

Part II – Probability and Measure (Incomplete)

Based on lectures by E. Brieuillard

Notes taken by Bhavik Mehta

Michaelmas 2017

Contents

0	Introduction	2
0.1	Course structure	2
0.2	Historical motivation	2
1	Lebesgue measure	3
2	Abstract Measure Theory	8
2.1	Measurable functions	13
3	Integration	16
4	Probability Theory	25
4.1	Independence	26
5	Type of convergence of random variables	33
6	L^p spaces	36
6.1	Hilbert spaces	38

0 Introduction

0.1 Course structure

- Week 1: Lebesgue measure
- Week 2: Abstract measure theory
- Week 3: Integration
- Week 4: Foundations of probability theory
- Week 5: L^p spaces
- Week 6: Modes of convergence
- Week 7: Fourier transform and Gaussians
- Week 8: Ergodic Theory

0.2 Historical motivation

Suppose we have a subset $E \subset \mathbb{R}^d$.

1. What does it mean to measure this subset?

In one dimension, we have some intuition of length, and in two and three dimensions we are familiar with the notions of surface area and volume.

2. Does it make sense to measure every subset?

This seems reasonable, but it turns out that assigning a measure to every subset can lead to logical contradictions.

3. What is a measure? It should be a function defined on subsets, in particular some assignment $E \mapsto m(E) \in \mathbb{R}$. It should satisfy some properties:

- Non-negativity: $m(E) \geq 0$ for all E
- Empty set: $m(\emptyset) = 0$
- Additivity: $m(E \sqcup F) = m(E) + m(F)$ for any two disjoint sets E and F
- Normalisation: $m([0, 1]^d) = 1$
- Translation invariant: $m(E + x) = m(E)$ for all E and all $x \in \mathbb{R}^d$

It's possible to construct pathological 'measures' satisfying all these axioms and defined on *all* subsets of \mathbb{R}^d , but they won't be 'nice'. When mathematicians construct such measures, they usually do so on a restricted class of sets, otherwise this leads to contradictions.

If $d \geq 3$, it can be shown that there is no $m : P(\mathbb{R}^d) \rightarrow [0, \infty)$ that is also rotation invariant. This is referred to as the Hausdorff–Banach–Tarski paradox. Namely, if we take $B = B(0, 1) = \{ \mathbf{x} \in \mathbb{R}^d : x_1^2 + \dots + x_d^2 \leq 1 \}$, then there is a partition

$$B = X_1 \sqcup \dots \sqcup X_k \sqcup Y_1 \sqcup \dots \sqcup Y_k$$

and isometries $g_1, \dots, g_k, h_1, \dots, h_k$ such that

$$\bigcup g_i X_i = B = \bigcup h_i Y_i.$$

1 Lebesgue measure

Definition (Jordan measure of a box). Consider a box $B \subset \mathbb{R}^d$, given by $B = \prod_{i=1}^d I_i$, where $I_i = [a_i, b_i]$ are intervals in \mathbb{R} . From here, we can define the **Jordan measure** of the box by $m(B) := \prod_{i=1}^d |b_i - a_i|$.

Definition (Elementary subset). An **elementary subset** of \mathbb{R}^d is a finite union of boxes.

Remark. Every **elementary set** can be written as a finite union of disjoint boxes. The family of elementary sets is stable under finite unions, finite intersections and set difference. The concern may arise that if the disjoint union can be taken in two different ways, then perhaps we could get different measures, but

$$\begin{aligned} E &= \bigsqcup_{i=1}^N B_i = \bigsqcup_{j=1}^M B'_j \\ \implies \sum_{i=1}^N m(B_i) &= \sum_{j=1}^M m(B'_j) \end{aligned}$$

This means that *makes sense* to define $m(E)$ as $\sum_1^N m(B_i)$, for elementary subsets.

Definition (Jordan measurable). A subset $E \subset \mathbb{R}^d$ is **Jordan measurable** if $\forall \epsilon > 0 \exists$ **elementary sets** A, B such that $A \subset E \subset B$ and $m(B \setminus A) < \epsilon$.

Remark. If E is **Jordan measurable**, then

$$\inf \{ m(B) \mid E \subset B, B \text{ elementary} \} = \sup \{ m(A) \mid A \subset E, A \text{ elementary} \}$$

Proof is left as an exercise for the reader.

Definition (Jordan measure). We define the **Jordan measure** of E as this supremum or infimum, and denote it by $m(E)$.

Exercise. m so defined satisfies all the axioms defined earlier.

Definition (Riemann integrable function). A function $f : [a, b] \rightarrow \mathbb{R}$ is **Riemann integrable** if all its Riemann sums converge. Formally, the integral $I(f) \in \mathbb{R}$ exists if $\forall \epsilon > 0$, we can find $\delta > 0$ such that for every partition P of $[a, b]$ of width $\tau(P) < \delta$, we have $|S(f, P) - I(f)| < \epsilon$, where we recall the following

A partition P given by $a = t_0 < t_1 < \dots < t_n = b$ has width

$$\tau(P) = \max_{0 \leq i < N} |t_{i+1} - t_i|$$

and

$$S(f, P) = \sum_{i=1}^{N-1} f(x_i)(t_{i+1} - t_i)$$

where $x_i \in [t_i, t_{i+1}]$.

Proposition. f is **Riemann integrable** if and only if

$$E^+ = \{ (x, t) \in \mathbb{R}^2 \mid 0 \leq t \leq f(x) \}$$

$$E^- = \{ (x, t) \in \mathbb{R}^2 \mid f(x) \leq t \leq 0 \}$$

are both **Jordan measurable**.

However the [Jordan measure](#) is not perfect. For instance, the complement of a [Jordan measurable](#) set is not Jordan measurable. Also, we can take an (infinite) union of Jordan measurable sets and produce a set which is not Jordan measurable. In addition, there are simple sets that are not Jordan measurable, and simple functions that are not [Riemann integrable](#).

Example.

(i)

$$\mathbb{1}_{\mathbb{Q}}(x) = \begin{cases} 1 & x \in \mathbb{Q} \\ 0 & \text{otherwise} \end{cases}$$

This is not Riemann integrable, as seen in earlier Analysis courses.

(ii) \mathbb{Q} or even $\mathbb{Q} \cap [0, 1]$ is not a Jordan measurable subset of \mathbb{R} . In fact no dense countable subset of an interval in \mathbb{R} can be Jordan measurable.

Another problem is with limits of functions: a pointwise limit of [Riemann integrable](#) functions is not always Riemann integrable.

Example. $f_n(x) = \mathbb{1}_{\frac{1}{n}\mathbb{Z} \cap [0,1]}$ is Riemann integrable, but $f(x) = \lim_{n \rightarrow 0} f_n(x) = \mathbb{1}_{\mathbb{Q} \cap [0,1]}$ is not.

In comparison to the [Jordan measure](#), the ideas of Lebesgue were to remove the containment $A \subset E$, and to allow a countable union of boxes instead of just a finite union.

We begin by defining m^* , the *Lebesgue outer measure*:

Definition (Lebesgue outer measure). For a subset $F \subseteq \mathbb{R}^d$, the **Lebesgue outer measure** m^* is given by

$$m^*(F) = \inf \left\{ \sum_{n \geq 1} m(B_n) \mid F \subset \bigcup_{n \geq 1} B_n, B_n \text{ are boxes} \right\}$$

This is defined for all subsets of \mathbb{R}^d , but it is not additive on all subsets.

Remark.

- If you let $m^{*,J}(F)$ (the Jordan outer measure) be defined similarly, but for finitely many boxes B_n instead of countably many, then $m^*(F) \leq m^{*,J}(F)$.
- But the inequality can be strict, for instance for $F = \mathbb{Q} \cap [0, 1]$ we have $m^*(F) = 0$ but $m^{*,J}(F) = 1$.

The [Lebesgue outer measure](#) satisfies

1. $m^*(\emptyset) = 0$
2. $m^*(E) \leq m^*(F)$ if $E \subseteq F$ (monotone)
3. $m^*\left(\bigcup_{i \geq 1} E_i\right) \leq \sum_{i \geq 1} m^*(E_i)$ (countable subadditivity)

Example (Vitali's example). Let E be a set of representations of the cosets of the subgroup $(\mathbb{Q}, +) \subset (\mathbb{R}, +)$. We can choose $E \subset [0, 1]$, such that

$$\forall x \in \mathbb{R} \exists! e \in E \text{ such that } x - e \in \mathbb{Q}.$$

So the family $\{E + r\}_{r \in \mathbb{Q}}$ is a disjoint family of subsets of \mathbb{R} , a partition. By translational invariance,

$$m^*(E + r) = m^*(E) \quad \forall r \in \mathbb{Q}$$

If m^* were additive, we can consider distinct $r_1, \dots, r_N \in \mathbb{Q} \cap [0, 1]$ so

$$m^*\left(\bigcup_{i=1}^N E + r_i\right) = Nm^*(E)$$

but

$$\begin{aligned} \bigcup_{i=1}^N E + r_i &\subseteq [0, 2] \\ \implies m^*\left(\bigcup_{i=1}^N E + r_i\right) &\leq m^*([0, 2]) \leq 2 \end{aligned}$$

So for any $N \in \mathbb{N}$, we have $Nm^*(E) \leq 2$, hence $m^*(E) = 0$. On the other hand, $[0, 1] \subseteq \bigcup_{r \in \mathbb{Q}} E + r$, so by countable subadditivity $1 = m^*([0, 1]) \leq \sum_{r \in \mathbb{Q}} m^*(E + r) = 0$, so we have a contradiction. This shows that m^* is not additive on all subsets.

Remark.

- This construction uses the axiom of choice to define E .
- We will soon define the Lebesgue measure of a Lebesgue measurable set as the [outer measure](#) of that set, and this will be additive on Lebesgue measurable sets. This means Vitali's set E was *not* Lebesgue measurable.

Definition (Lebesgue measurable subset). $E \subseteq \mathbb{R}^d$ is **Lebesgue measurable** if $\forall \epsilon > 0$, $\exists C := \bigcup_n B_n$, a countable union of boxes such that $m^*(C \setminus E) < \epsilon$ and $E \subseteq C$.

Example (Middle-thirds Cantor set). This is a compact subset C of $[0, 1]$, defined as follows. Start with the interval $I_0 = [0, 1]$ and remove the middle third $(\frac{1}{3}, \frac{2}{3})$, leaving $I_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. Then remove the middle third of each interval here, giving $I_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. Repeat, then define $C = \bigcap_{n \geq 0} I_n$. Equivalently, take a ternary expansion of $x \in [0, 1]$, and define C as the set of x which has *none* of its digits equal to 2.

Remark. I_n is a finite union of intervals, so it is an [elementary set](#). In particular, $m(I_n) = 2^n/3^n \rightarrow 0$ as $n \rightarrow \infty$, so C is [Jordan measurable](#) with measure 0. Every Jordan measurable set is [Lebesgue measurable](#) (clear from the definition).

We can define a 'fat' Cantor set which is [Lebesgue measurable](#) but not [Jordan measurable](#).

Remark. What is the [outer measure](#) of a Vitali set? $m^*(E) > 0$ by the argument given above, but we can create $E \subseteq [0, \frac{1}{n}]$, or inside any interval. So, $m^*(E)$ depends on the choice of E and could be arbitrarily small, but must be positive.

Remark. In the definition of the [Lebesgue outer measure](#), we used closed boxes, but open or half-open boxes would not change the definition (same for Jordan measure).

Remark. Every null set is [Lebesgue measurable](#).

Definition (Null set). A **null set** $E \subseteq \mathbb{R}^d$ is a subset E such that $m^*(E) = 0$.

We state two propositions which will be proved together.

Proposition 1. The family \mathcal{L} of [Lebesgue measurable](#) subsets of \mathbb{R}^d is stable under

- a) countable unions: if $E_n \in \mathcal{L}$ for all $n \in \mathbb{N}$, then $\bigcup_{n \geq 1} E_n \in \mathcal{L}$.
- b) complementation: if $E \in \mathcal{L}$, then $E^c \in \mathcal{L}$.
- c) countable intersections: if $E_n \in \mathcal{L}$ for all $n \in \mathbb{N}$, then $\bigcap_{n \geq 1} E_n \in \mathcal{L}$.

Proposition 2. Every closed (open) subset of \mathbb{R}^d is [Lebesgue measurable](#).

First note that part c) follows from parts a) and b), since $(\bigcap_n E_n)^c = \bigcup_n E_n^c$. We'll first show a).

Proof of Proposition 1a). Let $(E_n)_{n \geq 1}$ be a countable family in \mathcal{L} , and pick $\epsilon > 0$. By definition, $\exists C_n := \bigcup_{i \geq 1} B_i^{(n)}$ such that $m^*(C_n \setminus E_n) < \frac{\epsilon}{2^n}$. Then $\bigcup E_n \subseteq \bigcup C_n = \bigcup_{n,i} B_i^{(n)}$ is still a countable union of boxes.

Finally,

$$m^*\left(\bigcup B_i^{(n)} \setminus \bigcup E_n\right) \leq \sum_n m^*(C_n \setminus E_n) \leq \sum_{n \geq 1} \frac{\epsilon}{2^n} \leq \epsilon$$

by countable subadditivity of m^* , proving a). □

Next, prove a lemma which will help to prove [Proposition 2](#).

Lemma. Every open subset of \mathbb{R}^d is a countable union of open boxes.

Proof. For $r \in \mathbb{Q}_{\geq 0}$ and $s \in \mathbb{Q}^d$, let $B_{s,r} = \{x \in \mathbb{R}^d \mid |x_i - s_i| < \frac{1}{r} \text{ for } i = 1, \dots, d\}$. $(B_{s,r})_{s,r}$ form a countable family and any open set $U \subseteq \mathbb{R}^d$ is the union of those $B_{s,r}$'s it contains. □

Now we prove open and closed sets are Lebesgue measurable.

Proof of Proposition 2. By part a) of [Proposition 1](#) and the lemma, we see that every open set in \mathbb{R}^d is Lebesgue measurable. Now, let's show that closed sets are Lebesgue measurable.

It's enough to show that *compact* subsets of \mathbb{R}^d are in \mathcal{L} because every closed subset $F \subseteq \mathbb{R}^d$ is a countable union of compact sets:

$\mathbb{R}^d = \bigcup_{n \geq 1} A_n$, where A_n is an annulus $\{x \in \mathbb{R}^d \mid n-1 \leq \|x\| \leq n\}$, so we can write $F = \bigcup_{n \geq 1} (A_n \cap F)$, where $A_n \cap F$ is compact.

So, let $F \subseteq \mathbb{R}^d$ be a compact subset. By definition of $m^*(F)$, $\forall k \geq 1$, \exists a countable union of open boxes $\bigcup_n B_n^{(k)}$ such that $F \subseteq \bigcup_n B_n^{(k)}$ and $m^*(F) + \frac{1}{2^k} \geq \sum_n m(B_n^{(k)})$.

Note:

- Up to subdividing each $B_n^{(k)}$ into a finite number of smaller boxes, we can assume that each $B_n^{(k)}$ has diameter less than $\frac{1}{2^k}$.

- Without loss of generality we can assume that each box meets F .
- Finally, since F is compact there is a finite subcover, so we can assume that there are only finite many boxes at each step, that is, $F \subseteq \bigcup_{n=1}^{N_k} B_n^{(k)}$ for $N_k \in \mathbb{N}$.

Let $U_k = \bigcup_{n=1}^{N_k} B_n^{(k)}$, so $F \subseteq U_k$ for all k and F meets each box $B_n^{(k)}$. Then we must have $F = \bigcap_{k \geq 1} U_k$, because if $x \in \bigcap U_k$, for any k we can find some $x_k \in F$ such that x and x_k lie in the same box $B_n^{(k)}$, and so $\|x - x_k\|_\infty \leq \frac{1}{2^k}$. F is compact, and $x_k \in F$ has a limit point, so $x \in F$.

We need to show that $m^*(U_k \setminus F)$ tends to 0, because this implies that F is Lebesgue measurable.

Claim that if A, B are two disjoint compact subsets of \mathbb{R}^d then $m^*(A \cup B) = m^*(A) + m^*(B)$. This is clear from definition as we can choose disjoint covers by open boxes.

Apply this to $A = F$ and $B = \overline{U_k} \setminus U_{k+1}$ to get

$$\begin{aligned} m^*(\overline{U_k} \setminus U_{k+1}) + m^*(F) &\leq m^*(\overline{U_k} \setminus U_{k+1} \cup F) \\ &\leq m^*(\overline{U_k}) \\ &= m^*(U_k) \\ &\leq m^*(F) + \frac{1}{2^k} \end{aligned}$$

so $m^*(U_k \setminus U_{k+1}) \leq \frac{1}{2^k}$. Since $F = \bigcap_k U_k$, by countable subadditivity of m^* , we get

$$\begin{aligned} m^*(U_k \setminus F) &\leq \sum_{i \geq k} m^*(U_i \setminus U_{i+1}) \\ &\leq \sum_{i \geq k} \frac{1}{2^i} \\ &\leq \frac{1}{2^{k-1}} \rightarrow 0. \end{aligned} \quad \square$$

Finally, we show that the complement of a set in \mathcal{L} is in \mathcal{L} .

Proof of Proposition 1b). We start with $E \in \mathcal{L}$. By definition, $\forall n$, there is a countable family of open boxes C_n with $E \subseteq C_n$ and $m^*(C_n \setminus E) < \frac{1}{n}$.

Taking complements, we see $C_n^c \subseteq E^c$ and $C_n \setminus E = E^c \setminus C_n^c$, and note C_n is open and C_n^c is closed.

By Proposition 2, C_n^c is Lebesgue measurable, and by part a) of Proposition 1, so is $\bigcup_n C_n^c$. But $\bigcup_n C_n^c \subseteq E^c$ and

$$m^*\left(E^c \setminus \bigcup_n C_n^c\right) \leq m^*(E^c \setminus C_n^c) = m^*(C_n \setminus E) < \frac{1}{n}$$

Hence $m^*(E^c \setminus \bigcup_n C_n^c) = 0$, so

$$E^c = \underbrace{\bigcup_n C_n^c}_{\in \mathcal{L}} \cup \underbrace{E^c \setminus \bigcup_n C_n^c}_{\text{is null hence is in } \mathcal{L}}$$

so by Proposition 1 part a), $E \in \mathcal{L}$. □

2 Abstract Measure Theory

Definition (σ -algebra). Let X be a set. A family \mathcal{A} of subsets of X which contains the empty set \emptyset and is stable under countable unions and complementation is called a σ -algebra.

Remark. σ stands for ‘countable’.

Definition (Measurable space). A **measurable space** is a couple (X, \mathcal{A}) where X is a set and \mathcal{A} is a σ -algebra of subsets of X .

Definition (Measure). A **measure** on (X, \mathcal{A}) is a function $\mu : \mathcal{A} \rightarrow [0, \infty]$ such that $\mu(\emptyset) = 0$ and μ is countably additive. That is, if $\{A_n\}$ is a pairwise disjoint countable family of subsets from \mathcal{A} , then

$$\mu\left(\bigsqcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mu(A_n)$$

Remark. Sometimes people call a **set function** a function μ from a family of subsets of X to $[0, \infty]$ such that $\mu(\emptyset) = 0$.

Definition (Measure space). If X is a set, \mathcal{A} a σ -algebra on X and μ a **measure** on \mathcal{A} then (X, \mathcal{A}, μ) is called a **measure space**.

Definition (Lebesgue measure). If $E \in \mathcal{L}$ (E is **Lebesgue measurable**), we define the **Lebesgue measure** of E as $m(E) = m^*(E)$, the **outer measure** of E .

We’ve already shown $(\mathbb{R}^d, \mathcal{L})$ is a **measurable space**, but we would like to show that $(\mathbb{R}^d, \mathcal{L}, m)$ is a **measure space**. To do this, it only remains to show countable additivity of the **Lebesgue measure** on **Lebesgue measurable sets**, so it would be useful to know a little more about them. So, let’s prove a small lemma which helps with this.

Lemma. If E is a **Lebesgue measurable** subset of \mathbb{R}^d then

- (1) $\forall \epsilon > 0, \exists U \subseteq \mathbb{R}^d$ an open set with $E \subseteq U$ and $m^*(U \setminus E) < \epsilon$
- (2) $\forall \epsilon > 0, \exists F \subseteq E$ a closed set with $m^*(E \setminus F) < \epsilon$.

Proof. Note (2) follows from (1) applied to E^c : we get $E^c \subseteq U$ and $m^*(U \setminus E^c) \leq \epsilon$. So, set $F = U^c$ and direct calculation gives $U \setminus E^c = E \setminus F$.

So let’s prove (1). Since E is **Lebesgue measurable**, $\forall \epsilon > 0, \exists \bigcup B_n$ with $E \subseteq \bigcup B_n$ such that $m^*(\bigcup B_n \setminus E) < \epsilon$. Recall we can take the B_n to be open boxes, then just set $U = \bigcup_n B_n$. \square

With that done, let’s state and prove our proposition.

Proposition. m is countably additive on \mathcal{L} , hence $(\mathbb{R}^d, \mathcal{L}, m)$ is a **measure space**.

Proof. Recall we need to show additivity for a countable family of pairwise disjoint **Lebesgue measurable** subsets. We do this by proving a series of special cases, getting more general each time.

In turn, we prove additivity for

- (i) Two compact subsets
- (ii) Countably many compact subsets
- (iii) Countably many bounded subsets
- (iv) Countably many measurable subsets

In each case we can take the sets pairwise disjoint.

- (i) For two disjoint compact subsets $A, B \subseteq \mathbb{R}^d$, we've seen $m^*(A) + m^*(B) = m^*(A \cup B)$, as required.
- (ii) Suppose now $\{E_n\}_n$ is a family of pairwise disjoint compact subsets of \mathbb{R}^d . By iterating (i),

$$m^*\left(\bigcup_{i=1}^N E_i\right) = \sum_{i=1}^N m^*(E_i)$$

so,

$$\sum_{i=1}^N m^*(E_i) \leq m^*\left(\bigcup_{i=1}^{\infty} E_i\right) \leq \sum_{i=1}^{\infty} m^*(E_i)$$

Let $N \rightarrow \infty$, and therefore

$$\sum_{i=1}^{\infty} m^*(E_i) = m^*\left(\bigcup_{i=1}^{\infty} E_i\right)$$

- (iii) By (2) of the lemma, $\exists F_n \subseteq E_n$ with $m^*(E_n \setminus F_n) < \epsilon/2^n$ and F_n closed. But E_n is bounded by assumption, so F_n is closed and bounded, hence compact. E_n 's are pairwise disjoint, so the F_n 's are also. Then $m^*(E_n) \leq m^*(F_n) + m^*(E_n \setminus F_n)$ and

$$\begin{aligned} \sum_{n=1}^{\infty} m^*(E_n) &\leq \sum_{n=1}^{\infty} m^*(F_n) + \epsilon \sum_{n=1}^{\infty} \frac{1}{2^n} \\ &= m^*\left(\bigcup_{n=1}^{\infty} F_n\right) + \epsilon \\ &\leq m^*\left(\bigcup_{n=1}^{\infty} E_n\right) + \epsilon \end{aligned}$$

This holds $\forall \epsilon > 0$, so $\sum_{n \geq 1} m^*(E_n) \leq m^*(\bigcup_{n=1}^{\infty} E_n)$ and equality follows because the other inequality holds by countable subadditivity.

- (iv) In the general case, we reduce to the bounded case:

Let $A_m = \{x \in \mathbb{R}^d \mid m-1 \leq \|x\| < m\}$, so we can write $\mathbb{R}^d = \bigcup_{m \geq 1} A_m$, with A_m compact. Apply the previous case to the countable family $(A_m \cup E_n)_{n,m}$. \square

Remark. Of course there are many other measures on $(\mathbb{R}^d, \mathcal{L})$, for example $f \in C_c(\mathbb{R}^d)$, $f \geq 0$, $\mu_f(E) := \int_{\mathbb{R}^d} f(x) \mathbb{1}_E(x) dx$.

Proposition. On $(\mathbb{R}^d, \mathcal{L})$, there is a unique measure μ which is

- Translation invariant: $\forall E \in \mathcal{L}, \forall x \in \mathbb{R}^d, \mu(E + x) = \mu(E)$
- $\mu([0, 1]^d) = 1$.

Proof. We'll skip this for now, and see another proof later. \square

Remark. The countable additivity here is crucial. There are lots of finitely additive measures of $(\mathbb{R}^d, \mathcal{L})$ which are translation invariant and unrelated to m . There are even such μ with $\mu(\mathbb{R}^d) < \infty$.

In an [earlier lemma](#) we started to characterise [Lebesgue measurable sets](#), and we can easily extend this further.

Lemma. If $E \in \mathcal{L}$,

- \exists a sequence U_n of open sets with $U_{n+1} \subseteq U_n$ and $E \subseteq U_n$ such that $m(\bigcap U_n \setminus E) = 0$.
- \exists a sequence F_n of closed sets with $F_n \subseteq F_{n+1}$ and $F_n \subseteq E$ such that $m(E \setminus \bigcup F_n) = 0$.

So, we can write

$$E = \bigcap_n U_n \setminus N_1$$

$$E = \bigcup_n F_n \cup N_2$$

where N_1 and N_2 are [null sets](#).

Remark.

- A countable intersection of open sets is called a G_δ -set.
- A countable union of closed sets is called a F_σ -set.
- So every Lebesgue measurable set is the difference of a G_δ -set and a null set, alternatively it is the union of a F_σ -set and a null set.

Definition (Boolean algebra). If X is a set, a **Boolean algebra** of subsets of X is a family \mathcal{B} of subsets of X which contains \emptyset and X and is stable under finite union and complementation.

Remark.

- A σ -algebra is a [Boolean algebra](#) that is also stable under countable unions.
- Every finite Boolean algebra is a σ -algebra.

Example.

- Trivial Boolean algebra, $\mathcal{B} = \{\emptyset, X\}$.
- Discrete Boolean algebra, $\mathcal{B} = \mathcal{P}(X)$.
- The elementary Boolean algebra, with $\mathcal{B} = \{\text{finite unions of boxes in } \mathbb{R}^d, \text{ or complements of such}\}$.

- Take a partition X into finitely many pieces $X = \bigsqcup_{i=1}^N P_i$.

$$\mathcal{B} := \left\{ A \subseteq X \mid \exists I \subseteq \{1, \dots, N\}, A = \bigcup_{i \in I} P_i \right\}.$$

This is called the atomic Boolean algebra.

- X is a topological space, and \mathcal{B} = finite unions of sets of the form $U \cap F$, with U open and F closed. This is called the Boolean algebra of constructible subsets of X .
- Null algebra,

$$\mathcal{B} = \{ E \subseteq \mathbb{R}^d \mid m^*(E) = 0 \text{ or } m^*(E^c) = 0 \}.$$

Remark. The intersection of any number of Boolean algebras on the same set X is again a Boolean algebra. The same holds for σ -algebras.

Definition (Generator of an algebra).

- If \mathcal{F} is a family of subsets of X , the Boolean algebra **generated** by \mathcal{F} is the intersection of all Boolean algebras containing \mathcal{F} .
- Same for σ -algebras: we denote this by $\sigma(\mathcal{F})$.

Remark. The Boolean algebra generated by a family \mathcal{F} of subsets of X is the family of sets of the form

$$\bigcup_{\text{finite}} \bigcap_{\text{finite}} F \text{ or } F^c$$

where $F \in \mathcal{F}$.

Note a finite intersection of a finite union of sets from a family \mathcal{F} is always a finite union of finite intersections of sets from \mathcal{F} . But, this is no longer true if you change finite to countable:

$$\bigcap_{i \in \mathbb{N}} \bigcup_{j \in \mathbb{N}} E_j^{(i)} = \bigcup_{\substack{i \mapsto j_i \\ \mathbb{N} \rightarrow \mathbb{N}}} \bigcap_{i \in \mathbb{N}} E_{j_i}^{(i)}$$

and there are more than countably many functions $\mathbb{N} \rightarrow \mathbb{N}$.

Definition (Borel σ -algebra). The σ -algebra generated by the family of boxes in \mathbb{R}^d is called the **Borel σ -algebra**. Elements of the Borel σ -algebra are called **Borel sets**.

Remark. It is also the **σ -algebra** generated by all open subsets of \mathbb{R}^d (or by all closed subsets), because every open set is a countable union of boxes.

More generally, if X is an arbitrary topological space, the **Borel** σ -algebra of X is the σ -algebra generated by open subsets of X . We denote it by $\mathcal{B}(X)$.

Note, we've shown that every open subset $E \subseteq \mathbb{R}^d$ is **Lebesgue measurable**, hence $\mathcal{B}(\mathbb{R}^d) \subseteq \mathcal{L}$. We might ask if we have equality. No. One can show that $|\mathcal{B}(\mathbb{R}^d)| = 2^{|\mathbb{N}|}$. But, remember that every null set is in \mathcal{L} , and the middle-thirds Cantor set is null.

Definition (Null set).

- **Null sets** of a measure space (X, \mathcal{A}, μ) are subsets in \mathcal{A} with μ -measure 0.

- The **sub-null** sets are the subsets of X contained in a null set.

Definition (Complete algebra). The family \mathcal{A}^* of subsets of X of the form $A \cup N$ or $A \setminus N$ where $A \in \mathcal{A}$ and N is a **sub-null** set (with respect to μ) is called the **completion**. If all sub-null sets are in \mathcal{A} , then \mathcal{A} is called **complete**.

Proposition. The completion forms a σ -algebra.

Example. The completion of the **Borel σ -algebra** $\mathcal{B}(\mathbb{R}^d)$ is \mathcal{L} .

We can prove basic properties of measures on an arbitrary measure space.

Proposition (Monotone convergence for sets). If (X, \mathcal{A}, μ) is a measure space,

- (a) Upwards monotone convergence for sets: If we have $E_1 \subseteq E_2 \subseteq \cdots \subseteq E_n \subseteq \cdots$, with $E_i \in \mathcal{A}$, then

$$\lim_n \mu(E_n) = \sup_n \mu(E_n) = \mu \left(\bigcup_n E_n \right)$$

- (b) Downwards monotone convergence: If we have $E_1 \supseteq E_2 \supseteq \cdots \supseteq E_n \supseteq \cdots$ with $E_i \in \mathcal{A}$ and $\mu(E_1) < \infty$, then

$$\lim_n \mu(E_n) = \inf_n \mu(E_n) = \mu \left(\bigcap_n E_n \right)$$

Proof. This follows from countable additivity of μ , say for b) consider

$$E_1 = \bigsqcup_{i \geq 1} E_i \setminus E_{i+1} \sqcup \bigcap_n E_n. \quad \square$$

Remark. To be able to extend a **measure** μ on a **Boolean algebra** \mathcal{B} to $\sigma(\mathcal{B})$, μ has to be σ -additive: If $B_n \in \mathcal{B}$ disjoint, and $\bigcup B_n \in \mathcal{B}$, then

$$\mu \left(\bigcup B_n \right) = \sum_{n \geq 1} \mu(B_n).$$

Definition (σ -finite). A **measure** μ on X is called **σ -finite** if X is a countable union of subsets each with finite μ -measure.

Theorem (Carathéodory's Extension Theorem). If \mathcal{B} is a **Boolean algebra** of subsets of X and μ is a countably additive **measure** on \mathcal{B} , then μ extends to a measure on $\sigma(\mathcal{B})$.

Moreover, the extension is unique if μ is **σ -finite**.

The theorem has two parts - existence of the extension, and uniqueness. Prove existence first:

These proofs are non examinable, and are as such currently incomplete in these notes. They may be completed at some later date or removed entirely.

Proving existence of the extension. For an arbitrary subset $A \subseteq X$, define the following quantity:

$$\mu^*(A) = \inf \left\{ \sum_{n \geq 1} \mu(B_n) \mid A \subseteq \bigcup_{n \geq 1} B_n, \text{ each } B_n \in \mathcal{B} \right\}$$

analogous to the **Lebesgue outer measure** defined in terms of the **Jordan measure**. We aim to show this is the required extension of μ . \square

Proving uniqueness of the extension. □

Exercise.

- Recall $\mathcal{B}^* = \{ E \subseteq X \mid E \text{ is } \mu^* \text{-measurable} \}$, where E is μ^* measurable means that

$$\forall A \subseteq X \quad \mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c).$$

- Every subset in \mathcal{B}^* is of the form $E = A \cup N$ where $A \in \sigma(\mathcal{B})$ and $\mu^*(N) = 0$.
- Show \mathcal{B}^* is the completion of $\sigma(\mathcal{B})$ with respect to μ^* .
- When \mathcal{B} is the [elementary Boolean algebra](#) of \mathbb{R}^d then $\sigma(\mathcal{B})$ is the Borel σ -algebra, $\mathcal{B}^* = \mathcal{L} = \sigma$ -algebra of [Lebesgue measurable](#) sets.

Remark. This means the [Carathéodory extension theorem](#) gives an alternative way to construct the σ -algebra of Lebesgue measurable sets in \mathbb{R}^d and the Lebesgue measure.

There is a much stronger uniqueness theorem:

Theorem. Suppose (X, \mathcal{A}) is a [measurable space](#), and μ_1, μ_2 are two finite measures on (X, \mathcal{A}) such that $\mu_1(X) = \mu_2(X)$. Suppose further that $\mu_1(A) = \mu_2(A)$ for all $A \in \mathcal{F}$ where \mathcal{F} is a subfamily of \mathcal{A} which generates \mathcal{A} and is stable under finite intersections. Then $\mu_1 = \mu_2$.

Lemma (Dynkin's Lemma). Let \mathcal{A} be a σ -algebra and \mathcal{F} a family of subsets in \mathcal{A} which generates \mathcal{A} and is stable under finite intersections and contains \emptyset (such \mathcal{F} is sometimes called a π -system). Let \mathcal{C} be a subfamily of \mathcal{A} , which contains \mathcal{F} and is stable under complementation and countable disjoint unions. Then $\mathcal{C} = \mathcal{A}$.

Proof. Omitted here. □

We can prove the uniqueness theorem using [Dynkin's Lemma](#):

Proof. Observe that if $\mathcal{C} = \{ A \in \mathcal{A} \mid \mu_1(A) = \mu_2(A) \}$, then \mathcal{C} satisfies the assumptions of Dynkin's lemma, so $\mathcal{C} = \mathcal{A}$. So $\mu_1 = \mu_2$. □

2.1 Measurable functions

Definition (Real-valued measurable function). Let (X, \mathcal{A}) be a [measurable space](#) and $f : X \rightarrow \mathbb{R}$ a real-valued function. We say that f is \mathcal{A} -**measurable** (or just measurable if it is clear what the σ -algebra in question is) if

$$\forall t \in \mathbb{R}, \quad \{ x \in X \mid f(x) < t \} \in \mathcal{A}.$$

Remark. If f is \mathcal{A} -measurable, then $f^{-1}(B) \in \mathcal{A}$ for all [Borel subsets](#) $B \subseteq \mathbb{R}$. Indeed, f^{-1} is functorial with respect to \cup, \cap and complementation, so $\{ B \subset \mathbb{R} \mid f^{-1}(B) \in \mathcal{A} \}$ is a [\$\sigma\$ -algebra](#). If f is measurable by definition, it contains $(-\infty, t) \forall t$ so it contains all intervals, hence all Borel sets.

An important caveat is that the pre-image under f of a Lebesgue measurable subset of \mathbb{R} which isn't Borel may not be in \mathcal{A} .

More generally,

Definition (Measurable function). If (X, \mathcal{A}) and (Y, \mathcal{B}) are measurable spaces, $f : X \rightarrow Y$ is said to be **measurable** if

$$f^{-1}(B) \in \mathcal{A} \quad \forall B \in \mathcal{B}.$$

Remark.

- To check that $f : X \rightarrow Y$ is measurable, it is enough to check that $f^{-1}(B) \in \mathcal{A}$ for $B \in \mathcal{F}$, a generating family of \mathcal{B} .
- We can easily check that the composition of measurable functions is measurable.
- For (X, \mathcal{A}) a measurable space, $f : X \rightarrow \mathbb{R}^d$ is \mathcal{A} -measurable iff each coordinate function f_i , $i = 1, \dots, d$ is \mathcal{A} -measurable. Indeed,

$$f^{-1} \left(\prod_{i=1}^d (a_i, b_i) \right) = \bigcap_{i=1}^d \{x \in X \mid a_i < f_i(x) < b_i\}.$$

Proposition. If f, g are \mathcal{A} -measurable $X \rightarrow \mathbb{R}$ then so are $f+g, fg, df$ $d \in \mathbb{R}$. In particular, the set of \mathcal{A} -measurable functions on X is a vector space stable under multiplication (an \mathbb{R} -algebra).

Proof. Example sheet. □

Remark. Every continuous function $\mathbb{R} \rightarrow \mathbb{R}$ is measurable. Indeed, $\{x \mid f(x) < t\}$ is open hence [Borel](#).

Proposition. If (f_n) is a family of measurable functions $X \rightarrow \mathbb{R}$, then $\limsup_{n \rightarrow \infty} f_n$, $\liminf_{n \rightarrow \infty} f_n$, $\sup f_n$ and $\inf f_n$ are all measurable.

Proof. Example sheet. □

Remark.

- If $X = \bigsqcup_{i=1}^N P_i$ is a partition of X and \mathcal{A} is the [atomic Boolean algebra](#) associated with the partition, then $f : X \rightarrow \mathbb{R}$ is \mathcal{A} -measurable if and only if f is constant on each P_i . Equivalently,

$$f = \sum_{i=1}^N a_i \mathbb{1}_{P_i}, \quad a_i \in \mathbb{R}.$$

The intuition of this is that f being \mathcal{A} -measurable is a property that corresponds to ‘ $f(x)$ depends only on which subsets from \mathcal{A} x belongs to’.

- $E \in \mathcal{A} \iff \mathbb{1}_E$ is \mathcal{A} -measurable for every $E \subseteq X$.

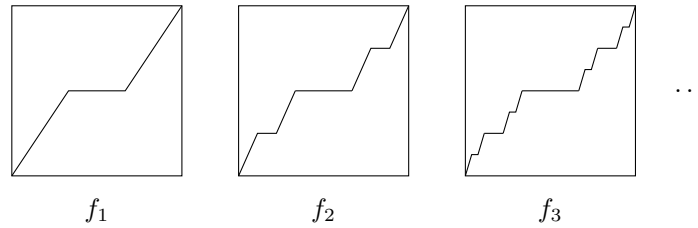
Definition (Borel function). If X is a topological space, a **Borel function** on X is a function which is $\mathcal{B}(X)$ measurable, where $\mathcal{B}(X)$ is the [Borel \$\sigma\$ -algebra](#), the σ -algebra generated by open sets.

Remark.

- Caveat: If $f : X \rightarrow \mathbb{R}$ is \mathcal{A} -measurable then $f^{-1}(B) \in \mathcal{A}$ for all Borel-measurable subsets $B \subseteq \mathbb{R}$, but $f^{-1}(E)$ for E Lebesgue measurable may not be in \mathcal{A} .

- The image of a Borel set under a continuous function may not be Borel, (can find examples in \mathbb{R}^2 under projection).

Example. The **Devil's staircase** is a non-decreasing function $\varphi : [0, 1] \rightarrow [0, 1]$ with $\varphi(0) = 0$, $\varphi(1) = 1$ which is constant on each component of $[0, 1] \setminus C$, where C is the [middle-thirds Cantor set](#).



and we set $\varphi = \lim_{n \rightarrow \infty} f_n$, which converges uniformly.

Exercise.

- (1) Show that there exists a continuous increasing function $\varphi : [0, 1] \rightarrow [0, 1]$, $\varphi(0) = 0$, $\varphi(1) = 1$ and a subset $F \subseteq [0, 1]$ of [Lebesgue measure](#) zero such that $\varphi(F)$ has > 0 measure.
- (2) Show that (1) gives rise to a counterexample explaining the caveat above. Hint: show that $\exists A \subseteq \varphi(F)$ non-Lebesgue measurable, and set $f = \varphi^{-1}$, $E = f(A)$. Hint: Show that every subset of \mathbb{R} of > 0 measure contains a non-Lebesgue measurable subset (modify Vitali's construction).

3 Integration

Take (X, \mathcal{A}, μ) a [measure space](#). We will define the integral

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

of a function with respect to μ . First we define $\mu(f)$ when f is a simple function, i.e. a function of the form

Definition (Simple function). $f : X \rightarrow \mathbb{R}$ is called a **simple function** if f can be written as

$$f = \sum_{i=1}^N a_i \mathbb{1}_{E_i},$$

a linear combination of indicator functions with each $E_i \in \mathcal{A}$.

Definition (Integral of a simple function).

$$\mu(f) := \sum_{i=1}^N a_i \mu(E_i)$$

This definition makes sense - see example sheet 2.

Definition (Integral of a nonnegative function). If $f : X \rightarrow \mathbb{R}$ has values in $[0, +\infty]$ and is [A-measurable](#), we define

$$\mu(f) := \sup \{ \mu(g) \mid g \text{ a simple function, } g \leq f \}$$

where $g \leq f$ simply means $g(x) \leq f(x) \forall x$.

Finally, if $f : X \rightarrow \mathbb{R}$ is any \mathcal{A} -measurable function then we set

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

Then clearly $f = f^+ - f^-$ and $|f| = f^+ + f^-$.

Definition (Integrable function). f is μ -**integrable** if $\mu(f^+) < \infty$ and $\mu(f^-) < \infty$ and if this is the case we define

$$\mu(f) = \mu(f^+) - \mu(f^-).$$

Theorem (Monotone convergence theorem). Let (f_n) be a sequence of [A-measurable](#) functions such that

$$\forall x \in X, \quad 0 \leq f_1(x) \leq f_2(x) \leq \dots \leq f_n(x) \leq \dots$$

Let $f(x) = \lim_{n \rightarrow \infty} f_n(x) \in [0, \infty]$.

Then,

$$\mu(f_n) \xrightarrow{n \rightarrow \infty} \mu(f).$$

Proof. Clearly $\mu(f_n) \leq \mu(f)$ and $\mu(f_n) \leq \mu(f_{n+1})$, $\forall n \in \mathbb{N}$. For any $\epsilon > 0$, let g be a [simple function](#) $g \leq f$ and let

$$E_n = \{x \in X \mid f_n(x) \leq (1 - \epsilon)g(x)\}.$$

Note that $E_n \in \mathcal{A}$ and $E_n \subseteq E_{n+1}$ and $X = \bigcup_{n \geq 1} E_n$. So, we can apply [upwards monotone convergence for sets](#) to the [measure](#) $m_g : \mathcal{A} \rightarrow \mathbb{R}$ given by

$$m_g : E \mapsto \mu(\mathbb{1}_E \cdot g).$$

(Exercise: Check this is a measure on (X, \mathcal{A}) .) This gives $m_g(E_n) \uparrow_{n \rightarrow +\infty} m_g(X) = \mu(g)$.

$$m_g(E_n) = \int_X \mathbb{1}_{E_n}(x)g(x)d\mu(x) \leq \frac{1}{1 - \epsilon} \int f_n(x)d\mu(x) = \frac{\mu(f_n)}{1 - \epsilon}$$

so $\mu(g)(1 - \epsilon) \leq \liminf_n \mu(f_n) \leq \limsup_n \mu(f_n) \leq \mu(f)$. Now this holds for all g simple with $0 \leq g \leq f$, so taking sup over all such g we get

$$\mu(f)(1 - \epsilon) \leq \liminf_n \mu(f_n) \leq \limsup_n \mu(f_n) \leq \mu(f).$$

This holds $\forall \epsilon > 0$, so we get $\mu(f) = \liminf_n \mu(f_n) = \limsup_n \mu(f_n)$. □

Corollary. Take f, g are [μ-integrable](#) functions on (X, \mathcal{A}, μ) . Then

- $\forall \alpha, \beta \in \mathbb{R}$, $\alpha f + \beta g$ is μ -integrable with $\mu(\alpha f + \beta g) = \alpha \mu(f) + \beta \mu(g)$
- if $f \geq 0$, then $\mu(f) \geq 0$
- if $f \geq 0$ and $\mu(f) = 0$, then $f = 0$ a.e.

Remark. The term a.e. means almost everywhere. A property $P(x)$ holds a.e. means that $\{x \in X \mid P(x) \text{ does not hold}\}$ is (sub-)null set for μ .

Proof.

- We know linearity holds for simple functions, so we will reduce to this case using the [Monotone Convergence Theorem](#).

First writing

$$\begin{aligned} f &= f^+ - f^- \\ g &= g^+ - g^- \end{aligned}$$

by setting $f^+ = \max(f, 0)$ and $f^- = (-f)^+$, we can reduce to the case when both f and g are non-negative.

Finally, every non-negative measurable function is a pointwise limit of simple functions. Indeed with

$$\begin{aligned} f_n(x) &:= n \wedge \frac{1}{2^n} [2^n f(x)] \\ \forall n \in \mathbb{N}, \forall x \in X, f_n(x) &\leq f(x) \quad \text{and} \quad \lim_{n \rightarrow \infty} f_n(x) = f(x) \end{aligned}$$

where $[x]$ refers to the integer part of x and \wedge means minimum. So now just apply the Monotone Convergence Theorem to the sequence (f_n) .

- Recall $\mu(f)$ was defined as

$$\sup \{ \mu(g) \mid 0 \leq g \leq f, g \text{ simple} \}$$

So it's clear that $f \geq 0 \implies \mu(f) \geq 0$.

- If $f \geq 0$ and $\mu(f) = 0$, consider

$$E_n = \left\{ x \in X \mid f(x) \geq \frac{1}{n} \right\}.$$

Note $E_n \in \mathcal{A}$ and $\bigcup_{n \geq 1} E_n \cup \{x \in X \mid f(x) = 0\} = X$. By countable subadditivity of μ , it is enough to show that $\forall n, \mu(E_n) = 0$. But

$$\frac{1}{n} \mathbb{1}_{E_n} \leq f, \text{ so } \frac{1}{n} \mu(E_n) \leq \mu(f).$$

As $\mu(f) = 0$, we get $\mu(E_n) = 0$. □

Remark. If f and g are measurable, f integrable and $f = g$ a.e., then g is integrable and $\mu(f) = \mu(g)$.

Theorem (Fatou's Lemma). If $(f_n)_n$ is a sequence of measurable functions on (X, \mathcal{A}, μ) such that $f_n \geq 0 \forall n$. Then

$$\mu(\liminf_{n \rightarrow \infty} f_n) \leq \liminf_{n \rightarrow \infty} \mu(f_n).$$

Remark. This inequality can be strict, for example in the following three types of moving bump.

- (a) $X = [0, \infty)$ with Lebesgue measure, $f_i := \mathbb{1}_{[n, n+1]}$ then $\lim f_n = 0$ but $\mu(f_n) = 1 \forall n$.
- (b) Take $f_n = \frac{1}{n} \mathbb{1}_{[0, n]}$, then again $\lim f_n = 0$ but $\mu(f_n) = 1 \forall n$.
- (c) $X = [0, 1]$ with Lebesgue measure, $f_n = n \mathbb{1}_{[\frac{1}{n}, \frac{2}{n}]}$, again $\lim f_n = 0$ but $\mu(f_n) = 1$.

Proof of Fatou's lemma. Let $g_n = \inf_{k \geq n} f_k$. Note

$$g_{n+1} \geq g_n \geq 0 \quad \forall n \implies \mu(g_n) \xrightarrow{n \rightarrow \infty} \mu(g_\infty)$$

where $g_\infty = \lim_{n \rightarrow \infty} g_n$, and

$$g_n \leq f_n \implies \mu(g_n) \leq \mu(f_n) \implies \lim \mu(g_n) \leq \liminf_n \mu(f_n).$$

Therefore $\mu(g_\infty) \leq \liminf_n \mu(f_n)$. But $g_\infty = \liminf_{n \rightarrow \infty} f_n$. □

Theorem (Lebesgue dominated convergence theorem). Take (X, \mathcal{A}, μ) a measure space. Let $(f_n)_n$ be a sequence of measurable functions on X .

- Assume that there exists a μ -integrable function g such that $|f_n| < g \forall n$.
- Assume further that $f_n(x)$ converges pointwise $\forall x \in X$ to $f(x)$.

Then

$$\mu(\lim_{n \rightarrow \infty} f_n) = \lim_{n \rightarrow \infty} \mu(f_n)$$

and $f = \lim f_n$ is μ -integrable.

Proof of Lebesgue's theorem. Note $|f_n| \leq g \implies |f| \leq g$, where $f = \lim f_n$. So, f is integrable and $\mu(|f|) \leq \mu(g) < \infty$.

Now observe that $g + f_n$ is ≥ 0 . So apply [to](#) $(g + f_n)_n$. So

$$\begin{aligned} \mu(\liminf_{n \rightarrow \infty} g + f_n) &\leq \liminf \mu(g + f_n) \\ \mu(g + f) &\leq \mu(g) + \liminf_n \mu(f_n) \end{aligned}$$

Cancelling $\mu(g)$ we get $\mu(f) \leq \liminf \mu(f_n)$. Now do the same with $-f_n$ in place of f_n , giving

$$\begin{aligned} \mu(-f) &\leq \liminf \mu(-f_n) \\ \mu(f) &\geq \limsup \mu(f_n). \end{aligned}$$

So,

$$\mu(f) \leq \liminf \mu(f_n) \leq \limsup \mu(f_n) \leq \mu(f)$$

and hence $\mu(f_n)$ converges to $\mu(f)$. □

Corollary (Exchange of \sum and \int).

- Let $(f_n)_{n \geq 1}$ be a sequence of measurable functions, and assume $f_n \geq 0 \forall n$. Then

$$\mu\left(\sum_{n \geq 1} f_n\right) = \sum_{n \geq 1} \mu(f_n).$$

- Let $(f_n)_{n \geq 1}$ be any sequence of measurable functions, and assume $\sum_{n \geq 1} |f_n|$ is μ -integrable, then

$$\mu\left(\sum_{n \geq 1} f_n\right) = \sum_{n \geq 1} \mu(f_n).$$

Proof.

- Take $g_N = \sum_{n \geq 1}^N f_n$, the partial sums, and just apply the [Monotone Convergence Theorem](#) to $(g_N)_{N \geq 1}$.
- Let $g = \sum_{n \geq 1} |f_n|$ and define g_N as above. Then $\forall N$, $|g_N| \leq g$, g is μ -integrable by assumption, and simply apply the [Dominated Convergence Theorem](#). □

Corollary (Exchange of $\frac{\partial}{\partial t}$ and \int). Let (X, \mathcal{A}, μ) be a measure space. Let

$$\begin{aligned} f : I \times X &\rightarrow \mathbb{R} \\ (t, x) &\mapsto f(t, x) \end{aligned}$$

where I is an open interval in \mathbb{R} . Assume

- $x \mapsto f(t, x)$ is μ -integrable $\forall t \in I$
- $t \mapsto f(t, x)$ is differentiable $\forall x \in X$
- (domination) There is some $g : X \rightarrow \mathbb{R}$, g μ -integrable such that

$$\left| \frac{\partial f}{\partial t}(t, x) \right| \leq g(x) \quad \forall x \in X \quad \forall t \in I.$$

Then $x \mapsto \frac{\partial f}{\partial t}(t, x)$ is μ -integrable $\forall t \in I$.

$$t \mapsto \int_X f(t, x) d\mu \text{ is differentiable and } \frac{\partial}{\partial t} \left(\int_X f(t, x) d\mu(x) \right) = \int_X \frac{\partial f}{\partial t}(t, x) d\mu(x).$$

Proof. This is a corollary of the [Dominated Convergence Theorem](#). Let $h_n \in \mathbb{R}$, $h_n \rightarrow 0$ as $n \rightarrow \infty$ and set

$$g_n(x) := \frac{f(t + h_n, x) - f(t, x)}{h_n}$$

Then $g_n \rightarrow \frac{\partial f}{\partial t}$ pointwise.

But $\exists \theta_n \in [t, t + h_n]$ such that

$$g_n(x) = \frac{\partial f}{\partial t}(t, x) \Big|_{t=\theta_n}.$$

Then $|g_n| \leq g$ and g is integrable, so just apply the [Dominated Convergence Theorem](#). \square

In the proof of the [Monotone Convergence Theorem](#), we've used the fact that if g is a simple function then $A \mapsto \mu(g\mathbb{1}_A)$ is a measure for $A \in \mathcal{A}$. Now this is clear, reducing to the case $g = \mathbb{1}_B$ by linearity, and it also holds $\forall g \geq 0$ measurable, but this is an exercise which may require the [Monotone Convergence Theorem](#).

Theorem (Egoroff's Theorem). Take (X, \mathcal{A}, μ) a [measure space](#). Assume $\mu(X) < \infty$. If $(f_n)_{n \geq 1}$ a sequence of measurable functions which converges pointwise to $f = \lim f_n$ then

$$\forall \epsilon > 0 \exists A \in \mathcal{A}, \quad \mu(A) > \mu(X) - \epsilon \\ \sup_{x \in A} |f_n(x) - f(x)| \xrightarrow{n \rightarrow \infty} 0.$$

Proof. Replacing f_n by

$$\sup_{k \geq n} |f_k(x) - f(x)|$$

we may assume that $f = 0$ and $0 \leq f_{n+1} \leq f_n$.

Then $f_n(x) \rightarrow 0$ for all $x \in X$, so $\forall k, \exists n_k(x) \in \mathbb{N}$ such that $0 \leq f_{n_k}(x) \leq \frac{1}{k}$. So $n_k(x) < \infty \quad \forall x$. So,

$$\bigcup_{1 \leq T < \infty} \{x \in X \mid n_k(x) \leq T\} = X$$

for any k , and hence

$$\mu(\{x \mid n_k \leq T\}) \xrightarrow{T \rightarrow \infty} \mu(x)$$

by Monotone Convergence for sets.

Let $T_k \in \mathbb{N}$ be such that $\mu(\{x \mid n_k \leq T\}) \leq \mu(X) - \frac{\epsilon}{2^k}$. Take

$$A = \bigcap_{k \geq 1} \{x \mid n_k(x) \leq T_k\},$$

by countable subadditivity of μ , we get $\mu(X \setminus A) \leq \sum_{k \geq 1} \frac{\epsilon}{2^k} \leq \epsilon$. But we have

$$\sup_{x \in A} f_n(x) \xrightarrow{n \rightarrow \infty} 0$$

because $\sup_A f_{T_k} \leq \frac{1}{k} \forall k$ and $n \mapsto \sup_A f_n$ is non-increasing. □

Theorem (Fundamental Theorem of Calculus). Take $f : [a, b] \rightarrow \mathbb{R}$ continuous and let

$$F(x) = \int_{[a, x]} f d\mu$$

(with μ the Lebesgue measure) then f is differentiable and $F'(x) = f(x) \forall x$.

Proof. Same as for the Riemann integral:

$$\frac{F(x+h) - F(x)}{h} = \int_x^{x+h} \frac{f(t)}{h} dt$$

but f is continuous so $\forall \epsilon \exists \delta$ such that $|f(t) - f(x)| < \epsilon$ if $|t - x| < \delta$ so

$$\left| \int_x^{x+h} \frac{f(t)}{h} dt - f(x) \right| \leq \int_x^{x+h} \frac{|f(t) - f(x)|}{h} dt \leq \epsilon$$

if $|h| < \delta$. □

Remark. There is a much more powerful version of the [Fundamental Theorem of Calculus](#), namely:

Theorem. Let $f : [a, b] \rightarrow \mathbb{R}$ be a μ -integrable function and take μ the Lebesgue measure. Let

$$F(x) := \int_{[a, x]} f d\mu.$$

Then F is differentiable [a.e.](#) with respect to μ and $F'(x) = f(x)$ μ -almost everywhere.

Theorem (Lebesgue differentiation theorem). Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a Lebesgue integrable function. Then

$$\frac{1}{\mu(B(x, r))} \int_{B(x, r)} f d\mu \xrightarrow{r \rightarrow 0} f(x) \quad \mu\text{-a.e.}$$

where $B(x, r) = \{y \in \mathbb{R}^d \mid \|x - y\| < r\}$.

Remark.

- It can be the case that $F'(x) = f(x)$ [almost everywhere](#) but not everywhere: take $f = \mathbb{1}_{[0, 1]}$, then $F(x) = \min\{x, 1\}$ for $x \geq 0$.

- The converse fails in general. Namely, if $F : [a, b] \rightarrow \mathbb{R}$ is continuous and differentiable almost everywhere, then it's not always the case that

$$F(x) - F(a) = \int_a^x F'(t) dt. \quad (*)$$

For instance, take $F(x) = x^2 \sin \frac{1}{x^2}$. $F'(x)$ exists $\forall x \in [0, 1]$ but it's not Lebesgue integrable on $[0, 1]$.

Another example is given by the Devil's Staircase: F is a non-decreasing continuous function $[0, 1] \rightarrow [0, 1]$, with $F(0) = 0$ and $F(1) = 1$ but $F'(x) = 0 \forall x \in (0, 1) \setminus C$, so $F'(x)$ exists and is 0 almost everywhere so $(*)$ fails for F .

- If F is assumed to be Lipschitz, then F is **a.e.** differentiable. This is the Rademacher differentiation theorem, and is not proved here. Note that $(*)$ will hold for such F . Indeed, $F'(x) = \lim_{n \rightarrow \infty} F_n(x)$, with $F_n(x) := n(F(x + \frac{1}{n}) - F(x))$. But if F is L -Lipschitz, then $|F_n(x)| \leq L \forall x$ so by the **Lebesgue Dominated Convergence theorem**,

$$\int_a^b F(x) dx = \int_a^b \lim_n F_n(x) dx = \lim \int_a^b F_n(x) = F(b) - F(a)$$

Theorem. If $U, V \subset \mathbb{R}^d$ are open sets and $\varphi : U \rightarrow V$ is a C^1 diffeomorphism and $f : V \rightarrow \mathbb{R}$ is Lebesgue measurable. Assume either $f \geq 0$ or f integrable. Then

$$\int_U f(\varphi(x)) |J_\varphi(x)| dx = \int_V f(y) dy$$

where $J_\varphi(x) = \det(d\varphi(x))$. (both dx and dy are Lebesgue measure).

Remark.

- We skip the proof. It can be proved first for a linear φ , then approximate on small boxes around each point.
- The theorem holds if φ is only assumed to be a Lipschitz homeomorphism.
- When $U = [a, b]$ this is essentially the **Fundamental Theorem of Calculus**.
- This shows that Lebesgue measure is invariant under all affine maps $x \mapsto Ax + b$, $b \in \mathbb{R}^d$, $A \in GL_d(\mathbb{R})$ with $\det A = 1$.

Definition (Product σ -algebra). Let (X, \mathcal{A}) and (Y, \mathcal{B}) be **measurable spaces**. On $X \times Y$ we define the **product σ -algebra**, denoted by $\mathcal{A} \otimes \mathcal{B}$ to be the σ -algebra **generated** by all $A \times B \subset X \times Y$, $A \in \mathcal{A}$, $B \in \mathcal{B}$.

Remark.

- The projection maps

$$\begin{aligned} \pi_X : X \times Y &\longrightarrow X & \text{and} & & \pi_Y : X \times Y &\longrightarrow Y \\ (x, y) &\longmapsto x & & & (x, y) &\longmapsto y \end{aligned}$$

are measurable when $X \times Y$ is endowed with $\mathcal{A} \otimes \mathcal{B}$.

$$A \in \mathcal{A} \quad \pi_X^{-1}(A) = A \times Y$$

- In fact $\mathcal{A} \otimes \mathcal{B}$ is the smallest σ -algebra on $X \times Y$ for which both projections are measurable.
- $\mathcal{B}(\mathbb{R}^{d_1+d_2}) = \mathcal{B}(\mathbb{R}^{d_1}) \otimes \mathcal{B}(\mathbb{R}^{d_2})$, in fact if X and Y are second-countable topological spaces then $\mathcal{B}(X \times Y) = \mathcal{B}(X) \otimes \mathcal{B}(Y)$, where $\mathcal{B}(X)$ is the [Borel \$\sigma\$ -algebra](#). But this is false in general.

Lemma. If $E \subset X \times Y$ belongs to $\mathcal{A} \otimes \mathcal{B}$ then

$$\begin{aligned} \forall x \in X \quad E_x &\in \mathcal{B} \\ \forall y \in Y \quad E^y &\in \mathcal{A} \end{aligned}$$

where

$$\begin{aligned} E_x &:= \{ y \in Y \mid (x, y) \in E \} \\ E^y &:= \{ x \in X \mid (x, y) \in E \} \end{aligned}$$

Proof. Set

$$\mathcal{C} := \{ E \subseteq X \times Y \mid E_x \in \mathcal{B} \text{ and } E^y \in \mathcal{A} \text{ for all } x \in X \text{ and } y \in Y \}.$$

We can see that \mathcal{C} is a σ -algebra, since $(E^c)_x = (E_x)^c$ and similarly for E^y , and

$$\left(\bigcup_n E_n \right)_x = \bigcup_n (E_n)_x.$$

Also it is clear that \mathcal{C} contains all subsets of the form $A \times B$ with $A \in \mathcal{A}$ and $B \in \mathcal{B}$. Hence $\mathcal{C} \supseteq \mathcal{A} \otimes \mathcal{B}$. \square

Using this lemma, we can prove (exercise) that $\mathcal{L}(\mathbb{R}^{d_1}) \otimes \mathcal{L}(\mathbb{R}^{d_2}) \subsetneq \mathcal{L}(\mathbb{R}^{d_1+d_2})$.

Proposition. Let (X, \mathcal{A}, μ) and (Y, \mathcal{B}, ν) be two σ -finite measure spaces. Then there is a unique [measure](#) on $(X \times Y, \mathcal{A} \otimes \mathcal{B})$ denoted by $\mu \otimes \nu$ such that

$$(\mu \otimes \nu)(A \times B) = \mu(A) \cdot \nu(B) \quad (*)$$

for all $A \in \mathcal{A}$, $B \in \mathcal{B}$. This is called the **product measure**.

Sketch of proof. Let \mathcal{B}_0 be the [Boolean algebra generated](#) by the product sets $A \times B$, $A \in \mathcal{A}$, $B \in \mathcal{B}$. Note \mathcal{B}_0 is made of disjoint finite unions of product sets, and $(*)$ defines a measure on \mathcal{B}_0 . One needs to check countable additivity and then we can apply the [Carathéodory extension theorem](#) to claim existence and uniqueness of the extension to $\sigma(\mathcal{B}_0) = \mathcal{A} \otimes \mathcal{B}$. \square

Theorem (Tonelli-Fubini theorem). Take (X, \mathcal{A}, μ) , (Y, \mathcal{B}, ν) be σ -finite measure spaces. Let f be a $\mathcal{A} \otimes \mathcal{B}$ -measurable function on $X \times Y$.

- Assume $f \geq 0$, then the functions

$$\begin{aligned} f_X(y) &:= \int_X f(x, y) d\mu(x) \\ f^Y(x) &:= \int_Y f(x, y) d\nu(y) \end{aligned}$$

and measurable with respect to \mathcal{B} and \mathcal{A} respectively, and

$$\int_{X \times Y} f(x, y) d\mu \otimes \nu(x, y) = \int_Y f_X(y) d\nu(y) = \int_X f^Y(x) d\mu(x). \quad (**)$$

- Now assume f is $\mu \otimes \nu$ -integrable. Then f_X is ν -integrable, f^Y is μ -integrable and $(**)$ holds.

Remark. When $f = \mathbb{1}_E$, $E \in \mathcal{A} \otimes \mathcal{B}$ then $x \mapsto \nu(E_x)$ and $y \mapsto \mu(E^y)$ are measurable and

$$\mu \otimes \nu(E) = \int_X \nu(E_x) d\mu = \int_Y \mu(E^y) d\nu$$

The theorem still holds if f is only assumed measurable with respect to the completion $(\mathcal{A} \otimes \mathcal{B})^*$.

Sketch of proof. Without loss of generality, we can assume that $\mu(X) < \infty$, $\nu(Y) < \infty$ by [σ-finiteness](#).

Write f as $\lim_n f_n$, a pointwise limit of simple functions, for instance

$$f_n(x) := n \wedge \frac{1}{2^n} [2^n f(x)]$$

which lets us reduce to the case for f simple, using the [monotone convergence theorem](#). By linearity we can reduce to the case where $f = \mathbb{1}_E$ for some $E \in \mathcal{A} \otimes \mathcal{B}$ so we're left to show

$$\mu \otimes \nu(E) = \int_X \nu(E_x) d\mu = \int_Y \mu(E^y) d\nu \quad (***)$$

Let $\mathcal{C} = \{E \in \mathcal{A} \otimes \mathcal{B} \mid \text{such that } (***) \text{ holds}\}$.

- $\forall A \in \mathcal{A}, \forall B \in \mathcal{B}$ we have $A \times B \in \mathcal{C}$, by definition of $\mu \otimes \nu$.
- \mathcal{C} is stable under complement (because μ, ν are finite, $(E_x)^c = (E^c)_x$).
- \mathcal{C} is stable under countable disjoint unions: $(\bigcup E_n)_x = \bigcup (E_n)_x$
- \mathcal{C} contains $\{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}$ which is stable under intersections.

So we can apply [Dynkin's lemma](#) to \mathcal{C} and conclude $\mathcal{C} = \mathcal{A} \otimes \mathcal{B}$. If f is only assumed integrable, write $f = f^+ - f^-$, apply the above to f^+ and f^- . \square

Definition (Radon measure). If X is a topological space and μ is a Borel measure on X , then μ is called **Radon** if $\mu(K) < \infty$ for any K a compact subset of X .

Let $C_c(X)$ be the space of compactly supported functions on X . Clearly every $f \in C_c(X)$ is μ -integrable when μ is Radon. Indeed $\mu(|f|) \leq \sup_X |f| \cdot \mu(\text{supp } f) < \infty$, where $\text{supp } f = \overline{\{x \in X \mid f(x) \neq 0\}}$. So, the map $T_\mu : C_c(X) \rightarrow \mathbb{R}$ is well defined and linear, defined by $f \mapsto \mu(f)$. Then T_μ is a linear functional such that $f \geq 0 \implies T_\mu(f) \leq 0$, and we say T_μ is positive.

Theorem (Riesz Representation Theorem). If X is locally compact and second countable (i.e. has a countable basis of open sets), then every positive functional on $C_c(X)$ is of the form T_μ for some unique Radon measure μ on X .

4 Probability Theory

We can apply our knowledge of measures to probability, by using a space Ω , \mathcal{F} a set of observable/measurable events which forms a σ -algebra and \mathbb{P} a **measure**. We also have a continuity axiom: for $A_n \in \mathcal{F}$, $A_{n+1} \subseteq A_n$ such that $\bigcap_n A_n = \emptyset$ then $\mathbb{P}(A_n) \xrightarrow{n \rightarrow \infty} 0$.

Formally,

Definition (Probability measure). Let (X, \mathcal{A}) be a **measurable space**. A **measure** μ on X with $\mu(X) = 1$ is called a **probability measure**.

Definition (Random variable). A **measurable function** $f : X \rightarrow \mathbb{R}$ is called a **random variable**.

Definition (Probability space). (X, \mathcal{A}, μ) is called a **probability space** when $\mu(X) = 1$.

Notation. In the context of probability, we use the following interchangeably:

$$X \longleftrightarrow \Omega, \quad \mathcal{A} \longleftrightarrow \mathcal{F}, \quad \mu \longleftrightarrow \mathbb{P}.$$

Definition (Distribution). A **random variable** X on $(\Omega, \mathcal{F}, \mathbb{P})$ determines a **measure** μ_X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ given by

$$\mu_X(A) = \mathbb{P}(X \in A)$$

for any $A \in \mathcal{B}(\mathbb{R})$, where

$$\mathbb{P}(X \in A) := \mathbb{P}(\{\omega \in \Omega \mid X(\omega) \in A\}).$$

This measure μ_X is called the **distribution** or the **law** of X .

Notation. The integral sign $\int d\mu$ is replaced by the expectation sign

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P}$$

and if $A \in \mathcal{F}$ then $\mathbb{E}(\mathbb{1}_A) = \mathbb{P}(A)$.

Example. Toss a coin N times, and let X be the number of heads. Then $\mathbb{P}(X = k) = \binom{N}{k} \frac{1}{2^N}$, a binomial distribution.

Note μ_X is the image of \mathbb{P} under $X : \Omega \rightarrow \mathbb{R}$.

Notation. $X \sim \mu_X$ means X is distributed according to μ_X .

Definition (Cumulative probability function). The function $F_X = \mathbb{P}(X \leq x) = \mu_X(-\infty, x]$ is the **cumulative probability function**.

Remark. F_X is non-decreasing right-continuous, determines μ_X , and

$$\begin{aligned} \lim_{x \rightarrow -\infty} F_X &= 0 \\ \lim_{x \rightarrow +\infty} F_X &= 1. \end{aligned} \tag{*}$$

Theorem. Conversely, given a non-decreasing function $F : \mathbb{R} \rightarrow [0, 1]$ which is right-continuous and satisfies (*), then there exists a unique probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $F(x) = \mu((-\infty, x])$ for all $x \in \mathbb{R}$.

Sketch proof.

- Define μ on intervals $(a, b]$ as $\mu((a, b]) = F(b) - F(a)$.
- Extend it to finite unions of intervals.
- Use [Carathéodory Extension Theorem](#). □

Definition (Lebesgue-Stieljes measure). Given F as in the theorem, the μ constructed is called the **Lebesgue-Stieljes** measure associated to F .

Definition (Density). If $F_X(x)$ has the form

$$\int_{-\infty}^x f(y) dy$$

with the Lebesgue measure, and $f \geq 0$ is Lebesgue measurable, then we say that F_X (or X , or μ_X) has a **density** with respect to Lebesgue measure, and f is called the **density function**.

4.1 Independence

Throughout this section, take $(\Omega, \mathcal{F}, \mathbb{P})$ a [probability space](#).

Definition (Independent events). A sequence of events $(A_n)_{n \geq 1}$ is said to be **(mutually) independent** if $\forall I \subseteq \mathbb{N}$, I a finite set,

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i).$$

Definition (Independent algebras). A sequence of sub- σ -algebras $(\mathcal{A}_n)_{n \geq 1}$ of \mathcal{F} is called **independent** if $\forall A_i \in \mathcal{A}_i \forall i \geq 1$, the family $(A_i)_{i \geq 1}$ is [independent](#).

Definition (Independent random variables). A sequence $(X_n)_{n \geq 1}$ of [random variables](#) is called **(mutually) independent** if the sequence $(\sigma(X_n))_{n \geq 1}$ is independent.

Notation. For X a [random variable](#),

$$\sigma(X) := X^{-1}(B(\mathbb{R})) = \sigma(\{\omega \in \Omega \mid X(\omega) \leq x\}_{x \in \mathbb{R}}),$$

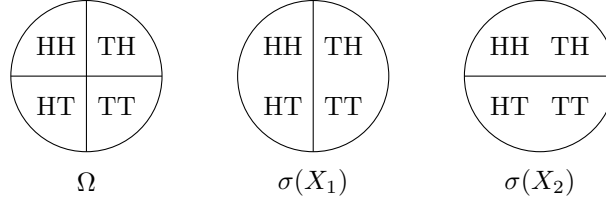
the smallest σ -algebra making X measurable.

Example.

- If X takes only finitely many values x_1, \dots, x_n , then $\sigma(X)$ is a **finite Boolean algebra**, whose atoms are $\{\omega \mid X(\omega) = x_i\}_{i=1, \dots, k}$.
- Use $\Omega = \{\text{Heads}, \text{Tails}\}^2$, representing two independent coin tosses X_1, X_2 .

$$\mathbb{P}(X_1 = \epsilon_1, X_2 = \epsilon_2) = \frac{1}{4} = \mathbb{P}(X_1 = \epsilon_1)\mathbb{P}(X_2 = \epsilon_2) \quad \forall \epsilon_1, \epsilon_2 \in \{H, T\}.$$

\mathcal{F} = discrete Boolean algebra on Ω , and $\mathcal{F} = 2^\Omega$.



$$\begin{aligned}\sigma(X_1) &= \{\{H\} \times \{H, T\}, \{T\} \times \{H, T\}, \emptyset, \Omega\} \\ \sigma(X_2) &= \{\{H, T\} \times \{H\}, \{H, T\} \times \{T\}, \emptyset, \Omega\}\end{aligned}$$

Lemma. Let $\mathcal{F}_i \subseteq \mathcal{A}_i$ be a subfamily of events which is stable under finite intersection, such that $\sigma(\mathcal{F}_i) = \mathcal{A}_i$. Then to check that $(\mathcal{A}_i)_{i \geq 1}$ forms an independent family it is enough to check that

$$\mathbb{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$$

holds $\forall I \subseteq \mathbb{N}$ finite and $\forall A_i \in \mathcal{F}_i$.

Example.

(a) Let (X_1, \dots, X_n) be **random variables**. They are **independent** \iff

$$\forall x_1, \dots, x_n \in \mathbb{R} \quad \mathbb{P}(\forall i = 1, \dots, n \ X_i \leq x_i) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i).$$

Indeed this is a special case of the lemma setting

$$\mathcal{F}_i = \{X_i^{-1}((-\infty, x]) \mid x \in \mathbb{R}\}$$

(b) If X_1, \dots, X_n take only a finite set E of values. (X_1, \dots, X_n) independent \iff

$$\forall e_1, \dots, e_n \in E \quad \mathbb{P}(\forall i \ X_i = e_i) = \prod_{i=1}^n \mathbb{P}(X_i = e_i).$$

Proof of lemma. For two σ -subalgebras \mathcal{A}_1 and \mathcal{A}_2 , let $A_2 \in \mathcal{A}_2$ be such that

$$\mathbb{P}(A \cap A_2) = \mathbb{P}(A)\mathbb{P}(A_2) \tag{*}$$

holds for all $A \in \mathcal{F}_1$. Look at the two measures

$$\begin{aligned}A &\longmapsto \mathbb{P}(A \cap A_2) \\ A &\longmapsto \mathbb{P}(A)\mathbb{P}(A_2).\end{aligned}$$

The two measures coincide on \mathcal{F}_1 , hence they coincide on \mathcal{A}_1 (by **Dynkin's lemma**). So (*) holds $\forall A \in \mathcal{A}_1$. Now just reverse the roles of \mathcal{A}_1 and \mathcal{A}_2 . \square

Example. $\Omega = [0, 1]$, $\mathbb{P} = \text{Lebesgue measure}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$ the **Borel σ -algebra**. Let $X_n(\omega)$ be the n th digit of the decimal expansion of $\omega \in [0, 1]$.

Claim $(X_n)_{n \geq 1}$ is **independent**. We have $X_n(\omega) \in \{0, 1, \dots, 9\}$. Let

$$\begin{aligned} I_{\epsilon_1, \dots, \epsilon_n} &= \{ \omega \in \Omega \mid X_1(\omega) = \epsilon_1, \dots, X_n(\omega) = \epsilon_n \} \\ &= \{ \omega \in [0, 1] \mid \lfloor 10^n \omega \rfloor = \epsilon_1 \cdots \epsilon_n \}. \end{aligned}$$

We need to check that

$$\begin{aligned} \mathbb{P}(X_1(\omega) = \epsilon_1, \dots, X_n(\omega) = \epsilon_n) &= \prod_{i=1}^n \mathbb{P}(X_i(\omega) = \epsilon_i) \\ &= \prod_{i=1}^n \frac{1}{10} = \frac{1}{10^n} \end{aligned}$$

but it is clear $\mathbb{P}(\omega \in I_{\epsilon_1, \dots, \epsilon_n}) = \frac{1}{10^n}$.

Remark.

- If f_1, \dots, f_n are **Borel measurable functions** $\mathbb{R} \rightarrow \mathbb{R}$ and X_1, \dots, X_n are **independent random variables**, then so are $f_1(X_1), \dots, f_n(X_n)$. So $\sigma(f(X)) \subseteq \sigma(X)$.
- **Independence** implies pairwise independence, but not conversely. Consider Bernstein's example: Take X, Y two independent coin tosses. $X = 1$ for heads, $X = 0$ for tails, same for Y . Set $Z = |X - Y|$, then we have pairwise independence but not independence.

Proposition. Let X_1, \dots, X_n be n **random variables**.

$$(X_1, \dots, X_n) \text{ are } \textbf{mutually independent} \iff \mu_{(X_1, \dots, X_n)} = \mu_{X_1} \otimes \cdots \otimes \mu_{X_n}.$$

Here $\mu_{(X_1, \dots, X_n)}$ is the **law or distribution** of (X_1, \dots, X_n) , and we say it's the joint law or joint distribution.

Proof.

$$\begin{aligned} \mu_{(X_1, \dots, X_n)} \left(\prod_{i=1}^n (-\infty, x_i] \right) &= \mathbb{P} \left((X_1, \dots, X_n) \in \prod_{i=1}^n (-\infty, x_i] \right) \\ &= \mathbb{P}(\forall i, X_i \leq x_i). \end{aligned}$$

$$\begin{aligned} \text{independence} &\iff \forall x_i \in \mathbb{R}, \quad \mathbb{P}(\forall i, X_i \leq x) = \prod_{i=1}^n \mathbb{P}(X_i \leq x_i) \\ &\iff \mu_{(X_1, \dots, X_n)} \left(\prod_{i=1}^n (-\infty, x_i] \right) = \prod_{i=1}^n \mu_{x_i}((-\infty, x_i]) \quad \square \end{aligned}$$

Proposition. Suppose X and Y are two **independent random variables**. Assume X, Y **integrable**, i.e. $\mathbb{E}(|X|) < \infty$, $\mathbb{E}(|Y|) < \infty$. Then XY is integrable and

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y). \quad (*)$$

Moreover $(*)$ holds also if X and Y are just assumed ≥ 0 .

Proof. Recall that X, Y are independent \iff the [joint law](#) $\mu_{(X,Y)}$ is the product $\mu_{(X,Y)} = \mu_X \otimes \mu_Y$ where $\mu_X =$ [law](#) of X and $\mu_Y =$ law of Y . So we can apply [Fubini](#): $\mathbb{E}(X) = \int_{\mathbb{R}} x d\mu_X$, $\mathbb{E}(Y) = \int_{\mathbb{R}} y d\mu_Y$.

$$\mathbb{E}(XY) = \int_{\mathbb{R} \times \mathbb{R}} xy d\mu_{(X,Y)}. \quad \square$$

Lemma (1st Borel-Cantelli lemma). Take $(\Omega, \mathcal{F}, \mathbb{P})$ a [probability space](#). Let $(A_n)_{n \geq 1}$ be a sequence of events. Assume $\sum_{n \geq 1} \mathbb{P}(A_n) < \infty$. Then $\mathbb{P}(\limsup A_n) = 0$.

Proof.

$$\begin{aligned} \limsup A_n &:= \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n \\ \mathbb{P}(\limsup A_n) &\leq \mathbb{P}\left(\bigcup_{n \geq m} A_n\right) \\ &\leq \sum_{n \geq m} \mathbb{P}(A_n) \rightarrow 0 \text{ as } m \rightarrow \infty. \end{aligned} \quad \square$$

Example. Take $\Omega = [0, 1]$. $\mathbb{P} =$ [Lebesgue](#), $\mathcal{F} =$ [Borel \$\sigma\$ -algebra](#). A number $\alpha \in \mathbb{R}$ is said to be very-well-approximated if $\exists \epsilon > 0$ such that there are infinitely many integers p, q such that

$$\left| \alpha - \frac{p}{q} \right| \leq \frac{1}{q^{2+\epsilon}}.$$

Claim: Lebesgue [a.e.](#) numbers α are not very-well-approximated. Proof: (exercise) Apply [Borel-Cantelli](#) to the events

$$A_q = \left\{ x \in [0, 1] \mid d(qx, \mathbb{Z}) < \frac{1}{q^{1+\epsilon}} \right\}.$$

Lemma (2nd Borel-Cantelli lemma). Conversely to the [first Borel-Cantelli lemma](#), assume that the events $(A_n)_{n \geq 1}$ are [independent](#) and $\sum_{n \geq 1} \mathbb{P}(A_n) = +\infty$. Then $\mathbb{P}(\limsup A_n) = 1$.

Remark. The conclusion fails without the assumption of independence, for instance take $A_n = [0, \frac{1}{n}] \subseteq \Omega = [0, 1]$, for which $\sum \mathbb{P}(A_n) = \sum \frac{1}{n} = \infty$, but $\limsup A_n = \{0\}$.

Proof.

$$\begin{aligned} \limsup A_n &= \bigcap_{m \geq 1} \bigcup_{n \geq m} A_m \\ (\limsup A_n)^c &= \bigcup_{m \geq 1} \bigcap_{n \geq m} (A_m)^c \end{aligned}$$

but $\forall m$,

$$\begin{aligned} \mathbb{P}\left(\bigcup_{n \geq m} A_n^c\right) &= \prod_{n \geq m} \mathbb{P}(A_n^c) = \prod_{n \geq m} (1 - \mathbb{P}(A_n)) \\ &\leq \prod_{n \geq m} e^{-\mathbb{P}(A_n)} = e^{-\sum_{n \geq m} \mathbb{P}(A_n)} \\ &\rightarrow e^{-\infty} = 0. \end{aligned}$$

□

Definition (Random process). A **random process** is an infinite family of $(X_n)_{n \geq 1}$ of real **random variables**, where n can be thought of as a time parameter. We use $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ = the smallest sub- σ -algebra of \mathcal{F} which makes all X_i , $i \leq n$ measurable. We also have $\mathcal{F}_{n+1} \supseteq \mathcal{F}_n$, the time filtration property.

Definition (Tail σ -algebra). The **tail σ -algebra** of $(X_n)_{n \geq 1}$ is defined as

$$\tau = \bigcap_{n \geq 1} \sigma(X_n, X_{n+1}, \dots).$$

Example. The following events belong to τ :

$$\{\limsup X_n \geq T\} \quad \{\omega \in \Omega \mid (X_n(\omega))_{n \geq 1} \text{ converges}\}.$$

Theorem (Kolmogorov 0-1 Law). Suppose $(X_n)_{n \geq 1}$ is a sequence of **independent random variables**. Then the **tail σ -algebra** τ is trivial ($\forall A \in \tau, \mathbb{P}(A) \in \{0, 1\}$).

Proof. Let $A \in \tau$ and let $B \in \sigma(X_1, \dots, X_n)$. Then A and B are independent, because $\sigma(X_1, \dots, X_n)$ and τ are independent. So $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$.

Now the measures

$$\begin{aligned} B &\mapsto \mathbb{P}(A \cap B) \\ B &\mapsto \mathbb{P}(A)\mathbb{P}(B) \end{aligned}$$

coincide on $\sigma(X_1, \dots, X_n)$ for all n , hence they coincide on $\sigma(X_1, \dots, X_n, \dots)$. But $\tau \subseteq \sigma(X_1, \dots, X_n, \dots)$. In particular $\mathbb{P}(A \cap A) = \mathbb{P}(A) \cdot \mathbb{P}(A)$ so $\mathbb{P}(A) = \mathbb{P}(A)^2$ hence $\mathbb{P}(A) \in \{0, 1\}$. □

Example.

- Take an i.i.d. space $(X_n)_{n \geq 1}$ of **random variables** such that $\forall T \geq 0, \mathbb{P}(X_1 < T) < 1$. Then almost surely $\limsup X_n = +\infty$.

Indeed,

$$\sum_{k=1}^n \mathbb{P}(X_k \geq T) = n\mathbb{P}(X_1 \geq T) = n(1 - \mathbb{P}(X_1 < T)) \rightarrow \infty$$

so the **second Borel-Cantelli lemma** applies and $\mathbb{P}(\limsup A_n) = 1$.

- Take $(\epsilon_n)_{n \geq 1}$ be i.i.d. random variables with $\mathbb{P}(\epsilon_1 = 1) = \mathbb{P}(\epsilon_1 = -1) = \frac{1}{2}$. Take $(a_n)_{n \geq 1}$ some sequence of real numbers, and ask when does $\sum_{n \geq 1} \epsilon_n a_n$ converge?

The **Kolmogorov 0-1 law** tells us that this happens with probability 0 or 1 depending only on the sequence $(a_n)_{n \geq 1}$. In fact, a theorem of Rademacher-Paley-Zygmund says

$$\sum_{n=1}^{\infty} \epsilon_n a_n \text{ converges a.s.} \iff \sum_{n=1}^{\infty} a_n^2 < \infty.$$

Take μ a **probability measure** on \mathbb{R} . Can you find a **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$ and a **random variable** X such that μ is the law of X ?

Yes: Take $\Omega = \mathbb{R}$, let \mathcal{F} be the Borel sets, and set $\mathbb{P} = \mu$, and $X(\omega) = \omega$.

Definition (Moment). If X is a [random variable](#), the k th **moment** of X is $\mathbb{E}(X^k)$ for $k \in \mathbb{N}$.

Some additional basic terminology: For X a [random variable](#), $\mathbb{E}(X)$ is called the mean, and $\mathbb{E}((X - \mathbb{E}X)^2)$ is the variance of X , denoted $\text{Var } X$.

Remark.

- $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}X)^2$
- If X, Y are independent then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

Theorem (Strong law of large numbers (under a [4th moment](#) assumption)). Let $(X_n)_{n \geq 1}$ be a sequence of [independent](#) and identically distributed (i.i.d.) [random variables](#) with common [law](#) μ . Assume $\int_{\mathbb{R}} x^4 d\mu(x) < \infty$ (i.e. $\mathbb{E}(|X_1|^4) < \infty$). Then

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{n \rightarrow \infty} \mathbb{E}(X_1)$$

almost surely.

Example. Take $\Omega = [0, 1]$ and $\mathbb{P} = \text{Lebesgue}$. Write $\omega = 0.\epsilon_1\epsilon_2\dots\epsilon_n\dots$ with $\epsilon_i \in \{0, \dots, 9\}$, the decimal expansion of ω . We observed that $(\epsilon_n(\omega))_{n \geq 1}$ were [independent](#) and uniformly distributed in $\{0, \dots, 9\}$. Set $X_n^{(i)}(\omega) = \mathbb{1}_{\epsilon_n(\omega)=i}$, these are i.i.d. [random variables](#) and $|X_n^{(i)}| \leq 1$, so finite fourth moment. So the [SLL](#) applies:

$$\frac{1}{n} \sum_{k=1}^n X_k^{(i)} = \frac{1}{n} \#\{k \leq n, \epsilon_k(\omega) = i\} \implies \mathbb{E}(X_1^{(i)}) = \frac{1}{10}.$$

So, for Lebesgue [almost every](#) ω the proportion of each digit is the same (a normal number).

Theorem (Borel). [Almost every](#) number is a normal number.

Lemma (Cauchy-Schwarz inequality). Take f, g square integrable functions on (X, \mathcal{A}, μ) then fg is integrable and

$$\left| \int fg d\mu \right| \leq \sqrt{\int f^2 d\mu} \sqrt{\int g^2 d\mu}.$$

Equivalently, take X, Y random variables such that $\mathbb{E}(X^2) < \infty$ and $\mathbb{E}(Y^2) < \infty$, then $\mathbb{E}(|XY|) \leq \sqrt{\mathbb{E}X^2 \mathbb{E}Y^2}$.

Proof. Look at $t \mapsto \mathbb{E}((X + tY)^2) = \mathbb{E}X^2 + t^2\mathbb{E}Y^2 + 2t\mathbb{E}(XY) \geq 0 \ \forall t \in \mathbb{R}$. So, $\Delta = 4(\mathbb{E}(XY))^2 - 4\mathbb{E}X^2\mathbb{E}Y^2 \leq 0$. \square

Proof of the strong law of large numbers. Without loss of generality, assume that $\mathbb{E}(X_1) = 0$, because we can replace X_n by $X_n - \mathbb{E}(X_1)$. This is allowed because $\mathbb{E}((X_n - \mathbb{E}X_1)^4) < \infty$. Indeed,

$$\begin{aligned} (X_n - \mathbb{E}X_1)^4 &= [(X_n - \mathbb{E}X_1)^2]^2 \\ &\leq (2X_n^2 + 2(\mathbb{E}X_1)^2)^2 \\ &\leq 8[X_n^4 + (\mathbb{E}X_1)^4]. \end{aligned}$$

So we have $\mathbb{E}(X_n) = 0 \forall n$.

Let $S_n = \frac{1}{n}(X_1 + \dots + X_n)$, so $S_n^4 = \frac{1}{n^4}(\sum_{i,j,k,l} X_i X_j X_k X_l)$. By independence, $\mathbb{E}(X_i X_j X_k X_l) = 0$ if

- i, j, k, l are distinct, or
- $k = l$ and $i \neq j \notin \{k, l\}$ or
- $k = l = j \neq i$.

So,

$$\begin{aligned} \mathbb{E}(S_n^4) &= \frac{1}{n^4} \left(\mathbb{E}X_1^4 + \dots + \mathbb{E}X_n^4 + 6 \sum_{i < j} \mathbb{E}X_i^2 \cdot \mathbb{E}X_j^2 \right) \\ &= \frac{1}{n^4} \left(n\mathbb{E}X_i^4 + 6 \frac{n(n-1)}{2} (\mathbb{E}(X_1^2))^2 \right) \\ &= \mathcal{O}\left(\frac{1}{n^2}\right). \end{aligned}$$

So in conclusion,

$$\sum_{n \geq 1} \mathbb{E}(S_n^4) = \mathbb{E} \left(\sum_{n \geq 1} S_n^4 \right) < \infty$$

So $\sum_{n \geq 1} S_n^4 < \infty$ a.s. $\implies \lim_{n \rightarrow \infty} S_n = 0$ almost surely. □

5 Type of convergence of random variables

Notation. $C_b(\mathbb{R}^d)$ is the space of continuous and bounded real functionals on \mathbb{R}^d .

Definition (Weak convergence). A sequence of [probability measures](#) on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, say $(\mu_n)_{n \geq 1}$ is said to **converge weakly** to a measure μ on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ if

$$\forall f \in C_b(\mathbb{R}^d) \quad \mu_n(f) \xrightarrow{n \rightarrow \infty} \mu(f).$$

Example.

- $\mu_n = \delta_{x_n}$ a Dirac mass,

$$\delta_x(A) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \quad \text{for } x \in \mathbb{R}^d$$

with $x_n \in \mathbb{R}^d$ with $x_n \rightarrow x \in \mathbb{R}^d$ then clearly $\mu_n \rightarrow \delta_x$ weakly.

- $\mu_n = \mathcal{N}(0, \sigma_x^2)$, the centered Gaussian with standard deviation σ_x . If $\sigma_x \rightarrow 0$ then $\mu_n \rightarrow \delta_0$ weakly.

Definition (Convergence). A sequence of \mathbb{R}^d -valued [random variables](#) $(X_n)_{n \geq 1}$ is said to converge to X

- (a) **almost surely** if for \mathbb{P} a.e. ω , $X_n(\omega) \xrightarrow{n \rightarrow \infty} X(\omega)$
- (b) **in probability** (or in measure) if $\forall \epsilon > 0$, $\mathbb{P}(\|X_n - X\| > \epsilon) \xrightarrow{n \rightarrow \infty} 0$.
- (c) **in distribution** if $\mu_{X_n} \xrightarrow{n \rightarrow \infty} \mu_X$ weakly, with μ_{X_n} the [law](#) of X_n on \mathbb{R}^d .

Remark. (a) \Rightarrow (b) \Rightarrow (c), but none of the reverse implications hold in general.

(a) \Rightarrow (b): $\mathbb{P}(\|X_n - X\| > \epsilon) = \mathbb{E}(\mathbb{1}_{\|X_n - X\| > \epsilon}) \xrightarrow{n \rightarrow \infty} 0$ by [Lebesgue Dominated Convergence](#).

(b) \Rightarrow (c): Let $f \in C_b(\mathbb{R}^d)$. f is uniformly continuous on compact subsets of \mathbb{R}^d . So $\forall \epsilon > 0$, $\exists \delta > 0$ such that if $\|x\| < 1/\epsilon$ and $\|x - y\| < \delta$, then $|f(x) - f(y)| < \epsilon$. So

$$\begin{aligned} |\mu_n(f) - \mu(f)| &\leq \mathbb{E}(|f(X_n) - f(X)|) \\ &\leq \mathbb{E}\left(\mathbb{1}_{\|X_n - X\| < \delta} \mathbb{1}_{\|X\| < \frac{1}{\epsilon}} |f(X_n) - f(X)|\right) \\ &\quad + 2\|f\|_\infty \mathbb{P}\left(\|X_n - X\| > \delta \text{ or } \|X\| > \frac{1}{\epsilon}\right) \\ &\leq \epsilon + 2\|f\|_\infty [\mathbb{P}(\|X_n - X\| > \delta) + \mathbb{P}(\|X\| > 1/\epsilon)] \end{aligned}$$

So

$$\limsup_n |\mu_n(f) - \mu(f)| \leq \epsilon + 2\|f\|_\infty \mathbb{P}(\|X\| > 1/\epsilon) \xrightarrow{\epsilon \rightarrow 0} 0.$$

When $d = 1$, (c) is equivalent to saying that the [distribution function](#) $F_{\mu_n}(X) \rightarrow F_\mu(x)$ for all $x \in \mathbb{R}$ where $F_\mu(x)$ is continuous. Recall

$$\begin{aligned} F_\mu(x) &:= \mu((-\infty, x]) \\ F_{\mu_n}(x) &:= \mu_n((-\infty, x]). \end{aligned}$$

For example, take $\mu_n = \delta_{1/n} \rightarrow \delta_0 = \mu$, so $F_{\mu_n} = \mathbb{1}_{x \geq \frac{1}{n}}$, but $F_\mu = \mathbb{1}_{x \geq 0}$. Then $F_{\mu_n}(0) = 0$ while $F_\mu(0) = 1$, so $F_{\mu_n}(0) \not\rightarrow F_\mu(0)$.

Proposition. If $(X_n)_{n \geq 1}$ converges in probability to X , then \exists a subsequence $(X_{n_k})_{k \geq 1}$ such that $X_{n_k} \xrightarrow[k \rightarrow \infty]{} X$ almost surely.

Proof. For all k we have

$$\begin{aligned} & \mathbb{P} \left(\|X_n - X\| > \frac{1}{k} \right) \xrightarrow{n \rightarrow \infty} 0 \\ \implies & \exists n_k \text{ such that } \mathbb{P} \left(\|X_{n_k} - X\| > \frac{1}{k} \right) \leq \frac{1}{2^k}. \end{aligned}$$

So

$$\sum \mathbb{P} \left(\|X_{n_k} - X\| > \frac{1}{k} \right) < \infty \implies \mathbb{P} \left(\limsup \|X_{n_k} - X\| > \frac{1}{k} \right) = 0.$$

by Borel-Cantelli. □

Definition (L^1 convergence). A sequence of \mathbb{R}^d -valued integrable random variables $(X_n)_{n \geq 1}$ converges in L^1 (or in mean) towards an integrable random variable X if

$$\mathbb{E}(\|X_n - X\|) \xrightarrow{n \rightarrow \infty} 0$$

which we'll write $X_n \xrightarrow{L^1} X$.

Remark.

- if $X_n \xrightarrow{L^1} X$ then $X_n \rightarrow X$ in probability. Indeed, $\forall \epsilon > 0$,

$$\epsilon \mathbb{P}(\|X_n - X\| > \epsilon) \leq \mathbb{E}(\|X_n - X\| \mathbb{1}_{\|X_n - X\| > \epsilon}) \leq \mathbb{E}(\|X_n - X\|).$$

- This demonstrates the Markov inequality,

$$\epsilon \mathbb{P}(|X| \geq \epsilon) \leq \mathbb{E}(|X|)$$

holds for any random variable X and any $\epsilon > 0$.

- The Markov inequality gives also the Chebyshev inequality,

$$\epsilon^2 \mathbb{P}(|X - \mathbb{E}X| > \epsilon) \leq \text{Var } X$$

holds for any random variable X and any $\epsilon > 0$.

In general, the converse does not hold, i.e. convergence in probability $\not\Rightarrow$ in L^1 .

Example. For example, take $\Omega = [0, 1]$, $\mathbb{P} = \text{Lebesgue}$ and $X_n = n \mathbb{1}_{[0, \frac{1}{n}]}$. Then $X_n \rightarrow 0$ in probability (since $\mathbb{P}(|X_n| > \epsilon) = \frac{1}{n}$) but $\mathbb{E}(X_n) = 1$ for any n .

If $\exists M \geq 0$ such that $\|X_n\| \leq M$ almost surely for all n , then the converse holds.

$$\begin{aligned} \mathbb{E}(\|X_n - X\|) &= \mathbb{E}(\|X_n - X\| \mathbb{1}_{\|X_n - X\| \leq \epsilon}) + \mathbb{E}(\|X_n - X\| \mathbb{1}_{\|X_n - X\| > \epsilon}) \\ &\leq \epsilon + 2M \mathbb{P}(\|X_n - X\| > \epsilon). \end{aligned}$$

so $\limsup \mathbb{E}(\|X_n - X\|) \leq \epsilon$.

Definition (Uniform integrability). A sequence of [integrable random variables](#) $(X_n)_{n \geq 1}$ is called **uniformly integrable** if

$$\sup_{n \geq 1} \mathbb{E}(\|X_n\| \mathbb{1}_{\|X_n\| > M}) \xrightarrow{M \rightarrow \infty} 0.$$

Remark.

- If $X_n = X$ for all n , and X is [integrable](#), then $(X_n)_n$ is [uniformly integrable](#):

$$\mathbb{E}(\|X\| \mathbb{1}_{\|X\| \leq M}) \uparrow \mathbb{E}(\|X\|)$$

as $M \rightarrow \infty$ by [monotone convergence](#), so

$$\mathbb{E}(\|X\| \mathbb{1}_{\|X\| > M}) \xrightarrow{M \rightarrow \infty} 0.$$

- Suppose the X_n s are dominated, i.e. there is some Y an integrable [random variable](#) such that $\|X_n\| \leq Y$ for all n . Then $(X_n)_n$ are [uniformly integrable](#):

$$\mathbb{E}(\|X_n\| \mathbb{1}_{\|X_n\| > M}) \leq \mathbb{E}(Y \mathbb{1}_{Y > M}) \xrightarrow{M \rightarrow \infty} 0.$$

- If $X_n \xrightarrow{L^1} X$, then $(X_n)_n$ is [uniformly integrable](#):

$$\begin{aligned} \mathbb{E}(\|X_n\| \mathbb{1}_{\|X_n\| > M}) &\leq \mathbb{E}(\|X_n - X\| \mathbb{1}_{\|X_n\| > M}) + \mathbb{E}(\|X\| \mathbb{1}_{\|X_n\| > M}) \\ &\leq \mathbb{E}(\|X_n - X\|) + \mathbb{E}(\|X\| \mathbb{1}_{\|X\| > M} \mathbb{1}_{\|X_n\| > M}) \\ &\quad + \mathbb{E}(\|X\| \mathbb{1}_{\|X\| < M} \mathbb{1}_{\|X_n\| > M}) \\ &\leq \mathbb{E}(\|X_n - X\|) + \mathbb{E}(\|X\| \mathbb{1}_{\|X\| > M}) + M \mathbb{E}(\mathbb{1}_{\|X_n\| \geq M}) \\ &\leq \mathbb{E}(\|X_n - X\|) + \epsilon M. \end{aligned}$$

Theorem. Let $(X_n)_{n \geq 1}$ be a sequence of [integrable random variables](#), and let Y be another random variable. Then (X_n) is [uniformly integrable](#) and [converges in probability](#) to X if and only if X is integrable and $X_n \xrightarrow{L^1} X$.

Proof of (\Rightarrow) . Suppose (X_n) is [uniformly integrable](#) and [converges to \$X\$ in probability](#). We've shown that there is a subsequence $(X_{n_k})_k$ such that $X_{n_k} \xrightarrow{k \rightarrow \infty} X$ almost surely.

$$\mathbb{E}(\|X\| \mathbb{1}_{\|X\| \geq M}) \leq \liminf_{n \rightarrow \infty} \mathbb{E}(\|X_{n_k}\| \mathbb{1}_{\|X_{n_k}\| > M}) = \epsilon M$$

by [Fatou's lemma](#).

So $\mathbb{E}(\|X\|) = \mathbb{E}(\|X\| \mathbb{1}_{\|X\| < M}) + \mathbb{E}(\|X\| \mathbb{1}_{\|X\| \geq M}) \leq M + \epsilon M < \infty$ so X is integrable. Hence **rest of the proof missing**. \square

Remark. This theorem subsumes the [Lebesgue Dominated Convergence theorem](#). If $\exists C > 0$ and some $p > 1$ such that $\mathbb{E}(\|X_n\|^p) \leq C$ for any n then $(X_n)_n$ is [uniformly integrable](#). Indeed

$$M^{p-1} \mathbb{E}(\|X_n\| \mathbb{1}_{\|X_n\| > M}) \leq \mathbb{E}(\|X_n\|^p) \leq C$$

so

$$\sup_n \mathbb{E}(\|X_n\| \mathbb{1}_{\|X_n\| > M}) \leq \frac{C}{M^{p-1}} \xrightarrow{M \rightarrow \infty} 0.$$

missing end of section

6 L^p spaces

Proposition (Hölder inequality). Let $p, q \in [1, \infty]$ such that $\frac{1}{p} + \frac{1}{q} = 1$ (taking $\frac{1}{\infty} := 0$). Let f, g be measurable functions on a measure space (X, \mathcal{A}, μ) . Then

$$\int_X |fg| d\mu \leq \sqrt[p]{\int_X |f|^p d\mu} \sqrt[q]{\int_X |g|^q d\mu}$$

with equality iff $\exists a, b \in \mathbb{R}$ not both 0 such that $a|f|^p = b|g|^q$ μ -almost everywhere.

Proof. Without loss of generality, assume $p, q \in (1, \infty)$ and $\int |f|^p d\mu = 1 = \int |g|^q d\mu$ (up to changing f, g by a scalar). The Young inequality for products gives

$$\forall a, b \geq 0 \quad ab \leq \frac{1}{p}a^p + \frac{1}{q}b^q$$

(follows from convexity of $x \mapsto -\log x$, since

$$\log(ta^p + (1-t)b^q) \geq t \log a^p + (1-t) \log b^q$$

and take $t = \frac{1}{p}$.) Write $|f(x)g(x)| \leq \frac{1}{p}|f(x)|^p + \frac{1}{q}|g(x)|^q$ and integrate, giving

$$\int |fg| d\mu \leq 1. \quad \square$$

Proposition (Minkowski inequality). Take $p \in [1, \infty]$ and f, g measurable functions on a measure space (X, \mathcal{A}, μ) . Then

$$\sqrt[p]{\int |f+g|^p d\mu} \leq \sqrt[p]{\int |f|^p d\mu} + \sqrt[p]{\int |g|^p d\mu}.$$

Proof. For $p = 1$, this follows by linearity of \int so take $p > 1$. Apply Hölder to $|f||f+g|^{p-1}$ and $|g||f+g|^{p-1}$:

$$\int |f+g|^p \leq \int (|f| + |g|)|f+g|^{p-1} \leq \left(\sqrt[p]{\int |f|^p} + \sqrt[p]{\int |g|^p} \right) \sqrt[q]{\int |f+g|^{q(p-1)}}$$

and $q(p-1) = p$. \square

Remark. We need to check $\int |f+g|^p < \infty$ for this proof to make sense. But this follows from

$$|f+g|^p \leq (2 \max\{|f|, |g|\})^p = 2^p \max\{|f|^p, |g|^p\} \leq 2^p(|f|^p + |g|^p)$$

so the integrability of $|f|^p$ and $|g|^p \implies \int |f+g|^p d\mu < \infty$.

Notation. Write

$$\mathcal{L}^p(X, \mathcal{A}, \mu) = \left\{ \mathcal{A}\text{-measurable functions } f \text{ on } X \text{ such that } \int_X |f|^p d\mu < \infty \right\}.$$

By the previous remark this is a real vector space.

Definition. The L^p -norm of f is defined by $\|f\|_p = \sqrt[p]{\int |f|^p d\mu}$ for $1 \leq p < \infty$.

The following hold:

- $\|\alpha f\|_p = |\alpha| \|f\|_p$ for all $\alpha \in \mathbb{R}$
- $\|f + g\|_p \leq \|f\|_p + \|g\|_p$ (from [Minkowski](#))

Remark. If $\|f\|_p = 0$ then $f = 0$ μ -a.e., so $\|\cdot\|_p$ is only a pseudo-norm on $\mathcal{L}^p(X, \mathcal{A}, \mu)$. Note

$$\begin{array}{ll} \text{if } f_1 = f_2 \text{ } \mu\text{-a.e.} & \text{then } f_1 + g_1 = f_2 + g_2 \text{ } \mu\text{-a.e.} \\ g_1 = g_2 \text{ } \mu\text{-a.e.} & f_1 g_1 = f_2 g_2 \text{ } \mu\text{-a.e.} \end{array}$$

so if we say that two functions f, g on (X, \mathcal{A}, μ) are equivalent $f \sim_\mu g$ if $f - g = 0$ μ -a.e. then this is an equivalence relation compatible with $+$, \times .

Definition (L^p space). Let $L^p(X, \mathcal{A}, \mu)$ be the set of equivalence classes of functions in $\mathcal{L}^p(X, \mathcal{A}, \mu)$. Define $\|[f]\|_p = \|f\|_p$. This is a real vector space and $\|[f]\|_p$ is a genuine norm of f .

When $p = \infty$ and f a measurable function,

$$\|f\|_\infty := \inf \{ t \geq 0 \mid |f(x)| \leq t \text{ for } \mu\text{-a.e. } x \}.$$

Proposition (Completeness of L^p spaces). $L^p(X, \mathcal{A}, \mu)$ for $p \in [1, \infty]$ is complete, i.e. every Cauchy sequence converges.

Proof. Let $(f_n)_{n \geq 1}$, $f_n \in \mathcal{L}^p(X, \mathcal{A}, \mu)$ and $\forall \epsilon \exists N$ with $\|f_n - f_m\|_p < \epsilon$ for all $n, m > N$. Pick a subsequence n_k such that $f_{n_{k+1}} - f_{n_k} \leq \frac{\epsilon}{2^k}$ so $\sum_{k \geq 1} \|f_{n_{k+1}} - f_{n_k}\| \leq \epsilon$. By [Minkowski](#),

$$\left\| \sum_{k=1}^K |f_{n_{k+1}} - f_{n_k}| \right\|_p \leq \epsilon \quad \forall k < \infty.$$

Let $K \rightarrow \infty$ so by [monotone convergence theorem](#) it holds also for $K = \infty$. Thus μ -a.e.

$$\sum_{k=1}^{\infty} |f_{n_{k+1}} - f_{n_k}|(x) < \infty$$

so for μ -a.e. x , $(f_{n_k}(x))_k$ converges (by completeness of \mathbb{R}), let $f(x)$ be the limit. Outside this set of x s, just set $f(x) = 0$.

Finally,

$$\|f_n - f\|_p = \left\| \lim_{k \rightarrow \infty} (f_n - f_{n_k}) \right\|_p \leq \liminf_{k \rightarrow \infty} \|f_n - f_{n_k}\|$$

by [Fatou's lemma](#) but since f_n is Cauchy this is $\leq \epsilon$ if n is large enough. \square

So, L^p space is a Banach space.

Remark (Approximation by simple functions). The vector space of [simple functions](#) of the form

$$\sum_{i=1}^N a_i \mathbb{1}_{A_i} \quad \text{where } a_i \in \mathbb{R}, A_i \in \mathcal{A} \text{ such that } \mu(A_i) < \infty$$

is dense in $L^p(X, \mathcal{A}, \mu)$ for all $p \in [1, \infty]$. Namely if $f \in L^p(X, \mathcal{A}, \mu)$ then let $f = f^+ - f^-$ and

$$f_n^+ = \frac{2^n f^+}{2^n} \wedge n.$$

So f_n^+ a simple function. Similarly define f_n^- . Then $f_n^+ \rightarrow f^+$ and $f_n^- \rightarrow f^-$ **a.e.** and $|f_n| < 2|f|$. So when $p < \infty$ by **dominated convergence** we get $\|f_n - f\|_p \xrightarrow{n \rightarrow \infty} 0$. When $p = \infty$ $\|f - f_n\| \leq \frac{1}{2^n}$.

In particular $L^\infty \cap L^p$ is dense in L^p , since the simple functions are contained in $L^\infty \cap L^p$.

Remark. Take $X = \mathbb{R}^d$, \mathcal{A} the **Borel σ -algebra** and $\mu = \text{Lebesgue}$. Then $C_c(\mathbb{R}^d)$ and $C_c^\infty(\mathbb{R}^d)$ are dense in $L^p(\mathbb{R}^d, \mathcal{A}, \mu)$ when $p \in [1, \infty)$. Recall $C_c(\mathbb{R}^d)$ is the set of continuous functions with compact support and $C_c^\infty(\mathbb{R}^d)$ are the smooth functions with compact support.

This fails for $p = \infty$, take for instance $f = \mathbb{1}_{[0,1]}$



then for any continuous f , $\|\mathbb{1}_{[0,1]} - f\|_\infty \geq \frac{1}{2}$.

6.1 Hilbert spaces

Definition (Hilbert space). A vector space \mathcal{H} over \mathbb{R} or \mathbb{C} is called a **Hilbert space** if it is endowed with an inner product

$$\begin{aligned} \mathcal{H} \times \mathcal{H} &\longrightarrow \mathbb{R} \text{ or } \mathbb{C} \\ (x, y) &\longmapsto \langle x, y \rangle \end{aligned}$$

such that

- $\langle x, x \rangle \in \mathbb{R}_+$ and $\langle x, x \rangle = 0 \implies x = 0$
- $\overline{\langle x, y \rangle} = \langle y, x \rangle$ for any $x, y \in \mathcal{H}$
- $\langle x + \lambda x', y \rangle = \langle x + y \rangle + \lambda \langle x', y \rangle \forall \lambda \in \mathbb{R} \text{ or } \mathbb{C}$

and \mathcal{H} is complete with respect to the norm $\|x\| := \sqrt{\langle x, x \rangle}$.

Example. • $\mathcal{H} = \mathbb{R}^d$ with $\langle x, y \rangle = \sum_{i=1}^d x_i y_i$

- $\mathcal{H} = \mathbb{C}^d$ with $\langle x, y \rangle = \sum_{i=1}^d x_i \overline{y_i}$
- $\mathcal{H} = L^2(X, \mathcal{A}, \mu)$ with $\langle f, g \rangle = \int_X f g d\mu$ as an \mathbb{R} vector space.

Lemma (Existence of orthogonal projection). Let \mathcal{H} be a **Hilbert space** and let V be a closed vector subspace. Then $\forall x \in \mathcal{H}$, $\exists! y \in V$ such that $d(x, V) = \|x - y\|$ (where $d(x, V) := \inf \{ \|x - z\| \mid z \in V \}$).

Proof. Pick a sequence $y_n \in V$ such that $\|x - y_n\| \rightarrow d(x, V)$. Then

$$\left\| x - \frac{y_n + y_m}{2} \right\|^2$$

□

Corollary. $H = V \oplus V^\perp$ where $V^\perp = V^\perp = \{y \in H \mid \langle x, y \rangle = 0 \forall x \in V\}$.

Proof.

- $V \cap V^\perp = \{0\}$ because $\langle x, x \rangle = 0 \implies x = 0$.
- $H = V + V^\perp$ because $x - y \in V^\perp$ if y is the projection of x to V given by the lemma. Indeed, $\forall z \in V$, $\|x - y - z\|^2 \geq \|x - y\|^2$ since $z + y \in V$, ie $\langle x - y - z, x - y - z \rangle \leq \|x - y\|^2$. Also, $\langle x - y - z, x - y - z \rangle = \|x - y\|^2 + \|z\|^2 + 2\operatorname{Re}\langle x - y, z \rangle$. So $\forall z \in V$,

$$\|z\|^2 + 2\operatorname{Re}\langle x - y, z \rangle \leq 0.$$

We have $\forall t > 0$.

$$t^2 \|z\|^2 + 2t \operatorname{Re}\langle x - y, z \rangle \leq 0t \|z\|^2 + 2\operatorname{Re}\langle x - y, z \rangle \leq 0$$

Take $t \rightarrow 0$, so

$$\operatorname{Re}\langle x - y, z \rangle \leq 0 \forall z \in V.$$

By taking $z \mapsto -z$ and $z \mapsto iz$, we see both $\operatorname{Re}\langle x - y, z \rangle = 0$ and $\operatorname{Im}\langle x - y, z \rangle = 0$, so $\langle x - y, z \rangle = 0$, and $x - y \in V^\perp$.

□

Let E be a Banach space (a complete normed vector space). E^* , the dual of E is defined as the set of bounded functionals on E : $\{l : E \rightarrow \mathbb{C} \mid \text{linear } \sup_{\|x\| \leq 1} |l(x)| < \infty\}$. E^* is a Banach space (completeness follows by that of \mathbb{C}) with norm $\|l\| = \sup_{\|x\| \leq 1} |l(x)|$.

Proposition (Riesz Representation Theorem). If $E = H$ is Hilbert space, then H is self-dual, ie $H \cong H^*$ namely the map $H \rightarrow H^*$, $y \mapsto (l_y : x \mapsto \langle x, y \rangle)$ is an isomorphism. In particular, $\forall l \in H^*$, $\exists y \in H$ such that $l(x) = \langle x, y \rangle$.

Proof.

- $V = \operatorname{Ker} l$. V is closed because l is bounded.
- So by previous corollary, $H = V \oplus V^\perp = \operatorname{Ker} l \oplus (\operatorname{Ker} l)^\perp$, say $l \neq 0$.
- Pick $x_0 \in (\operatorname{Ker} l)^\perp$ such that $l(x_0) = 1$.

Clearly x_0 is unique because if $l(x_1) = 1$ then $l(x_0 - x_1) = 0$, so $x_0 - x_1 \in (\operatorname{Ker} l)^\perp \cap \operatorname{Ker} l = \emptyset$. Now $l(y) - \langle y, \frac{x_0}{\|x_0\|^2} \rangle$ vanishes on $\operatorname{Ker} l$ and on $(\operatorname{Ker} l)^\perp = \mathbb{C}x_0$. So, $l(y) = \langle y, \frac{x_0}{\|x_0\|^2} \rangle$. □

Theorem (Jensen's inequality). Let $\Omega \subseteq \mathbb{R}^d$ be a convex set. Let $\phi : \Omega \rightarrow \mathbb{R}$ be a convex (continuous) function. Let X be an integrable \mathbb{R}^n valued random variable such that $X \in \Omega$ almost surely. Then

$$\mathbb{E}(\phi(X)) \geq \phi(\mathbb{E}X)$$

Remark.

- Ω convex means ...
- ϕ convex means ...

Proof.

□