S}^+\} \ge I(f).$$

Next, by monotonicity for I, if $g \in \mathcal{S}$ and $g \le f$, then $I(g) \le I(f)$. Hence

$$\sup\{I(g) : g \le f, g \in \mathcal{S}^+\} \le I(f).$$

This proves that $I(f) = \int f\, d\mu$ for $f \in \mathcal{S}^+$. Now if $f_1 \le f_2$ are non-negative and measurable, then

$$\int f_1\, d\mu = \sup\{I(g) : g \le f_1, g \in \mathcal{S}^+\} \le \sup\{I(g) : g \le f_2, g \in \mathcal{S}^+\} = \int f_2\, d\mu$$

proving the monotonicity. □

We cannot directly prove the linearity and try to obtain this from that of simple functions by taking limits.

Definition 5.4.3 We say $f_n \to f$ a.e. if there exists a null set A such that $f_n \to f$ everywhere outside of the null set A.

Theorem 5.4.3 (Fatou's lemma) *Let f_n be a sequence of non-negative measurable functions that converges almost-everywhere to a function f, then*

$$\int f d\mu \le \liminf_{n\to\infty} \int f_n d\mu.$$

Proof Let $E = \{x : f_n(x) \underset{n\to\infty}{\longrightarrow} f(x)\}$. At the expanse of replacing f_n by $\mathbf{1}_E\, f_n$ and f by $\mathbf{1}_E\, f$, we may assume without loss of generality that $f_n \to f$ pointwise.

Making this assumption it is sufficient to show that for any $g \in \mathcal{S}^+$ with $g \le f$,

$$\int g\, d\mu \le \liminf_{n\to\infty} \int f_n\, d\mu.$$

Step 1. Suppose that $\int g < \infty$, set

$$\mathcal{X}_0 = \{x : g(x) > 0\}.$$

Since g takes only a finite number of values $\mu(\mathcal{X}_0) < \infty$ and $M = \max g < \infty$ (We only need to restrict to $\mathcal{X}_0$ as $\int_{\mathcal{X}_0} g = \int_{\mathcal{X}} g$.)

Fix $\epsilon > 0$. Since $g \le f = \lim_{n\to\infty} f_n$, for any x there exists $N(x, \epsilon)$ s.t. for $n > N(x, \epsilon)$,

$$f_n(x) \ge (1-\epsilon) g(x).$$

Let

$$A_n = \{x \in \mathcal{X}_0 : \quad f_k(x) > (1-\epsilon) g(x), \qquad \forall k \ge n\}.$$

Then A_n is an increasing sequence and $\cup_n A_n = \mathcal{X}_0$. By the continuity property of measures,

$$\lim_{n\to\infty} \mu(\mathcal{X}_0 \setminus A_n) = 0.$$

Then for $k \ge n$,

$$(1-\epsilon) \int_{A_n} g \le \int_{A_n} f_k \le \int f_k.$$

Also,

$$(1-\epsilon) \int_{A_n} g = (1-\epsilon) \int g - (1-\epsilon) \int_{A_n^c} g.$$

Hence

$$(1-\epsilon) \int g - (1-\epsilon) \int_{A_n^c} g \le \int_{A_n} f_k.$$

Since $g \le M$, we may take ϵ to 0,

$$\int g - \int_{A_n^c} g \le \liminf_{k\to\infty} \int f_k,$$

for every n. Take $n \to \infty$ to conclude $\int g \le \liminf_{n\to\infty} \int f_n$.

Step 2. Suppose that $\int g = \infty$, since g take only a finite number of values, and $g \ge 0$, it is necessary that there exists a set of infinite measure on which g takes a positive value $a > 0$. Denote this set by A. Set

$$A_k = \{x : f_n(x) \ge \frac{1}{2} a, \qquad \forall n \ge k\}.$$

Then A_n is increasing. Since $\lim_{n\to\infty} f_n \geq g$, we have $\cup_n A_n = A$ and $\lim_{n\to\infty}\mu(A_n) = \mu(A) = \infty$. For all $n \geq k$,

$$\int f_n \geq \int \frac{1}{2} a \mathbf{1}_{A_k} = \frac{1}{2} a\mu(A_k).$$

Thus,

$$\liminf_{n\to\infty} \int f_n \geq \frac{1}{2} a\mu(A_k).$$

Take $k \to \infty$ $\liminf_{n\to\infty} f_n \geq \infty$, this completes the proof. □

Theorem 5.4.4 (Monotone convergence theorem) *If f_n is a sequence of non-negative measurable functions converging almost everywhere to f. Suppose that $f_n \leq f$ for all n. Then*

$$\int \lim_n f_n = \lim_n \int f_n.$$

Proof If $0 \leq g \leq f_n$ then $0 \leq g \leq f$, hence for every n, $\int f_n \leq \int f$.

$$\limsup_n \int f_n \leq \int f.$$

By Fatou's lemma,

$$\liminf \int f_n \geq \int \lim_n f_n = \int f,$$

completing the proof. □

The definition of integrals for non-negative functions using the value $\sup\{I(g) : 0 \leq g \leq f, g \in \mathcal{S}\}$ is not easy to use. The monotone convergence theorem allow us to use a sequence. Choose $f_n \in \mathcal{S}$ with $f_n \uparrow f$ (this exists by Theorem 4.7.3), then

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

The name 'Monotone convergence theorem' comes from the following Corollary.

Corollary 5.4.5 *If $0 \leq f_1 \leq f_2 \leq \dots$ is a sequence of increasing real valued measurable functions, and let $f = \sup_n f_n$. Then f is integrable if and only if $\sup_n \int f_n < \infty$, and then*

$$\int \sup_n f_n = \sup_n \int f_n.$$

This means in particular that

Corollary 5.4.6 *If f_n is a sequence of increasing simple functions with limit f, then*

$$\int f = \lim_{n\to\infty} \int f_n.$$

Proposition 5.4.7 *If f, g are non-negative and measurable, then the following holds:*
(1)[Linearity] for any $a, b \geq 0$,

$$\int (af + bg) = a \int f + b \int g.$$

(2) [Monotonicity] If $f \geq g$ then $\int f \geq \int g$.
(3) $\int f = 0$ if and only if $f = 0$ a.e.

Proof (1) Note that linearity holds on $\mathcal{S}$. Now we approximate f, g with positive simple increasing sequence of functions. Take $f_n \in \mathcal{S}^+, g_n \in \mathcal{S}^+$ such that

$$f_n \uparrow f, \quad g_n \uparrow g.$$

with $f_n \leq f$ and $g_n \leq g$. Then

$$\int (af_n + bg_n) d\mu = \int f_n d\mu + \int bg_n d\mu,$$

We apply the monotone convergence theorem to take the limit inside the integrals, proving the linearity.

(2) It is sufficient to show $f \geq 0$ a.e. then $\int f \geq 0$. This is true as the simple functions approximating f is non-negative.

(3) Let $f \geq 0$ a.e. If $f = 0$ a.e. it is clear that $\int f d\mu = 0$ (e.g. use monotonicity.)

Suppose that $\int f = 0$. Let $A_n = \{f \geq \frac{1}{n}\}$. Then $\{f > 0\} = \cup_{n=1}^{\infty} A_n$.

$$\mu(A_n) = \int \mathbf{1}_{A_n} d\mu \leq \int \mathbf{1}_{A_n} n f d\mu \leq n \int f d\mu = 0.$$

Thus

$$\mu(\{f > 0\} \leq \sum_{n=1}^{\infty} \mu(A_n) = 0,$$

concluding the proof. □

These shows that we can exchange summation (over a finite number of functions) with integration. For a sum of non-negative functions, this holds also with infinitely many summands, as to be explained below.

Corollary 5.4.8 (Exchange of order of summation and integration) *Suppose that f_n is a sequence of non-negative measurable functions, then*

$$\int \sum_{n=1}^{\infty} f_n = \sum_{n=1}^{\infty} \int f_n.$$

Proof By Proposition 5.4.7, finite sum can be taken out of the integral. Fatou's lemma allow us to take the limit.

$$\int \sum_{n=1}^{\infty} f_n = \int \lim_{N\to\infty} \sum_{n=1}^{N} f_n \overset{Monotone}{=} \lim_{N\to\infty} \int \sum_{n=1}^{N} f_n$$
$$\overset{linearity}{=} \lim_{N\to\infty} \sum_{n=1}^{N} \int f_n = \sum_{n=1}^{\infty} \int f_n,$$

concluding the proof. □

5.5 Integrable functions

Definition 5.5.1 Let $f : \mathcal{X} \to \mathbf{R}$ be measurable. If both f^+ and f^- are integrable we say that f is integrable, in which case we define its integral to be:

$$\int f = \int f^+ - \int f^-.$$

Note that f is integrable if and only if $|f|$ is integrable. A bounded measurable function is integrable with respect to a finite measure.

Definition 5.5.2 We denote by the class of integrable functions by L_1. The notation $L_1(\mathcal{X})$ is used to emphasize the space. Also $L_1(\mu)$, $L_1(\mathcal{X};\mu)$ are used to emphasize the measure.

Proposition 5.5.1 *If f, g are integrable functions,t then*

(1) (linearity) For any $a, b \in \mathbf{R}$,
$$\int (af + bg) = a \int f + b \int g.$$

(2) (Monotonicity) If $g \le f$ a.e., then $\int g \le \int f$.

(3) $\int |f| = 0$ if and only if $f = 0$ almost-everywhere.

(4) $|\int f\, d\mu| \le \int |f|\, d\mu$.

Proof Part one follows from splitting $f = f^+ - f^-$ and $g = g^+ - g^-$ and the same property for non-negative functions. If $g \le f$ then $g - f \ge 0$, so $\int (g - f) \ge 0$, monotonicity now follows from the linearity. The third property follows from Proposition 5.4.7 and the property $|f| \ge 0$. Finally, for (4), since $-|f| \le f \le |f|$, by monotonicity we deduce that

$$-\int |f|\, d\mu \le \int f\, d\mu \le \int |f|\, d\mu,$$

thus yielding the claim. □

Exercise 5.5.1 Assume that f, g are non-negative measurable functions and h is an integrable function on $\mathcal{X}$ sastisfying that $f \geq g + h$ μ-a.e. Show that

$$\int f\, d\mu \geq \int g\, d\mu + \int h\, d\mu.$$

Exercise 5.5.2 If f, g are real valued integrable functions, show the following statements hold:

1. If $\mu(A) = 0$ then $\int_A f\, d\mu = 0$.
2. If $\int_A f = 0$ for every measurable set A then $f = 0$ μ almost-everywhere.

Hint: For part (2), use Proposition 5.4.7 and choose a suitable A.

Exercise 5.5.3 Let $f, g : \mathcal{X} \to [0, \infty]$ be two non-negative measurable functions. We assume that $\int_A f\, d\mu = \int_A g\, d\mu$ for every measurable set A.

1. Show that if μ is also assumed to be finite, then necessarily $f = g$ μ-a.e. [1]
2. Show that if μ is σ-finite, then the conclusion $f = g$ μ-a.e. remains true.

Example 5.3 Recall that for $x_0 \in \mathcal{X}$, the **Dirac measure** at x_0 is defined by:

$$\delta_{x_0}(A) = \begin{cases} 0, & \text{if } x_0 \in A \\ 1, & \text{if } x_0 \notin A \end{cases} = \mathbf{1}_A(x_0).$$

Let $f : \mathcal{X} \to \mathbf{R}$ be measurable. Then f is integrable and $\int_{\mathcal{X}} f\; d\delta_{x_0} = f(x_0)$.

Example 5.4 Let $p : \mathcal{X} \to [0, \infty)$ be integrable. Set

$$\mu_p(A) := \int_A p(x)\mu(dx).$$

Then μ_p defines a measure.

Proof Firstly $\mu_p(\phi) = \int_\phi p\, \mu = 0$. Next if $A \cap B = \phi$, $\mathbf{1}_{A\cap B} = \mathbf{1}_A + \mathbf{1}_B$, So

$$\mu_p(A \cup B) = \int (\mathbf{1}_A + \mathbf{1}_B) p d\mu = \mu_p(A) + \mu_p(B).$$

Finally, if A_n is an increasing sequence of measurable sets with union A,

$$\lim_{n\to\infty} 1_{A_n} = \mathbf{1}_{\cup_{n=1}^\infty A_n} = \mathbf{1}_A,$$

[1] **Hint**: Beware in this case f and g may not be integrable, so we may not use the result from the previous exercise. One may rather consider, for all $a, b \in \mathbf{Q}$ such that $0 \leq a < b$, the measurable subset $A = \{f \leq a < b \leq g\}$, and show that $\mu(A) = 0$.

showing that

$$\mu_p(A) = \int 1_A p \, d\mu = \int \lim_{n\to\infty} 1_{A_n} \, p d\mu = \lim_{n\to\infty} \int 1_{A_n} p \, d\mu = \lim_{n\to\infty} \mu_p(A_n).$$

Hence the σ-additive property holds. □

Questions.
Let $x_0 \in \mathbf{R}$. Can one find a function $p : \mathbf{R} \to \mathbf{R}$ such that, for all $A \in \mathcal{B}(\mathbf{R})$, $\delta_{x_0}(A) = \int_A p \, d\lambda$?
Can one find a function $p : \mathbf{R} \to \mathbf{R}$ such that, for all $A \in \mathcal{B}(\mathbf{R})$, $\lambda(A) = \int_A p \, d\delta_{x_0}$?

Exercise 5.5.4 The answer are no and no. Prove it.

Two measures μ and ν on Ω are mutually singular if there are disjoint sets A_1 and A_2 such that $\Omega = A_1 \bigcup A_2$ with $\mu(A_1) = 0$ and $\nu(A_2) = 0$. This is denoted by $\mu \perp \nu$. The Lebesgue measure and any Dirac measures δ_x are singular.

Definition 5.5.3 We say that μ is absolutely continuous with respect to ν if whenever $\nu(A) = 0$, we have $\mu(A) = 0$. We write $\mu \ll \nu$.

Later we will see that if μ and ν are two finite measures such that $\mu \ll \nu$ then we can found a a density function p such that μ is given by integration of p with respect to ν. This is called the Radon-Nikodym Theorem.

5.6 Further limit theorems

Theorem 5.6.1 (Fatou's Lemma) *Suppose that f_n is a sequence of non-negative measurable functions, then*

$$\int \liminf_{n\to\infty} f_n \, d\mu \le \liminf_{n\to\infty} \int f_n \, d\mu.$$

Proof Firstly,

$$\liminf_{n\to\infty} f_n = \lim_{k\to\infty} \inf_{m\ge k} f_m.$$

Observe that $\inf_{m\ge k} f_m$ is an increasing sequence and $\inf_{m\ge k} f_m \le f_k$. So

$$\begin{aligned}\int \liminf_{n\to\infty} f_n &= \int \lim_{k\to\infty} \inf_{m\ge k} f_m \\ &= \liminf_{k\to\infty} \int \inf_{m\ge k} f_m \le \liminf_{k\to\infty} \int f_k.\end{aligned}$$

This completes the proof. □

Theorem 5.6.2 (Dominated convergence theorem) *Suppose that f_n is a sequence of measurable functions converging to a function f almost everywhere. Suppose that there exists an integrable function g such that $|f_n| \leq |g|$, then f_n and f are integrable w.r.t. μ, and*

(1)
$$\lim_{n\to\infty} \int f_n \, d\mu = \int \lim_{n\to\infty} f_n \, d\mu.$$

(2)
$$\lim_{n\to\infty} \int |f_n - f| \, d\mu = 0.$$

Proof Since g is integrable and $|f_n| \leq |g|$ μ a.e., the f_n are integrable as well. Moreover as $f_n \to f$ a.e., we deduce that $|f| \leq g$ a.e., so f is also μ-integrable. Then statement (1) follows from statement (2), as

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

Let us prove (2). By the triangle inequality,

$$|f_n - f| \leq |f_n| + |f| \leq 2|g| \qquad \text{a.e.,}$$

hence the measurable function $2|g| - |f_n - f|$ is non-negative and converges to $2|g|$ a.e. Therefore, by Fatou's lemma

$$2 \int g \, d\mu \leq \liminf_{n\to\infty} \int (2|g| - |f_n - f|) \, d\mu = 2 \int g \, d\mu - \limsup_{n\to\infty} \int |f_n - f| \, d\mu.$$

Since $\int g \, d\mu < \infty$, we deduce that $\limsup\limits_{n\to\infty} \int |f_n - f| \, d\mu \leq 0$, and the claim follows. □

Exercise 5.6.1 Prove part (2) of the dominated convergence theorem.

Definition 5.6.1 A sequence of measurable functions f_n is said to converge to f in L_1 if

$$\lim_{n\to\infty} \int |f_n - f| \, d\mu = 0.$$

Exercise 5.6.2 Show that if $f_n \to f$ in L_1 then,

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

Definition 5.6.2

1. A measurable functions with the property $\int |f|^p < \infty$ are said to be in L_p where $1 \leq p < \infty$.
2. If there exists a measurable function f such that $\int |f_n - f|^p d\mu \to 0$, then we say f_n converges to f is L_p, in which case f is also in L_p.

5.7 Integrals depending on a parameter

As before, let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space, and let E be a metric space. In many cases we are integrating a function $f : E \times \mathcal{X} \to \mathbf{R} \ni (t,x) \to f(t,x)$ with respect to the variable x, and are interested in the behaviour of $\int_{\mathcal{X}} f(t,x)\, d\mu(x)$ when the value of t changes.

Proposition 5.7.1 *Let $f : E \times \mathcal{X} \to \mathbf{R}$ be a function, and let $t_0 \in E$. We assume that:*

1. *For all $t \in E$, the function $x \mapsto f(t,x)$ is measurable from $(\mathcal{X}, \mathcal{F})$ to $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$.*
2. *$\mu(dx)$-a.e., the map $t \mapsto f(t,x)$ is continuous at the point t_0.*
3. *there exists $g : \mathcal{X} \mapsto [0,\infty]$ integrable such that, for all $t \in E$, $|f(t,x)| \le g(x)$ $\mu(dx)$-a.e.*

Let $F : E \mapsto \mathbf{R}$ given, for $t \in E$, by $F(t) = \int_{\mathcal{X}} f(t,x)\, \mu(dx)$. Then the function F is well-defined and continuous at t_0.

Proof Since for all $t \in E$, $|f(t,x)| \le g(x)$ μ-dx a.e. where g is integrable, we deduce that $x \mapsto f(t,x)$ is integrable, hence $F(t)$ is well-defined. Moreover, if $(t_n)_{n\ge1}$ is a sequence in E converging to t_0, then, in view of asssumption 2., $f(t_n,x) \to f(t_0,x)$ $\mu(dx)$-a.e. Moreover, by Assumption 3, for all $n \ge 1$, $|f(t_n,x)| \le g(x)$. Since g is integrable, by the dominated convergence theorem, $\int f(t_n,x)\mu(dx) \underset{n\to\infty}{\longrightarrow} \int f(t_0,x)\mu(dx)$, i.e. $F(t_n) \underset{n\to\infty}{\longrightarrow} F(t_0)$. The continuity property for F follows. □

Assume now that $E = I$ is an open interval $\mathbf{R}$. The next proposition provides sufficient conditions for the function $t \mapsto \int_{\mathcal{X}} f(t,x) d\mu(x)$ to be differentiable at a point $t_0 \in I$.

Proposition 5.7.2 *Let $f : I \times \mathcal{X} \to \mathbf{R}$ be a function, and let $t_0 \in I$. We assume that:*

1. *For all $t \in E$, the function $x \mapsto f(t,x)$ is integrable with respect to μ.*
2. *$\mu(dx)$-a.e., the map $t \mapsto f(t,x)$ admits a derivative at the point t_0, which we denote by $\frac{\partial f}{\partial t}(t_0,x)$.*
3. *there exists an integrable function $g : \mathcal{X} \mapsto [0,\infty]$ such that, for all $t \in E$, we have*

$$|f(t,x) - f(t_0,x)| \le g(x)\,|t - t_0|, \qquad \mu(dx) - a.e. \tag{5.1}$$

Then the function $F : I \to \mathbf{R}$ given, for $t \in I$, by $F(t) = \int_{\mathcal{X}} f(t,x)\, \mu(dx)$, is differentiable at t_0, and

$$F'(t_0) = \int_{\mathcal{X}} \frac{\partial f}{\partial t}(t_0,x)\, d\mu(x).$$

Remark 5.7.3 Under the assumptions of the above Proposition, we thus have

$$\frac{\partial}{\partial t}\left(\int_{\mathcal{X}} f(t,x)\,d\mu(x)\right)\Big|_{t=t_0} = \int_{\mathcal{X}} \frac{\partial f}{\partial t}(t_0,x)\,d\mu(x),$$

i.e., we may interchange the derivation and the integration.

Proof By Assumption 1., $F(t)$ is well-defined for all $t \in I$. Moreover, by Assumption 2., for all sequence of points $t_n \in I$ converging to t_0, we have

$$\frac{f(t_n,x)-f(t_0,x)}{t-t_0} \underset{n\to\infty}{\longrightarrow} \frac{\partial f}{\partial t}(t_0,x), \qquad \mu(dx) - \text{a.e.}$$

Moreover, Assumption 3. yields the domination property

$$\left|\frac{f(t_n,x)-f(t_0,x)}{t-t_0}\right| \le g(x), \qquad \mu(dx) - \text{a.e.}$$

holding for all $n \ge 1$. Since g is integrable, by the Dominated Convergence Theorem

$$\frac{F(t_n)-F(t_0)}{t_n-t_0} = \int_{\mathcal{X}} \frac{f(t_n,x)-f(t_0,x)}{t-t_0}\,d\mu(x) \underset{n\to\infty}{\longrightarrow} \int_{\mathcal{X}} \frac{\partial f}{\partial t}(t_0,x)\,d\mu(x),$$

which yields the claim. □

Often the following, weaker version of the previous theorem proves convenient. It is less general, but its assumptions are quicker to check and are very often satisfied in most applications.

Proposition 5.7.4 *Let $f : I \times \mathcal{X} \to \mathbf{R}$ be a function. We assume that:*

1. *For all $t \in E$, the function $x \mapsto f(t,x)$ is integrable with respect to μ.*

2'. *$\mu(dx)$-a.e., the map $t \mapsto f(t,x)$ is differentiable on I, with derivative at t denoted by $\frac{\partial f}{\partial t}(t,x)$.*

3'. *there exists an integrable function $g : \mathcal{X} \mapsto [0,\infty]$ such that, $\mu(dx)$-a.e, we have*

$$\forall t \in I, \qquad \left|\frac{\partial f}{\partial t}(t,x)\right| \le g(x).$$

Then the function $F : I \to \mathbf{R}$ given, for $t \in I$, by $F(t) = \int_{\mathcal{X}} f(t,x)\,\mu(dx)$, is differentiable on I, and

$$F'(t) = \int_{\mathcal{X}} \frac{\partial f}{\partial t}(t,x)\,d\mu(x), \qquad t \in I.$$

Proof By Assumptions 2'. and 3'. and the mean value theorem, for any given $t_0 \in I$, the inequality (5.1) is satisfied at t_0. Hence the result follows by Proposition 5.7.2. . □

Here are two important examples of applications of the above three propositions

Example 5.5 (Convolution) Let $\varphi : \mathbf{R} \to \mathbf{R}$ be a Lebesgue-integrable function, and let $\varrho : \mathbf{R} \to \mathbf{R}$ be a bounded continuous function. Then the convolution $\varrho * \varphi : \mathbf{R} \to \mathbf{R}$, defined by

$$\varrho * \varphi(t) = \int \varrho(t-x)\, \varphi(x)\, \lambda(dx)$$

is continuous on $\mathbf{R}$. Moreover, if ϱ is C^1 and ϱ' is bounded, then $\varrho * \varphi$ is C^1 and $(\varrho * \varphi)' = \varrho' * \varphi$.

Example 5.6 (Fourier transform) Let $\varphi : \mathbf{R} \to \mathbf{R}$ be a Lebesgue-integrable function. Its Fourier transform $\hat{\varphi} : \mathbf{R} \to \mathbf{C}$ is defined by

$$\hat{\varphi}(\xi) = \int_{\mathbf{R}} e^{i\xi x} \varphi(x)\, \lambda(dx), \qquad \xi \in \mathbf{R}.$$

Then $\hat{\varphi}$ is continuous. Moreover, if

$$\int_{\mathbf{R}} |x|\, |\varphi(x)|\, \lambda(dx) < \infty$$

then $\hat{\varphi}$ is C^1, and for all $\xi \in \mathbf{R}$, $\hat{\varphi}'(\xi) = \int_{\mathbf{R}} ix e^{i\xi x} \varphi(x)\, \lambda(dx)$.

Remark 5.7.5 The above three propositions give criteria for regularity (continuity and differentiability) of the function $F : t \mapsto \int_{\mathcal{X}} f(t,x)\, \mu(dx)$. One may also ask for criteria of integrability: this question will be tackled in Chapter 7 below with Fubini's Theorems.

5.8 Examples and exercises

Exercise 5.8.1 If a bounded function $f : [a,b] \to \mathbf{R}$ is Riemann integrable, then it is Lebesgue measurable and Lebesgue integrable. Furthermore, the integrals have the common value:

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

Hint. Check that the integrals agree on functions of the form:

$$\sum_{i=1}^{n} \alpha_i \mathbf{1}_{(c_i, d_i)}$$

where (c_i, d_i) are disjoint intervals.

Let Δ be the dyadic partition of $[a,b]$: $t_j = \frac{(b-a)}{2^n}$. Consider

$$f_n(t) = \sum_j \sup\{f(x) : x \in [t_i, t_{i+1}]\} \mathbf{1}_{(t_i, t_{i+1}]}(t).$$

Then f_n is measurable and converges to f a.e.

Exercise 5.8.2 1. Show that if for two measures μ and ν, $\int f d\mu = \int f d\nu$ for all $f \in \mathcal{S}$, then $L_1(\mu) = L_1(\nu)$, and $\int f d\mu = \int f d\nu$ for all integrable functions.

2. Let E be a measurable set and μ_E the restriction of μ to the trace σ-algebra $\mathcal{F}_E = \{E \cap A : A \in \mathcal{F}\}$. Show that
$$\int_E f d\mu = \int_{\mathcal{X}} f d\mu_E.$$
Also $f \in L_1(\mu_E)$ if and only if $f\mathbf{1}_E \in L_1(\mu)$.

(1) By the construction of integrals, $\int f d\mu = \int f d\nu$ for all non-negative functions. Thus $\int |f| d\mu = \int |f| d\nu$, which are either both finite or both infinite. This implies that $L_1(\mu) = L_1(\nu)$ and the integrals are the same for they are defined by the integrals of their positive and negative parts.

(2) If g is elementary, so is $g\mathbf{1}_E$ and $\int g\mathbf{1}_E d\mu = \int g d\mu_E$ for all elementary $g \in \mathcal{S}$. Let f_n be an increasing sequence of simple functions converging to f, then $f_n\mathbf{1}_E \uparrow f\mathbf{1}_E$. Hence
$$\int_E f d\mu = \lim_{n\to\infty} \int f_n \mathbf{1}_E \, d\mu = \lim_{n\to\infty} \int f_n \, d\mu_E = \int f d\mu_E.$$

Exercise 5.8.3 Suppose that $p : [a, b] \to \mathbf{R}$ is a non-negative continuous function. Set
$$F(x) = F(a) + \int_a^x p(t)dt.$$
Let $f : [a, b] \to \mathbf{R}$ be a continuous function. Describe the Lebesgue-Stilejes measure μ_F and the integral
$$\int f \, d\mu_F$$
in terms of Riemann integral.

Sketch. Let $A \subset \mathcal{B}([a, b])$. Then,
$$\int \mathbf{1}_A \, d\mu_F = \mu_F(A) = \int_c^d p(t)dt = \int_{[a,b]} \mathbf{1}_A \, p(t) \, dt.$$
By linearity for any simple function f,
$$\int f d\mu_F = \int_a^b f(t)p(t)dt.$$
By Fatou's lemma we can pass this to non-negative functions, and then by linearity to integrable functions:
$$\int_{[a,b]} f d\mu_F = \int_a^b f(t)p(t)dt.$$

Example 5.7 Let $f_n = n\mathbf{1}_{[0, \frac{1}{n}]}$ Then, $f_n \to 0$ everywhere except at 0.
$$\int f_n d\lambda = \int_0^n dx = 1.$$
Hence we cannot exchange the order of taking limit with taking integration in this case.

5.9 The Monotone Class Theorem for Functions *

Proposition 5.9.1 (The Monotone Class Theorem for Functions) *Let $(\Omega, \mathcal{F})$ be a measure space. Let $\mathcal{C}$ be a π system with $\sigma(\mathcal{C}) = \mathcal{F}$. Let $\mathcal{H}$ be a* **vector space** *of functions from Ω to* **R***, with the following property:*

1. *$1 \in \mathcal{H}$, and $\mathbf{1}_A \in \mathcal{H}$ for every $A \in \mathcal{C}$.*
2. *(Monotone class property) If $f_n \in \mathcal{H}$ is an increasing sequence of non-negative functions with $f(x) = \sup_n f(x)$ finite for every x (resp. with $f = \sup_n f$ bounded), then $f \in \mathcal{H}$.*

Then $\mathcal{H}$ contains the set of all real valued (resp. all bounded) $\mathcal{F}$-measurable functions.

Proof Let

$$J = \{B \in \mathcal{F} : \mathbf{1}_B \in \mathcal{H}\}.$$

By assumption, $1 = 1_\Omega \in \mathcal{H}$, and $J \supset \mathbf{C} \cup \{\Omega\}$. If $A \subset B$, $A, B \in J$ then $\mathbf{1}_{B \setminus A} = \mathbf{1}_B - \mathbf{1}_A \in \mathcal{H}$ by the vector space property of $\mathcal{H}$. Thus $B \setminus A \in J$. Let $A_n \in J$ be an increasing sequence of sets then by condition 2,

$$\mathbf{1}_{\cup_{n=1}^\infty A_n} = \lim_{n \to \infty} \mathbf{1}_{A_n} \in \mathcal{H}.$$

Hence $\cup_{n=1}^\infty A_n \in J$ and J is a λ-system. It therefore contains $\mathcal{F}$. This means all indictor functions are in $\mathcal{H}$ and simple functions are in J.

Suppose H is closed under monotone limit of functions with bounded limit. If f is bounded positive and measurable there is an sequence of positive simple functions f_n converging to f, thus $f \in \mathcal{H}$. If f is not positive let $f = f^+ - f^-$ to conclude.

Suppose H is closed under monotone limit of functions with finite limit. If f is positive and measurable there is an sequence of positive simple functions f_n converging to f, thus $f \in \mathcal{H}$. Again, if f is not positive let $f = f^+ - f^-$ to conclude. □

Remark 5.9.2 From this we see a meta theorem (means the theorem is likely to hold, of course one needs to give a proof): If a property concerning integration of functions holds for all indicator functions, then it holds for all measurable functions (use Fatou's lemma to obtain the monotone property) or it holds for all bounded measurable functions (use dominated convergence theorem to obtain the monotone property).

5.9.1 Lebesgue Integration

In this section let $\mathcal{X} = \mathbf{R}$, $\mathcal{F}$ the completion of the Borel σ algebra, i.e. consists of Lebesgue measurable sets, and μ the Lebesgue measure λ. The restriction of λ to any measurable subset is also denoted by the same letter.

A function $f : \mathbf{R} \to \mathbf{R}$ is said to be Lebesgue measurable if it is measurable with respect to $\mathcal{F}/\mathcal{B}(\mathbf{R})$. Borel measurable functions are of course Lebesgue measurable. In this section by a measurable function we refer to a Lebesgue measurable function.

Definition 5.9.1 Integrals with respect to a Lebesgue measure, are called Lebesgue integrals.

Exercise 5.9.1 Let I be a Lebesgue measurable set. Suppose that $f : I \to \mathbf{R}$ is bounded.

1. If f is Lebesgue measurable, show that
$$\inf\left\{\int h\, d\lambda, \quad h \ge f, h \in \mathcal{S}.\right\} = \sup\left\{\int g\, d\lambda : g \le f, g \in \mathcal{S}\right\} \tag{5.2}$$
2. Show that if (5.2) holds, then f is Lebesgue measurable.

5.10 The pushed forward measure

Suppose we are given two measurable spaces $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$. Let μ be a measure on $\mathcal{X}$ and $f : X \to Y$ be a measurable function. Denote by $f_*\mu$ the **pushed forward (induced) measure** on $(Y, \mathcal{B})$, i.e.
$$f_*(\mu)(B) = \mu\{x : f(x) \in B\}.$$
The right hand side $\mu(f^{-1}(B))$, so we given B the measure of its pre-image.

Proposition 5.10.1 *Let $\varphi : \mathcal{Y} \to \mathbf{R}$ a measurable function, we have*
$$\int_{\mathcal{X}} \varphi \circ f \; d\mu = \int_{\mathcal{Y}} \varphi \; d(f_*\mu).$$
This is in the sense that φ is integrable with respect to $f_*\mu$ if and only if $\varphi \circ f$ is integrable with respect to μ.

Proof This holds for indicator functions of measurable sets by the definition of pushed forward measures. Apply monotone convergence theorem on both sides to see that those functions with the desired property is a monotone class. Apply the monotone class theorem to conclude. □

Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability measure. Let $(S, \mathcal{G})$ be a measurable space. A measurable function from $X : \Omega \to S$ is called a random variable, S is called its state space. If $\varphi : S \to \mathbf{R}$ and $X : \Omega \to S$ are measurable, then
$$\mathbf{E}[\varphi(X)] := \int_\Omega \varphi \circ X \; dP = \int_S \varphi \; dX_*(\mathbf{P}).$$

Exercise 5.10.1 If both Y, Y' are real valued integrable functions on $(\Omega, \mathcal{F}, \mathbf{P})$ with $\int_A Y = \int_A Y'$ for every measurable set A, then $Y = Y'$ almost surely.

Proof Let $A = \{Y > Y'\}$. By symmetry it is sufficient to proves that $\mathbf{P}(A) = 0$. Thus we only need to prove the following statement: for any $Y \ge 0$, $\int_A Y = 0$ for any $A \in \mathcal{F}$, then $Y = 0$ almost surely. For this just note that if $\{Y \ne 0\} = \{Y > 0\}$ has positive measure then for some $a > 0$, $\{Y > a\}$ has positive measure (for otherwise $\mathbf{P}(Y > 0) = \lim_{n\to\infty} \mathbf{P}(Y > \frac{1}{n}) = 0$), and then

$$\int_A Y d\mathbf{P} \ge a\mathbf{P}(Y > a) > 0,$$

this contradicts the assumption. Hence $\mu(\{Y \ne 0\}) = 0$,completting the proof. $\square$

Example 5.8 Let $f : \mathcal{X} \to \mathcal{Y}$. Given a measure on $\mathcal{Y}$, can we use f and μ to define a measure on $\mathcal{X}$ in a meaningful way? The answer is no in general. There is no good sensible way for this construction, for otherwise you should be able to pull back a direct measure, what that would be if $f : \{1, 2\} \to \mathbf{R}$ $f(1) = 0$ and $f(2) = 0$? If $f : \mathbf{R} \to \mathbf{R}$ is injective, one can define a measure from a measure on the target space, it would be the pushed forward measure for the inverse map.

Definition 5.10.1 If f is a measurable map from $\mathcal{X}$ to $\mathcal{X}$, we say μ is **invariant** by f if $f_*(\mu) = \mu$.

Example 5.9 Let δ_x be the **Dirac measure** on $\mathbf{R}^n$. Then for any transformation $T : \mathbf{R}^n \to \mathbf{R}^n$, $T_*(\delta_x) = \delta_{Tx}$. The δ-measure is not invariant under rotations (unless $x = 0$), nor by translation.

5.11 Appendix: Riemann integrals

If a function $f : [a, b] \to \mathbf{R}$ is Riemann integrable, it is Lebesque integrable and the two integrals agree. In this and the next section we review Riemann integrals and Riemann-Stieltjes integrals.

I do not intend to cover this in class. However you might find this helpful for comparing several notions of integration theory.

We know how to compute the areas of a triangle and a polygon. We then define the integral underneath a continuous curve by rectangle approximation. If $y = f(x)$, $x \ge 0$ is the curve, the approximation leads to $\int_0^t f(s)ds$, this is the Riemann integral $\int_a^b f(x)dx$. (Riemann integrals are covered in M2PM1: Real Analysis)

The Riemann integral of f is defined as follows. Take a partition $\Delta : a = x_0 < a_1 < \cdots < a_n = b$. Let M_i and m_i be respectively the supremum and infimum values of f on the interval $[a_{i-1}, a_i]$. Define

$$U(\Delta, f) = \sum_{i=1}^n M_i(a_i - a_{i-1}), \qquad L(\Delta, g) = \sum_{i=1}^n m_i(a_i - a_{i-1}).$$

$$U(f) = \inf_\Delta U(\Delta, f), \qquad L(f) = \sup_\Delta U(\Delta, f).$$

Definition 5.11.1 A bounded function $f : [a,b] \to \mathbf{R}$ is Riemann integrable, if $U(f) = L(f)$ and the common value is the Riemann integral of f on $[a,b]$.

A continuous function is Riemann integrable.

Theorem 5.11.1 *A bounded function f is Riemann integrable if and only if for every $\epsilon > 0$ there exists a partition such that $U(f,\Delta) - L(f,\Delta) < \epsilon$.*

Observe that the oscillation of f on $[a_{i-1}, a_i]$ is $\mathrm{Osc}(f) = M_i - m_i$,

$$U(f,\Delta) - L(f,\Delta) = \sum \mathrm{Osc}(f, [a_{i-1}, a_i])(a_i - a_{i-1}).$$

Define the Riemann sum:

$$R(f,\Delta) = \sum_{i=1}^{n} f(x_i^*)(a_i - a_{i-1}).$$

Theorem 5.11.2 *Suppose a bounded function $f : [a,b] \to \mathbf{R}$ is Riemann integrable. If Δ_n is a sequence of partitions with mesh goes to zero, then their Riemann sum $R(f,\Delta_n)$ converges to $\int_a^b f(x)dx$.*

This can be proved using Theorem 5.11.1.

Remark 5.11.3 If $f : [0,\infty) \to \mathbf{R}$ is measurable and Riemann integrable on every interval $[0,n]$ where $n \in \mathbf{N}$, then f is Lebesgue integrable on $[0,\infty)$ if and only if

$$\lim_{N\to\infty} \int_0^N |f(x)|dx < \infty.$$

If so, $\int_0^\infty f(x)dx = \int_{[0,\infty)} f d\lambda$.

The concept 'Lebesgue integrable' does not allow cancellation, while improper Riemann integration allows it. $\frac{\sin(x)}{x}$ is improper Riemann integrable, bot not Lebesgue integrable.

5.12 Appendix: Riemann-Stieltjes integrals*

Riemann-Stieltjes integrals aree defined in parallel to Riemann integrals. Suppose : $f[a,b] \to \mathbf{R}$ is a bounded function and g a monotone increasing function on $[a,b]$, for example $g(x) = x$ in which we case we are back to Riemann integrals.

Write $\Delta g_i = g(a_i) - g(a_{i-1})$. Define

$$U(\Delta, f, g) = \sum_{i=1}^{n} M_i \Delta g_i, \qquad L(\Delta, f, g) = \sum_{i=1}^{n} m_i \Delta g_i.$$

$$U(f,g) = \inf_{\Delta} U(\Delta, f, g), \qquad L(f,g) = \sup_{\Delta} U(\Delta, f, g).$$

Definition 5.12.1 If $U(f,g) = L(f,g)$, we say f is Riemann-Stieltjes integrable (with respect to g), we denote $f \in \mathbb{D}(g)$. The common value is denoted by $\int_a^b f dg$.

Theorem 5.12.1 (Integrability Criterion) *Suppose f is a bounded function and g a monotone increasing function on $[a,b]$. Then f is Riemann-Stieltjes integrable if and only if for any $\epsilon > 0$, there exists a partition Δ of $[a,b]$ such that*

$$U(\Delta, f, g) - L(\Delta, f, g) < \epsilon.$$

Corollary 5.12.2 *Suppose f is a bounded function with only a finite number of discontinuity points, and g a monotone increasing function on $[a,b]$, continuous at every point where f is discontinuous. Then f is Riemann-Stieltjes integrable.*

Proposition 5.12.3 *Suppose that $f, f_1, f_2 : [a,b] \to \mathbf{R}$ are bounded, Riemann-Stieltjes integrable with respect to a monotone increasing function. $g : [a,b] \to \mathbf{R}$. Let $c \in \mathbf{R}$.*

1. *If $f \in \mathbb{D}(g)$, then for any $c \in \mathbf{R}$, $cf \in \mathbb{D}(g)$and $\int_a^b (cf) dg = c \int_a^b f dg$.*
2. *$f_1 + f_2 \in \mathbb{D}(g)$, and $\int_a^b (f_1 + f_2) dg = \int_a^b f_1 dg + \int_a^b dg$.*
3. *If $f_1 \le f_2$ then $\int_a^b f_1 dg \le \int_a^b f_2 dg$.*
4. *If $f \in \mathbb{D}(g)$ and $c > 0$ is a constant, then $f \in \mathbb{D}(cg)$ and $\int_a^b f d(cg) = c \int_a^b f dg$.*
5. *If $f \in \mathbb{D}(g) \cap \mathbb{D}(g')$, then $f \in \mathbb{D}(g + g')$,*
$$\int_a^b f d(g + g') = \int_a^b f dg + \int_a^b ddg'.$$

Definition 5.12.2 If $g = g_1 - g_2$, where monotone increasing functions we define

$$\int f d(g_1 - g_2) = \int f dg_1 - \int f dg_2.$$

Riemann-Stieljes integrals do not exists if the integrator and the integrand have the same point of discontinuity. That $\int_a^b f dF$, $\int_b^c f dF$ exist in Riemann-Stieljes sense do not necessarily imply that $\int_a^c f dF$ is Riemann-Stieltjes integrable.

Proposition 5.12.4 *If $g : [a,b] \to \mathbf{R}$ is continuous and $F : [a,b] \to \mathbf{R}$ has bounded variation then the Riemann-Stieltjes integral $\int_a^b g dF$ and $\int_a^b F dg$ exist. If F is furthermore absolutely continuous and the integral equals to the corresponding Lebesgue integral:*

$$\int_a^b g dF = \int_a^b g F' dx.$$

5.13 Appendix: Functions with bounded variation*

The most commonly used integrators are functions of bounded variations. Functions of bounded variations are differences of increasing functions.

Given an interval $[a, b]$, define $P([a, b])$ to be the set of all partitions of $[a, b]$:

$$P([a, b]) = \{\underline{t} = (t_0, t_1, \dots, t_n) : a = t_0 < t_1 < \cdots < t_{n-1} < t_n = b, n \in N\}.$$

Let $\Delta\underline{t} = \max_{0 \le i \le n-1}(t_{i+1} - t_i)$ be the modulus of the partition. Write

$$\sum_{[u,v]\subset\Delta} |f(u) - f(v)| = \sum_{j=1}^{n} |f(t_j) - f(t_{j-1})|.$$

Definition 5.13.1 The total variation of a function f on $[a, b]$ is defined by

$$|f|_{TV}([a, b]) = \sup_{\Delta\in P([a,b])} \sum_{[u,v]\subset\Delta} |f(u) - f(v)|.$$

If this number is finite we say that f has bounded total variation over $[a, b]$.

The function $x \sin(\frac{1}{x})$ does not have bounded variation over any interval containing 0.

Note that f could have bounded total variation over $[a, b]$ without having a bounded variation over **R** (e.f. $f(x) = \sin x$).

Theorem 5.13.1 *A. function is of bounded variation on $[a, b]$ if and only if f is the difference of two monotone real-valued functions on $[a, b]$.*

Proof Let $f = f_1 - f_2$, where f, g are increasing functions, then

$$|f|_{TV}([a, b]) = \sup_{\Delta\in P([a,b])} \sum_{[u,v]\subset\Delta} |f_1(u) - f_1(v) - (f_2(u) - f_2(v))|.$$

□

For $f : [a, b] \to \mathbf{R}$, define a function:

$$|f|_{TV}(x) = |f|_{TV}([a, x]).$$

Example 5.10 1. If $f : \mathbf{R} \to \mathbf{R}$ is an increasing function, $|f|_{TV}(x) = f(x) - f(-\infty)$.

2. If f is locally Lipschitz continuous, then f is of bounded variation on any finite time interval.

3. The set of bounded variation function form a vector space.

If $f : \mathbf{R} \to \mathbf{R}$ is an increasing then it has left and right derivatives at every point and there are only a countable number of points of discontinuity. Furthermore f is differentiable almost everywhere.

Theorem 5.13.2 *Let $F : \mathbf{R} \to \mathbf{R}_+$ be a function of bounded variation, right continuous with $F(-\infty) = 0$. Then*

1. *F determines a Borel measure μ on* **R**.
2. *The Borel measure μ, with distribution function F, is absolutely continuous with respect to the Lebesgue measure dx if and only if*

$$F(x) = \int_0^x F'(t)dt.$$

3. *It is singular with respect to dx if and only if $F' = 0$.*

Chapter 6

L^p spaces

Lebesgue spaces, also known as L^p spaces, are a family of Banach spaces that play a fundamental role in functional analysis. Their construction, which relies heavily on Lebesgue's theory of integration, is outlined here. Throughout this chapter, we consider a fixed measure space $(\mathcal{X}, \mathcal{B}, \mu)$.

6.1 Holder's inequality

Let $f : \mathcal{X} \to \mathbf{R}$ measurable. For all $p \in [1, \infty)$, we set

$$\|f\|_p := \left(\int |f|^p d\mu \right)^{\frac{1}{p}} \in [0, \infty].$$

Moreover, for $p = \infty$, we set

$$\|f\|_\infty := \inf\{M \in [0, \infty] : |f| \leq M \; \mu - \text{a.e.}\} \in [0, \infty].$$

Remark 6.1.1 Note that $|f| \leq \|f\|_\infty$ μ-a.e. Indeed, by definition of $\|f\|_\infty$, for all $n \geq 1$

$$\mu\left(\left\{ |f| \geq \|f\|_\infty + \frac{1}{n} \right\} \right) = 0$$

so that

$$\mu(\{|f| > M\}) = \mu\left(\bigcup_{n \geq 1} \left\{ |f| \geq \|f\|_\infty + \frac{1}{n} \right\} \right) = 0.$$

So indeed $|f| \leq \|f\|_\infty$ μ-a.e, and $\|f\|_\infty$ is thus the smallest constant that bounds $|f|$ μ-a.e. We also call it the essential supremum of f.

Exercise 6.1.1 Show that, if μ is a probability measure, then as $p \uparrow \infty$, $\|f\|_p \to \|f\|_\infty$.

Remark 6.1.2 For all $p \in [1, \infty]$, $\| \cdot \|_p$ is homogeneous, i.e. for all $f : \mathcal{X} \to \mathbf{R}$ measurable and all $a \in \mathbf{R}$, $\|af\|_p = |a| \|f\|_p$.

Theorem 6.1.3 (Hölder's inequality (Cauchy-Schwarz if $p = q = 2$)) *Let $1 \leq p, q \leq \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$ (with the convention $\frac{1}{\infty} = 0$). Then, for all $f, g : \mathcal{X} \to \mathbf{R}^d$ measurable,*

$$\int |f \cdot g| d\mu \leq \|f\|_p \, \|g\|_q.$$

Proof If $p = 1$ and $q = \infty$, then $|fg| \leq |f| \, \|g\|_\infty$ μ-a.e., hence

$$\int |fg| d\mu \leq \int |f| \, \|g\|_\infty d\mu = \|f\|_1 \|g\|_\infty,$$

so the claim follows. The case where $p = \infty$ and $q = 1$ is symmetric, so we now focus on the case $p, q \in (1, \infty)$.

If $\|f\|_p = 0$, then $\int |f|^p \, d\mu = 0$, whence $f = 0$ μ -a.e.. Hence we also have $fg = 0$ μ-a.e., so $\int |fg| \, d\mu = 0$. Similarly, the inequality holds if $\|g\|_q = 0$, hence we may assume $\|f\|_p, \|g\|_q > 0$. By homogeneity, at the expanse of considering

$$\tilde{f} = \frac{f}{\|f\|_p}, \qquad \tilde{g} = \frac{g}{\|g\|_p}$$

we may even assume $\|f\|_p = \|g\|_q = 1$. We now claim the following statement, known as Young's inequality:

$$\forall x, y \geq 0, \qquad xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

To prove that claim, note that the case where x or y vanishes is straightforward, so we may assume $x, y > 0$. Then, setting $u = x^p$ and $v = y^q$, by convexity of the exponential, we have

$$u^{1/p} \, v^{1/q} \leq \frac{u}{p} + \frac{v}{q},$$

which yields the claim. Now, by Young's inequality, we have

$$\int |f| \, |g| \, d\mu \leq \int \left(\frac{|f|^p}{p} + \frac{|g|^q}{q} \right) d\mu = \frac{\|f\|_p^p}{p} + \frac{\|g\|_q^q}{q} = \frac{1}{p} + \frac{1}{q} = 1 = \|f\|_p = \|g\|_q,$$

which yields the requested inequality. □

6.2 Definition and basic properties of the L^p spaces

For $p \in [1, \infty]$, we set

$$\mathcal{L}_p = \{ f : \mathcal{X} \to \mathbf{R}^d : \text{ measurable}, \|f\|_p < \infty \}.$$

That $\mathcal{L}_p$ is a vector space is clear from the elementary inequality $|a+b|^p \leq C_p|a|^p + C_p|b|^p$, for all $p \in [1, \infty)$.

Question: Does $\|f\|_p$ define a norm on $\mathcal{L}_p$? As we saw it is homogeneous, and it also satisfies the triangle inequality which follows from the following:

Theorem 6.2.1 (Minkowski' inequality) *Let $f, g \in \mathcal{L}_p$ where $1 \leq p \leq \infty$. Then, $f + g \in \mathcal{L}_p$ and*

$$\|f+g\|_p \leq \|f\|_p + \|g\|_p.$$

Proof If $p = 1$ or $p = \infty$, the inequality follows at once from the triangle inequality, so we restrict our attention to the case $p \in (1, \infty)$. We may also assume WLOG that $\|f+g\|_p > 0$. Let $q \in (1, \infty)$ such that $\frac{1}{p} + \frac{1}{q} = 1$, then $p = 1 + \frac{p}{q}$. Hence

$$\begin{aligned}
\|f+g\|_p^p &= \int |f+g|^p \, d\mu = \int |f+g|^{1+\frac{p}{q}} \, d\mu \\
&= \int |f+g| \, |f+g|^{\frac{p}{q}} \, d\mu \\
&\leq \int |f| \, |f+g|^{\frac{p}{q}} \, d\mu + \int |g| \, |f+g|^{\frac{p}{q}} \, d\mu
\end{aligned}$$

where we used the triangle inequality in the last line. Using Hölder's inequality, noting that

$$\left(\int |f+g|^{\frac{p}{q} \cdot q} \, d\mu \right)^{\frac{1}{q}} = \left(\int |f+g|^p \, d\mu \right)^{\frac{1}{q}} = \|f+g\|_p^{\frac{p}{q}},$$

we therefore obtain

$$\|f+g\|_p^p \leq \|f\|_p \|f+g\|_p^{\frac{p}{q}} + \|g\|_p \|f+g\|_p^{\frac{p}{q}},$$

whence $\|f+g\|_p^{p(1-\frac{1}{q})} \leq \|f\|_p + \|g\|_p$. Since $p(1 - \frac{1}{q}) = p\frac{1}{p} = 1$, the claim follows. □

Unfortunately, $\|\cdot\|_p$ falls short of defining a norm on $\mathcal{L}_p$ because of the following observation: a function $f \in \mathcal{L}_p$ satisfies $\|f\|_p = 0$ iff $f = 0$, μ a.e. In general, a vast collection of non-zero functions may vanish μ a.e.

Example 6.1 If $(\mathcal{X}, \mathcal{F}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$ and $\mu = \delta_0$, then $f = 0$ μ-a.e. iff $f(0) = 0$.

To fix this degeneracy, we introduce an equivalence relation on functions as follows:

Definition 6.2.1 Two functions $f, g : \mathcal{X} \to \mathbf{R}$ are considered to be in the same equivalent class, and we write $f \sim g$, if $f = g$ μ-a.e.

Note that this equivalence relation is compatible with the normed vector space structure of $\mathcal{L}_p$. More precisely, if $f \sim f'$ and $g \sim g'$, then for all $a, b \in \mathbf{R}$, $af + bg \sim af' + bg'$. Moreover, if $f \sim f'$, then $\|f\|_p = \|f'\|_p$. This allows us to introduce a normed vector space L^p as follows:

Definition 6.2.2 For $1 \leq p \leq \infty$, the space $L^p(\mathcal{X}, \mathcal{F}, \mu)$ (often written $L^p(\mu)$, or just L^p) is the space of equivalence classes of measurable functions f such that $f \in \mathcal{L}_p$.

Loosely speaking,

$$L^p = \{f : f : \mathcal{X} \to \mathbf{R} \text{ measurable}, \ \|f\|_p < \infty\},$$

but keeping in mind that $f = g$ in L^p as soon as $f = g$ μ-a.e.

Proposition 6.2.2 *The L^p spaces are vector spaces, they are Banach spaces under the norm $\|\cdot\|_{L_p}$.*

That L^p is a Banach space means that it is complete for the norm $\|\cdot\|_{L_p}$, i.e. for every Cauchy sequence $f_n \in L^p$, there exists $f \in L^p$ such that

$$\lim_{n\to\infty} \|f_n - f\|_p = 0.$$

Proof We have already seen that $\|\cdot\|_p$ is homogeneous and verifies the triangle inequality (by Minkowski'sinequality). Moreover, note that $\|f\|_p = 0$ if and only if $|f| = 0$ μ a.e., i.e. if and only if $f = 0$ in L^p. Hence $\|\cdot\|_p$ indeed defines a norm on L^p.

We recall a useful fact: a normed vector space is a Banach space if and only if any absolutely convergent sequence is convergent. Let $f_n \in L^p$ with $\sum_{n=1}^{\infty} \|f_n\|_p = M < \infty$. Our aim is to show that there exists $f \in L^p$ such that $\sum_{k=1}^{n} f_k \underset{n\to\infty}{\longrightarrow} f$ in L^p. We first assume that $p < \infty$. We define

$$g_n(x) = \sum_{k=1}^{n} |f_k(x)|.$$

Then, by Minkowski's inequality, for all $n \geq 1$, $\|g_n\|_p \leq \sum_{k=1}^{n} \|f_k\|_p \leq M$. Hence, invoking the monotone convergence theorem, we obtain

$$\int \left(\sum_{k=1}^{\infty} |f_k(x)|\right)^p d\mu = \int \lim_{n\to\infty} |g_n|^p d\mu = \lim_{n\to\infty} \int |g_n|^p d\mu \leq M^p < \infty$$

Hence the first integral above is finite, so $\sum_{k=1}^{\infty} |f_k(x)| < \infty$ μ-a.e., i.e. the series $\sum_{k=1}^{\infty} |f_k(x)|$ converges absolutely in $\mathbf{R}$ $\mu(dx)$-a.e Then we may define $f(x) = \sum_{k=1}^{\infty} f_k(x)$ whenever the series converges absolutely, and $f(x) = 0$ otherwise. Then $f : \mathcal{X} \to \mathbf{R}$ is measurable. Set $h_n(x) = \sum_{k=1}^{n} f_k(x)$. Then $|h_n(x)| \leq \sum_{k=1}^{\infty} |f_k(x)| \in L^p$. By the dominated convergence theorem (since $|f_n - f|^p \leq 2^p(\sum_{k=1}^{\infty} |f_k(x)|)^p \in L^1$), $|f_n - f|^p$ converges in L_1, hence $f_n \to f$ in L_p. This completes the proof when $p < \infty$. When $p = \infty$, we have, μ-a.e.,

$$\sum_{k=1}^{\infty} |f_k(x)| \leq \sum_{k=1}^{\infty} \|f_k\| < \infty,$$

so again the series $\sum_{k=1}^{\infty} f_k(x)$ converges absolutely in $\mathbf{R}$ $\mu(dx)$-a.e. Then we may define $f(x) = \sum_{k=1}^{\infty} f_k(x)$ whenever the series converges absolutely, and $f(x) = 0$ otherwise. We have the bound $|f| \leq \sum_{k=1}^{\infty} \|f_k\|_\infty < \infty$ μ-a.e., so $f \in L^\infty$. Moreover, $\mu(dx)$-a.e.,

$$|\sum_{k=1}^{n} f_k(x) - f(x)| \leq \sum_{k=n+1}^{\infty} |f_k(x)| \leq \sum_{k=n+1}^{\infty} |f_k|_\infty,$$

so that $|\sum_{k=1}^{n} f_k - f|_\infty \leq \sum_{k=n+1}^{\infty} |f_k|_\infty \underset{n\to\infty}{\longrightarrow} 0$, i.e. $\sum_{k=1}^{n} f_k$ converges to f in L^∞. $\square$

Remark 6.2.3 In the proof above we actually established a stronger fact: if the series $\sum_{k=1}^{n} f_k$ converges absolutely in L^p, there exists $f \in L^p$ such that $\sum_{k=1}^{n} f_k \underset{n\to\infty}{\longrightarrow} f$ μ-a.e. *and* in L^p. In particular, we deduce the following, useful result.

Corollary 6.2.4 *Let* $p \in [1, \infty)$. *If a sequence* $(f_n)_{n\geq 1}$ *converges to* f *in* L^p*, we can extract a subsequence* $(f_{\varphi(n)})_{n\geq 1}$ *converging to* f μ*-a.e.*

Proof Since $\|f_n - f\|_p \underset{n\to\infty}{\longrightarrow} 0$, we can extract a subsequence $(f_{\varphi(n)})_{n\geq 1}$ such that $\|f_{\varphi(n)} - f\|_p \leq 2^{-n}$, so that the series $\sum_{k=1}^{n}(f_{\varphi(k)} - f_{\varphi(k-1)})$ converges absolutely in L^p (where we set $f_{\varphi(0)} := 0$). Thus, by the previous remark, $f_{\varphi(n)} = \sum_{k=1}^{n}(f_{\varphi(k)} - f_{\varphi(k-1)})$ converges μ-a.e. to its limit in L^p, which is f. $\square$

Exercise 6.2.1 If $p = +\infty$, and if a sequence $(f_n)_{n\geq 1}$ converges to f in L^∞, then $(f_n)_{n\geq 1}$ converges to f μ-a.e. (without need to extract a subsequence).

6.3 Density results

When we need to prove a certain property for all functions in L^p, where $p \in [1, \infty]$, it is often convenient to prove the property on a smaller set of functions, and argue by a density argument that the property remains true on the whole space L^p. We state here some fundamental density results.

Theorem 6.3.1 *Let* $p \in [1, \infty]$. *If* $f \in L^p(\mathcal{X})$ *then there exists a sequence of simple functions* f_n *such that* $\lim_{n\to\infty} \|f_n - f\|_p = 0$. *In words,* $\mathcal{S} \cap L^p(\mathcal{X})$ *is dense in* $L^p(\mathcal{X})$.

Proof First assume that $p < \infty$. We may assume that $f \geq 0$ (standard methods passes to $f = f^+ - f^-$). Let $f_n \leq f$ be a sequence of non-negative increasing simple functions converging to f. Define

$$g_n = f^p - (f - f_n)^p.$$

Then $g_n \to f^p$, and g_n is non-negative and increasing , by the monotone convergence theorem,

$$\lim_{n\to\infty} \int \Big(f^p - (f - f_n)^p\Big) d\mu = \int f^p d\mu$$

i.e.

$$\lim_{n\to\infty}\int (f-f_n)^p d\mu = 0,$$

and the claim follows. If $p=\infty$, then $|f|\le \|f\|_\infty<\infty$ μ-a.e. Setting $\tilde f = \mathbf{1}_{\{|f|\le\|f\|_\infty\}}$, then $f=\tilde f$ μ-a.e. Moreover $\tilde f$ is bounded, so there exists a sequence simple functions f_n converging to f uniformly. Therefore

$$\|f_n-f\|_\infty = \|f_n-\tilde f\|_\infty \underset{n\to\infty}{\longrightarrow} 0,$$

thus concluding the proof. □

Corollary 6.3.2 *Let $1\le p<\infty$, then $L^p(\mathbf{R}^d)$ is a separable metric space.*

Proof The subset D of simple functions $\sum_{i=1}^n a_i\mathbf{1}_{A_i}$ where $A_i=\Pi_{j=1}^d(c_{i,j},d_{i,j})$ with $a_i,c_{i,j},d_{i,j}\in Q$ is a countable subset of $L^p(\mathbf{R}^d)$. We also claim that D is dense in $L^p(\mathbf{R}^d)$. Let us prove this claim. For convenience of notation, let us assume that $d=1$ (the general case is similar). We know that the subspace $\mathcal{S}\cap L^p(\mathbf{R})$ is dense in $L^p(\mathbf{R})$. Note that $\mathcal{S}\cap L^p(\mathbf{R})$ is the subspace of functions of the form $\sum_{i=1}^n a_i\,\mathbf{1}_{A_i}$, where $n\ge 1$, $a_i\in\mathbf{R}\setminus\{0\}$, and $A_i\in\mathcal{B}(\mathbf{R})$ with $\lambda(A_i)<\infty$. Since $\mathbf{Q}$ is a dense subset of $\mathbf{R}$, the claim will follow upon showing that if $A\in\mathcal{B}(\mathbf{R})$ satisfies $\lambda(A)<\infty$, then

$$\forall\epsilon>0,\quad \exists f\in D,\quad \|\mathbf{1}_A-f\|_p\le\epsilon. \tag{6.1}$$

Note moreover that for all $A\in\mathcal{B}(\mathbf{R})$ such that $\lambda(A)<\infty$, by the Dominated Convergence Theorem, $\mathbf{1}_{A\cap(-N,N)} \underset{N\to\infty}{\longrightarrow} \mathbf{1}_A$ in $L^p(\mathbf{R})$, so it is enough to prove (6.1) holds for all subset $A\in\mathcal{B}((-N,N))$, for $N\ge 1$.

Given $N\ge1$, let us therefore consider $\mathcal{A}$, the collection of subsets $A\in\mathcal{B}((-N,N))$ satisfying (6.1). $\mathcal{A}$ is a λ-system, indeed:

- $\mathbf{1}_{(-N,N)}\in D$, hence $(-N,N)\in\mathcal{A}$
- if $A,B\in\mathcal{A}$ and $A\subset B$, then for all $\epsilon>0$ there exist $f,g\in D$ such that $\|\mathbf{1}_A-f\|_p<\epsilon/2$, and $\|\mathbf{1}_B-g\|_p<\epsilon/2$. Since $\mathbf{1}_{B\setminus A}=\mathbf{1}_B-\mathbf{1}_A$, it follows by the Minkowski inequality that $\|\mathbf{1}_{B\setminus A}-(g-f)\|_p<\epsilon$; since $g-f\in D$, this shows that $B\setminus A\in\mathcal{A}$.
- if $(A_n)_{n\ge1}$ is a non-decreasing sequence of elements of $\mathcal{A}$, and $A=\cup_{n\ge1}A_n$, then $\mathbf{1}_{A_n}\underset{n\to\infty}{\longrightarrow}\mathbf{1}_A$ in L^p by the Dominated Convergence Theorem. Hence, for all $\epsilon>0$, there exists a $n\ge1$ such that $\|\mathbf{1}_{A_n}-\mathbf{1}_A\|_p<\epsilon/2$. Since $A_n\in\mathcal{A}$, there exists $f\in D$ such that $\|\mathbf{1}_{A_n}-f\|_p<\epsilon/2$. By Minkowski's inequality, we deduce that $\|\mathbf{1}_A-f\|_p<\epsilon$, so $A\in\mathcal{A}$.

Thus $\mathcal{A}$ is a λ-system, and it contains $\mathcal{C}:=\{(c,d): c,d\in\mathbf{Q}\cap[-N,N], c<d\}$. Since $\mathcal{C}$ is a π-system and $\sigma(\mathcal{C})=\mathcal{B}((-N,N))$, by the $\pi-\lambda$ Theorem, we deduce that $\mathcal{A}=\mathcal{B}((-N,N))$. Hence D is indeed a dense countable subset $L^p(\mathbf{R}^d)$, which is therefore separable.

□

Remark 6.3.3 Beware that $L^\infty(\mathbf{R}^d)$ is *not* separable.

Theorem 6.3.4 *Let $1 \le p < \infty$. The space $C_c^\infty(\mathbf{R}^d)$ of C^∞ functions on $\mathbf{R}^d$ with compact support is a dense subspace of $L^p(\mathbf{R}^d)$.*

Proof First note that $C_c^\infty(\mathbf{R}^d) \subset L^p(\mathbf{R}^d)$. Indeed, if $f \in C_c^\infty(\mathbf{R}^d)$, then there exists $N \ge 1$ such that f is supported in $[-N,N]^d$. Since moreover f is continuous, it is therefore bounded. Thus

$$\int |f|^p \, d\lambda = \int |f|^p \, \mathbf{1}_{[-N,N]^d} \, d\lambda \le \|f\|_\infty^p \, \lambda([-N,N]^d) < \infty,$$

so $f \in L^p(\mathbf{R}^d)$. We now prove that $C_c^\infty(\mathbf{R}^d)$ is dense in $L^p(\mathbf{R}^d)$. By the proof of Corollary 6.3.2, the subset D of simple functions $\sum_{i=1}^n a_i \mathbf{1}_{A_i}$ where $A_i = \Pi_{j=1}^d (c_{i,j}, d_{i,j})$ with $a_i, c_{i,j}, d_{i,j} \in \mathbf{Q}$ is dense in $L^p(\mathbf{R}^d)$. Hence it suffices to show that for any subset A of the form $\Pi_{j=1}^d (c_j, d_j)$, with $c_j, d_j \in \mathbf{Q}$, there exists a sequence of functions $f_n \in C_c^\infty(\mathbf{R}^d)$ such that $f_n \underset{n\to\infty}{\longrightarrow} \mathbf{1}_A$ in L^p. We now recall the following useful fact: there exists a non-decreasing sequence of functions $\varphi_n \in C_c^\infty(\mathbf{R})$ such that

- $0 \le \varphi_n \le \mathbf{1}_{(0,1)}$ for all $n \ge 1$
- $\varphi_n \underset{n\to\infty}{\longrightarrow} \mathbf{1}_{(0,1)}$ pointwise from below.

Given $A = \Pi_{j=1}^d (c_j, d_j)$, with $c_j, d_j \in \mathbf{Q}$, and $c_j < d_j$, we now define $f_n \in C_c^\infty(\mathbf{R}^d)$ by setting

$$f_n(x) = \Pi_{j=1}^d \, \varphi_n\left(\frac{x_j - c_j}{d_j - c_j}\right), \qquad x = (x_1, \ldots, x_d) \in \mathbf{R}^d.$$

Then we have

- $0 \le f_n \le \mathbf{1}_A$ for all $n \ge 1$
- $f_n \underset{n\to\infty}{\longrightarrow} \mathbf{1}_A$ pointwise from below.

By the Dominated Convergence Theorem, we deduce that $f_n \underset{n\to\infty}{\longrightarrow} \mathbf{1}_A$ in L^p, as requested. $\square$

6.4 The case of finite measures

Here we state some fundamental properties specific to the case when μ is a finite measure.

Proposition 6.4.1 *If μ is a finite measure, then for all $r, r' \in [1,\infty]$ with $r < r'$, $L^{r'}(\mu) \subset L^r(\mu)$.*

Proof If $r' = \infty$, and $f \in L^\infty(\mu)$, then $f \leq \|f\|_\infty$ μ a.e., hence for all $r \in [1, \infty)$,

$$\int |f|^r \, d\mu \leq \|f\|_\infty^r \, \mu(\mathcal{X}) < \infty$$

so $f \in L^r(\mu)$, thus proving the claim. Let us now assume $r' < \infty$. Let $p = \frac{r'}{r}$ and $q = \frac{r'}{r'-r}$, so that p, q are two conjugate numbers in $(1, \infty)$. If $f \in L^{r'}(\mu)$, by Hölder's inequality

$$\|f\|_r = \left(\int |f|^r \, 1 \, d\mu\right)^{1/r} \leq \left(\int |f|^{pr} \, d\mu\right)^{1/pr} \left(\int 1 \, d\mu\right)^{1/qr} = \|f\|_{r'} \, \mu(\mathcal{X})^{\frac{r'-r}{rr'}} < \infty,$$

so $f \in L^r(\mu)$, thus proving the claim. □

Remark 6.4.2 In the special case where μ is a *probability* measure, the proof above shows that we furthermore have $\|f\|_r \leq \|f\|_{r'}$ for all $r' > r \geq 1$ and all $f \in L^{r'}(\mu)$. In probabilistic notations, $E(|X|^r)^{1/r} \leq E(|X|^{r'})^{1/r'}$ for any random variable $X \in L^{r'}(\mu)$.

Terminology: If X is a random variable on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$, for all $p \in [1, \infty)$, $E(|X|^p)$ is called the p-th moment of X.

We will now state an inequality that plays an important role in probability theory: Jensen's inequality. We first recall the definition of a convex function.

Definition 6.4.1 A function $\varphi : \mathbf{R}^d \to \mathbf{R}$ is convex if

$$\varphi(tx + (1-t)y) \leq t\varphi(x) + (1-t)\varphi(y)$$

for any $x, y \in \mathbf{R}^d$ and for any $t \in [0, 1]$.

The following are convex functions on $\mathbf{R}^d$: $\varphi(x) = |x|$, $|x|^p$, $p > 1$. If $d = 1$, e^x is convex, and so are $\varphi(x) = x^p$ for $p > 1$. Moreover, if φ is twice differentiable, φ is convex if and only if $\varphi'' \geq 0$.

Theorem 6.4.3 (Jensen's Inequality) *If $\varphi : \mathbf{R} \to \mathbf{R}_+$ is a convex function and μ is a probability measure, then for all $f \in \mathcal{L}^1(\mathcal{X}, \mathcal{F}, \mu)$,*

$$\varphi\left(\int f d\mu\right) \leq \int \varphi(f) \, d\mu.$$

Proof It is a fact that if φ is convex then there exists a constant $c \in \mathbf{R}$ such that

$$\forall y \in \mathbf{R}, \qquad \varphi(y) \geq \varphi\left(\int f d\mu\right) + c\left(y - \int f d\mu\right).$$

Take $y = f(x)$,

$$\varphi(f(x)) \geq \varphi\left(\int f d\mu\right) + c\left(f(x) - \int f d\mu\right).$$

Integrate this over with respect to the probability measure μ (see Exercise 5.5.1) to conclude. □

Example 6.2 Let $x_1, \dots x_n$ be points in **R** and suppose that $\mu = \sum_{i=1}^n \lambda_i \delta_{x_i}$ is a probability measure on **R**. Then apply Jensen to $f(x) = x$ and any convex φ:

$$\varphi(\sum_{i=1}^n \lambda_i x_i) \leq \sum_{i=1}^n \lambda_i \varphi(x_i).$$

6.4.1 Mode of convergence

To wrap up we note the commonly used notions of convergence. Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space with σ-finite measure, and $f_n, f : \mathcal{X} \to$ **R** be Borel measurable functions.

Definition 6.4.2 We say f_n converges to f in L_p if

$$\int |f_n - f|^p d\mu \to 0.$$

Definition 6.4.3 We say f_n converges to f in measure if for any $\epsilon > 0$,

$$\lim_{n\to\infty} \mu(x : |f_n(x) - f(x)| > \epsilon) = 0.$$

Definition 6.4.4 We say f_n converges to f a.s. if there exists a null set N and for any $x \notin N$,

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

Remark 6.4.4 If $f_n \to f$ in L_p it converges in measure.

Chapter 7

Product measures and Fubini's Theorem

Throughout this section let $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ be two measure spaces. A (measurable) product set is of the form $A \times B$ where $A \in \mathcal{F}$ and $B \in \mathcal{G}$. Denote by C the collection of products measurable set.

$$C = \{A \times B: \ A \in \mathcal{F}, B \in \mathcal{G}\}.$$

The product σ-algebra is generated by such sets.

$$\mathcal{F} \otimes \mathcal{G} = \sigma(C).$$

Can we assign a measure to $\mathcal{F} \otimes \mathcal{G}$ that is consistent to μ and ν?

If $\mathcal{X} = \{a_1, \ldots, a_n\}$, we use the power set σ-algebra. If we assign any non-negative values to a_i, say p_i, this determines a measure on $2^{\mathcal{X}}$ (no relations is needed on $\{p_i\}$, as intersections of singletons are emptysets, the union of two singletons is no longer a singleton). The power σ-algebra has precisely 2^n elements. Every measure on the finite space $\mathcal{X}$, with the power σ-algebra, is determined uniquely by their values of the singleton sub-sets.

Let $\mathcal{Y} = \{b_1, \ldots, b_m\}$, then $\mathcal{X} \times \mathcal{Y} = \{(a_i, b_j): \quad 1 \leq i \leq n, 1 \leq j \leq m\}$, and every element of $\mathcal{X} \times \mathcal{Y}$ is contained in $2^{\mathcal{X}} \otimes 2^{\mathcal{Y}}$. Thus $2^{\mathcal{X}} \otimes 2^{\mathcal{Y}} = 2^{\mathcal{X} \times \mathcal{Y}}$. Hence by assigning a value to every $\{(a_i, b_j)\}$, this is a product set, this determines a measure on the product σ-algebra. Given μ on $\mathcal{X}$ and ν on $\mathcal{Y}$, we define the product measure by

$$\mu \times \nu(\{(a_i, b_j)\}) = \mu(\{a_i\})\nu(\{b_j\}).$$

Similarly if $\mathcal{X}$ is given by a partition $\{A_1, \ldots, A_n\}$, and $\mathcal{Y}$ is given a partition $\{B_1, \ldots, B_m\}$, and if we set $\mathcal{F} = \sigma(\{A_1, \ldots, A_n\})$ and $\mathcal{G} = \sigma(\{B_1, \ldots, B_m\})$. Then there are 2^n and 2^m elements in $\mathcal{F}$ and $\mathcal{G}$ respectively. If we assign values $\mu(A_i)$, this determine a measure on $\mathcal{X}$ (because the intersections of the A_i are emptyset, no relation between these numbers are neeeded for extending these to every element of $\mathcal{F}$). Every measure on $\mathcal{F}$ is obtained in this way. Also $\mathcal{F} \otimes \mathcal{G} = \sigma(\{A_i \times B_j\})$. It is clear that

$A_i \times B_j$ is a partition of $\mathcal{X} \times \mathcal{Y}$. Thus this is totally analogous to the finite state spaces discussed earlier. Furthermore given μ, ν on $\mathcal{F}$ and $\mathcal{G}$ respectively, the product measure on $\mathcal{F} \otimes \mathcal{G}$ will be defined by

$$\mu \times \mathcal{G}(A_i \times B_j) = \mu(A_i)\nu(B_j).$$

If $\mathcal{F}$ and $\mathcal{G}$ are not finite σ-algebras, they are uncountable, the construction for the product measure becomes much more involved.

7.1 Product measures

Recall the definition of elementary family in Definition 2.2.4.

Lemma 7.1.1 *The collection C of product sets is an elementary family. The collection $\mathcal{A}$ of a finite number of disjoint unions of product sets is an algebra.*

Proof The empty set is in C. Also,

$$(A \times B) \cap (A' \times B') = (A \cap A') \times (B \cap B'), \qquad (A \times B)^c = (A \times B^c) \cup (A^c \times B) \cup (A^c \times B^c).$$

Thus, C is closed under taking complement, and the complement of a set is the disjoint union of a finite number of product sets. By exercise 2.2.4, the collection of finite number of disjoint union of sets from C is an algebra. □

Let us define a function $\mu \times \nu : \mathcal{A} \to [0, \infty]$ by the following rules: If $E = \cup_{j=1}^{n}(A_i \times B_j)$ where $A_i \in \mathcal{F}$ and $B_i \in \mathcal{G}$, then we want to define

$$\mu \times \nu(E) = \sum_{j=1}^{n} \mu(A_i)\nu(B_i). \tag{7.1}$$

We must show that this is independent of the expression of E into disjoint unions of product sets.

Lemma 7.1.2 *Let $E = A \times B \in C$. Suppose that $E = \cup_{j=1}^{\infty}(A_j \times B_j)$ where $A_i \in \mathcal{F}$ and $B_i \in \mathcal{G}$, where $\{A_j \times B_j\}$ are disjoint sets. then*

$$\mu(A)\nu(B) = \sum_{j=1}^{\infty} \mu(A_j)\nu(B_i).$$

Proof Observe that $\mathbf{1}_{A\times B}(x, y) = \mathbf{1}_A(x)\mathbf{1}_B(y)$ for $x \in \mathcal{X}, B \in \mathcal{Y}$. Also,

$$\mathbf{1}_{A\times B}(x, y) = \sum_{i=1}^{\infty} \mathbf{1}_{A_i \times B_i}(x, y) = \sum_{i=1}^{\infty} \mathbf{1}_{A_i}(x)\mathbf{1}_{B_i}(y).$$

Holding y fixed, we may integrate $\mathbf{1}_A(x)\mathbf{1}_B(y) = \sum_{i=1}^{\infty} \mathbf{1}_{A_i}(x)\mathbf{1}_{B_i}(y)$ with respect to μ.

$$\int_{\mathcal{X}} \mathbf{1}_A(x)\,\mathbf{1}_B(y)\,d\mu(x) = \int_{\mathcal{X}} \sum_{i=1}^{\infty} \mathbf{1}_{A_i}(x)\mathbf{1}_{B_i}(y)d\mu(y).$$

Use Corollary 5.4.8 for exchanging the summation and integration,

$$\mu(A)\mathbf{1}_B = \sum_{i=1}^{\infty} \mu(A_i)\mathbf{1}_{B_i}.$$

We use the convention

$$0 \cdot \infty = 0.$$

(So if $\mu(A) = \infty$, then the left hand side is ∞ if $y \in B$ and 0 otherwise. The right hand side is interpreted in the same way. Observe that $B_i \subset B$ and $A_i \subset A$.)

Integrating both sides from the above line w.r.t. ν we obtain $\mu(A)\nu(B) = \sum_{i=1}^{\infty} \mu(A_i)\nu(B_i)$, completing the proof. □

Lemma 7.1.3 *Suppose that* $\bigcup_{i=1}^{\infty}(A_i \times B_i)$*,* $\bigcup_{j=1}^{\infty}(C_j \times D_j)$ *are disjoint unions and*

$$\bigcup_{i=1}^{\infty}(A_i \times B_i) = \bigcup_{j=1}^{\infty}(C_j \times D_j),$$

then

$$\sum_{i=1}^{\infty} \mu(A_i)\nu(B_i) = \sum_{j=1}^{\infty} \mu(C_j)\nu(D_j).$$

Proof Let $E = \bigcup_{j=1}^{\infty}(A_i \times B_i)$. Then,

$$\begin{aligned}
\bigcup_{i=1}^{\infty}(A_i \times B_i) &= (\bigcup_{j=1}^{\infty}(A_i \times B_i)) \cap E \\
&= \bigcup_{i=1}^{\infty}\bigcup_{j=1}^{\infty}(A_i \times B_i) \cap (C_j \times D_j) \\
&= \bigcup_{i=1}^{\infty}\bigcup_{j=1}^{\infty}(A_i \cap C_j) \times (B_i \cap D_j).
\end{aligned}$$

The last is also a union of disjoint sets.

For each i, we apply the previous lemma to

$$E_i = A_i \times B_i = \bigcup_{j=1}^{\infty}(A_i \cap C_j) \times (B_i \cap D_j)$$

to see

$$\mu(A_i \times B_i) = \sum_{j=1}^{\infty} \mu(A_i \cap C_j)\nu(B_i \cap D_j)$$

So,

$$\sum_{i=1}^{\infty} \mu(A_i \times B_i) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu(A_i \cap C_j)\nu(B_i \cap D_j).$$

Similarly

$$\sum_{j=1}^{\infty} \mu(C_j \times D_j) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \mu(A_i \cap C_j)\nu(B_i \cap D_j).$$

This finishes the proof. □

The map (7.1) defines a pre-measure on $\mathcal{A}$. Firstly $\mu \times \nu(\phi) = 0$, secondly it has finite additive property on $\mathcal{A}$. Furthermore, suppose that $E = \bigcup_{i=1}^{\infty} E_i$, where $E_i \in \mathcal{A}$ are disjoint, with the property that $E \in \mathcal{A}$, we show that $\mu(E) = \sum_{i=1}^{\infty} \mu(E_i)$. On one hand, each E_i is a finite disjoint union of product sets, we pull these product sets , that composes of E_i, together: they are all disjoint. Relabel these disjoint product sets we see $E = \bigcup_{i=1}^{\infty} A_i \times B_i$. Re-arrange it, we see that $\mu(E) = \sum \mu_i(E_i)$.

With this we define an outer measure μ^*, by (3.1):

$$\mu^*(E) = \inf\left\{\sum_{j=1}^{\infty} \mu \times \nu(E_j) : E \subset \bigcup_{j=1}^{\infty} E_j, \quad E_j \in \mathcal{A}\right\}.$$

By the Caratheodory theorem, the outer measure μ^* is a measure on $\mathcal{F} \otimes \mathcal{G}$ and agrees with $\mu \times \nu$ on $\mathcal{A}$. This measure on $\mathcal{F} \otimes \mathcal{G}$ will be called the product measure. If μ and ν are σ-finite, there exists a unique measure such that the measure of product sets is the product of the measures on each factor.

If μ, ν are finite, so is $\mu \times \nu(\mathcal{X} \times \mathcal{Y}) = \mu(\mathcal{X})\nu(\mathcal{Y}) < \infty$. If they are both σ-finite, so is $\mu \times \nu$. Indeed, there exist $A_n \uparrow \mathcal{X}$ and $B_n \uparrow \mathcal{Y}$ such that $\mu(A_n) < \infty$, $\nu(B_n) < \infty$ for every n. Now $\{A_i \times b_j\}$ is an increasing sequence converging to $\mathcal{X} \times \mathcal{Y}$ with finite measure, hence $\mu \times \nu$ is σ-finite.

Remark: Since a countable union of measures of $\mathcal{A}$ is a countable union of product sets,

$$\mu^*(E) = \inf\left\{\sum_{j=1}^{\infty} \mu(A_j)\nu(A_j) : E \subset \bigcup_{j=1}^{\infty} A_j \times B_j, \quad A_j \in \mathcal{F}, B_j \in \mathcal{F} \in \mathcal{A}\right\}.$$

Remark 7.1.4 Similar constructions can be used to obtain a product measure on a finite number of copies of product spaces and their product σ-algebras.

Let us take for example three measure spaces $(\mathcal{X}_i, \mathcal{F}_i, \mu_i)$. Recall that

$$\otimes_{i=1}^{3} \mathcal{F}_i = \sigma(\{\Pi_{i=1}^{3} A_i : A_i \in \mathcal{F}_i,\ i = 1, 2, 3\}).$$

We define $\mu_1 \times \mu_2 \times \mu_3$ by

$$\mu_1 \times \mu_2 \times \mu_3(A_1 \times A_2 \times A_3) = \Pi_{i=1}^3 \mu_i(A_i).$$

Proposition 7.1.5 *Let $(\mathcal{X}_i, \mathcal{F}_i, \mu_i)$ be measurable spaces. We have,*

$$(\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3 = \otimes_{i=1}^3 \mathcal{F}_i.$$

If μ_i are σ-finite, then

$$(\mu_1 \times \mu_2) \times \mu_3 = \mu_1 \times \mu_2 \times \mu_3.$$

Proof It is clear that, $\left\{\Pi_{i=1}^3 A_i : A_i \in \mathcal{F}_i,\ i = 1, 2, 3\right\} \subset \{E \times C : E \in \mathcal{F}_1 \otimes \mathcal{F}_2, C \in \mathcal{F}_3\}$. Hence

$$\otimes_{i=1}^3 \mathcal{F}_i \subset (\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3.$$

On the other hand, for any $C \in \mathcal{F}_3$,

$$\{E \times C : E \in \mathcal{F}_1 \otimes \mathcal{F}_2\} \subset \otimes_{i=1}^3 \mathcal{F}_i.$$

so

$$\{E \times C : E \in \mathcal{F}_1 \otimes \mathcal{F}_2, C \in \mathcal{F}_3\} \subset \cup_{C \in \mathcal{F}_3} \{E \times C : \mathbf{E} \in \otimes_{i=1}^2 \mathcal{F}_i\} \subset \otimes_{i=1}^3 \mathcal{F}_i.$$

Thus

$$(\mathcal{F}_1 \otimes \mathcal{F}_2) \otimes \mathcal{F}_3 = \sigma(\{E \times C : E \in \mathcal{F}_1 \otimes \mathcal{F}_2, C \in \mathcal{F}_3\}) \subset \otimes_{i=1}^3 \mathcal{F}_i.$$

Now define $\mu_1 \times \mu_2$ on $\mathcal{F}_1 \otimes \mathcal{F}_2$. Then, $(\mu_1 \times \mu_2) \times \mu_3$ and $\mu_1 \times \mu_2 \times \mu_3$ agree on the product sets. These product set determines σ-finite measures. $\square$

Example 7.1 Let λ be the Lebesgue measure, the completion of the product measure on $\mathbf{R} \times \mathbf{R}$ is called the Lebesgue measure on $\mathbf{R}^2$. This is denoted by the same letter or by λ^2. Our question is: given $f : \mathbf{R}^2 \to \mathbf{R}$ measurable, what is

$$\int_{\mathbf{R}^2} f d\lambda^2?$$

Remark 7.1.6 If each factor $(\mathcal{X}, \mathcal{F}, \mu)$ and $(\mathcal{Y}, \mathcal{G}, \nu)$ are complete, (i.e. the σ-algebra is complete with respect to the given measure), the product σ-algebra is often NOT complete with respect to the product measure. We can of course complete the product σ-algebra under the product measure.

Exercise 7.1.1 Give an example of a Borel measure on $\mathbf{R}^2$ (i.e. on the Borel σ-algebra) that is not a product measure.

7.1.1 Sections of subsets of product spaces

If $E \subset \mathcal{X} \times \mathcal{Y}$, for $x \in \mathcal{X}$, we define the section

$$E_x = \{y \in \mathcal{Y} : (x,y) \in E\}.$$

Similarly for $y \in \mathcal{Y}$ we define

$$E^y = \{x \in \mathcal{X} : (x,y) \in E\}.$$

Proposition 7.1.7 *If $E \in \mathcal{F} \otimes \mathcal{G}$, then*

1. *$E_x \in \mathcal{G}$ for every $x \in \mathcal{X}$ and $\nu(E_x)$ is measurable.*
2. *$E^y \in \mathcal{F}$ for every $y \in \mathcal{Y}$ and $\mu(E^y)$ is measurable.*
3. *Also,*
$$\mu \times \nu(E) = \int_{\mathcal{X}} \nu(E_x)\mu(dx) = \int_{\mathcal{Y}} \mu(E^y)\nu(dy).$$

Proof (1) We first assume that the measures are finite. Let

$$\mathbb{D} = \{E \in \mathcal{F} \otimes \mathcal{G} \ : \ \text{conclusion holds}\}.$$

Observe that

$$(A \times B)_x = \begin{cases} B, & x \in A, \\ \phi, & x \notin A. \end{cases}$$

Take μ measure of the above,

$$\nu((A \times B)_x) = \begin{cases} \nu(B), & x \in A, \\ 0, & x \notin A. \end{cases}$$

Integrate w.r.t. x, we see

$$\nu((A \times B)_x)\mu(dx) = \nu(B)\mu(A) = \mu \times \nu(A \times B).$$

A similar conclusion holds for the y-sections. This means $\mathbb{D}$ contains the set of products sets, which is a π-system.

If $A \subset B$, $A, B \in \mathbb{D}$ then

$$(A \setminus B)_x = A_x \setminus B_x \in \mathcal{G} \qquad (A \setminus B)^y = A^y \setminus B^y \in \mathcal{F}.$$

$$\int \nu((A \setminus B)_x)\mu(dx) = \int \nu(A_x)d\mu(x) - \int \nu(B_x)d\mu(y),$$

A similar conclusion for the y-sections. Hence $A \setminus B \in \mathbb{D}$. If $A_n \in \mathbb{D}$ is an increasing sequence,

$$(\bigcup_n A_n)_x = \bigcup (A_n)_x \in \mathcal{G}.$$

Thus $\mathbb{D}$ is a λ system and thus equals $\mathcal{F} \otimes \mathcal{G}$, as it contains a generating set which is a π-system. This concludes the finite measure case.

(2) Suppose that the measures are σ-finite, we take $A_i \times B_i$ increasing with finite measure, then apply the proposition to $E \cap (A_i \times B_i)$, we see that

$$(E \cap (A_i \times B_i))_x = \begin{cases} E_x \cap B_i, & \text{if } x \in A_i \\ \phi, & \text{otherwise,} \end{cases}$$

is measurable for every x, then so is $E_x = \cup_i (E \cap (A_i \times B_i))_x$. Similarly conclusions for the y-sections. Since,

$$\mu \times \nu(E \cap (A_i \times B_i)) = \int_{\mathcal{X}} \nu((E \cap (A_i \times B_i))_x)\mu(dx) = \int_{\mathcal{Y}} \mu((E \cap (A_i \times B_i))_x^y)\nu(dy),$$

we may take $i \to \infty$ and use the monotone convergence theorem to conclude. □

Proposition 7.1.8 *If $f : \mathcal{X} \times \mathcal{Y} \to \mathbf{R}$ is measurable, then so are the following functions:*

$$f_x = f(x, \cdot) : \mathcal{Y} \to \mathbf{R}, \qquad f^y = f(\cdot, y) : \mathcal{X} \to \mathbf{R}$$

are measurable for every $x \in \mathcal{X}$ and for every $y \in \mathcal{Y}$.

Proof For any $B \in \mathcal{B}(\mathbf{R})$,

$$(f^y)^{-1}(B) = \{x : f(x, y) \in B\} = (f^{-1}(B))^y \in \mathcal{F},$$
$$(f_x)^{-1}(B) = \{y : f(x, y) \in B\} = (f^{-1}(B))_x \in \mathcal{G}.$$

This conclusion follows from Proposition 7.1.7. □

7.2 Fubini's Theorem

Theorem 7.2.1 (The Fubini-Tonelli Theorem) *Let μ and ν be σ-finite measures.*

(1) Suppose $f : \mathcal{X} \times \mathcal{Y} \to [0, \infty]$ is measurable. Then $g(x) = \int f_x(y) d\nu(y)$ and $h(y) = \int f^y(x) d\mu(x)$ are both non-negative and measurable. Furthermore,

$$\int f d(\mu \times \nu) = \int_{\mathcal{X}} \left(\int_{\mathcal{Y}} f(x, y) d\nu(y) \right) d\mu(x) = \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} f(x, y) d\mu(x) \right) d\nu(y). \qquad (7.2)$$

(2) Suppose $f \in L_1(\mu \times \nu)$, then

(a) $f_x \in L_1(\nu)$ for a.e. x

(b) $f^y \in L_1(\mu)$ for a.e. $y \in \mathcal{Y}$.

(c) The almost everywhere defined functions $g_1(x) = \int f_x(y)d\nu(y) \in L_1(\mu)$ and $g_2(y) = \int f^y(x)d\mu(x) \in L_1(\nu)$ are both integrable.

(d) Furthermore, (7.2) holds.

Proof This theorem holds for characteristic functions, and therefore holds for simple functions by non-linearity. We prove (1) first. Let $f^n \in \mathcal{S}^+$ such that $f^n \uparrow f$. Then

$$\begin{aligned} g_n(x) &= \int f^n(x,y)d\nu(y) \uparrow g(x) = \int f(x,y)d\nu(y), \\ h_n(y) &= \int f^n(x,y)d\mu(x) \uparrow h(y) = \int f(x,y)d\mu(x). \end{aligned}$$

So g, h are measurable and

$$\int f_n d(\mu \times \nu) = \int_{\mathcal{X}} \left(\int_{\mathcal{Y}} f_n(x,y)d\nu(y) \right) d\mu(x) = \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} f_n(x,y)d\mu(x) \right) d\nu(y).$$

Apply the monotone convergence theorem to conclude (1).

Also, $\int f d(\mu \times \nu) < \infty$ implies that $\int_{\mathcal{Y}} f_n(x,y)d\nu(y)$ is finite for a.e. x and $\int_{\mathcal{X}} f_n(x,y)d\mu(x)$ is finite for a.e. y.

For a general integrable function f, apply this to f^+ and f^-, and use the statement above to conclude.

□

Remark 7.2.2 For simplicity we remove the brackets and write

$$\begin{aligned} \int_{\mathcal{X}} \int_{\mathcal{Y}} f(x,y)d\nu(y)d\mu(x) &= \int_{\mathcal{X}} \left(\int_{\mathcal{Y}} f(x,y)d\nu(y) \right) d\mu(x), \\ \int_{\mathcal{X}} \int_{\mathcal{Y}} f(x,y)d\nu(y)d\mu(x) &= \int_{\mathcal{Y}} \left(\int_{\mathcal{X}} f(x,y)d\nu(y) \right) d\mu(x). \end{aligned}$$

Corollary 7.2.3 *Fubini's theorem holds if we replace $\mu \times \nu$ by its completion, provided that μ, ν are complete.*

To see this we note that if f is measurable with respect to the completion of $\mathcal{F} \otimes \mathcal{G}$, then there exists an $\mathcal{F} \otimes \mathcal{G}$ -measurable function $\hat{f}$ such that $f = \hat{f}$ except on a null set E, which by enlarging we may assume to be $\mathcal{F} \times \mathcal{G}$ measurable. Then $\mu(E_x) = 0$ for a.e. x, $\nu(E^y) = 0$ for a.e. y. Since μ, ν are complete, they are measurable. Thus, we may apply Fubini's theorem to $\hat{f}$,

$$\int_{\mathcal{X} \times \mathcal{Y}} f \; d\mu \times \nu = \int_{\mathcal{X} \times \mathcal{Y}} \hat{f} \; d\mu \times \nu = \int \int \hat{f}(x,y) \; d\mu(x)d\nu(y) = \int \int \hat{f}(x,y) \; d\nu(y)d\mu(x).$$

We know that $\hat{f}(x, \cdot)$ is measurable for almost every x and $f(x, \cdot)$ differs at most on E_x, a null set. Similarly we conclude $f(\cdot, y)$ is measurable a..e. y, and also they are in L_1, the conclusion follows.

Exercise 7.2.1 Let $\mathcal{X} = \{1, \ldots, N\}$ be a finite state space. Write down the the Fubini-Tonelli Theorem specific to this case.

Given a measure μ on $\mathcal{X} \times \mathcal{Y}$. We say μ_i are the marginals of μ if

$$\mu(\mathcal{X} \times B) = \mu_2(B), \quad \mu(A \times \mathcal{Y}) = \mu_1(A), \quad \forall A \in \mathcal{F}, B \in \mathcal{G}.$$

We also say that μ is a coupling of μ_1 and μ_2.

Exercise 7.2.2 Given an example of a Borel measure μ on $\mathbf{R}^2$, two Borel measures μ_i on $\mathbf{R}$ such that μ is not the product measure $\mu_1 \times \mu_2$, but μ_i are marginals of μ. Given another set of examples of measures such that $\mu = \mu_1 \times \mu_2$ but μ_i are not the Lebesgue measure.

Note that if f is a non-negative measurable function, $\hat{\mu}(A) := \int_A f \, d\mu$ is a measure.

Exercise 7.2.3 For $f : \mathbf{R} \to \mathbf{R}$ given by $f(x) = x^2 + 1$ and λ the Lebesgue measure, compute the pushed forward measure. $f_*\lambda$ and compute $\int_{[1,2]} e^{\sqrt{x-1}} \, d(f_*\lambda)$.

Exercise 7.2.4 Let $f(x, y) = \frac{(x^2-y^2)}{(x^2+y^2)^2}$ for $x \in [0, 1]$ and $y \in (0, 1]$, also $f(0, 0) = 0$. Show that $f \notin L_1([0, 1] \times [0, 1])$.

7.3 Product measures and independence

The notion of product measure admits an important interpretation in probability theory with the notion of independence. We first recall the following definitions, for which we assume that we are given a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. We will say that $\mathcal{A}$ is a sub-σ-algebra of $\mathcal{F}$ if it is a σ-algbera contained in $\mathcal{F}$.

Definition 7.3.1

1. We say that two sub-σ-algebras $\mathcal{A}$ and $\mathcal{B}$ of $\mathcal{F}$ are independent if any $A \in \mathcal{A}$ and any $B \in \mathcal{B}$ are independent, i.e. they satisfy $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.
2. Two random variables X and Y on $(\Omega, \mathcal{F}, \mathbf{P})$ are independent if $\sigma(X)$ is independent of $\sigma(Y)$.
3. A random variable Y is said to be independent of a sub-σ-algebra $\mathcal{A}$ of $\mathcal{F}$, if $\sigma(Y)$ and $\mathcal{A}$ are independent.

Thus, if $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ are two measurable spaces, and if $X : \Omega \to \mathcal{X}$ and $Y : \Omega \to \mathcal{Y}$ are two random variables, then X and Y are independent if and only if, for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$,

$$\mathbf{P}(X \in A, Y \in B) = \mathbf{P}(X \in A) \, \mathbf{P}(Y \in B). \tag{7.3}$$

Let us denote by P_X (resp. P_Y) the probability distribution of X (resp. Y): this is a probability measure on $(\mathcal{X}, \mathcal{A})$ (resp. $(\mathcal{Y}, \mathcal{B})$). Let us further denote by $P_{(X,Y)}$ the probability distribution of the random variable $(X, Y) : \Omega \to \mathcal{X} \times \mathcal{Y}$: this is a probability measure on $(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B})$.

Theorem 7.3.1 *The random variables X and Y are independent if and only if $P_{(X,Y)} = P_X \times P_Y$.*

Proof X and Y are independent if and only if, for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$, the equality (7.3) holds, and this equality can be written $P_{(X,Y)}(A \times B) = P_X(A)\, P_Y(B)$. Thus X and Y are independent if and only if $P_{(X,Y)}$ satisfies the characteristic property of the product measure $P_X \times P_Y$, whence the claim. □

Corollary 7.3.2 *If X and Y are independent, then for all measurable functions $f : \mathcal{X} \to [0, \infty]$ and $g : \mathcal{Y} \to [0, \infty]$, we have*

$$\mathbf{E}(f(X)\, g(Y)) = \mathbf{E}(f(X))\, \mathbf{E}(g(Y)).$$

Proof By the transfer lemma, we have

$$\mathbf{E}(f(X)\, g(Y)) = \int_{\mathcal{X} \times \mathcal{Y}} f(x)\, g(y) dP_{(X,Y)}(x, y).$$

But since X and Y are independent, $P_{(X,Y)} = P_X \times P_Y$, so by the Fubini-Tonnelli Theorem we obtain

$$\begin{aligned} \mathbf{E}(f(X)\, g(Y)) &= \int_{\mathcal{X}} f(x)\, dP_X(x) \int_{\mathcal{Y}} g(y) dP_Y(y) \\ &= \mathbf{E}(f(X))\, \mathbf{E}(g(Y)) \end{aligned}$$

where we used the transfer lemma to obtain the last equality. □

Exercise 7.3.1 Assume that $f : \mathcal{X} \to \mathbf{R}$ and $g : \mathcal{Y} \to \mathbf{R}$ are measurable functions such that the random variables $f(X)$ and $g(Y)$ are integrable. Show that if X and Y are independent, then the random variable $f(X)\, g(Y)$ is integrable and

$$\mathbf{E}(f(X)\, g(Y)) = \mathbf{E}(f(X))\, \mathbf{E}(g(Y)).$$

Chapter 8

Radon-Nikodym Theorem

In this chapter we compare two measure, discuss the concepts of singular and absolutely continuity.

8.1 Singular and absolutely continuity

Let $(\mathcal{X}, \mathcal{F})$ be a measurable space.

Definition 8.1.1 We say that a measure μ is absolutely continuous with respect to ν if whenever $\nu(A) = 0$, we have $\mu(A) = 0$.

We write $\mu \ll \nu$. Given any two finite measures μ_1, μ_2 we can find a reference measure (e.g. $\mu = \mu_1 + \mu_2$, such that $\mu_1 \ll \mu$ and $\mu_2 \ll \mu$).

Definition 8.1.2 Two measures μ and ν on $(\mathcal{X}, \mathcal{F})$ are mutually singular if there are disjoint sets A_1 and A_2 such that $\mathcal{X} = A_1 \cup A_2$ such that $\mu(A_1) = 0$ and $\nu(A_2) = 0$. This is denoted by $\mu \perp \nu$.

Remark 8.1.1 Note that the above property is equivalent to the existence of a measurable subset B satisfying $\mu(B^c) = \nu(B) = 0$.

The Lebesgue measure and any Dirac measures δ_x are singular.

Exercise 8.1.1 If ν is a measure on $(\mathcal{X}, \mathcal{G})$ and $p : \mathcal{X} \to [0, \infty)$ is an integrable function. Show that

$$\mu(A) = \int_A p(y)\nu(dy)$$

defines a measure and $\mu \ll \nu$. We write $\mu = p\,\nu$, or $d\mu = p\,d\nu$.

Example 8.1 The Gaussian measure $N(0,1)$ on $\mathbf{R}^1$ is absolutely continuous w.r.t. the Lebesgue measure.

$$N(0,1)(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{|x|^2}{2}} dx.$$

Example 8.2 Let $\mathcal{X} = [0,1]$ and P the Lebesgue measure . Define $Q(A) = 2\int_A x^2 dx$ so

$$\frac{dQ}{dP}(x) = 2x^2.$$

Define Q_1 by

$$\frac{dQ_1}{dP} = 2\mathbf{1}_{[0,\frac{1}{2}]}.$$

Then $Q_1 \ll P$ and P is not absolutely continuous with respect to Q_1. Define Q_2 by

$$\frac{dQ_2}{dP} = 2\mathbf{1}_{[\frac{1}{2},1]}.$$

The two measures Q_1 and Q_2 are singular.

The following lemma will be useful in the sequel:

Lemma 8.1.2 *Let μ, $\tilde{\mu}$ and ν three measures on a measurable space $(\mathcal{X}, \mathcal{F})$ such that $\mu \perp \nu$ and $\tilde{\mu} \perp \nu$. Then there exists a measurable subset B such $\mu(B^c) = \tilde{\mu}(B^c)$ while $\nu(B) = 0$.*

Proof Since $\mu \perp \nu$, there exist measurable disjoint susbets A_1 and A_2 such that $A_1 \cup A_2 = \mathcal{X}$ and $\mu(A_1) = \nu(A_2) = 0$. Since moreover $\tilde{\mu} \perp \nu$, there also exist measurable disjoint susbets $\tilde{A}_1$ and $\tilde{A}_2$ such that $\tilde{A}_1 \cup \tilde{A}_2 = \mathcal{X}$ and $\tilde{\mu}(\tilde{A}_1) = \nu(\tilde{A}_2) = 0$. Setting $B = A_2 \cup \tilde{A}_2$, it follows that $\mu(B^c) = \mu(A_1 \cap \tilde{A}_1) \leq \mu(A_1) = 0$, and similarly $\tilde{\mu}(B^c) = 0$, while $\nu(B) \leq \nu(A_1) + \nu(\tilde{A}_1) = 0$. □

Exercise 8.1.2 Let ν be a finite measure. Sow that $\nu \ll \mu$ if and only if the following holds: for any $\epsilon > 0$ there exists $\delta > 0$ such that whenever $\mu(E) < \delta$ then $\nu(E) < \epsilon$.

Exercise 8.1.3 Show that if $f \in L_1(\mu)$ is a non-negative function, then for every $\epsilon > 0$ there exists $\delta > 0$ such that whenever $\mu(A) < \delta$ we have

$$\int_E f d\mu < \epsilon.$$

Note that $\nu(A) = \int_A d\mu$ defines a measure with $\nu \ll \mu$.

Exercise 8.1.4 Give examples of pairs of measure μ, ν with $\mu \ll \nu$; then give examples such that $\mu \perp \nu$.

8.2 Signed measure

Definition 8.2.1 Let $(\mathcal{X}, \mathcal{F})$ be a measurable space. A signed measure on $(\mathcal{X}, \mathcal{F})$ is a map $\mu : \mathcal{F} \to \mathbf{R}$ satisfying the following properties:

1. $\mu(\phi) = 0$,
2. if $(A_j)_{j \geq 1}$ is a sequence of disjoint measurable sets, then the series $\sum_{j \geq 1} \mu(A_j)$ is absolutely convergent in **R** and
$$\mu(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j).$$

Remark 8.2.1 We stress that, in the definition of a signed measure μ, we impose μ to have *finite* real values (this is sometimes refered to as a finite signed measure). We thus allow for $A \in \mathcal{F}$ such that $\mu(A) < 0$, but not for $\mu(A) \in \{\pm\infty\}$.

Terminology: When talking about genuine measures (as introduced in Chapter 3), we will often use the expression "positive measures" to avoid confusion with signed measures.

Example 8.3 If ν is a positive measure, and $f \in L^1(\mathcal{X}, \mathcal{F}, \nu)$, then $\mu(A) := \int_A f \, d\nu$ defines a signed measure.

Definition 8.2.2 A measurable set P is said to be positive for μ if for any measurable $F \subset P$, we have $\mu(F) \geq 0$. In other words μ restricts to a (positive) measure on P. A measurable set N is said to be negative for μ if for any $F \subset N$, we have $\mu(F) \leq 0$; in other words $(-\mu)$ restricts to a (positive) measure on N.

Exercise 8.2.1 If P_n is a sequence of positive sets for μ, then so is $\cup_{n=1}^{\infty} P_n$.

Note that each $P_n \setminus \bigcup_{k=1}^{n-1} P_n$ is positive for μ.

Lemma 8.2.2 *If $A \in \mathcal{F}$ satisfies $\mu(A) < 0$, then there exists a negative subset $N \subset A$ such that $\mu(N) < 0$.*

Proof Let $\delta_1 := \sup\{\mu(B) : B \in \mathcal{F}, B \subset A\}$. Note that $\phi \subset A$, so that the collection of sets over which we are taking the sup is non-empty, and $\delta_1 \geq \mu(\phi) = 0$. Thus $\delta_1 \in [0, \infty]$. By definition of δ_1, we can find $B_1 \in \mathcal{F}$ such that $\mathcal{B}_1 \subset A$ and $\mu(B_1) \geq \min(\frac{\delta_1}{2}, 1)$. Continuing in this way we define, by induction over n, a sequence of non-negative numbers δ_n and a sequence of disjoint measurable subsets B_n by setting
$$\delta_n := \sup\left\{\mu(B) : B \in \mathcal{F},\, B \subset A \setminus \bigcup_{i=1}^{n-1} B_i\right\}$$

and by taking a measurable subset $B_n \subset A \setminus \left(\cup_{i=1}^{n-1} B_i\right)$ such that $\mu(B_n) \geq \min(\frac{\delta_n}{2}, 1)$. Note in particular that $\mu(B_n) \geq 0$. We define

$$N := A \setminus \left(\bigcup_{i=1}^{\infty} B_i \right).$$

Note that, since the B_i are disjoint, $\mu(\cup_{i=1}^{\infty} B_i) = \sum_{i=1}^{\infty} \mu(B_i) \geq 0$. Hence

$$\mu(N) = \mu(A) - \mu(\cup_{i=1}^{\infty} B_i) \leq \mu(A) < 0.$$

We finally show that N is a negative set. Let B be a measurable subset of N. Then, for all $n \geq 1$, $B \subset N \subset A \setminus \left(\cup_{i=1}^{n-1} B_i\right)$ so, by the definition of δ_n, $\mu(B) \leq \delta_n$. But $\delta_n \to 0$ as $n \to \infty$: indeed, since the B_n are disjoint, the series $\sum_{n \geq 1} \mu(B_n)$ is absolutely convergent, so in particular $\mu(B_n) \to 0$ when $n \to \infty$; since $0 \leq \min(\frac{\delta_n}{2}, 1) \leq \mu(B_n)$, the claim follows. Hence $\mu(B) \leq 0$. Thus N is a negative set. □

Theorem 8.2.3 (The Hahn decomposition theorem) *If μ is a signed measure on $(\mathcal{X}, \mathcal{F})$, then there exists a partition of $\mathcal{X}$: $\mathcal{X} = P \cup N$ and $P \cap N = \phi$, such that P is positive and N is negative for μ.*

If $\mathcal{X} = P' \cup N'$ is another such partition then $\mu(P \Delta P') = 0$ and $\mu(N \Delta N') = 0$.

Proof Let

$$c := \inf\{\mu(N) : N \in \mathcal{F} \text{ negative set}\}.$$

Note that ϕ is a negative set, so the collection of subsets over which we are taking the inf is non-empty, and $c \leq \mu(\phi) = 0$. Hence $c \in [-\infty, 0]$. Let (N_n) be a sequence of negative sets such that $\mu(N_n) \to c$ as $n \to \infty$, and let

$$N := \bigcup_{n \geq 1} N_n.$$

Note that N is a negative set as a countable union of negative sets. Moreover, for all $n \geq 1$, $N \setminus N_n \subset N$, so that $\mu(N) - \mu(N_n) = \mu(N \setminus N_n) \leq 0$. Thus $c \leq \mu(N) \leq \mu(N_n)$ and, sending $n \to \infty$, we deduce that $\mu(N) = c$. In particular, we deduce that $c > -\infty$. Let $P := N^c$. We show that P is a positive set. If not, there exists $A \subset P$ with $\mu(A) < 0$. By Lemma 8.2.2, one can then find a negative set $\tilde{A} \subset A$ such that $\mu(\tilde{A}) < 0$. Then the set $\tilde{N} := N \cup \tilde{A}$ is negative, and, since N and $\tilde{A}$ are disjoint,

$$\mu(\tilde{N}) = \mu(N) + \mu(\tilde{A}) < \mu(N) = c,$$

which contradicts the definition of c. Therefore P is indeed a positive set, and (N, P) gives the a partition of $\mathcal{X}$ into a negative and a positive subset.

If $\mathcal{X} = P' \cup N'$ is another such partition, then

$$P \Delta P' = (P \cap (P')^c) \cup (P^c \cap P') = (P \cap N') \cup (N \cap P') = N \Delta N'.$$

In particular $P \Delta P' \subset P \cup P'$ and, since $P \cup P'$ is a positive subset, we deduce that $\mu(P \Delta P') \geq 0$. But we also have $P \Delta P' \subset N \cup N'$ which, since $N \cup N'$ is a negative subset, implies $\mu(P \Delta P') \leq 0$. So $\mu(P \Delta P') = 0$, and therefore $\mu(N \Delta N') = \mu(P \Delta P') = 0$. □

Theorem 8.2.4 (Jordan decomposition theorem) *If μ is a signed measure on $(\mathcal{X}, \mathcal{F})$, then there exist uniquely finite positive measures μ^+ and μ^- such that*

$$\mu = \mu^+ - \mu^-, \qquad \mu^+ \perp \mu^-.$$

Proof Take $\mu^+ = \mu(P \cap \cdot)$, $\mu^- = -\mu(N \cap \cdot)$ where $\mathcal{X} = P \cup N$ as in Hahn decomposition theorem. To see it is unique suppose $\mu = Q_1 - Q_2$ where Q_1 and Q_2 are two finite positive measures such that $Q_1 \perp Q_2$. Then $\mathcal{X} = E_1 \cup E_2$, with disjoint subsets E_1, E_2 such that $Q_1(E_2) = 0$ and $Q_2(E_1) = 0$. Then E_1 and E_2 provide another Hahn decompostion, by its uniqueness we conclude. $\square$

We can talk about a signed measure being absolutely continuous with respect to a (positive) measure.

Definition 8.2.3 If μ is a signed measure on $(\mathcal{X}, \mathcal{F})$, we set $|\mu| = \mu^+ + \mu^-$. $|\mu|$ is a finite positive measure, is called the total variation of μ.

Definition 8.2.4 If ν is a (positive) measure, we say $\mu \ll \nu$ if $\mu(A) = 0$ whenever $\nu(A) = 0$ where $A \in \mathcal{F}$. On the other hand, we say that $\mu \perp \nu$ if there are two disjoint sets $A_1, A_2 \in \mathcal{F}$ such that $\mathcal{X} = A_1 \cup A_2$, and $\nu(A_2) = 0$ while $\mu(B \cap A_1) = 0$ for all measurable subset B.

Exercise 8.2.2

1. Show that $\mu \ll \nu$ if and only if $|\mu| \ll \nu$. Further show that this holds if and only if $\mu^+ \ll \nu$ and $\mu^- \ll \nu$.
2. Show that $\mu \perp \nu$ if and only if $|\mu| \perp \nu$. Further show that this holds if and only if $\mu^+ \perp \nu$ and $\mu^- \perp \nu$.

Definition 8.2.5 If μ is a signed measure we set $L_1(\mu) = L_1(\mu^+) \cap L_1(\mu^-)$, and for $f \in L_1(\mu)$, we set

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

Lemma 8.2.5 *Suppose that μ and ν are finite (positive) measures. Then either $\mu \perp \nu$ or there exists $\epsilon > 0$ and a measurable set E such that $\nu(E) > 0$ and E is a positive set for $\mu - \epsilon\nu$.*

Proof By Hahn's decomposition theorem, there exists a partition $\mathcal{X} = P_k \cup N_k$ such that P_k is a positive set and N_k a negative set for the signed measure $\mu - \frac{1}{k}\nu$.

Let $P = \bigcup_k P_k$ and $N = \cap_k N_k$. Then N, a subset of N_k, is negative for every $\mu - \frac{1}{k}\nu$, in particular

$$\mu(N) - \frac{1}{k}\nu(N) \leq 0.$$

Taking $k \to \infty$, we see that $\mu(N) \leq \lim_{k\to\infty} \frac{1}{k}\nu(N) = 0$, so that $\mu(N) = 0$. If μ, ν are not mutually singular, ν must charge the complement of N which is P, i.e. we must have $\nu(P) > 0$. Then, since P is a countable union of P_k's we have $\nu(P_k) > 0$ for some k. We take $\epsilon = 1/k$ and $E = P_k$ which is positive for $\mu - \frac{1}{k}\nu$. $\square$

8.3 Radon-Nikodym Theorem

Use with caution the following (for it might be confused with $\ll$, I used this in proofs.) We use $\mu \leq \nu$ to indicate that $\mu(A) \leq \nu(A)$ for every measurable set A.

Theorem 8.3.1 (Lebesgue-Radon-Nikodym Theorem) *Let μ and ν be two σ-finite positive measures on $(\mathcal{X}, \mathcal{F})$. Then*

(1) [Lebesgue decomposition] there exist, uniquely, σ-finite positive measures μ_1 and μ_2 with the property that $\mu = \mu_1 + \mu_2$, $\mu_1 \perp \nu$, and $\mu_2 \ll \nu$.

(2) [Radon-Nikodym Theorem] There exists a measurable function $D\colon \mathcal{X} \to [0, \infty]$ such that $\mu_2(A) = \int_A D\, d\nu$ for all $A \in \mathcal{A}$. Any two such functions agree ν almost-everywhere.

(3) If μ is finite, then D is integrable with respect to ν.

Definition 8.3.1 We denote D by $\frac{d\mu}{d\nu}$ and call it the Radon-Nikodym derivative of μ with respect to ν.

Proof *Step 1.* We first assume that μ and ν are finite positive measures. Let us define

$$\mathcal{H} = \left\{ f : \mathcal{X} \to [0, \infty] : \int_A f d\nu \leq \mu(A) \text{ for all } A \in \mathcal{F} \right\}.$$

We choose the ‘largest f’ from $\mathcal{H}$ as follows. Let

$$M = \sup \left\{ \int_{\mathcal{X}} f d\nu : f \in \mathcal{H} \right\}.$$

This is possible since $\mathcal{H}$ is not empty: $f \equiv 0 \in \mathcal{H}$. Also any $f \in \mathcal{H}$ has finite integral $\int f d\nu \leq \mu(\mathcal{X})$ so $M < \infty$.

Next we see, If $f, g \in \mathcal{H}$ then $\max(f, g) \in \mathcal{H}$. Let $E = \{f < g\}$.

$$\int_A \max(f, g) d\nu = \int_{E \cap A} g d\nu + \int_{E^c \cap A} f d\nu \leq \mu(A \cap E) + \mu(A \cap E^c) = \mu(A).$$

There exists a sequence of functions $f_n \in \mathcal{H}$ such that

$$\lim_n \int f_n d\mu = M.$$

Set

$$g_n = \max\{f_1, \ldots, f_n\} \in \mathcal{H}.$$

This is an increasing sequence and

$$\int_{\mathcal{X}} f_n \leq \int_{\mathcal{X}} g_n \leq M.$$

By the monotone convergence theorem, $\int \lim_n \uparrow g_n d\mu = \lim_n \int g_n d\mu = M$.

Set

$$D = \lim_n \uparrow g_n,$$

Then $D \in \mathcal{H}$. Indeed, it is measurable and for any $A \in \mathcal{F}$,

$$\int_A D d\nu = \int \mathbf{1}_A D d\nu = \lim_n \int \mathbf{1}_A\, g_n d\nu \leq \mu(A).$$

We show that $\mu - D d\nu$ is singular w.r.t. ν. If not there exists $\epsilon > 0$ and E measurable such that $\nu(E) > 0$ with $\mu - D d\nu - \epsilon \mathbf{1}_E d\nu > 0$ on E, i.e. for every $A \in \mathcal{F}$,

$$\int_{A \cap E} (D + \epsilon \mathbf{1}_E) d\nu \leq \mu(A \cap E).$$

Then

$$\int_A (D + \epsilon \mathbf{1}_E) d\nu = \int_{A \cap E} (D + \epsilon \mathbf{1}_E) d\nu + \int_{A \cap E^c} D d\nu \leq \mu(A \cap E) + \mu(A \cap E^c) = \mu(A)$$

and $(f + \epsilon \mathbf{1}_E) \in \mathcal{H}$ and its integral is larger than M, which contradicts with M being the largest integral among all $f \in \mathcal{H}$.

Uniqueness of the Lebesgue decomposition: if $(\tilde{\mu}_1, \tilde{\mu}_2)$ is another Lebesgue decomposition of μ wrt ν, then $\mu_1 + \mu_2 = \tilde{\mu}_1 + \tilde{\mu}_2$, so we have the following equality between signed measures $\mu_1 - \tilde{\mu}_1 = \tilde{\mu}_2 - \mu_2$. Since $\mu_1 \perp \nu$ and $\tilde{\mu}_1 \perp \nu$, by Lemma 8.1.2 there exists $B \in \mathcal{F}$ such that $\mu_1(B^c) = \tilde{\mu}_1(B^c) = 0$ while $\nu(B) = 0$. Hence, for all $A \in \mathcal{F}$,

$$(\mu_1 - \tilde{\mu}_1)(A) = (\mu_1 - \tilde{\mu}_1)(A \cap B) = (\tilde{\mu}_2 - \mu_2)(A \cap B) = 0,$$

where, to obtain the last equality, we used that $0 \leq \nu(A \cap B) \leq \nu(B) = 0$ which implies $\nu(A \cap B) = 0$ and hence $\mu_2(A \cap B) = \tilde{\mu}_2(A \cap B) = 0$ as $\tilde{\mu}_2$ and μ_2 are absolutely continuous w.r.t. ν. So $\mu_1(A) = \tilde{\mu}_1(A)$. Since A was arbitrary, $\mu_1 = \tilde{\mu}_1$, and therefore $\mu_2 = \tilde{\mu}_2$ as well.

Step 2. Suppose μ and ν are σ-finite (positive) measures. Then there exists E_j disjoint with $\mathcal{X} = \cup_j E_j$, $\mu(E_i) < \infty$ and $\nu(E_j) < \infty$. Let $\mu^j = \mu(E_j \cap \cdot)$ and $\nu^j = \nu(E_j \cap \cdot)$. By the previous argument, for all $j \geq 1$, there exist two finite positive measures μ_1^j and μ_2^j, and a measurable function $D_j : \mathcal{X} \to [0, \infty]$ such that $\mu^j = \mu_1^j + \mu_2^j$, $\mu_1^j \perp \nu^j$ and $d\mu_2^j = D\, d\nu$. Then $\mu = \mu_1 + \mu_2$ where $\mu_1 = \sum_{j=1}^\infty \mu_1^j$ and $\mu_2 = \sum_{j=1}^\infty \mu_2^j$. Moreover, setting : $D = \sum_{j=1}^\infty \mathbf{1}_{E_j} D_j$; we have $d\mu_2 = D\, d\nu$: indeed, recalling that the D_j are non-negative measurable functions, for all $A \in \mathcal{F}$ we have

$$\mu_2(A) = \sum_{j=1}^\infty \int_A D_j\, d\nu^j = \sum_{j=1}^\infty \int_A \mathbf{1}_{E_j} D_j\, d\nu = \int_A \left(\sum_{j=1}^\infty \mathbf{1}_{E_j} D_j \right) d\,\nu = \int_A D\, d\,\nu.$$

Finally, $\mu_1 \perp \nu$. Indeed, for all $j \geq 1$, since $\mu_1^j \perp \nu^j$, there exists $A_j \in \mathcal{F}$ such that $\mu_1^j(A_j^c) = 0$ while $\nu^j(A_j) = 0$. Let $A := \cup_{j \geq 1}(E_j \cap A_j)$. Note that, since E_j form a partition of $\mathcal{X}$, this is a disjoint union,

so
$$\nu(A) = \sum_{j=1}^{\infty} \nu(E_j \cap A_j) = \sum_{j=1}^{\infty} \nu^j(A_j) = 0.$$
Moreover, $A^c = \cup_{j\geq 1}(E_j \cap A_j^c)$, so that
$$\mu_1(A^c) = \sum_{j=1}^{\infty} \mu_1(E_j \cap A_j^c) = \sum_{j=1}^{\infty} \mu_1^j(E_j \cap A_j^c) \leq \sum_{j=1}^{\infty} \mu_1^j(A_j^c) = 0,$$
where, for the second equality, we used that $\mu_1^k(E_j) \leq \mu^k(E_j) = 0$ for all $k \neq j$. Therefore $\mu_1 \perp \nu$ as claimed.

Uniqueness of the Lebesgue decomposition: If $(\tilde{\mu}_1, \tilde{\mu}_2)$ is another Lebesgue decomposition for μ with respect to ν, then for all $j \geq 1$, the measures $\tilde{\mu}_1(E_j \cap \cdot)$ and $\tilde{\mu}_2(E_j \cap \cdot)$ provide a Lebesgue decomposition of μ^j with respect to ν^j. By uniqueness in the finite case, we deduce that $\tilde{\mu}_1(E_j \cap \cdot) = \mu_1^j$ and $\tilde{\mu}_2(E_j \cap \cdot) = \mu_2^j$. Hence, for $i = 1, 2$
$$\tilde{\mu}_i = \sum_{j\geq 1} \tilde{\mu}_i(E_j \cap \cdot) = \sum_{j\geq 1} \mu_i^j = \mu_i,$$
and uniqueness follows.

Step 3. There remains to show that if D and $\tilde{D}$ are two non-negative measurable functions on $\mathcal{X}$ such that $d\mu_2 = D\, d\nu$ and $d\mu_2 = \tilde{D}\, d\nu$, then $D = \tilde{D}$ ν-a.e. But in that case we have $\int_A D\, d\nu = \int_A \tilde{D}\, d\nu$ for all $A \in \mathcal{F}$, and the result follows from Exercise 5.5.3.

□

Exercise 8.3.1 (1) In Step 1 of the proof, why did we not take the density to be $D(x) = \sup_{f\in\mathcal{H}} f(x)$?[1]

(2) If we have a finite state space and discrete σ-algebra, construct the function D directly.

(3) If one has a finite σ-algebra, construct this function D directly.

Terminology: We use 'essentially unique' to say that if D_1 and D_2 are two possible choices for the density of μ with respect to ν, then $D_1 = D_2$ ν-a.e.

We now state a version of Lebesgue-Radon-Nikodym Theorem corresponding to the case of a signed measure.

Theorem 8.3.2 (Lebesgue-Radon-Nikodym Theorem for a signed measure) *Let ν be a σ-finite positive measures, and μ be a signed measure on $(\mathcal{X}, \mathcal{F})$. Then,*

[1]Hint: Let A be a non-measurable set, let us consider the collections of functions indexed by A as follows:
$$f_y(x) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases} = \mathbf{1}_{\{y\}}(x).$$
Is $\sup_{y\in A} f(y)$ measurable?

(1) *[Lebesgue decomposition] there exist, uniquely, two signed measures μ_1 and μ_2 with the property that $\mu = \mu_1 + \mu_2$, $\mu_1 \perp \nu$, $\mu_2 \ll \nu$.*

(2) *[Radon-Nikodym Theorem] There exists an integrable function $D\colon \mathcal{X} \to \mathbf{R}$ such that $\mu_2(A) = \int_A D\, d\nu$ for all $A \in \mathcal{F}$. Any two such functions agree ν almost-everywhere.*

(3) *If μ is a positive measure then $D \geq 0$ ν-a.e.*

Proof The existence of μ_1, μ_2 and D follows by applying the previous Lebesgue-Radon-Nikodym Theorem to the positive measures μ^+ and μ^-. To prove uniqueness of the Lebesgue Decomposition, assume that we have $\mu = \mu_1 + \mu_2 = \tilde{\mu}_1 + \tilde{\mu}_2$ with $\mu_1 \perp \nu$ and $\tilde{\mu}_1 \perp \nu$, while $\mu_2 \ll \nu$ and $\tilde{\mu_2} \ll \nu$. In particular $|\mu_1| \perp \nu$ and $|\tilde{\mu}_1| \perp \nu$, hence by Lemma 8.1.2 there exists a measurable susbet B such that $|\mu_1|(B^c) = |\tilde{\mu}_1|(B^c) = 0$, while $\nu(B) = 0$. Then, for all measurable subset A,

$$\mu_1(A) = \mu_1(A \cap B) + \mu_1(A \cap B^c) = \mu_1(A \cap B),$$

where we used that $|\mu_1(A \cap B^c)| \leq |\mu_1|(A \cap B^c) \leq |\mu_1|(B^c) = 0$ to obtain the second equality. But since $\nu(A \cap B) = 0$ and $\mu_2 \ll \nu$, we also have $\mu_2(A \cap B) = 0$. Hence the above equality yields $\mu_1(A) = \mu(A \cap B)$. By symmetry we also have $\tilde{\mu}_1(A) = \mu(A \cap B)$, so $\mu_1 = \tilde{\mu}_1$, and therefore we also obtain $\mu_2 = \mu - \mu_1 = \mu - \tilde{\mu}_1 = \tilde{\mu}_2$. Finally, the uniqueness ν-a.e. of the function D follows from Exercise 5.5.3, while statement (3) follows from the uniqueness of the density D in the Radon-Nikodym Theorem for positive measures. □

Definition 8.3.2 Also in this case, we denote D by $\frac{d\mu}{d\nu}$ and call it the Radon-Nikodym derivative of μ with respect to ν. Note that in this case, D corresponds to a unique element of $L^1(\mathcal{X}, \mathcal{F}, \nu)$.

Exercise 8.3.2 Define a Borel measure on $\mathbf{R}$ by $\mu(A) = \int_A p(x)dx + \mathbf{1}_A(2\pi)$ where p is a continuous function vanishing everywhere on the complement of $[-1, 1]$. Is μ absolutely continuous with respect to the Lebesgue measure dx? Write down its Lebesgue decomposition w.r.t. dx

Proposition 8.3.3 *Let μ_1, μ_2, μ_3 be σ -finite measures, positive measures.*

1. *Suppose $\mu_1 \ll \mu_2$. If $g \in L_1(\mu_1)$ then $g\frac{d\mu_1}{d\mu_2} \in L_1(\mu_2)$ and*

$$\int g d\mu_1 = \int g \frac{d\mu_1}{d\mu_2} d\mu_2.$$

2. *If $\mu_1 \ll \mu_2$, $\mu_2 \ll \mu_3$ then $\mu_1 \ll \mu_3$, and*

$$\frac{d\mu_1}{d\mu_3} = \frac{d\mu_1}{d\mu_2}\frac{d\mu_2}{d\mu_3}, \qquad \mu_3 - \textit{ almost-everywhere.}$$

Proof (1) If $g = \mathbf{1}_A$ and $\mu_1(A) < \infty$, we know by the definition,

$$\mu_1(A) = \int_A \frac{d\mu_1}{d\mu_2} d\mu_2.$$

So the claim holds. This extends to simple functions by linearity, to non-negative functions by the monotone convergence theorem and to any L_1 functions by linearity again.

(2) The transitivity of the absolute continuity is clear(check!). Use part (1), first suppose that $\frac{d\mu_1}{d\mu_2} \in L_1(\mu_2)$, we see

$$\mu_1(A) = \int_A \frac{d\mu_1}{d\mu_2} d\mu_2 \overset{g=\frac{d\mu_1}{d\mu_2}}{=} \int_A \frac{d\mu_1}{d\mu_2} \frac{d\mu_2}{d\mu_3} d\mu_3.$$

Then we assume μ_1 is a positive σ-finite measure. break down $\mathcal{X} = \cup E_j$ where $\mu_1(E_j) < \infty$ and E_j are disjoint measurable sets. We apply the argument, on each E_j, then sum them up. For a signed measure we break it down to $m_1 = \mu^+ - \mu^-$ and to each we apply the above conclusion. $\square$

Definition 8.3.3 If $\mu \ll \nu$ and $\nu \ll \mu$ we say that they are equivalent. In this case they have the same set of measure zero's.

Proposition 8.3.4 *If μ and ν are equivalent then $\frac{d\mu}{d\nu} = \frac{1}{\frac{d\nu}{d\mu}}$ a.e.*

Chapter 9

Conditional Expectations

Let us take a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The σ-algebra $\mathcal{F}$ represents the set of all observation and measurements one can make on a physical system. From observations (representing partial information) we would like to make predictions on the value of an observables, an observable is a measurable function (a random variable). The observed information is represented by a sub-σ-algebra $\mathcal{G}$. Suppose $\mathcal{G} = \sigma(Y)$ where Y is a random variable (this represents information obtained by for example an experiment). If a random variable $X : \Omega \to \mathbf{R}$ is furthermore measurable w.r.t. $\mathcal{G}$, then there exists a Borel measurable function $\varphi : \mathbf{R} \to \mathbf{R}$, such that $X = \varphi(Y)$, thus the observable X is determined by the information in $\mathcal{G}$ (i.e. by Y). If X is not measurable, can we predict its value by $\mathcal{G}$, what is the best approximation for it?

We denote by $\mathbf{E}X$ or $\mathbf{E}(X)$ the integral of X, called the (mathematical) expectation of X:

$$\mathbf{E}X = \int_\Omega X dP.$$

This is the best predicted value for X given $\mathcal{F}$. If X is real valued, $\mathbf{E}X$ minimizes, among all constants, the value $\min_c \mathbf{E}(X-c)^2$.

9.1 Conditional expectation and probabilities

Definition 9.1.1 Let X be a real-valued random variable on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{E}|X| < \infty$ and let $\mathcal{G}$ be a sub σ-algebra of $\mathcal{F}$. Then the **conditional expectation** of X with respect to $\mathcal{G}$ is an integrable, $\mathcal{G}$-measurable random variable X' such that

$$\int_A X(\omega)\,\mathbf{P}(d\omega) = \int_A X'(\omega)\,\mathbf{P}(d\omega)\,, \tag{9.1}$$

for every $A \in \mathcal{G}$. We denote this by $X' = \mathbf{E}(X \mid \mathcal{G})$.

Proposition 9.1.1 *With the notations as above, there exists a unique element conditional expectation $X' = \mathbf{E}(X \mid \mathcal{G})$ exists and is unique as an element of $L^1(\Omega, \mathcal{G}, \mathbf{P})$. If moreover $X \geq 0$ a.s., then $E(X|\mathcal{G}) \geq 0$ a.s.*

Proof Denote by P also the restriction of $\mathbf{P}$ to $\mathcal{G}$ and define the signed measure Q on $(\Omega, \mathcal{G})$ by

$$Q(A) = \int_A X(\omega)\,\mathbf{P}(d\omega), \qquad \forall A \in \mathcal{G}.$$

Then Q is a signed measure on $(\Omega, \mathcal{G})$ which is absolutely continuous with respect to P. Its density with respect to P, given by the Radon-Nikodym theorem, is then a $\mathcal{G}$-measurable functions and is the required conditional expectation. The uniqueness follows from the uniqueness statement in the Radon-Nikodym theorem. Finally, if $X \geq 0$ a.s., then Q is non-negative measure, so $E(X|\mathcal{G}) \geq 0$ a.s. by the Radon-Nikodym theorem for positive measures. □

Using the Monotone Class Theorem we obtain:

Proposition 9.1.2 *For all bounded $\mathcal{G}$-measurable function g.*

$$\int_\Omega gX\,dP = \int_\Omega g\mathbf{E}(X|\mathcal{G})\,dP. \tag{9.2}$$

Example 9.1
- If $\mathcal{G} = \{\phi, \Omega\}$, verify that $\mathbf{E}(X|G) = \mathbf{E}(X)$. Remember that X' being $\{\phi, \Omega\}$-measurable means that the pre-image under X' of an arbitrary Borel set is either ϕ or Ω, so the conditional expectation is a constant.
- If X is $\mathcal{G}$-measurable, then $\mathbf{E}(X|\mathcal{G}) = X$ a.e.

Example 9.2 If the only information we know is whether a certain event B happened or not, then it should by now be intuitively clear that the conditional expectation of a random variable X with respect to this information is given by $\mathbf{E}(X|B)$, if B happened and by $\mathbf{E}(X|B^c)$ if B didn't happen. It is a straightforward exercise that the conditional expectation of X with respect to the σ-algebra $\mathcal{F}_B = \{\phi, \Omega, B, B^c\}$ is indeed given by

$$\mathbf{E}(X \mid \mathcal{F}_B)(\omega) = \begin{cases} \mathbf{E}(X|B) & \text{if } \omega \in B \\ \mathbf{E}(X|B^c) & \text{otherwise.} \end{cases}$$

Proof. Denote by Y the right hand side. Then, $\mathbf{E}(\mathbf{1}_B Y) = \mathbf{E}(\mathbf{E}(X|B)\mathbf{1}_B) = \mathbf{E}(X\mathbf{1}_B)$. Similarly $\mathbf{E}(\mathbf{1}_{B^c} Y) = \mathbf{E}(\mathbf{1}_{B^c} X)$, and hence also $\mathbf{E}(X) = \mathbf{E}(Y)$.

If Y is a measurable function, we denote

$$\mathbf{E}(X|Y) = \mathbf{E}(X|\sigma(Y)). \tag{9.3}$$

Exercise 9.1.1 Suppose that $\{A_1, \dots, A_n\}$ is a partition of Ω by measurable subsets: the A_i are disjoint and their union is Ω. We further assume that $\mathbf{P}(A_i) > 0$ for every i. Let $\mathcal{G}$ be generated by them. Then

$$\mathbf{E}(X|\mathcal{G}) = \sum_{i=1}^n \mathbf{E}(X|A_i)\mathbf{1}_{A_i}.$$

If Y a.s. takes only countably many values s_i, each of which are taken with positive probability, then $\{Y = s_i\}$ is a partition of Ω and

$$\mathbf{E}(X|Y) = \sum_i \mathbf{E}(X|Y = s_i)\mathbf{1}_{Y=s_i}.$$

9.1.1 Conditioning on a random variable

In many situations, we will describe $\mathcal{F}'$ as the σ-algebra generated by another random variable $Y : \Omega \to \mathcal{X}$. By the factorization lemma there exists $\varphi : \mathcal{X} \to \mathbf{R}$ such that $\mathbf{E}(X|Y) = \varphi(Y)$ a.s. If two functions φ and φ' satisfy this relation, then $\varphi = \varphi'$ on a set of $\mathcal{B}(\mathcal{X})$ of full measure (measured by the law of Y which is denoted by $\mathbf{P}_Y$). To see this just note that if $A = \{\omega : \varphi(Y(\omega) = \tilde{\varphi}(Y(\omega)\}$ and $B = \{x : \varphi(x) = \tilde{\varphi}(x)\}$, then $A = Y^{-1}(B)$ and $\mathbf{P}_Y(B) = \mathbf{P}(Y^{-1}(B))$. And we do know that by the uniqueness fo the conditional expectation, $\mathbf{P}\Big(\{\omega : \varphi(Y(\omega)) = \tilde{\varphi}(Y(\omega)\}\Big) = 1$.

Notation. We denote by $\mathbf{E}(X|Y = y)$ the measurable function $\varphi : \mathcal{X} \to \mathbf{R}$ such that $\mathbf{E}(X|Y) = \varphi(Y)$.

$\mathbf{E}(X|Y) = \psi(Y)$ a.s. If two functions φ and $\tilde{\varphi}$ satisfy this relation, by the uniqueness fo the conditional expectation,

$$\mathbf{P}\Big(\{\omega : \varphi(Y(\omega)) = \tilde{\varphi}(Y(\omega)\}\Big) = 1.$$

Write $Y_*\mu$ for the pushed forward measure on $\mathbf{C}$. Then

$$(Y_*\mu)(\{\varphi = \varphi'\}) = 1.$$

(Just note that if $A = \{\omega : \varphi(Y(\omega)) = \tilde{\varphi}(Y(\omega))\}$ and $B = \{x : \varphi(x) = \tilde{\varphi}(x)\}$, then $A = Y^{-1}(B)$ and $(Y_*\mu)(B) = \mathbf{P}(Y^{-1}(B))$ by the definition).

Definition 9.1.2 We denote by $\mathbf{E}(X|Y = y)$ the measurable function $\varphi : \mathcal{X} \to \mathbf{R}$ s.t. $\mathbf{E}(X|Y) = \varphi(Y)$.

Example 9.3 Take a measurable function $Y : \Omega \to E = \{y_1, y_2, \dots, y_n\}$. Define

$$A_i = Y^{-1}(\{y_i\}) = \{\omega : Y(\omega) = y_i\}.$$

Then $\sigma(Y) = \sigma\{A_i\}$ and $A_i \cap A_j = \phi$ if $i \neq j$. Let $X \in L_1(\Omega, \mathcal{F}, P)$. Define

$$\varphi(y_i) = \begin{cases} \frac{\mathbf{E}(\mathbf{1}_{A_i}X)}{P(A_i)}, & \text{if } P(A_i) \neq 0 \\ 0, & \text{if } P(A_i) = 0 \end{cases}.$$

Define

$$\mathbf{E}(X|A_i) = \frac{\mathbf{E}(X\mathbf{1}_{A_i})}{P(A_i)}, \quad \text{if } P(A_i) \neq 0,$$

otherwise define it to be zero. Then

$$\varphi(y) = \sum_{i=1}^{n} \mathbf{E}(X|A_i)\mathbf{1}_{A_i}(y), \qquad \varphi(Y(\omega)) = \sum_{i=1}^{n} \mathbf{E}(X|A_i)\, 1_{\{Y=y_i\}}.$$

Then $\mathbf{E}(X|Y)(\omega) = \varphi(Y(\omega))$. Indeed for any $A_k \in \sigma(Y)$,

$$\int_{A_k} \varphi(Y)dP = \int_{A_k} \left(\sum_{i=1}^{n} \mathbf{E}(X|A_i)\, 1_{A_i}\right) dP = \int_{A_k} \mathbf{E}(X|A_k)\, dP = \int_{A_k} X dP.$$

We could also define the function $\mathbf{E}(X|Y=y)$ as follows.

Definition 9.1.3 Define $\mathbf{P}_Y(A) = \mathbf{P}(Y \in A)$, the probability distribution of Y. Let X be a real-valued random variable which is integrable or non-negative. We denote by $\mathbf{E}(X \mid Y = y)$ any random function such that for any B Borel measurable,

$$\int_B \mathbf{E}(X \mid Y = y)\, \mathbf{P}_Y(dy) = \int_{\{\omega: Y(\omega) \in B\}} X d\mathbf{P}.$$

When conditioning on a number of variables $Y_1, \dots, Y_n$, the following notation is also used:

$$\mathbf{E}(X|Y_1, \dots, Y_n) = \mathbf{E}(X|\sigma(Y_1) \vee \dots \vee \sigma(Y_n)). \tag{9.4}$$

9.2 Properties

Conditional expectation has very nice properties, they behave almost like integration with respect to a (family of) measures.

Proposition 9.2.1

- *(linearity) If X, Y are integrable and a, b are numbers, then*

$$\mathbf{E}(aX + bY|\mathcal{G}) = a\mathbf{E}(X|\mathcal{G}) + b\mathbf{E}(Y|G), a.s.$$

- $|\mathbf{E}(X|\mathcal{G})| \le \mathbf{E}(|X||\mathcal{G})$ *a.s, and as a consequence* $\mathbf{E}(|\mathbf{E}(X|\mathcal{G})|) \le \mathbf{E}(|X|)$.
- *(monotonicity) If X and Y are integrable and $X \le Y$ a.s., then $\mathbf{E}(X|\mathcal{G}) \le \mathbf{E}(Y|G)$ a.s.*

Proof The linearity follows from uniqueness of conditional expectations. (ii) is a direct consequence of the Radon-Nikodyme theorem. □

Lemma 9.2.2 *If X_n is a sequence of integrable r.v. such that $\mathbf{E}|X_n - X| \to 0$ for some integrable r.v. X, then $\mathbf{E}|\mathbf{E}(X_n|\mathcal{G}) - \mathbf{E}(X|\mathcal{G})| \to 0$.*

Proof We have $|\mathbf{E}(X_n - X\big|\mathcal{G})| \leq \mathbf{E}(|X_n - X|\,\big|\,\mathcal{G})$, and the latter converges in L_1 as

$$\mathbf{E}(\mathbf{E}(|X_n - X|\,\big|\,\mathcal{G})) = \mathbf{E}|X_n - X| \to 0,$$

completing the proof. □

9.3 Conditional expectation of a non-negative random variable

So far, we have only defined conditional expectations of integrable random variables. In analogy with integrals/expectations, it would be natural to have a corresponing notion for non-negative random variables.

Let $X : \Omega \to [0, \infty]$ be a non-negative r.v., and let $\mathcal{G} \subset \mathcal{F}$ be a sub-σ-algebra.

Theorem 9.3.1 *There exists a $\mathcal{G}$-measurable r.v. Y with values in $[0, \infty]$ s.t.*

$$\forall A \in \mathcal{G}, \qquad \mathbf{E}(\mathbf{1}_A X) = \mathbf{E}(\mathbf{1}_A Y).$$

Such a r.v. is unique up to a $\mathbf{P}$-null set.

We denote the r.v. provided by the Theorem by $\mathbf{E}(X|\mathcal{G})$, and call it the conditional expectation of X given $\mathcal{G}$.

Proof For all $n \geq 1$, the truncated r.v. $\min(X, n)$ is integrable, hence $\mathbf{E}(\min(X, n)|\mathcal{G})$ is well-defined. Moreover, by montonicity of the conditional expectation, almost-surely $\mathbf{E}(\min(X, n)|\mathcal{G})$ is non-decreasing in n, hence the r.v. $Y := \lim_{n\to\infty} \uparrow \mathbf{E}(\min(X, n)|\mathcal{G})$ is well-defined on an event of probability 1 (we can extend it arbitrarily by 0 outside of this event). Then, for all $A \in \mathcal{G}$, invoking the Monotone Convergence Theorem, we have

$$\mathbf{E}(\mathbf{1}_A Y) = \lim_{n\to\infty} \mathbf{E}(\mathbf{1}_A \mathbf{E}(\min(X, n)|\mathcal{G})) = \lim_{n\to\infty} \mathbf{E}(\mathbf{1}_A \min(X, n)) = \mathbf{E}(\mathbf{1}_A X).$$

Thus Y has the requested property. Now, if $\tilde{Y}$ is another non-negative $\mathcal{G}$-measurable r.v. such that $\mathbf{E}(\mathbf{1}_A X) = \mathbf{E}(\mathbf{1}_A \tilde{Y})$ for all $A \in \mathcal{G}$, then we have $\mathbf{E}(\mathbf{1}_A Y) = \mathbf{E}(\mathbf{1}_A \tilde{Y})$ for all $A \in \mathcal{G}$. By Exercise 5.5.3, we deduce that $Y = \tilde{Y}$ a.s. □

Remark 9.3.2 If a non-negative random variable X is also integrable, the uniqueness statement of the above Theorem ensures the two notions for $\mathbf{E}(X|G)$ actually coincide a.s.

Proposition 9.3.3 $\mathbf{E}(X|\mathcal{G})$ *is, up to a* $\mathbf{P}$*-null set, the only non-negative r.v. such that, for all non-negative* $\mathcal{G}$*-measurable r.v.* Z*,* $\mathbf{E}(Z\,X) = \mathbf{E}(Z\,\mathbf{E}(X|\mathcal{G}))$*.*

Proof The statement is true for $Z = \mathbf{1}_A$ for $A \in \mathcal{G}$, and extends to any non-negative r.v. Z by linearity and by the Monotone Convergence Theorem (thorugh approximating Z by non-negative simple functions). The uniqueness statement follows at once by the uniqueness part of Theorem 9.3.1. □

Example 9.4

1. If $\mathcal{G} = \{\phi, \Omega\}$, then $\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X)$ a.s. (note that $\mathbf{E}(X) \in [0, \infty]$ in this non-negative case).
2. If X is $\mathcal{G}$-measurable, then $\mathbf{E}(X|\mathcal{G}) = X$ a.s.

Proposition 9.3.4 *Let* $X, Y : \Omega \to [0, \infty]$ *be two non-negative r.v., and let* $\mathcal{G} \subset \mathcal{F}$ *be a sub-*σ*-algebra.*

1. *(Linearity) For all* $a, b \geq 0$*,* $\mathbf{E}(aX + bY|\mathcal{G}) = a\mathbf{E}(X|\mathcal{G}) + b\mathbf{E}(Y|\mathcal{G})$ *a.s.*
2. $\mathbf{E}(\mathbf{E}(X|\mathcal{G})) = \mathbf{E}(X)$
3. *(Monotonicity) If* $X \leq Y$ *a.s., then* $\mathbf{E}(X|\mathcal{G}) \leq \mathbf{E}(Y|\mathcal{G})$ *a.s.*

Proof The first two properties follow at once from the characteristic property of the conditional expectation of a non-negative r.v., so we focus on the third one. If $X \leq Y$ a.s., then for all $n \geq 1$, $\min(X, n) \leq \min(Y, n)$ a.s. Hence, by linearity of the conditional expectation of integrable random variables, $\mathbf{E}(\min(X, n)|\mathcal{G}) \leq \mathbf{E}(\min(Y, n)|\mathcal{G})$ a.s. Sending $n \to \infty$ yields the result. □

Exercise 9.3.1 Assume that X, Y are non-negative random variables, Z is an integrable random variable, and $X \leq Y + Z$ a.s. Show that

$$\mathbf{E}(X|\mathcal{G}) \leq \mathbf{E}(Y|\mathcal{G}) + \mathbf{E}(Z|\mathcal{G}) \quad \text{a.s.}$$

9.4 Some further properties

We now state some further properties that are specific to conditional expectations. We first recall some definitions.

Definition 9.4.1

1. We say B is independent of $\mathcal{G}$ if $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$ for every $A \in \mathcal{G}$.
2. We say two σ-algebras $\mathcal{F}$ and $\mathcal{G}$ are independent is any $A \in \mathcal{F}$ is independent of any $B \in \mathcal{G}$.
3. Two random variables are independent if $\sigma(X)$ is independent of $\sigma(Y)$.
4. A random variable is said to be independent of a σ-algebra $\mathcal{G}$, if $\sigma(Y)$ and $\mathcal{G}$ are independent.

Proposition 9.4.1 *Let $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{F}'$ be a sub-σ-algebra of $\mathcal{F}$. Then the following hold*

1. $\mathbf{E}(X) = \mathbf{E}(\mathbf{E}(X|\mathcal{F}'))$.
2. *If X is $\mathcal{F}'$ measurable, then $\mathbf{E}(X|\mathcal{F}') = X$.*
3. *(Tower property) If $\mathcal{G}_1 \subset \mathcal{G}_2$,*
$$\mathbf{E}(X|\mathcal{G}_1) = \mathbf{E}(\mathbf{E}(X|\mathcal{G}_1)|\mathcal{G}_2)) = \mathbf{E}(\mathbf{E}(X|\mathcal{G}_2)|\mathcal{G}_1)).$$
4. *(Without any information, the best estimate is expectation) If X is independent of $\mathcal{G}$,*
$$\mathbf{E}(X|\mathcal{G}) = \mathbf{E}(X), \quad a.e.$$

Proof (1) To see this set $A = \Omega$ in equation (9.1.1). (2) follows from uniqueness.

(3) If $\mathcal{G}_1 \subset \mathcal{G}_2$, $\mathbf{E}(X|\mathcal{G}_1)$ is $\mathcal{G}_2$-measurable, so $\mathbf{E}(X|\mathcal{G}_1) = \mathbf{E}(\mathbf{E}(X|\mathcal{G}_1)|\mathcal{G}_2)$. One can also show by the definition that
$$\mathbf{E}(X|\mathcal{G}_1) = \mathbf{E}(\mathbf{E}(X|\mathcal{G}_2)|\mathcal{G}_1)).$$
(4) For any $A \in \mathcal{G}$, $\int_A \mathbf{E}(X|\mathcal{G})dP = \mathbf{E}(X\mathbf{1}_A) = \mathbf{E}(X)P(A) = \int_A \mathbf{E}(X)\,dP$. Since $\mathbf{E}(X)$ is a constant, hence $\mathcal{G}$-measurable r.v., the conclusion follows. □

The following result is a generalisation of Property 4. of the previous Proposition.

Proposition 9.4.2 *Suppose that $\sigma(X) \vee \mathcal{G}$ is independent of $\mathcal{A}$, then $\mathbf{E}(X|\mathcal{A} \vee \mathcal{G}) = \mathbf{E}(X|\mathcal{G})$.*

Proof Let $A \in \mathcal{A}, B \in \mathcal{G}$,
$$\int_{A\cap B} XdP = \mathbf{E}(X\mathbf{1}_B)P(A),$$
$$\int_{A\cap B} \mathbf{E}\{X \mid G\}dP = \int_A \mathbf{E}\{X\mathbf{1}_B|\mathcal{G}\}dP = P(A)\mathbf{E}(X\mathbf{1}_B).$$
Since $\{A \cap B\}$ forms a π-system, and we can prove that
$$C = \{D \in \mathcal{G} \vee \mathcal{A} : \int_D XdP = \int_D \mathbf{E}\{X|\mathcal{G}\}dP\}$$
is a λ system. Hence $C = \mathcal{G} \vee \mathcal{A}$. □

Exercise 9.4.1 Let $X, Y \in L_1(\Omega, \mathcal{F}, P)$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. If X is $\mathcal{G}$ measurable, $XY \in L_1$ then
$$\mathbf{E}(XY|\mathcal{G}) = X\mathbf{E}(Y|\mathcal{G}).$$
This means we take out 'what is known'.

9.5 Convergence Theorems

The important convergence theorems of integration admit an analog statement for conditional expectations, which we now state. Let henceforth $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.

Theorem 9.5.1 (Conditional Monotone Convergence Theorem) *If $(X_n)_{n\geq 1}$ is a sequence of non-negative random variables such that $X_n \uparrow X$, then $\mathbf{E}(X_n\mathcal{G}) \uparrow \mathbf{E}(X|\mathcal{G})$ a.s.*

Proof (1) Note that a.s. $\mathbf{E}(X_n|\mathcal{G})$ is non-negative and non-decreasing in n, hence it admits a limit Y, which is $\mathcal{G}$-measurable. Then, by the Monotone Convergence Theorem, for all $A \in \mathcal{G}$,

$$\mathbf{E}(\mathbf{1}_A X_n) \uparrow \mathbf{E}(\mathbf{1}_A X), \qquad \mathbf{E}(\mathbf{1}_A \mathbf{E}(X_n|\mathcal{G})) \uparrow \mathbf{E}(\mathbf{1}_A Y).$$

Since $\mathbf{E}(\mathbf{1}_A X_n) = \mathbf{E}(\mathbf{1}_A \mathbf{E}(X_n|\mathcal{G}))$ for all $n \geq 1$, sending $n \to \infty$ we deduce that $\mathbf{E}(\mathbf{1}_A X) = \mathbf{E}(\mathbf{1}_A Y)$. This being true for all $A \in \mathcal{G}$, and since Y is $\mathcal{G}$-measurable, we deduce that $\mathbf{E}(X|\mathcal{G}) = Y = \lim_{n\to\infty} \mathbf{E}(X_n|\mathcal{G})$ a.s. □

Lemma 9.5.2 (Conditional Fatou Lemma) *If $(X_n)_{n\geq 1}$ is a sequence of non-negative r.v., then*

$$\mathbf{E}(\liminf_{n\to\infty} X_n|\mathcal{G}) \leq \liminf_{n\to\infty} \mathbf{E}(X_n|\mathcal{G}) \quad \textit{a.s.}$$

Proof By the Conditional Monotone Convergence Theorem

$$\mathbf{E}(\inf_{n\geq m} X_n|\mathcal{G}) \underset{n\to\infty}{\longrightarrow} \mathbf{E}(\liminf_{n\to\infty} X_n|\mathcal{G}) \quad \text{a.s.}$$

But, for all $m \geq 1$ and $p \geq m$, by monotonicty of the conditional expectation

$$\mathbf{E}(\inf_{n\geq m} X_n|\mathcal{G}) \leq \mathbf{E}(X_p|\mathcal{G}) \quad \text{a.s.}$$

whence

$$\mathbf{E}(\inf_{n\geq m} X_n|\mathcal{G}) \leq \inf_{p\geq m} \mathbf{E}(X_p|\mathcal{G}) \quad \text{a.s.}$$

Sending $n \to \infty$ in the above inequality yields the result. □

Theorem 9.5.3 (Conditional Dominated Convergence Theorem) *If $(X_n)_{n\geq 1}$ is a sequence of random variables such that*

- *X_n converges a.s. to a random variable X*
- *there exists an integrable non-negative r.v. Y such that, for all $n \geq 1$, $|X_n| \leq Y$ a.s.*

Then $\mathbf{E}(X_n|\mathcal{G}) \underset{n\to\infty}{\longrightarrow} \mathbf{E}(X|\mathcal{G})$ a.s. and in L^1.

Proof First note that the assumptions guaranteee that the r.v. X_n and X are all integrable, so their conditional expectations are well-defined. At the expanse of modifying X_n, X and Y on a null-set, we may assume that $X_n \underset{n\to\infty}{\longrightarrow} X$ pointwise, and that $|X_n| \leq Y$ for all $n \geq 1$. Then for all $n \geq 1$, $Z_n := 2Y - |X_n - X|$ is a non-negative r.v. By the conditional Fatou Lemma applied to Z_n,

$$2\mathbf{E}(Y|\mathcal{G}) \leq \liminf_n (2\mathbf{E}(Y|\mathcal{G}) - \mathbf{E}(|X_n - X||\mathcal{G})) \quad \text{a.s.}$$

whence $\limsup_n \mathbf{E}(|X_n - X||\mathcal{G}) \leq 0$ a.s., i.e. $\mathbf{E}(|X_n - X||\mathcal{G}) \underset{n\to\infty}{\longrightarrow} 0$ a.s. Therefore

$$|\mathbf{E}(X_n|\mathcal{G}) - \mathbf{E}(X|\mathcal{G})| \leq \mathbf{E}(|X_n - X|\mathcal{G}) \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.}$$

which entails the a.s. convergence statement. For the convergence in L^1, note that

$$\mathbf{E}(|\mathbf{E}(X_n|\mathcal{G}) - \mathbf{E}(X|\mathcal{G})|) \leq \mathbf{E}(\mathbf{E}(|X_n - X|\mathcal{G})) = \mathbf{E}(|X_n - X|)$$

and $\mathbf{E}(|X_n - X|) \underset{n\to\infty}{\longrightarrow} 0$ by the (usual) Dominated Convergence Theorem. □

9.6 Conditional expectation of square-integrable random variables

In this section we will see an interesting interpretation of the conditional expectation of a square-integrable r.v. in terms of an orthogonal projection. We first need a conditional version of Jensen's inequality.

9.6.1 Conditional Jensen inequality

Proposition 9.6.1 (Conditional Jensen Inequality) *Let $\varphi : \mathbf{R} \to \mathbf{R}_+$ be a convex function. Then*

$$\varphi(\mathbf{E}(X|\mathcal{G})) \leq \mathbf{E}(\varphi(X)|\mathcal{G}), \quad \textit{a.s.}$$

Proof Since φ is convex, for all $x \in \mathbf{R}$, we have

$$\forall y \in \mathbf{R}, \qquad \varphi(y) - \varphi(x) \geq \varphi'_+(x)(y - x), \tag{9.5}$$

where

$$\varphi'_+(x) := \lim_{\substack{h\to 0\\ h>0}} \frac{\varphi(x+h) - \varphi(x)}{h} \in \mathbf{R}$$

is the right-hand derivative of φ at x. In particular, we have $\varphi'_+(x) := \lim_{n\,\infty} \frac{\varphi(x+1/n)-\varphi(x)}{1/n}$: since φ is a convex, hence continuous function on $\mathbf{R}$, this shows that $x \mapsto \varphi'_+(x)$ is a pointwise limit of Borel measurable functions, hence it is Borel measurable. Applying for each $\omega \in \Omega$ the inequality (9.5) with $x = \mathbf{E}(X|\mathcal{G})(\omega)$ and $y = X(\omega)$ yields the inequality

$$\varphi(X) - \varphi(\mathbf{E}(X|\mathcal{G})) \geq Z\,(X - \mathbf{E}(X|\mathcal{G})), \tag{9.6}$$

where Z is the random variable $\varphi'_+(\mathbf{E}(X|\mathcal{G}))$. Note that, since $\mathbf{E}(X|\mathcal{G})$ is $\mathcal{G}$ measurable and $x \mapsto \varphi'_+(x)$ is Borel measurable, Z is $\mathcal{G}$-measurable by the composition rule. At this stage it would be tempting to take $\mathbf{E}(\cdot|\mathcal{G})$ in both sides of (9.6) to conclude, however a caveat is that the random variable in the right-hand side is neither non-negative nor integrable in general.

To circumvent this difficulty, we proceed with a truncation argument. By (9.6), for all $n \geq 1$, we have

$$\varphi(X) \geq \varphi(\mathbf{E}(X|\mathcal{G}))\mathbf{1}_{\{|Z| \leq n\}} + Z\mathbf{1}_{\{|Z|\leq n\}}\,(X - \mathbf{E}(X|\mathcal{G})),$$

where now the random variable $Z\mathbf{1}_{\{|Z|\leq n\}}\,(X - \mathbf{E}(X|\mathcal{G}))$ is bounded by $n(|X| + |\mathbf{E}(X|\mathcal{G})|)$, so it is integrable as X and $\mathbf{E}(X|\mathcal{G})$ are. By Exercise 9.3.1, we therefore obtain

$$\mathbf{E}(\varphi(X)|\mathcal{G}) \geq \varphi(\mathbf{E}(X|\mathcal{G}))\mathbf{1}_{\{|Z|\leq n\}} + \mathbf{E}(Z\mathbf{1}_{\{|Z|\leq n\}}\,(X - \mathbf{E}(X|\mathcal{G}))|\mathcal{G}) \quad \text{a.s.},$$

where we used the fact that the first term in the right-hand side is $\mathcal{G}$-measurable. Now, since $Z\mathbf{1}_{\{|Z|\leq n\}}$ is $\mathcal{G}$-measurable, by Exercise (9.4.1), the second term in the right-hand side equals

$$Z\mathbf{1}_{\{|Z|\leq n\}}\,\mathbf{E}(X - \mathbf{E}(X|\mathcal{G})|\mathcal{G}) = 0 \quad \text{a.s.}$$

Thus for all $n \geq 1$, we have $\mathbf{E}(\varphi(X)|\mathcal{G}) \geq \varphi(\mathbf{E}(X|\mathcal{G}))\mathbf{1}_{\{|Z|\leq n\}}$ a.s. Since $\mathbf{1}_{\{|Z|\leq n\}} \underset{n\to\infty}{\longrightarrow} 1$, by taking $n \to \infty$ we obtain the claim. $\square$

Corollary 9.6.2 *For $p \geq 1$, $\|\mathbf{E}(X|\mathcal{G})\|_p \leq \|X\|_p$. In particular, if $X \in L^p$, then $\mathbf{E}(X|\mathcal{G}) \in L^p$.*

Proof It suffices to apply the conditional Jensen inequality with $\varphi(x) = |x|^p$, $x \in \mathbf{R}$. $\square$

9.6.2 Conditional expectation as orthogonal projection

We have seen above that L_p, the space of L_p integrable and $\mathcal{F}$-measurable random variables, is complete under the norm L_p, and simple functions are dense in L_p. Furthermore L_2 is a Hilbert space, we denote it by $L_2(\Omega, \mathcal{F}, P)$. If $\mathcal{G}$ is a sub-σ algebra, the sub-space of $L_2(\Omega, \mathcal{F}, P)$ that are measurable with respect to $\mathcal{G}$, which we denote by $L_2(\Omega, \mathcal{G}, P)$ is a closed sub-space of L_2.

Since $L_2(\Omega, \mathcal{F}, P)$ is a Hilbert space and $L_2(\Omega, \mathcal{G}, P)$ is a closed subspace of L_2, let π denote the orthogonal projection defined by the projection theorem (§II.2 Functional Analysis [**?**]),

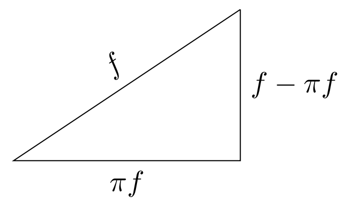
$$\pi : L_2(\Omega, \mathcal{F}, P) \to L_2(\Omega, \mathcal{G}, P).$$

Theorem 9.6.3 *If $X \in L_2(\Omega, \mathcal{F}, P)$ then $\mathbf{E}(X|\mathcal{G}) = \pi(X)$.*

Proof * Let $f \in L_2(\Omega, \mathcal{F}, P)$. Then for any $h \in L_2(\Omega, \mathcal{G}, P)$,

$$\langle f - \pi f, h \rangle_{L_2(\Omega,\mathcal{F},P)} = 0.$$

This is, $\int_\Omega fhdP = \int_\Omega \pi fhdP$.



Let $A \in \mathcal{G}$ and take $h = \mathbf{1}_A$ to see that $\pi f = \mathbf{E}(f|\mathcal{G})$. □

Corollary 9.6.4 *$\mathbf{E}(X)$ is the constant that minimizes $\mathbf{E}(X - c)^2$ and $\mathbf{E}(X|\mathcal{G})$ minimizes $\mathbf{E}(X - Y)^2$ where Y is a $\mathcal{G}$-measurable random variable.*

Remark 9.6.5 The projection πX is the unique element of $L_2(\Omega, \mathcal{G}, P)$ such that

$$\mathbf{E}|X - \pi X|^2 = \min_{Y \in L_2(\Omega,\mathcal{G},P)} \mathbf{E}|X - Y|^2.$$

Remark 9.6.6 The conditional expectation defines on L_2 can be extended to L_1 to give the conditional expectation, as defined earlier. Indeed, simple functions and bounded functions are in L_2. Let $f \in L_1$ with $f \geq 0$. Let f_n be a sequence of bounded functions (increasing with n) converging to f pointwise. Then πf_n is well defined and

$$\int_A f_n dP = \int_A \pi f_n dP, \qquad A \in \mathcal{G}.$$

Since f_n increases with n and is positive; so does πf_n and $\lim_{n\to\infty} \pi f_n$ exists. We may exchange limits and integration:

$$\int_A f dP = \lim_{n\to\infty} \int_A f_n dP = \int_A \lim_{n\to\infty} \pi(f_n) dP.$$

Thus we define

$$\mathbf{E}(f|\mathcal{G}) = \lim_{n\to\infty} \pi(f_n)$$

For $f \in L_1$ let $f = f^+ - f^-$ and define $\mathbf{E}(f|\mathcal{G}) = \mathbf{E}(f^+|\mathcal{G}) - \mathbf{E}(f^-|\mathcal{G})$.

9.7 Useful Inequalities and exercises*

Exercise 9.7.1 (Chebychev's inequality/ Markov inequality) If X is an L_p random variable then for $a \geq 0$,

$$P(|X| \geq a) \leq \frac{1}{a}\mathbf{E}|X|^p.$$

9.7.1 Exercises

Exercise 9.7.2 Suppose that Y_1, Y_2 be random variables on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ taking values in two measurable spaces $(\mathcal{X}_1, \mathcal{F}_1)$ and $(\mathcal{X}_2, \mathcal{F}_2)$ respectively. Let X be an integrable random variable on $(\Omega, \mathcal{A}, \mathbf{P})$. Show that the statement that a $\sigma(Y_1) \vee \sigma(Y_2)$-measurable random variable Z is the conditional expectation of X with respect to $\sigma(Y_1) \vee \sigma(Y_2)$ is equivalent to one of the following statements

1. For any $A \in \sigma(Y_1)$ and $B \in \sigma(Y_2)$,
$$\int_{A \cap B} Z \, d\mathbf{P} = \int_{A \cap B} X \, d\mathbf{P},$$
2. For any $g_i : \mathcal{X}_i \to \mathbf{R}$ Borel measurable and bounded,
$$\mathbf{E}(g_1(Y_1) g_2(Y_2) Z) = \mathbf{E}(g_1(Y_1) g_2(Y_2) X),$$

Exercise 9.7.3 Let $h : E \times E \to \mathbf{R}$ be a function, where $(E, \mathcal{F})$ is some measurable space. Let X, Y be random variables on a probability space $(\Omega, \mathcal{A}, \mathbf{P})$ with state space E such that $h(X, Y) \in L_1$. We assume that X and Y are independent. Let $H(y) = \mathbf{E}(h(X, y))$, for $y \in E$. Show that
$$\mathbf{E}(h(X, Y) | \sigma(Y)) = H(Y).$$

Exercise 9.7.4 Let $X : \Omega \to \mathcal{X}_1$ be $\mathcal{F}' \subset \mathcal{F}$ measurable and let $Y : \Omega \to \mathcal{X}_2$ be independent of $\mathcal{F}'$. Let $\Phi : \mathcal{X}_1 \times \mathcal{X}_2 \to \mathbf{R}$ be measurable and such that $\mathbf{E}|\Phi(X, Y)| < \infty$. For $x \in \mathcal{X}_1$ define $g(x) = \mathbf{E}[\Phi(x, Y)] = \int_\Omega \Phi\Big(x, Y(\omega)\Big) \mathbf{P}(d\omega)$. Then
$$\mathbf{E}\Big(\Phi(X, Y) \;\Big|\; \mathcal{F}'\Big) = g(X).$$

Proof Let $g_i : \mathcal{X}_i \to \mathbf{R}$ be bounded measurable. Let $\Phi(x, y) = g_1(x) g_2(y)$.
$$\mathbf{E}\big(g_1(X) \mathbf{E}(g_2(Y) | \mathcal{F}')\big) = g_1(X) \mathbf{E}\big(g_2(Y) \| \mathcal{F}'\big) = g_1(X) \mathbf{E} g_2(Y) = \tilde{\Phi}(X).$$

So $g_1(X) g_2(Y) \in \mathcal{H}$. where $\mathcal{H} \stackrel{def}{=} \Big\{\Phi : \qquad \mathbf{E}\big(\Phi(X, Y) \mid \mathcal{F}'\big) = \tilde{\Phi}(X)\Big\}$. The earlier statement shows that if $A \times B \in \mathcal{C}$, where $\mathcal{C} = \{A \times B : A \in \mathcal{B}(\mathcal{X}_1), B \in \mathcal{B}(\mathcal{X}_2)\}$. That $\mathcal{C}$ is a π system follows from $A \times B \cap C \times D = (A \cap C) \times (B \cap D)$. Finally, the π-system $\mathcal{C}$ generates $\mathcal{B}(\mathcal{X}_1) \times \mathcal{B}(\mathcal{X}_2)$ by the $\pi - \lambda$ theorem, and then we conclude the statement holds for any bounded measurable functions Φ, by approximating Φ with simple functions, and thus $\mathcal{H}$ contains all bounded Borel measurable functions from $\mathcal{X}_1 \times \mathcal{X}_2 \to \mathbf{R}$. □

Exercise 9.7.5 Prove that if X is an integrable random variable such that (X, Y) has a joint density $p(x, y)$, then for any Borel measurable set A and $x \in \mathbf{R}$,
$$P(X \in A | Y = y) = \frac{\int_A p(x, y) dx}{\int_{-\infty}^{\infty} p(x, y) dx}.$$

9.8 Conditional probability

We define similarly the concept of **conditional probability**.

Definition 9.8.1 Let $B \in \mathcal{F}$, define

$$\mathbf{P}(B|\mathcal{F}') := \mathbf{E}(\mathbf{1}_B|\mathcal{F}').$$

This is called the conditional probability of B given $\mathcal{F}'$.

Remark 9.8.1 ** For every $B \in \mathcal{F}$, this identity $\mathbf{P}(B \mid \mathcal{F}') = \mathbf{E}(\mathbf{1}_B \mid \mathcal{F}')$ holds outside of a null set. If B_i is a sequence of disjoint sets, there is a common set of null sets for every B_i, and so

$$\mathbf{P}(\bigcup_{i=1}^{\infty} B_i \mid \mathcal{F}') = \mathbf{E}\Big(\sum_i \mathbf{1}_{B_i} \mid \mathcal{F}'\Big) = \sum_i \mathbf{P}(B|\mathcal{F}')$$

holds a.e. To make $\mathbf{P}(\cdot \mid \mathcal{F}')$ a probability measure, the σ-additive property must hold for all sequences of disjoint unions, that is an unaccountable number of sequences in general, and one has to be able to choose a common set of null sets for all of them. If this can be done we have a family of probability measures $\mathbf{P}(\cdot|\mathcal{F}')(\omega)$, called regular conditional probabilities, and then

$$\mathbf{E}(X|\mathcal{F}')(\omega) = \int_\Omega X(\omega')\mathbf{P}(d\omega'|\mathcal{F}')(\omega), \quad a.e..$$

See Thm 7.1 in 'Probability measures on metric spaces' by K.R. Parthasarthy: regular conditional probabilities exist for complete separable metric spaces and their Borel σ-algebras.

9.8.1 Finite σ-algebras

For finite σ-algebras, we can construct a family of probability measures P^ω, so the conditional expectations is integration w.r.t. this family, by which we mean $\mathbf{E}(X|\mathcal{G})(\omega) = \int X dP^\omega$.

Let $A, B \in \mathcal{F}$. If $P(B) > 0$. define

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

Let $X : \Omega \to \mathbf{R}$, define,

$$\mathbf{E}(X|B) = \frac{\mathbf{E}(X\mathbf{1}_B)}{P(B)}.$$

If $\{B_1, \dots, B_n\}$ is a partition of Ω with $P(B_i) > 0$ for all i, we define $\mathcal{G} = \sigma\{B_i, i = 1, \dots, n\}$. Then for $\omega \in B_i$,

$$\mathbf{E}(X|\mathcal{G})(\omega) = \mathbf{E}(X|B_i)(\omega).$$

Fix ω. For $A \in \mathcal{F}$, define

$$P(A|\mathcal{G})(\omega) = \sum_{i=1}^{n} P(A|B_i)\mathbf{1}_{B_i}(\omega).$$

This is a probability measure, denoted by $P(d\omega'|\mathcal{G})(\omega)$. And

$$\mathbf{E}(X|\mathcal{G})(\omega) = \int X(\omega')P(d\omega'|\mathcal{G})(\omega).$$

For a large class of general σ-algebra, we can indeed construct a family of probability measures, called conditional probabilities.

Chapter 10

Mastery Material: Ergodic Theory

The mastery material are basics of ergodic theory. These are: definitions of measure preserving transformations, invariant sets, ergodic measures, Poincaré's Recurrence Theorem and Birkhoff's ergodic Theorem (and their proofs). These can be found in many textbooks.

- An Introduction to Infinite Ergodic Theory, by Jon Aaronson, published by the Americam Mathematical Society. In library, 517.518.1AAR
- Ergodic Theory and Dynamical systems by Yves Coudene (Chapter 2)

  https://link.springer.com/content/pdf/10.1007
- Introduction to ergodic theory, by Peter Walters
- Ergodic theory and information by Billingsley.

For your convenience, some basic concepts are given below, **this is not the full scope of the reading material.** Please refer to the references for a comprehensive study.

10.1 Basic Concepts

Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a measure space with the measure σ-finite.

Definition 10.1.1 A map $T : \mathcal{X} \to \mathcal{X}$ is a measure preserving transformation (map) if the pushed forward measure by T is the same as μ, i.e.

$$T_*\mu = T.$$

In other words $\mu(T^{-1}(A)) = \mu(A)$ for every measurable set A. We also say T preserves μ, also μ is invariant under T, and also μ is an invariant measure for T.

A Dirac measure δ_a is invariant under T if T leaves the point a fixed.

Definition 10.1.2 1. The invariant σ-algebra of a transformation T is

$$\mathcal{I} = \{A \in \mathcal{F} : T^{-1}(A) = A\}.$$

2. A function $f : \mathcal{X} \to \mathbf{R}$ is invariant under T if $f(Tx) = f(x)$ for every $x \in \mathcal{X}$.

It is easy to see that $\mathcal{I}$ is a σ-algebra.

Theorem 10.1.1 (Poincaré Recurrence Theorem) *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a finite measure space and let $T : \mathcal{X} \to \mathcal{X}$ be a measure preserving transformation. Let $A \in \mathcal{F}$. Then for almost every point in A, there exists some $n \geq 1$ (and hence an infinitely many n) such that $T^n(x) \in A$.*

If μ is a probability measure the celebrated Birkhoff's ergodic theorem states that the time average equals to spatial average.

Theorem 10.1.2 (Birkhoff's ergodic theorem) *Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space and T preserves measure. Then for any $f \in L_1$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} f(T^j x) = \mathbf{E}(f|\mathcal{I})(x), \quad a.e.$$

A transformation is said to be non-singular if $T_*\mu \ll \mu$. Let $T : \mathcal{X} \to \mathcal{X}$ preserves a probability measure μ. We say T is ergodic (we also say μ is ergodic) if whenever $A \in \mathcal{I}$ then $\mu(A) = 0$ or $\mu(A^c) = 0$.

10.1.1 Example: circle rotation

The Lebesgue measure on the unit circle S^1 is defined to be the pushed forward measure on the Lebesgue measure on $[0, 1)$ to S^1 by the map $\alpha \mapsto e^{2\pi\alpha i}$. A rotation is $x \mapsto xe^{2\pi\theta i}$ where $\theta \in \mathbf{R}$.

We are going to actually identify the unit circle S^1 with the quotient space $\mathbf{R}/Z$, so $x \sim y$ if $x = n+y$, where n is an integer. So S^1 is $[0, 1]$ with 0 and 1 glued together. From now on this is what we use.

Definition 10.1.3 Let $\theta \in S^1$ define $T_\theta : S^1 \to S^1$ by $T_\theta(\alpha) = \theta + \alpha$ mod 1. This is a called a rotation of the circle.

The Lebesgue measure is invariant under any circle rotation.

Chapter 11

Exercises from the problem sheets

11.1 Problem Sheet 1: Measurable sets

Exercise 11.1.1 Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and let $(A_n)_{n\geq 1}$ be a sequence of measurable sets. Show that the sets lim inf A_n and lim sup A_n are measurable, where

$$\text{lim inf } A_n = \bigcup_{N\geq 1} \bigcap_{n\geq N} A_n, \qquad \text{lim sup } A_n = \bigcap_{N\geq 1} \bigcup_{n\geq N} A_n.$$

Exercise 11.1.2 If $f : \Omega \to \mathcal{X}$ is a map and $\mathcal{B}$ is a σ-algebra over $\mathcal{X}$, prove that the collection of pre-image sets

$$\sigma(f) := \{f^{-1}(A) : A \in \mathcal{B}\}$$

is a σ-algebra.

Exercise 11.1.3 Let Ω be a set. Below we endow **R** with the Borel σ-algebra $\mathcal{B}$(**R**).

1. If $X : \Omega \to \mathbf{R}$ is a constant function, what is $\sigma(X)$?
2. If $X : \Omega \to \{0, 1\}$, and $\{0, 1\}$ is endowed with the discrete σ-algebra, what is $\sigma(X)$?
3. We say that $X : \Omega \to \mathbf{R}$ is an elementary function if it takes a finite number of values. If X is an elementary function, what is $\sigma(X)$?

Exercise 11.1.4 Give an example of an algebra which is not a σ-algebra.

Exercise 11.1.5 If $\mathcal{F}$ is a σ-algebra over a set X, and $A \subset X$, show that $\{A \cap B : B \in \mathcal{F}\}$ is a σ-algebra over A.

Exercise 11.1.6 Show that the intersection of an arbitrary family of σ-algebras is a σ-algebra. If $\mathcal{C}$ is any collection of subsets of a set $\mathcal{X}$, show that there always exists a smallest σ-algebra containing $\mathcal{C}$.

Exercise 11.1.7 Provide counter-examples to show the following:

1. If $\mathcal{A}$ and $\mathcal{B}$ are two σ-algebras on $\mathcal{X}$, $\mathcal{A} \cup \mathcal{B}$ is in general not a σ-algebra on $\mathcal{X}$.

2. If $\mathcal{A}$ (resp. $\mathcal{B}$) is a σ-algebras on $\mathcal{X}$ (resp. $\mathcal{Y}$), the collection of subsets

$$\{A \times B : \; A \in \mathcal{A}, B \in \mathcal{B}\}$$

   is in general not a σ-algebra on $\mathcal{X} \times \mathcal{Y}$.

3. If $(\mathcal{X}, \mathcal{A})$ is a measurable space, $\mathcal{Y}$ is a set, and $f : \mathcal{X} \to \mathcal{Y}$ is a map, the collection of subsets $\{f(A) : A \in \mathcal{A}\}$ is in general not a σ-algebra on $\mathcal{Y}$.

Exercise 11.1.8 Let $\mathcal{F}_1$ and $\mathcal{F}_2$ be two σ-algebras over a set $\mathcal{X}$. Show that $\mathcal{F}_1 \vee \mathcal{F}_2$, that is the σ-algebra generated by $\mathcal{F}_1 \cup \mathcal{F}_2$, can equivalently be characterised by the expressions:

- $\mathcal{F}_1 \vee \mathcal{F}_2 = \sigma\{A \cup B \,|\, A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2\}$,

- $\mathcal{F}_1 \vee \mathcal{F}_2 = \sigma\{A \cap B \,|\, A \in \mathcal{F}_1 \text{ and } B \in \mathcal{F}_2\}$.

Exercise 11.1.9 Show that:

1. the Borel σ-algebra over **R** is generated by closed intervals.

2. the set of irrational numbers is Borel measurable.

3. K is Borel measurable, where $K \subset [0, 1]$ is the standard Cantor set defined as follows. One first defines recursively a sequence $K_n, n \geq 1$ of subsets of $[0, 1]$ that are unions of 2^n disjoint closed intervals: we set $K_0 = [0, 1]$ and, for $n \geq 0$, we define K_{n+1} by removing the central third of each interval composing K_n. Thus

$$K_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right],$$

$$K_2 = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right],$$

   etc. Then K is defined by setting $K := \bigcap_{n \geq 0} K_n$.

Exercise 11.1.10 A collection of subsets $\mathcal{E}$ is an elementary family if

- $\phi \in \mathcal{E}$;
- If $A, B \in \mathcal{E}$, then $A \cap B \in \mathcal{E}$;
- if $A \in \mathcal{E}$ then A^c is a finite union of disjoint sets from $\mathcal{E}$.

Show that if $\mathcal{E}$ is an elementary family, then the collection $\mathcal{A}$ of finite disjoint unions of members of $\mathcal{E}$ is an algebra.

* **Exercise 11.1.11** The goal is to show that $\mathcal{B}(\mathbf{R}^2) = \mathcal{B}(\mathbf{R}) \otimes \mathcal{B}(\mathbf{R})$.

1. Let $\mathcal{A}$ be the collection of subsets defined by
$$\mathcal{A} := \{A \in \mathcal{B}(\mathbf{R}) : A \times V \in \mathcal{B}(\mathbf{R}^2) \text{ for all open subset } V \subset \mathbf{R}\}.$$
Show that $\mathcal{A} = \mathcal{B}(\mathbf{R})$.
2. Let $\mathcal{B}$ be the collection of subsets defined by
$$\mathcal{B} := \{B \in \mathcal{B}(\mathbf{R}) : A \times B \in \mathcal{B}(\mathbf{R}^2) \text{ for all } A \in \mathcal{B}(\mathbf{R})\}.$$
Show that $\mathcal{B} = \mathcal{B}(\mathbf{R})$.
3. Deduce that $\mathcal{B}(\mathbf{R}) \otimes \mathcal{B}(\mathbf{R}) \subset \mathcal{B}(\mathbf{R}^2)$.
4. We recall that any open subset of $\mathbf{R}^2$ is a countable union of subsets of the form $U \times V$, where U, V are open subsets of $\mathbf{R}$. Deduce that $\mathcal{B}(\mathbf{R}^2) \subset \mathcal{B}(\mathbf{R}) \otimes \mathcal{B}(\mathbf{R})$ and conclude.

Exercise 11.1.12 Let $\mathcal{X}$ be a set and let $\mathcal{C}$ be a non-empty collection of subsets of $\mathcal{X}$. Lucy claims: "For any $A \in \sigma(\mathcal{C})$, there must exist a countable sub-collection $\mathcal{D} \subset \mathcal{C}$ such that $A \in \sigma(\mathcal{D})$". Do you agree with Lucy? Prove your claim or give a counter example.

11.2 Problem Sheet 2: Measures

Exercise 11.2.1 Show that the countable additive property of a measure is equivalent to additive and continuous from below, i.e. if $(A_n)_{n\geq 1}$ is a monotone increasing sequence of sets then
$$\lim_{n\to\infty} \mu(A_n) = \mu(\bigcup_{n=1}^{\infty} A_n).$$

Exercise 11.2.2 We say that a measure μ on $(\mathbf{Z}, 2^{\mathbf{Z}})$ is invariant under translation if, for any $A \subset \mathbf{Z}$ and $x \in \mathbf{Z}$, $\mu(x + A) = \mu(A)$. Find all finite measures on $(\mathbf{Z}, 2^{\mathbf{Z}})$ which are invariant under translation.

Exercise 11.2.3 Consider the measure space $(\mathbf{R}, \mathcal{B}(\mathbf{R}), \lambda)$, with λ the Lebesgue measure on $\mathbf{R}$.

1. Show that, for any $x \in \mathbf{R}$, we have $\lambda(\{x\}) = 0$
2. John claims : "One can write $\mathbf{R}$ as the disjoint union, over all $x \in \mathbf{R}$, of the singleton $\{x\}$, and therefore:
$$\lambda(\mathbf{R}) = \sum_{x\in\mathbf{R}} \lambda(\{x\}) = \sum_{x\in\mathbf{R}} 0 = 0."$$
What is wrong with John's argument?

Exercise 11.2.4 Show that any countable set is Borel measurable and has Lebesgue measure zero.

Exercise 11.2.5 Let $(\mathcal{X}, \mathcal{F}, \mu)$ be a probability space and $E \in \mathcal{F}$.

1. Show that the set of $B \in \mathcal{F}$ with the properties $\mu(E \cap B) = \mu(E)\mu(B)$ is a λ -system.
2. We assume that there exists a π-system $\mathcal{C}$ such that $\sigma(\mathcal{C}) = \mathcal{F}$ and that, for all $B \in \mathcal{C}$, $\mu(E \cap B) = \mu(E)\mu(B)$. Show that the same equality holds for all $B \in \mathcal{F}$.

Exercise 11.2.6 Let $\mathcal{X}$ and $\mathcal{Y}$ be two sets. We assume that $\mathcal{X} = A_1 \cup A_2$ and $\mathcal{Y} = B_1 \cup B_2$, where both unions are disjoint. We endow $\mathcal{X}$ and $\mathcal{Y}$ with the σ-algebras $\mathcal{A} = \sigma(\{A_1, A_2\})$ and $\mathcal{B} = \sigma(\{B_1, B_2\})$ respectively, and we endow $\mathcal{X} \times \mathcal{Y}$ with the product σ-algebra $\mathcal{A} \otimes \mathcal{B}$ which is defined as $\sigma(\{A \times B,\ A \in \mathcal{A}, B \in \mathcal{B}\})$.

1. Show that $\mathcal{A} \otimes \mathcal{B}$ is given by the collection of sets written as a disjoint union
$$\bigcup_{(i,j)\in K} A_i \times B_j,$$
where $K \subset \{1, 2\} \times \{1, 2\}$. Deduce that a measure m on $\mathcal{A} \otimes \mathcal{B}$ is uniquely determined by its values on the product sets $A_i \times B_j$, for $i, j = 1, 2$.
2. Let a_1, a_2 and b_1, b_2 be positive numbers. We define a measure μ on $\mathcal{X}$ by $\mu(A_i) = a_i$, $i = 1, 2$, as well as a measure ν on $\mathcal{Y}$ by $\nu(B_j) = b_j$, $j = 1, 2$. Under which conditions on a_1, a_2, b_1, b_2 does there exist a measure m on the product σ-algebra $\mathcal{A} \times \mathcal{B}$ with respective marginals μ and ν, i.e. such that
$$m(A \times \mathcal{Y}) = \mu(A), \qquad m(\mathcal{X} \times B) = \nu(B),$$
for all $A \in \mathcal{A}$ and $B \in \mathcal{B}$? When these conditions are fulfilled, describe all such measures.

Exercise 11.2.7 Let $(\mathcal{X}, \mathcal{A})$ be a measurable space such that $\{x\} \in \mathcal{A}$ for all $x \in \mathcal{X}$. We say that $x \in \mathcal{X}$ is a point atom for μ if $\mu(\{x\}) > 0$. Furthermore, μ is said to be diffuse if it does not have any point atom, while it is said to be discrete if $\mu(S) = 0$ for any set S which does not contain any of its point atoms. Give a few examples of diffuse measures and discrete measures.

Exercise 11.2.8 Let μ be a Borel measure on $\mathbf{R}^n$. We set

$$S := \{x \in \mathbf{R}^n,\ \mu(B(r,x)) > 0 \text{ for all } r > 0\}.$$

Show that:

1. S is a closed subset of $\mathbf{R}^n$,
2. $\mu(S^c) = 0$,
3. any strict closed subset F of S satisfies $\mu(S \setminus F) > 0$.

S is called the *support* of the measure μ.

Exercise 11.2.9 Let λ be the Lebesgue measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Show that λ is regular, in the sense that, for all $A \in \mathcal{B}(\mathbf{R})$,

$$\begin{aligned}\lambda(A) &= \inf\{\lambda(U) : U \text{ open set, } A \subset U\}\\ &= \sup\{\lambda(K) : K \text{ compact set, } K \subset A\}.\end{aligned}$$

Exercise 11.2.10 Let $\mathcal{A}$ be a σ-algebra over a set $\mathcal{X}$. We say that two elements x and y of $\mathcal{X}$ are equivalent if

$$\forall A \in \mathcal{A}, \qquad x \in A \iff y \in A.$$

1. Show that this is indeed an equivalence relation on $\mathcal{X}$.
2. Show that, for all $x \in \mathcal{X}$, the equivalence class $[x]$ of x is given by $\bigcap\limits_{A \in \mathcal{A}:\, x \in A} A$.
3. (a) Give a description of the equivalence classes when $(\mathcal{X}, \mathcal{A}) = (\mathbf{R}, \mathcal{B}(\mathbf{R}))$, where $\mathcal{B}(\mathbf{R})$ is the Borel σ-algebra over $\mathbf{R}$.
   (b) Give an example of $(\mathcal{X}, \mathcal{A})$ for which the equivalence classes are not all singletons.
4. * Show that there does not exist an infinitely countable σ-algebra, i.e. that any σ algebra is either finite or uncountably infinite (**Hint**: show that if $\mathcal{A}$ is countable, then $[x] \in \mathcal{A}$ for all $x \in \mathcal{X}$, and that for each $A \in \mathcal{A}$, we have $A = \bigcup_{x \in A}[x]$; conclude by a contradiction argument that $\mathcal{A}$ cannot be countably infinite).

Exercise 11.2.11 Let $T : \Omega \to \Omega$ by a bijection and let $\mathcal{I} = \{A \subset \Omega : T^{-1}(A) = A\}$.

1. Show that $\mathcal{I}$ is a σ-algebra. This is the invariant σ-algebra for T.
2. Show that if $X : \Omega \to \mathbf{R}$ is a function satisfying $X = X \circ T$, then X is measurable w.r.t. $\mathcal{I}$.
3. Let $n \in \mathbf{N}$, let $X(\omega) = \sum_{i=1}^n b_i \mathbf{1}_{A_i}$ where b_i are real numbers and $A_i \in \mathcal{I}$. Show that $X = X \circ T$.

4. If $X : \Omega \to \mathbf{R}$ is $\mathcal{I}$-measurable, show that $X = X \circ T$. [1]

11.3 Problem Sheet 3: measurable functions, integrals

Exercise 11.3.1 Are the following statements true or false? Give a proof or provide a counter-example.

1. If two non-decreasing, right-continuous functions on $\mathbf{R}$ only differ by a constant, then they have the same Lebesgue-Stieltjes measure. What about the converse?
2. Let μ and ν be two measures on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, and let $\mathcal{C}$ be a π-system such that $\sigma(\mathcal{C}) = \mathcal{B}(\mathbf{R})$. If μ and ν agree on all elements of $\mathcal{C}$, then $\mu = \nu$.
3. Let $(\mathcal{X}_1, \mathcal{B}_1)$ and $(\mathcal{X}_2, \mathcal{B}_2)$ be measurable spaces, and $f : \mathcal{X}_1 \to \mathcal{X}_2$ a measurable map. If μ is a *probability* measure on $\mathcal{B}_1$, then $f_*(\mu)$ is a *probability* measure on $\mathcal{B}_2$.
4. Any continuous function from $\mathbf{R}$ to $\mathbf{R}$ is Borel measurable.
5. Any non-decreasing function from $\mathbf{R}$ to $\mathbf{R}$ is Borel measurable.
6. Let f be a differentiable function from $(0, 1)$ to $\mathbf{R}$. Prove that f' is Borel measurable.

Exercise 11.3.2 Let $(\mathcal{X}_1, \mathcal{F}_1)$ and $(\mathcal{X}_2, \mathcal{F}_2)$ be measurable spaces.

1. Show that if $\mathcal{F}_1$ is given by the power set of $\mathcal{X}_1$, then any map from $\mathcal{X}_1$ to $\mathcal{X}_2$ is measurable.
2. We assume that $\mathcal{F}_2$ contains all singletons $\{y\}$, for $y \in \mathcal{X}_2$. Show that if $\mathcal{F}_1 = \{\phi, \mathcal{X}_1\}$, then a measurable map from $\mathcal{X}_1$ to $\mathcal{X}_2$ is necessarily constant.

* **Exercise 11.3.3** Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and Ω a set. Let $Y : \Omega \to \mathcal{X}$ be a measurable function and consider the measurable space $(\Omega, \sigma(Y))$. Show that $X : \Omega \to \mathbf{R}$ is $\sigma(Y)$-measurable if and only if there exists a measurable function $f : \mathcal{X} \to \mathbf{R}$ such that $X = f \circ Y$. (**Hint:** recall that one can approximate pointwise any measurable function by a sequence of simple functions).

Exercise 11.3.4 Let $\mathcal{Y}$ be a set and $(\mathcal{Y}_i, \mathcal{B}_i)_{i \in I}$ be a family of measurable spaces. For all $i \in I$ we are given a map $f_i : \mathcal{Y} \to \mathcal{Y}_i$.

1. We endow $\mathcal{Y}$ with $\sigma(f_i, i \in I)$, the smallest σ-algebra for which all the f_i are measurable. Show that $\sigma(f_i, i \in I)$ coincides with $\sigma(\cup_{i \in I} \sigma(f_i))$.

[1] Hint: show that if

$$X_n(\omega) = \sum_{j=-\infty}^{\infty} \frac{j}{2^n} \mathbf{1}_{\left\{\omega : X(\omega) \in [\frac{j}{2^n}, \frac{j+1}{2^n})\right\}}(\omega).$$

then $X_n = X_n \circ T$. Show that X_n is an increasing sequence with $X_n \uparrow X$ pointwise (by this we mean $X_n(\omega) \uparrow X(\omega)$ for every ω.

2. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space and $g : \mathcal{X} \to \mathcal{Y}$. Show that g is measurable from $(\mathcal{X}, \mathcal{A})$ to $(\mathcal{Y}, \sigma(f_i, i \in I))$ if an only if, for all $i \in I$, $f_i \circ g$ is measurable from $\mathcal{X}$ to $\mathcal{Y}_i$.

Exercise 11.3.5 We recall that if $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ are two σ-algebras, the product σ-algebra $\mathcal{A} \otimes \mathcal{B}$ is defined as the σ-algebra over $\mathcal{X} \times \mathcal{Y}$ generated by sets of the form $A \times B$, with $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Moreover, we denote by $p_X : X \times Y \to X$ and $p_Y : X \times Y \to Y$ the canonical projections on X and Y.

1. Show that $\mathcal{A} \otimes \mathcal{B}$ coincides with $\sigma(p_X, p_Y)$, the smallest σ-algebra for which both p_X and p_Y measurable.

2. Let $(\mathcal{Z}, \mathcal{C})$ be a measurable space, and let $f : \mathcal{Z} \to \mathcal{X} \times \mathcal{Y}$. Show that f is measurable from $(\mathcal{Z}, \mathcal{C})$ to $(\mathcal{X} \times \mathcal{Y}, \mathcal{A} \otimes \mathcal{B})$ if and only if the maps $p_X \circ f : (\mathcal{Z}, \mathcal{C}) \to (\mathcal{X}, \mathcal{A})$ and $p_Y \circ f : (\mathcal{Z}, \mathcal{C}) \to (\mathcal{Y}, \mathcal{B})$ are measurable.

3. * Show that $\mathcal{B}(\mathbf{R}^2) = \mathcal{B}(\mathbf{R}) \otimes \mathcal{B}(\mathbf{R})$ (**Hint:** recall that any open subset of $\mathbf{R}^2$ is of the form $\cup_{n \geq 1} I_n \times J_n$, where I_n and J_n are open intervals of $\mathbf{R}$).

* **Exercise 11.3.6** Let $(\mathcal{X}, \mathcal{A})$ be a measurable space. We consider a sequence of Borel measurable maps $f_n : \mathcal{X} \to \mathbf{R}$, $n \geq 1$.

1. Show that the set
$$\{x \in \mathcal{X} : (f_n(x))_{n \geq 1} \text{ converges in } \mathbf{R}\}$$
is measurable.

2. Show that if $(f_n)_{n \geq 1}$ converges pointwise, that is, for all $x \in \mathbf{R}$, $(f_n(x))_{n \geq 1}$ converges in $\mathbf{R}$, then the map $\lim_{n \to \infty} f_n$ is Borel measurable from $(\mathcal{X}, \mathcal{A})$ to $\mathbf{R}$.

3. Let $a \in \mathbf{R}$. Prove the Borel measurability of the map $g : \mathcal{X} \to \mathbf{R}$ defined by
$$g(x) := \begin{cases} \lim_{n \to \infty} f_n(x) & \text{if } (f_n(x))_{n \geq 1} \text{ converges in } \mathbf{R} \\ a & \text{otherwise.} \end{cases}$$

Exercise 11.3.7 Let $[a, b]$ be an interval. A function $f : [a, b] \to \mathbf{R}$ is called Lebesgue measurable if it is measurable from $([a, b], \overline{\mathcal{B}([a, b])})$ to $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Suppose that I assigns a real number to every bounded Lebesgue measurable function $f : [a, b] \to \mathbf{R}$, which we denote by $I(f)$. We assume that I has the following properties:

1. (Linearity) If f, g are bounded Lebesgue measurable and $c, d \in \mathbf{R}$, we have $I(cf + dg) = cI(f) + dI(g)$.

2. (Monotonicity) If $f \leq g$ then $I(f) \leq I(g)$

3. For any Borel measurable subset A of $[a, b]$, $I(\mathbf{1}_A) = \lambda(A)$ where λ is the Lebesgue measure.

Show that if f is bounded and Riemann integrable, then $I(f) = \int_a^b f(x)dx$, where the latter is the Riemann integral of f. [2]

Exercise 11.3.8 For any non-decreasing, right-continuous function $F : \mathbf{R} \to \mathbf{R}$, we shall denote by μ_F the Lebesgue-Stieltjes measure associated with μ_F.

1. Show that if μ is a finite Borel measure on **R**, there exists a unique non-decreasing, right-continuous, bounded function F satisfying $F(-\infty) = 0$, and such that $\mu = \mu_F$. Give an expression for F in terms of μ.

2. Let $F : \mathbf{R} \to \mathbf{R}$ be a non-constant, non-decreasing, right-continuous, bounded function satisfying $F(-\infty) = 0$. Let $a := F(+\infty) \in (0, +\infty)$. We define $G : (0, a) \to \mathbf{R}$ by

$$G(u) := \inf\{x \in \mathbf{R} : F(x) \geq u\}, \qquad u \in (0, a).$$

   (a) Show that G is a well-defined, non-decreasing function on $(0, a)$. Deduce in particular that it is Borel measurable.

   (b) Show that, for all $x \in \mathbf{R}$ and $u \in (0, a)$,

$$G(u) \leq x \iff F(x) \geq u.$$

   (c) Let λ denote the Lebesgue measure on $(0, a)$, and let $G_\star(\lambda)$ denote the pushed forward measure through G. Show that, for any $x \in \mathbf{R}$, $G_*(\lambda)((-\infty, x]) = F(x)$. Deduce that $G_*(\lambda) = \mu_F$.

Exercise 11.3.9 Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space with a non-zero measure μ, and let $f : E \to \mathbf{R}$ be a Borel measurable function. Show that, for all $\epsilon > 0$, there exists $A \in \mathcal{A}$ with $\mu(A) > 0$ such that, for all $x, y \in A$, $|f(x) - f(y)| < \epsilon$.

Exercise 11.3.10 (Egoroff's Theorem) Let $(\mathcal{X}, \mathcal{F})$ be a measurable space endowed with a finite measure μ, and let $(f_n)_{n\geq 1}$ be a sequence of measurable functions from $\mathcal{X}$ to **R**. We assume that there exists a measurable function $f : \mathcal{X} \to \mathbf{R}$ such that $f_n \underset{n\to\infty}{\longrightarrow} f$ pointwise on $\mathcal{X}$.

1. Let $k \geq 1$. Sow that

$$\mu\left(\bigcap_{N\geq 1} \bigcup_{n\geq N} \left\{x \in \mathcal{X} : |f_n(x) - f(x)| \geq \frac{1}{k}\right\}\right) = 0.$$

[2] Hint: Given a partition of $[a, b]$, let $m_i = \inf\{f(x) : x \in [x_{i-1}, x_i)\}$, for which simple function g, $I(g) = \sum_{i=1}^n m_i(x_i - x_{i-1})$. Review the definition for f to be Riemann integrable.

Deduce that, for all $\delta > 0$, there exists a $N \geq 1$ such that

$$\mu\left(\bigcup_{n\geq N}\left\{x \in \mathcal{X} : |f_n(x) - f(x)| \geq \frac{1}{k}\right\}\right) \leq \delta.$$

2. * Deduce that, for all $\epsilon > 0$, there exists a $A \in \mathcal{F}$ such that $\mu(\mathcal{X} \setminus A) \leq \epsilon$ and f_n converges to f uniformly on f.

3. * Can one choose such a A with full measure, i.e. such that $\mu(\mathcal{X} \setminus A) = 0$? Prove your statement or give a counter-example.

11.4 Problem Sheet 4: limit theorems, computation of integrals

Exercise 11.4.1 Are the following statements true or false? When appropriate, give a proof or provide a counter-example.

1. Any continuous function on **R** is integrable with respect to the Lebesgue measure.

2. Any continuous function on $[0, 1]$ is integrable with respect to the Lebesgue measure.

3. If a Borel measurable function $f : \mathbf{R} \to \mathbf{R}$ is such that $\int_{\mathbf{R}} f \, d\lambda = 0$, then $f = 0$ almost-everywhere.

4. If f_n, f are measurable real-valued functions on a measure space $(\mathcal{X}, \mathcal{A}, \mu)$, and $f_n \uparrow f$ as $n \to \infty$, then $\int f_n d\mu \uparrow \int f d\mu$ as $n \to \infty$.

5. If $(f_n)_{n\geq 1}$ is a sequence of nonnegative measurable functions on a measure space $(\mathcal{X}, \mathcal{A}, \mu)$ such that $\sup_{n\geq 1} \int f_n d\mu < \infty$, and if $f_n \underset{n\to\infty}{\longrightarrow} f$ pointwise, then $\int f d\mu < \infty$.

Exercise 11.4.2 If f, g are real valued integrable functions on a measure space $(\mathcal{X}, \mu)$, show the following statements hold:

1. If $\mu(A) = 0$ then $\int_A f \, d\mu = 0$.

2. If $\int_A f \, d\mu = 0$ for every measurable set A then $f = 0$ μ almost-everywhere.

3. $|\int f \, d\mu| \leq \int |f| \, d\mu$

Exercise 11.4.3 (Markov's inequality) Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space and let f be a nonnegative, measurable function on $\mathcal{X}$. For all $M > 0$, show that $\int f \, d\mu \geq \int f \, \mathbf{1}_{\{f \geq M\}} \, d\mu$, and deduce that

$$\mu(\{f \geq M\}) \leq \frac{\int f \, d\mu}{M}.$$

Exercise 11.4.4 Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space.

1. Show that if $f_n \to f$ in $L^1(\mathcal{X}, \mathcal{A}, \mu)$, then
$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$

2. Let $(f_n)_{n\geq 1}$ be a sequence of nonnegative integrable functions converging μ-a.e. to an integrable function f. We assume that
$$\lim_{n\to\infty} \int f_n \, d\mu = \int f \, d\mu.$$
Show that $f_n \to f$ in $L^1(\mathcal{X}, \mathcal{A}, \mu)$ (**Hint**: first show that $\lim_{n\to\infty} \int (f - f_n)^+ \, d\mu = 0$).

Exercise 11.4.5 Let f be a Borel measurable function on **R**, and let $y \in$ **R**. Let λ be the Lebesgue measure on **R**.

1. Show that if f is non-negative, then
$$\int f(x+y) \, d\lambda(x) = \int f(x) \, d\lambda(x).$$

2. In general, show that f is integrable if and only if $x \to f(x+y)$ is integrable on **R**, and if so their integrals coincide.

Exercise 11.4.6 Let $[a, b]$ be an interval of **R**.

1. If $\varphi : [a, b] \to$ **R** is continuous, show that the function $F : [a, b] \to$ **R** defined by
$$F(x) = \int_{[a,x]} \varphi(x) \, d\lambda(x), \qquad x \in [a, b]$$
is differentiable on $[a, b]$, and $F' = \varphi$.

2. * If $f : [a, b] \to$ **R** is a differentiable function with bounded derivative, show that
$$\int_{[a,b]} f'(x) d\lambda(x) = f(b) - f(a).$$

Exercise 11.4.7 (Uniform continuity of integrals) Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space, and let f be a non-negative integrable function on $\mathcal{X}$. Let $\epsilon > 0$ be fixed.

1. Show that there exists a $M > 0$ such that
$$\int f \, \mathbf{1}_{\{f \geq M\}} \, d\mu \leq \frac{\epsilon}{2}$$

2. Deduce that there exists a $\delta > 0$ such that, for all $A \in \mathcal{A}$ with $\mu(A) \leq \delta$,

$$\int_A f\, d\mu \leq \epsilon.$$

3. Let $f : \mathbf{R} \to \mathbf{R}$ be a Borel measurable function. If f is integrable with respect to the Lebesgue measure, show that the function $F : \mathbf{R} \to \mathbf{R}$ defined by

$$F(x) = \int_{(-\infty,x]} f\, d\lambda, \qquad x \in \mathbf{R}$$

is uniformly continuous on $\mathbf{R}$.

Exercise 11.4.8 (Convergence in measure) Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a measure space with $\mu(\mathcal{X}) < \infty$, and let f_n, $n \geq 1$, and f be measurable functions from $(\mathcal{X}, \mathcal{A})$ to $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. We say that f_n converges to f in measure if, for all $\epsilon > 0$,

$$\mu(\{|f_n - f| > \epsilon\}) \underset{n\to\infty}{\longrightarrow} 0.$$

1. Using Markov's inequality (Ex.3), show that if $\int |f_n - f|\, d\mu \underset{n\to\infty}{\longrightarrow} 0$, then f_n converges to f in measure. Show, with a counter-example, that the converse is wrong.

2. Show that if f_n converges to f μ-a.e., then f_n converges to f in measure. Show, with a counter-example, that the converse is wrong.

3. * Assume that f_n converges to f in measure, and that there exists a non-negative integrable function $g : \mathcal{X} \to \mathbf{R}$ such that $|f_n| \leq g$ μ-a.e., for all $n \geq 1$.

   (a) Show that $|f| \leq g$ μ-a.e..

   (b) Using Exercise 7.2, show that

   $$\int |f_n - f|\, d\mu \underset{n\to\infty}{\longrightarrow} 0.$$

Exercise 11.4.9 Let I be a bounded Lebesgue measurable set. Suppose that $f : I \to \mathbf{R}$ is bounded.

1. Show that if f is Lebesgue measurable, then

$$\inf\left\{\int h\, d\lambda, \quad h \geq f, h \in \mathcal{S}\right\} = \sup\left\{\int g\, d\lambda : g \leq f, g \in \mathcal{S}\right\} \tag{11.1}$$

2. Show that if (11.1) holds, f is Lebesgue measurable. [3]

[3] Hint: extract a non-decreasing sequence of simple functions whose supremum is f.

11.5 Problem Sheet 5: Fubini's Theorems

Exercise 11.5.1 Let $\mathcal{X} = \{1, \ldots, N\}$ be a finite state space. We consider the measure space $(\mathcal{X}, \mathcal{A}, m)$, where $\mathcal{A} = 2^{\mathcal{X}}$ is the discrete σ-algebra and $m = \sum_{i=1}^{N} \delta_i$ is the counting measure on $(\mathcal{X}, \mathcal{A})$.

1. What can you say about the sigma-algebra $\mathcal{A} \otimes \mathcal{A}$ on $\mathcal{X} \times \mathcal{X}$? And about $m \otimes m$?
2. Write down the Fubini-Tonelli Theorem for the specific case of a measurable function on the measure space $(\mathcal{X} \times \mathcal{X}, \mathcal{A} \otimes \mathcal{A}, m \otimes m)$.

Exercise 11.5.2 Given a measure μ on $\mathcal{X} \times \mathcal{Y}$. We say μ_i are the marginals of μ if

$$\mu(\mathcal{X} \times B) = \mu_2(B), \quad \mu(A \times \mathcal{Y}) = \mu_1(A), \quad \forall A \in \mathcal{F}, B \in \mathcal{G}.$$

We also say that μ is a coupling of μ_1 and μ_2.

Give an example of a Borel measure μ on $\mathbf{R}^2$, two Borel measures μ_i on $\mathbf{R}$ such that μ is not the product measure $\mu_1 \times \mu_2$, but μ_i are marginals of μ. Give another set of examples of measures such that $\mu = \mu_1 \times \mu_2$ but μ_i are not the Lebesgue measure.

Exercise 11.5.3 Consider the measure space $(\mathbf{R}^2, \mathcal{B}(\mathbf{R}^2), \lambda)$, where λ is the Lebesgue measure on $\mathbf{R}^2$. Let f be the function on $\mathbf{R}^2$ given by

$$f(x,y) = \begin{cases} \sqrt{\frac{1-y}{x-y}} & \text{if } 0 \le y < x \text{ and } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Show that f is integrable and compute $\int f \, d\lambda$.

Exercise 11.5.4 Consider the measure space $([0,1]^2, \mathcal{B}([0,1]^2), \lambda)$, where λ is the Lebesgue measure on $[0,1]^2$. Let $\alpha \in \mathbf{R}$. and, for all $(x,y) \in [0,1]^2$, let

$$f(x,y) = \begin{cases} \frac{1}{|x-y|^\alpha} & \text{if } x \neq y \\ 0 & \text{otherwise} \end{cases}$$

Compute $\int_{[0,1]\times[0,1]} f d\lambda$. For which values of α is f integrable?

Exercise 11.5.5 Let $f(x,y) = e^{-xy} - 2e^{-2xy}$, for $(x,y) \in [0,1] \times [1,+\infty)$. Show that the integrals $\int_{[0,1]} \int_{[1,+\infty)} f(x,y) \, d\lambda(y) \, d\lambda(x)$ and $\int_{[1,+\infty)} \int_{[0,1]} f(x,y) \, d\lambda(x) \, d\lambda(y)$ exist but do not coincide. Deduce therefrom that f is not integrable on $[0,1] \times [1,+\infty)$.

Exercise 11.5.6 Let $f : (-1,1)^2 \to \mathbf{R}$ be the function defined by

$$f(x,y) = \begin{cases} \frac{xy}{(x^2+y^2)^2} & \text{if } (x,y) \in (-1,1)^2 \setminus \{(0,0)\} \\ 0 & \text{if } (x,y) = (0,0). \end{cases}$$

Show that the iterated integrals $\int\int f(x,y)\,d\lambda(x)\,d\lambda(y)$ and $\int\int f(x,y)\,d\lambda(y)\,d\lambda(x)$ exist and coincide, but that f is not integrable on $(-1,1)^2$.

Exercise 11.5.7 We consider the measurable space $([0,1],\mathcal{B}([0,1]))$. Let μ be the counting measure on $[0,1]$, which assigns to each $A\in\mathcal{B}([0,1])$ its number of elements $\#A\in\{0,1,\dots,+\infty\}$. Let also λ be Lebesgue measure on $[0,1]$. Let $\Delta=\{(x,x),\ x\in[0,1]\}$.

1. Show that $\Delta\in\mathcal{B}([0,1]^2)$.
2. John claims: " The function $\mathbf{1}_\Delta$ is a non-negative Borel measurable function on $[0,1]^2$, therefore by Fubini's Theorem we have
$$\int_{[0,1]}\int_{[0,1]}\mathbf{1}_\Delta(x,y)\,d\lambda(y)\,d\mu(x)=\int_{[0,1]}\int_{[0,1]}\mathbf{1}_\Delta(x,y)\,d\mu(x)\,d\lambda(y)\text{''}.$$
Is John's statement correct? If not, what is wrong with his argument?

Exercise 11.5.8 Let $f:\mathbf{R}\to\mathbf{R}_+$ be a measurable, non-negative function. Show that
$$\int_{\mathbf{R}_+}\lambda(\{x:f(x)\geq y\})\,d\lambda(y)=\int f\,d\lambda.$$
Deduce that, for all $g:\mathbf{R}\to\mathbf{R}$ measurable and all $p\geq1$,
$$p\int_{\mathbf{R}_+}\lambda(\{x:|g(x)|\geq a\})\,a^{p-1}d\lambda(a)=\int|g|^p\,d\lambda.$$

Exercise 11.5.9 (A nice application of Fubini) If $f:\mathbf{R}\to\mathbf{R}$ is a measurable function, show that $\lambda(\{x\in\mathbf{R}:f(x)=y\})=0$ for λ a.e. $y\in\mathbf{R}$.

Exercise 11.5.10

1. Show that the function $x\mapsto\frac{\sin(x)}{x}$ is Borel measurable from $(0,\infty)$ to $\mathbf{R}$, but not integrable with respect to λ.
2. * Show that, for all $x>0$,
$$\int_{(0,\infty)}e^{-tx}\,dt=\frac{1}{x},$$
and deduce that
$$\lim_{n\to\infty}\int_{(0,n)}\frac{\sin(x)}{x}=\frac{\pi}{2}.$$

Exercise 11.5.11 (Volume of the unit ball in $\mathbf{R}^d$) For all $d\geq1$ and $r\geq0$, let
$$B_d(r)=\{x\in\mathbf{R}^d:\sum_{i=1}^d x_i^2\leq r^2\}$$
be the closed ball of radius r in $\mathbf{R}^d$ centered at 0, and let $V_d(r)=\lambda^d(B_d(r))$, where λ^d is the Lebesgue measure on $\mathbf{R}^d$.

1. Prove that, for all $r \geq 0$, $V_d(r) = r^d V_d(1)$.

2. * Show that, for all $d \geq 3$. $V_d(1) = V_{d-2}(1) \int_{B_2(1)} (1 - x^2 - y^2)^{\frac{d-2}{2}} \, d\lambda(x) \, d\lambda(y)$. Using a polar change coordinates, deduce that $V_d(1) = V_{d-2}(1) \frac{2\pi}{d}$.

3. We denote by Γ the Gamma function defined for all $x > 0$ by $\Gamma(x) = \int_{(0,\infty)} t^{x-1} e^{-t} \, d\lambda(t)$, and recall that $\Gamma(x+1) = x\Gamma(x)$ for all $x > 0$, and that $\Gamma(1/2) = \sqrt{\pi}$. Show that for all $d \geq 1$,

$$V_d(1) = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2} + 1)}.$$

11.6 Problem Sheet 6: the Radon-Nikodym Theorems

Exercise 11.6.1 In the following let $(\mathcal{X}, \mathcal{A})$ be a fixed measurable space. Are the following statements true or false? Give a proof of provide a counter-example

1. If μ and ν are two measures on $(\mathcal{X}, \mathcal{A})$ such that $\mu \leq C\nu$ for some $C > 0$, then $\mu \ll \nu$. How about the converse?

2. If μ and ν are two measures on $(\mathcal{X}, \mathcal{A})$, there always exists a measure ξ such that $\mu \ll \xi$ and $\nu \ll \xi$.

3. If m is the counting measure on $\mathcal{X}$, then every measure μ on $(\mathcal{X}, \mathcal{A})$ is absolutely continuous with respect to m.

4. On $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, the measure $\mathbf{1}_{[0,1]}\lambda$ is absolutely continuous with respect to λ.

5. On $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$, the measure λ is absolutely continuous with respect to $\mathbf{1}_{[0,1]}\lambda$.

Exercise 11.6.2 Let μ and ν be two finite measures on a measurable space $(\mathcal{X}, \mathcal{A})$, and let D be a non-negative measurable function on $\mathcal{X}$. Show that the following conditions are equivalent:

1. $\mu \ll \nu$ and $\frac{d\mu}{d\nu} = D$ ν-a.e.

2. For all non-negative measurable function f on $\mathcal{X}$, $\int f \, d\mu = \int f D \, d\nu$

3. For all bounded measurable function f on $\mathcal{X}$, $\int f \, d\mu = \int f D \, d\nu$.

Exercise 11.6.3 Let μ and ν be two σ- finite measures on a measurable space $(\mathcal{X}, \mathcal{A})$ such that $\mu \ll \nu$. On which condition on $\frac{d\mu}{d\nu}$ do we also have $\nu \ll \mu$? In that case, give an expression for $\frac{d\nu}{d\mu}$.

Exercise 11.6.4 1. Let μ be a Borel measure on $\mathbf{R}^2$ admitting a density h with respect to the Lebesgue measure, where

$$h(x,y) = f(x)\,g(y), \qquad (x,y) \in \mathbf{R}^2,$$

where f and g are two non-negative Borel measurable functions on $\mathbf{R}$. Show that μ is the product of two measures on $\mathbf{R}$. In what case is μ a finite measure, resp. a probability measure?

2. Let X and Y be two real-valued random variables admitting a joint density f with respect to the Lebesgue measure λ^2 on $\mathbf{R}^2$:

$$\forall A \in \mathcal{B}(\mathbf{R}^2), \qquad \mathbf{P}((X,Y) \in A) = \int_A f\, d\lambda^2.$$

   (a) Show that X and Y both admit a density, denoted by f_X and f_Y respectively, with respect to the Lebesgue measure λ on $\mathbf{R}$, that is, for all $A \in \mathcal{B}(R)$

$$\mathbf{P}(X \in A) = \int_A f_X\, d\lambda, \qquad \mathbf{P}(Y \in A) = \int_A f_Y\, d\lambda.$$

   Give an expression for f_X and f_Y in terms of f.

   (b) Under which condition on f are X and Y independent?

Exercise 11.6.5 For all $h \in \mathbf{R}$, let $N(h,1)$ be the probability measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ given by

$$N(h,1)(A) = \int_A \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-h)^2}{2}}\, dx.$$

Show that the measures $N(h,1)$ and $N(0,1)$ are equivalent, and compute the corresponding Radon-Nikodym derivatives.

Exercise 11.6.6 Let ν be a finite measure. Show that $\nu \ll \mu$ if and only if the following holds: for any $\epsilon > 0$ there exists $\delta > 0$ such that whenever $\mu(E) < \delta$ then $\nu(E) < \epsilon$.

Exercise 11.6.7 On $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ we consider the Lebesgue measure λ and the counting measure μ which assigns, to every set $A \in \mathcal{B}(\mathbf{R})$, the number $\#A \in \{0, 1, \ldots, +\infty\}$ of elements of A.

1. Show that $\lambda \ll \mu$

2. Show that there does not exist a measurable function $f : \mathbf{R} \to \mathbf{R}_+$ such that, for all $A \in \mathcal{B}(\mathbf{R})$, $\lambda(A) = \int_A f\, d\mu$. Does this contradict the Radon-Nikodym theorem?

Exercise 11.6.8 Recall that $x \in \mathbf{R}$ is a point atom of a measure μ on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$ if $\mu(\{x\}) > 0$. Recall also that μ is said to be discrete if $\mu(A^c) = 0$, where A is the set of point atoms of μ, while μ is said to be diffuse if it does not have any point atom.

Let μ be a finite measure on $(\mathbf{R}, \mathcal{B}(\mathbf{R}))$. Show that there exists a unique decomposition $\mu = \mu_a + \mu_d + \mu_s$ where:

- μ_d is a discrete measure
- $\mu_a(A) = \int_A f \, d\lambda$ for all $A \in \mathcal{B}(\mathbf{R})$, for some non-negative integrable function f
- μ_s is diffuse and singular with respect to λ.

Exercise 11.6.9 (The devil's staircase) We construct recursively a sequence $(f_n)_{n\geq 0}$ of non-decreasing, piecewise linear functions on $[0,1]$ such that $f(0) = 0$ and $f(1) = 1$ as follows. We set $f_0(x) = x$ for $x \in [0,1]$. For $n \geq 0$, given the piecewise linear function f_n, we construct f_{n+1} by replacing f_n, on each maximal subinterval $[u, v]$ where it is not constant, by the piecewise linear function such that $f_{n+1}(u) = f_n(u)$, $f_{n+1}(v) = f_n(v)$, and f_{n+1} is identically equal to $(f_n(u)+f_n(v))/2$ on $[\frac{2u+v}{3}, \frac{u+2v}{3}]$. Thus, for $n = 1$,

$$f_1(x) = \begin{cases} (3x)/2 & \text{if } x \in [0, 1/3] \\ 1/2 & \text{if } x \in [1/3, 2/3] \\ (3x)/2 - 1/3 & \text{if } x \in [2/3, 1]. \end{cases}$$

1. Show that $|f_{n+1}(x) - f_n(x)| \leq 2^{-n}$ for all $n \geq 0$ and $x \in [0,1]$. Deduce therefrom that f_n converges uniformly on $[0,1]$ to a continuous, non-decreasing function f.
2. Let μ be the Lebesgue-Stieltjes measure on $[0,1]$ associated with f. Show that μ is a probability measure, and that it is diffuse (i.e. $\mu(\{x\}) = 0$ for all $x \in [0,1]$).
3. * Show that μ is singular with respect to the Lebesgue measure on $[0,1]$. [4]

11.7 Problem Sheet 7: Conditional expectations

Exercise 11.7.1 Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space and X an integrable real-valued random variable. Suppose that Y takes distinct values $\{s_1, \ldots, s_n\}$ with positive probability. Give an explicit formula for $\mathbf{E}(X|Y)$.

Exercise 11.7.2 We fix a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space, and Z a random variable with values in $\mathcal{X}$. Show that for every non-negative real-valued random variable X, there exists a measurable map $\varphi : (\mathcal{X}, \mathcal{A}) \to \mathbf{R}_+$ such that, for all measurable function $h : \mathcal{X} \to \mathbf{R}_+$, we have

$$\mathbf{E}(h(Z)X) = \mathbf{E}(h(Z)\varphi(Z)).$$

[4] Hint: Show that $\mu(K) = 1$ while $\lambda(K) = 0$, where K is the Cantor set.

Exercise 11.7.3 Let $X, Y \in L_1(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G}$ a sub-σ-algebra of $\mathcal{F}$. Show that if X is $\mathcal{G}$ measurable and $XY \in L_1$ then

$$\mathbf{E}(XY|\mathcal{G}) = X\mathbf{E}(Y|\mathcal{G}).$$

This means we take out 'what is known'.

Exercise 11.7.4 Let X, Y be two real-valued random variables defined on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. For all $A \in \mathcal{B}(\mathbf{R})$, we define

$$\mathbf{P}(X \in A|Y = y) := \mathbf{E}(\mathbf{1}_{\{X\in A\}}|Y = y), \qquad y \in \mathbf{R}.$$

Prove that if (X, Y) has joint density $p(x, y)$ w.r.t. the Lebesgue measure on $\mathbf{R}^2$, then for any $A \in \mathcal{B}(\mathbf{R})$ and a.e. $y \in \mathbf{R}$,

$$\mathbf{P}(X \in A|Y = y) = \frac{\int_A p(x, y)dx}{\int_{-\infty}^{\infty} p(x, y)dx}.$$

Exercise 11.7.5 Let X, Y be integrable real-valued random variables such that $\mathbf{E}(X) = \mathbf{E}(Y)$. Let

$$\mathcal{A} = \{A \in \mathcal{F} : \mathbf{E}(X\mathbf{1}_A) = \mathbf{E}(Y\mathbf{1}_A)\}.$$

Show that $\mathcal{A}$ is a λ-system.

Exercise 11.7.6 Show that if Y and Y' are integrable real-valued random variables on $(\Omega, \mathcal{F}, \mathbf{P})$ with $\mathbf{E}(Y) = \mathbf{E}(Y')$, and $\mathbf{E}(\mathbf{1}_A Y) = \mathbf{E}(\mathbf{1}_A Y')$ for all element A of a π-system $\mathcal{C}$ such that $\sigma(\mathcal{C}) = \mathcal{F}$, then $Y = Y'$ almost-surely.

Exercise 11.7.7 Let $(\Omega, \mathcal{F}, \mathbf{P})$ be a probability space, $\mathcal{F}_i$ sub-σ-algebras of $\mathcal{F}$ and denote by $\vee_{i=1}^n \mathcal{F}_i$ the σ-algebra generated by $\cup_{i=1}^n \mathcal{F}_i$.
Let $\mathcal{C} = \{\cap_{i=1}^n A_i : A_1 \in \mathcal{F}_1, \dots, A_n \in \mathcal{F}_n\}$.

(1) Show that $\vee_{i=1}^n \mathcal{F}_i = \sigma(\mathcal{C})$.

(2) Let X be an integrable real-valued random variable. If Y is $\vee_{i=1}^n \mathcal{F}_i$-measurable, $\mathbf{E}(Y) = \mathbf{E}(X)$, and $\mathbf{E}(\mathbf{1}_A X) = \mathbf{E}(\mathbf{1}_A Y)$ for all $A \in \mathcal{C}$, show that $Y = \mathbf{E}(X| \vee_{i=1}^n \mathcal{F}_i)$.

Exercise 11.7.8 (Conditional variance) Let X and Y be two real-valued random variables such that Y is square-integrable. We call conditional variance of Y given X, and denote by Var$(Y|X)$, the random variable $\mathbf{E}((Y - \mathbf{E}(Y|X))^2|X)$.

1. Provide an alternative expression for Var$(Y|X)$ in terms of $\mathbf{E}(Y^2|X)$ and $\mathbf{E}(Y|X)$.

2. In which case do we have Var$(Y|X) = 0$ a.s.? What is Var$(Y|X)$ if X and Y are independent?

3. Show that, for all $f : \mathbf{R} \to \mathbf{R}$ Borel measurable such that $f(X)$ is square-integrable

$$\mathbf{E}((Y - f(X))^2) = \mathbf{E}(\mathrm{Var}(Y|X)) + \mathbf{E}((\mathbf{E}(Y|X) - f(X))^2). \qquad (*)$$

   Which functions f minimise $\mathbf{E}((Y - f(X))^2)$?

4. By choosing an appropriate function f in $(*)$, prove that

$$\mathrm{Var}(Y) = \mathbf{E}(\mathrm{Var}(Y|X)) + \mathrm{Var}(\mathbf{E}(Y|X)).$$

Exercise 11.7.9 ** Show by a counter-example that $\mathbf{E}(\mathbf{E}(X \mid \mathcal{F}_2) \mid \mathcal{F}_1) = \mathbf{E}(\mathbf{E}(X \mid \mathcal{F}_1) \mid \mathcal{F}_2)$ a.s. is *not* true in general. [5]

[5] Example: divide a square, our Ω, into 4 squares by drawing a horizontal line and a vertical line through it. We denote by A_1, A_2, A_3, A_4 the sub-squares. Let $\mathcal{F}_1 = \sigma(\{A_1 \bigcup A_2, A_3 \bigcup A_4\})$ be generated by the horizontal partition and $\mathcal{F}_2 = \sigma(\{A_1 \bigcup A_3, A_2 \bigcup A_4\})$ the vertical partition. Let X be a random variable taking values in a_i on A_i respectively for $i = 1, 2, 3, 4$. Compute $\mathbf{E}(\mathbf{E}(X|\mathcal{F}_1)|F_2)$.