Notes for the course of Functional Analysis, PhD in Statistics

Annalisa Cesaroni

Academic Year 2024-2025

Abstract

These notes are intended for the first year students of the PhD course in Statistics, at University of Padova. They are not exhaustive, nor complete, but they could serve as a basis of the study of the arguments presented during the course of Functional Analysis. The topics are presented in a quite informal way, trying to reach also students without a specific preparation in mathematics. Only few proofs are provided and for the others bibliographical references are provided. At the end of each section some exercises are proposed, more or less simple to solve. In the appendix there are the (sketchy) solutions to the problems.

Contents

1 Measure theory and integration

1.1 Measure space

We fix a set X and we define $\mathcal{P}(X)$ the set of all subsets of X.

Definition 1.1. $\Sigma \subset \mathcal{P}(X)$ *is a* $\sigma-$*algebra on* X *if*

- *it is closed by complement, that is if* $A \in \Sigma$ *then* $X \backslash A \in \Sigma$,
- *it is closed by countable union, that is if* $(A_i)_i$ *is a sequence of elements in* Σ *then* $\cup_{i=1}^{\infty} A_i \in \Sigma$.

Let $C \subseteq \mathcal{P}(X)$, *then* $\Sigma(\mathcal{C})$, *the* σ*-algebra generated by* $\mathcal{C}$ *is the smallest* $\sigma-$*algebra which contains all the elements in* $\mathcal{C}$ *(and then all countable intersections and countable unions of elements in* $\mathcal{C}$*).*

The smallest possible σ-algebra on X is given by $\Sigma = \{\emptyset, X\}$, and the largest possible σ-algebra on X is $\Sigma = \mathcal{P}(X)$.

Definition 1.2. $\mathcal{B}(\mathbb{R})$ *is the* σ*-algebra on* $\mathbb{R}$ *generated by all the intervals* $\mathcal{C} = \{(a, b) \mid a, b \in \mathbb{R}\}$. $\mathcal{B}(\mathbb{R}^N)$ *is the* σ*-algebra on* $\mathbb{R}^N$ *generated by all the pluri-rectangulars* $\mathcal{C} = \{\Pi_{i=1}^N (a_i, b_i) \mid a_i, b_i \in \mathbb{R}\}$.

Remark 1.3. Note that $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ also when $\mathcal{C} = \{(a, b] \mid a, b \in \mathbb{R}\}$, since $(a, b) = \cup_n \left(a, b - \frac{1}{n}\right]$, or when $\mathcal{C} = \{[a, b) \mid a, b \in \mathbb{R}\}$, since $(a, b) = \cup_n \left[a + \frac{1}{n}, b\right)$, or when $\mathcal{C} = \{[a, b] \mid a, b \in \mathbb{R}\}$ again because $(a, b) = \cup_n \left[a + \frac{1}{n}, b - \frac{1}{n}\right]$. Analogously $\sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$ when $\mathcal{C} = \{(a, +\infty) \mid a \in \mathbb{R}\}$, since $(a, b] = (a, +\infty) \cap (-\infty, b]$, and $(-\infty, b] = \mathbb{R} \backslash (b, +\infty)$ and so on.

Definition 1.4. *Let* Σ *be a* σ*-algebra on* X. *A function* $\mu : \Sigma \to [0, +\infty]$ *is a measure if*

- $\mu(\emptyset) = 0$,
- *it is* σ*-additive, that is if* $(A_i)_i$ *is a sequence of elements in* Σ *with* $A_i \cap A_j = \emptyset$ *for* $i \neq j$ *then* $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{+\infty} \mu(A_i)$.

(X, Σ, μ) *is called a measure space.*

If $\mu(X) < +\infty$, *then* μ *is a finite measure (a probability measure if* $\mu(X) = 1$*). Usually measure spaces with probability measures are denoted with* Ω *(in place of* X*), the* σ*-algebra is* $\mathcal{F}$ *(in place of* Σ*) and the measure is* $\mathbb{P}$ *(in place of* μ*).*

If $X = \cup_i A_i$, *with* $\mu(A_i) < +\infty$ *for all* i, μ *is* σ*-finite.*

If $X = \mathbb{R}^n$, $n \geqslant 1$ *and* $\Sigma = \mathcal{B}(\mathbb{R}^n)$, *then* μ *is called a Borel measure.*

Example 1.5. Let $x_0 \in \mathbb{R}$, and define the measure on $\mathcal{P}(\mathbb{R})$ as $\delta_{x_0}(A) = \begin{cases} 1 & x_0 \in A \\ 0 & x_0 \notin A \end{cases}$.

Then δ_{x_0} is called Dirac measure centered at x_0.

Proposition 1.6 (Monotonicity, subadditivity, continuity)**.** *Let μ be a measure on Σ. Then*

(i) *if $A \subset B$, $A, B \in \Sigma$, then $\mu(A) \leqslant \mu(B)$ (monotonicity with respect to inclusion);*

(ii) *if $(A_i)_i$ is a sequence of elements in Σ then $\mu(\cup_{i=1}^{\infty} A_i) \leqslant \sum_{i=1}^{+\infty} \mu(A_i)$;*

(iii) *if $(A_i)_i$ is a sequence of elements in Σ with $A_i \subseteq A_{i+1}$ then $\mu(\cup_{i=1}^{\infty} A_i) = \lim_{i\to+\infty} \mu(A_i)$;*

(iv) *if $(A_i)_i$ is a sequence of elements in Σ with $A_i \supseteq A_{i+1}$ and $\mu(A_{i_0}) < +\infty$ for some i_0, then $\mu(\cap_{i=1}^{\infty} A_i) = \lim_{i\to+\infty} \mu(A_i)$.*

Proof. (i) Observe that $B = A \cup (B\backslash A)$, so by σ-additivity $\mu(B) = \mu(A) + \mu(B\backslash A) \geqslant \mu(A)$.

(ii) Let $B_1 = A_1$ and $B_i = A_i \backslash \cup_{k=1}^{i-1} A_k$ then B_i are disjoint and

$$\mu(\cup_i A_i) = \mu(\cup_i B_i) = \sum_{i=1}^{+\infty} \mu(B_i) \leqslant \sum_{i=1}^{+\infty} \mu(A_i).$$

(iii) Let $B_1 = a_1$ and $B_i = A_i \backslash A_{i-1}$ then

$$\mu(\cup_i A_i) = \mu(\cup_i B_i) = \sum_{i=1}^{+\infty} \mu(B_i) = \lim_{n\to+\infty} \sum_{i=1}^{n} \mu(B_i) = \mu(A_n).$$

(iv) Let $F_i = A_{i_0} \backslash A_i$ for $i > i_0$. Then $\mu(A_{i_0}) = \mu(F_i) + \mu(A_i)$, $F_i \subseteq F_{i+1}$ and $\cup_i F_i = A_{i_0} \backslash \cap_i A_i$. Therefore by 1), we get

$$\mu(A_{i_0}) = \mu(\cap_i A_i) + \lim_i \mu(F_i) = \mu(\cap_i A_i) + \lim_i (\mu(A_{i_0}) - \mu(A_i))$$

and we cancel $\mu(A_{i_0})$ from both sides.

□

Definition 1.7. *Let (X, Σ, μ) a measure space. The completion of Σ with respect to μ is the σ-algebra*

$$\mathcal{M} = \{A \subseteq X \mid \exists B, C \in \Sigma, \mu(C) = 0, B \subseteq A, A\backslash B \subseteq C\}.$$

Definition 1.8. *Let (X, Σ, μ) a measure space. A property holds* **almost everywhere** *if there exists $N \in \Sigma$ with $\mu(N) = 0$ such that the property holds for all $x \in X\backslash N$.*

Proposition 1.9. *Let Σ be a σ-algebra on X and $\mu : \Sigma \to [0, +\infty]$ with $\mu(\varnothing) = 0$. Then they are equivalent:*

(i) *μ is σ-additive: if $(A_i)_i$ is a sequence of elements in Σ with $A_i \cap A_j = \varnothing$ for $i \neq j$ then $\mu(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{+\infty} \mu(A_i)$,*

(ii) *μ is additive: if $A, B \in \Sigma$ and $A \cap B = \varnothing$ then $\mu(A \cap B) = \mu(A) + \mu(B)$*
and
μ is countable subadditive: if $(A_i)_i$ is a sequence of elements in Σ then $\mu(\cup_{i=1}^{\infty} A_i) \leqslant \sum_{i=1}^{+\infty} \mu(A_i)$;

(iii) *μ is additive: if $A, B \in \Sigma$ and $A \cap B = \varnothing$ then $\mu(A \cap B) = \mu(A) + \mu(B)$*
and
μ is continuous on increasing sequence of sets: if $(A_i)_i$ is a sequence of elements in Σ with $A_i \subseteq A_{i+1}$ then $\mu(\cup_{i=1}^{\infty} A_i) = \lim_{i\to+\infty} \mu(A_i)$.

Proof. The fact that (i) implies (ii) and that (i) implies (iii) has been proved in Proposition 1.6. We prove that (ii) implies (i). We consider a sequence $(A_i)_i$ of elements in Σ with $A_i \cap A_j = \varnothing$ for $i \neq j$. Then by (ii) we get that $\mu(\cup_{i=1}^{\infty} A_i) \leqslant \sum_{i=1}^{+\infty} \mu(A_i)$. On the other hand by additivity and monotonicity (which is a consequence of additivity) we get that for every n, $\mu(\cup_{i=1}^{\infty} A_i) \geqslant \mu(\cup_{i=1}^{n} A_i) = \sum_{i=1}^{n} \mu(A_i)$. Sending $n \to +\infty$ we conclude $\mu(\cup_{i=1}^{\infty} A_i) \geqslant \sum_{i=1}^{+\infty} \mu(A_i)$.

We prove that (iii) implies (i). We consider a sequence $(A_i)_i$ of elements in Σ with $A_i \cap A_j = \varnothing$ for $i \neq j$. We define $B_i = \cup_{j=1}^{i} A_j$. Then $\cup_i B_i = \cup_i A_i$. Note that by additivity $\mu(B_i) = \sum_{j=1}^{i} \mu(A_j)$ and that $B_1 \subseteq B_2 \subseteq B_3 \dots$. Therefore by (iii) and additivity we get

$$\mu(\cup_{i=1}^{\infty} A_i) = \mu(\cup_{i=1}^{\infty} B_i) = \lim_{i \to +\infty} \mu(B_i) = \lim_{i \to +\infty} \sum_{j=1}^{i} \mu(A_j) = \sum_{j=1}^{+\infty} \mu(A_j).$$

□

1.2 Borel measures on $\mathbb{R}$ and cumulative distribution functions

Let $F : \mathbb{R} \to \mathbb{R}$ be an increasing function which is right continuous, that is $\lim_{x \to a^+} F(x) = F(a)$. We define for all $a, b \in \mathbb{R}$,

$$\mu_F(a, b] = F(b) - F(a) \qquad \mu_F(\varnothing) = 0.$$

Then for every set $C \subset \mathbb{R}$ we define

$$\mu_F^*(C) = \inf\{\sum_i F(b_i) - F(a_i) \mid C \subseteq \cup_i (a_i, b_i]\}.$$

Note that since F is increasing, we get that for sequences $a_1 < b_1 < a_2 < b_2 < \cdots < a_i < b_i < a_{i+1} < b_{i+1} \dots$, we obtain

$$\mu_F^*(\cup_i (a_i, b_i]) = \sum_i F(b_i) - F(a_i).$$

Observe that if we define $\mathcal{C} = \{(a, b], a, b \in \mathbb{R}\}$, then $\Sigma(\mathcal{C}) = \mathcal{B}(\mathbb{R})$. Note that if $F_1 = F_2 + c$ for some constant then $\mu_{F_1}^* = \mu_{F_2}^*$. Also the viceversa is true: if $\mu_{F_1}^* = \mu_{F_2}^*$, then $F_1 = F_2 + c$ for some constant c.

Remark 1.10. Note that F monotone increasing implies that $\mu_F(a, b] \geqslant 0$, and moreover, since F is right continuous, then

$$\mu_F\left(\cup_n (a + 1/n, b]\right) = \mu_F(a, b] = F(b) - F(a) = F(b) - \lim_n F(a + 1/n) = \lim_n \mu_F(a + 1/n, b].$$

Reasoning as before, it is possible to see that, at least when restricted to $\mathcal{C}$, there holds that μ_F has positive values, is additive and is continuous with respect to increasing sequences of sets (which is enough to get σ-additivity if μ_F is defined on a σ-algebra, see Proposition 1.9).

We recall that F is monotone increasing and then $\lim_{x \to +\infty} F(x) = \sup F$ and $\lim_{x \to -\infty} F(x) = \inf F$ (we say that if $F(\mathbb{R})$ is unbounded from above, $\sup F = +\infty$ and if $F(\mathbb{R})$ is unbounded from below, $\inf F = -\infty$).

We may extend μ_F^* to intervals obtained by unions and intersections of elements in $\mathcal{C}$, and using additivity and continuity. In particular we get

$$\begin{aligned}
\mu_F^*(a,+\infty) &= \mu_F^*\left(\cup_n (a,a+n]\right) = \lim_n F(a+n) - F(a) = \sup F - F(a) \\
\mu_F^*(-\infty,b] &= \mu_F^*\left(\cup_n (b-n,b]\right) = \lim_n F(b) - F(b-n) = F(b) - \inf F \\
\mu_F^*(a,b) &= \mu_F^*\left(\cup_{n\geqslant n_0} (a,b-1/n]\right) = \lim_n F(b-1/n) - F(a) = \lim_{x\to b^-} F(x) - F(a) \\
\mu_F^*(-\infty,b) &= \mu_F^*\left((-\infty,b-1] \cup (b-1,b)\right) = \mu_F^*((-\infty,b-1]) + \mu_F^*((b-1,b)) \\
&= \lim_{x\to b^-} F(x) - F(b-1) + F(b-1) - \inf F = \lim_{x\to b^-} F(x) - \inf F \\
\mu_F^*[a,b) = &= \mu_F^*[(a-1,b)\backslash(a-1,a)] = \mu_F^*(a-1,b) - \mu_F^*(a-1,a) \\
&= \lim_{x\to b^-} F(x) - F(a-1) - \lim_{x\to a^-} F(x) + F(a-1) = \lim_{x\to b^-} F(x) - \lim_{x\to a^-} F(x) \\
\mu_F^*[a,b] &= \mu_F^*[[a,b+1)\backslash(b,b+1)] = \mu_F^*[a,b+1) - \mu_F^*(b,b+1) \\
&= F(b) - \lim_{x\to a^-} F(x) \\
\mu_F^*[a,+\infty) &= \sup_F - \lim_{x\to a^-} F(x).
\end{aligned}$$

Note that

$$\begin{aligned}
\mu_F^*(\mathbb{R}) &= \mu_F^*\left(\cup_n (a-n,b+n]\right) = \lim_n F(b+n) - F(a-n) = \sup F - \inf F \\
\mu_F^*(\{a\}) &= \mu_F^*\left((c,a]\backslash(c,a)\right) \\
&= \mu_F^*((c,a]) - \mu_F^*((c,a)) = F(a) - F(c) - (\lim_{x\to a^-} F(x) - F(c) = F(a) - \lim_{x\to a^-} F(x).
\end{aligned}$$

Theorem 1.11. (i) *There exists a unique Borel measure μ_F which coincides with μ_F^* on intervals $(a,b]$. This measure is σ-finite and it is finite if and only if* $\sup F - \inf F < +\infty$.

(ii) *Given a Borel measure on $\mathbb{R}$ which is σ-finite, there exists F monotone increasing and right continuous such that $\mu = \mu_F$. F is unique up to addition of constants: that is if $\mu = \mu_F = \mu_G$ then there exists $c \in \mathbb{R}$ such that $F(x) = G(x) + c$ for all x. t*

Proof. (i) The proof is based on the Caratheodory criterion, and we refer to [2, Theorem 1.14, Theorem 1.16]. As for the σ- finiteness it is sufficient to observe that $\mu_F(-n,n] = F(n) - F(-n) < +\infty$ and $\mathbb{R} = \cup_n(-n,n]$. Moreover, since $\mu_F(\mathbb{R}) = \sup F - \inf F$, we conclude that F is finite iff $\sup_F - \inf F < +\infty$.

(ii) We want to construct F. Put $F(0) = 0$ and

$$F(x) = \begin{cases} \mu(0,x] & x > 0 \\ -\mu(x,0] & x < 0. \end{cases}$$

Observe that if $b > a \geqslant 0$, $F(b) - F(a) = \mu(0,b] - \mu(0,a] = \mu(0,b]\backslash(0,a] = \mu(a,b] \geqslant 0$, if $0 \geqslant b > a$, then $F(b) - F(a) = -\mu(b,0] + \mu(a,0] = \mu(a,0]\backslash(b,0] = \mu(a,b] \geqslant 0$ and finally if $a < 0 < b$, then $F(b) - F(a) = \mu(0,b] + \mu(a,0] = \mu(a,b] \geqslant 0$. So F is increasing.

We check that it is right continuous. First of all observe that for $a > 0$, $\lim_{x\to a^+} F(x) = \lim_n F(a+1/n) = \lim_n \mu(0,a+1/n] = \mu(\cap_n(0,a+1/n]) = \mu(0,a] = F(a)$. If $a = 0$ $\lim_{x\to 0^+} F(x) = \lim_n F(1/n) = \lim_n \mu(0,1/n] = \mu(\cap_n(0,1/n]) = \mu(\varnothing) = 0 = F(0)$. Finally if $a < 0$, then $\lim_{x\to a^+} F(x) = \lim_n F(a+1/n) = -\lim_n \mu(a+1/n,0] = -\mu(\cup_n(a+1/n,0]) = -\mu(a,0] = F(a)$.

Finally we already checked that $\mu(a,b] = F(b) - F(a)$ and then we conclude that $\mu = \mu_F$.

Assume now that there exists a right continuous increasing function G such that $\mu = \mu_G$. Then for $x > 0$, $F(x) = \mu(0,x] = \mu_G(0,x] = G(x) - G(0)$ and for $x < 0$ then $F(x) = -\mu(x,0] = \mu_G(x,0] = -(G(0) - G(x)) = G(x) - G(0)$. So, this implies that $F(x) = G(x) - G(0)$ (for $x = 0$ this is trivially verified).

□

Definition 1.12. *Let μ be a finite Borel measure. The function $F(x)$ associated to the measure μ and normalized in order to have* $\inf F = 0$ *is called the cumulative distribution function of the measure μ. It is easy to check that $F(x) := \mu(-\infty, x]$.*

1.3 The Lebesgue measure on $\mathbb{R}$ and $\mathbb{R}^n$.

Definition 1.13. *Let $F(x) = x$ for all x, then $\overline{\mu}_F$ is called* **Lebesgue measure**. *We indicate with $\mathcal{L}$. We denote with $\mathcal{M}(\mathbb{R})$ the completion of $\mathcal{B}(\mathbb{R})$ with respect to $\mathcal{L}$, and we call it the Σ-algebra of Lebesgue measurable sets.*

Proposition 1.14. *The Lebesgue measure*

(i) *associates to each interval its length,*

(ii) *is translation invariant, that is $\mathcal{L}(A + x) = \mathcal{L}(A)$ for all $x \in \mathbb{R}$, $A \in \mathcal{M}$,*

(iii) *is homogenous, that is $\mathcal{L}(\lambda A) = \lambda \mathcal{L}(A)$ for all $\lambda > 0$, $A \in \mathcal{M}$,*

(iv) *assigns measure* 0 *to points, and so also to countable sets (e.g. $\mathbb{Q}$),*

(v) *it is σ-finite, since $\mathbb{R} = \cup_{n\in\mathbb{N}}(-n, n)$ and $\mathcal{L}(-n, n) = 2n$.*

Proof. The proof is immediate by definitions and σ-additivity. Exercise. □

Measurable sets in $\mathbb{R}$ which contain at least one interval (they are called sets with non empty interior) have positive measure. On the other hand sets which are given by countable union of isolated points have measure zero. Nevertheless there are sets with empty interior in $\mathbb{R}$ (so that do not contain any interval) and with positive measure (almost full measure).

Example 1.15 (A set of positive measure which does not contain any interval)**.** Let (r_n) be an enumeration of $\mathbb{Q} \cap [0, 1]$ and fix $\varepsilon > 0$ small.

Set $A = \cup_n(r_n - \varepsilon 2^{-n}, r_n + \varepsilon 2^{-n})$. Then by subadditivity, $\mathcal{L}(A) \leqslant \sum_n 2\varepsilon 2^{-n} = 4\varepsilon$. Moreover $B = [0, 1]\backslash A$ is a set which does not contain any interval (otherwise it should contain some rational number but $\mathbb{Q} \cap [0, 1] \subseteq A$), and moreover $\mathcal{L}(B) \geqslant 1 - 4\varepsilon > 0$.

Not all the subsets of $\mathbb{R}$ are contained in $\mathcal{M}(\mathbb{R})$, so there are sets which are not measurable. This is due to the fact that if we want to define a measure μ on the intervals of $\mathbb{R}$ such that $\mu([0, 1]) = 1$, $\mu(A \cup B) = \mu(A) + \mu(B)$ if $A \cap B = \varnothing$ and $\mu(A) = \mu(B)$ if B can be obtained translating and rotating A, then the σ- algebra of measurable sets cannot be $\mathcal{P}(\mathbb{R})$.

Example 1.16 (A set which is not (Lebesgue) measurable)**.** We say that $x, y \in [0, 1]$ are equivalent if $x - y \in \mathbb{Q}$. Let $P \in [0, 1]$ a set such that P consists of exactly one representative point from each equivalence class (this set exists by the axiom of choice). In particular this means that if $p, p' \in P$, $p \neq p'$, then $p - p' \notin \mathbb{Q}$. We claim that P provides the required example of a non measurable set. We prove it by contradiction, showing that it is not possible for P to be measurable.

For each $q \in \mathbb{Q} \cap [0, 1]$, define

$$P_q = [(P+q)\cap[0,1)]\cup[(P+q)\backslash[0,1))-1] = \{p+q,\ p\in P\cap[0,1-q)\}\cup\{p+q-1,\ p\in P\cap[1-q,1)\}.$$

So P_q is obtained by considering $P + q$ and then shifting back of 1 unit the part of $P + q$ which is outside the interval $[0, 1)$.

First of all we observe that $\mathcal{L}(P) = \mathcal{L}(P_q)$. Indeed $[(P + q) \cap [0, 1)] \cap [(P + q)\backslash[0, 1)) - 1] = \varnothing$, since if $p + q \in [0, 1)$ for some $p \in P$ and $p' + q - 1 \in [0, 1)$ for some $p' \in P$, then necessarily $p + q \neq p' + q - 1$, since $p, p' \in [0, 1)$.

Moreover we observe that if $r \neq q \in \mathbb{Q} \cap [0, 1)$, then $P_r \cap P_q = \varnothing$. Indeed assume it is not true and $x \in P_r \cap P_q$, this means that $x = p + r = p' + q$, for some $p, p' \in P$ or $x = p + r = p' + q - 1$, or $x = p + r - 1 = p' + q$. In any case we get that $p - p' \in \mathbb{Q}$, which implies that $p = p'$ by definition of the set P and so $r = q$.

Finally we observe that $\cup_{q\in\mathbb{Q}\cap[0,1)} P_q = [0,1)$. Indeed take $x \in [0,1)$, then there exists $p \in P$ such that x is equivalent to P, which means that there exists $q \in \mathbb{Q}$ such that $x = p + q$. In particular this implies that $q \in (0,1]$ and $x \in P_q$.

We conclude by σ-additivity that

$$1 = \mathcal{L}([0,1)) = \mathcal{L}(\cup_{q\in\mathbb{Q}\cap[0,1)} P_q) = \sum_{q\in\mathbb{Q}\cap[0,1)} \mathcal{L}(P_q) = \sum_{q\in\mathbb{Q}\cap[0,1)} \mathcal{L}(P) = \begin{cases} 0 & \text{if } \mathcal{L}(P) = 0 \\ +\infty & \text{if } \mathcal{L}(P) > 0 \end{cases}$$

which is not possible.

It is possible to define the Lebesgue measures on $\mathbb{R}^n$ as the product measure of the Lebesgue measure on $\mathbb{R}$. It is a Borel maesure and we denote with $\mathcal{M}$ the Σ-algebra of Lebesgue measurable sets. We refer to [2, Section2.6].

Proposition 1.17. *The Lebesgue measure on* $\mathbb{R}^n$

(i) *associates to each set its volume,*

(ii) *is translation invariant, that is* $\mathcal{L}(A + x) = \mathcal{L}(A)$ *for all* $x \in \mathbb{R}^n$, $A \in \mathcal{M}$,

(iii) *is n-homogenous, that is* $\mathcal{L}(\lambda A) = \lambda^n \mathcal{L}(A)$ *for all* $\lambda > 0$, $A \in \mathcal{M}$, *in particular* $\mathcal{L}(B(0,r)) = r^n \mathcal{L}(B(0,1))$, *where* $B(0,r)$ *is the ball if radius* r *centered at* 0,

(iv) *it is* σ*-finite, since* $\mathbb{R}^n = \cup_{k\in\mathbb{N}} B(0,k)$ *and* $\mathcal{L}B(0,k) = k^n \mathcal{L}(B(0,1))$.

1.4 Measurable functions

Definition 1.18. *Let* (X, Σ, μ) *be a measure space, and let* $f : X \to \mathbb{R}$ *be a function. Then* f *is measurable if for all* $t \in \mathbb{R}$,

$$A(t) := \{x \in X \mid f(x) > t\} = f^{-1}(t, +\infty) \in \Sigma.$$

In particular we will be interested in the case in which $(X, \Sigma, \mu) = (\mathbb{R}^n, \mathcal{M}, \mathcal{L})$. *In this case saying that* $f : \mathbb{R}^n \to \mathbb{R}$ *is measurable is equivalent to require that for all* $A \in \mathcal{B}(\mathbb{R})$, $f^{-1}(A) \in \mathcal{M}$.

Example 1.19. Let $A \in \mathcal{M}$ and define the characteristic function of A as

$$\chi_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A. \end{cases}$$

Then χ_A is measurable. Indeed $A(t) = \varnothing$ for $t \geqslant 1$, $A(t) = \mathbb{R}^n$ for $t \leqslant 0$ and $A(t) = A$ for $t \in (0,1)$.

Example 1.20 (Random variables)**.** If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space (that is a measure space endowed with a probability measure), the measurable functions, that is functions $f : \Omega \to \mathbb{R}$ such that for all $t \in \mathbb{R}$, $A(t) := \{\omega \in \Omega \mid f(\omega) > t\} \in \mathcal{F}$, are called **random variables**. Usually random variables are indicated with X instead of f.

There is a notion of convergence of measurable functions which is quite used in probability.

Definition 1.21 (Convergence in measure)**.** *Let* f_n, f *be measurable functions defined on the measure space* (X, Σ, μ). *Then* f_n *converge to* f *in measure if for every* $\varepsilon > 0$

$$\lim_n \mu\{x \in X \mid |f_n(x) - f(x)| \geqslant \varepsilon\} = 0.$$

If we are in a probability space, this convergence is called **convergence in probability***, since it reads*

$$\lim_n \mathbb{P}\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| \geqslant \varepsilon\} = 0.$$

1.5 Integration with respect to the Lebesgue measure

Definition 1.22. *Let $k \geqslant 1$, $A_1, \dots A_K$ a finite family of disjoint sets in $\mathcal{M}$ and $c_1, \dots c_k > 0$. The function $\phi(x) = \sum_{i=1}^{k} c_i \chi_{A_i}(x)$ is called* **simple function***. It is a measurable (positive) function and we define its integral as*

$$\int_{\mathbb{R}^N} \phi(x)dx = \sum_{i=1}^{k} c_i \mathcal{L}(A_i).$$

Definition 1.23 (Lebesgue integral)**.** *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function such that $f(x) \geqslant 0$ for all x. Then*

$$\int_{\mathbb{R}^n} f(x)dx = \sup \left\{ \int_{\mathbb{R}^n} \phi(x)dx \mid \phi \text{ simple function with } \phi \leqslant f \right\}.$$

If f is not positive we define its positve part $f^+(x) = \max(f(x), 0)$ and its negative part $f^-(x) = \max(-f(x), 0)$ and we define

$$\int_{\mathbb{R}^n} f(x)dx = \int_{\mathbb{R}^n} f^+(x)dx - \int_{\mathbb{R}^n} f^-(x)dx.$$

Note that $\int_{\mathbb{R}^n} |f(x)|dx = \int_{\mathbb{R}^n} f^+(x)dx + \int_{\mathbb{R}^n} f^-(x)dx$.

Since $f^+ \leqslant |f|, f^- \leqslant |f|$, we have that

$$\left| \int_{\mathbb{R}^n} f(x)dx \right| < +\infty \quad \textit{iff} \quad \int_{\mathbb{R}^n} |f(x)|dx < +\infty.$$

We denote

$$L^1(\mathbb{R}^n) := \{f : \mathbb{R}^n \to \mathbb{R} \mid \ f \text{ is measurable and } \int_{\mathbb{R}^n} |f(x)|dx < +\infty\}.$$

If $A \in \mathcal{M}$, then we define

$$L^1(A) = \left\{ f : \mathbb{R}^n \to \mathbb{R} \,|f \text{ is measurable and } \int_{\mathbb{R}^n} |f(x)|\chi_A(x) = \int_A |f(x)|dx < +\infty \right\}.$$

Proposition 1.24. *The following properties hold.*

- *If $f = 0$ almost everywhere then $\int_{\mathbb{R}^n} f(x) = 0$. If $\int_{\mathbb{R}^n} |f(x)|dx = 0$ then $f = 0$ almost everywhere.*
- *If f, g are measurable functions such that $f = g$ almost everywhere, then $\int_{\mathbb{R}^n} f(x)dx = \int_{\mathbb{R}^n} g(x)dx$.*
- *If $f, g \in L^1(\mathbb{R}^n)$, $\alpha, \beta \in \mathbb{R}$, then $\int_{\mathbb{R}^n} \alpha f(x) + \beta g(x)dx = \alpha \int_{\mathbb{R}^n} f(x)dx + \beta \int_{\mathbb{R}^n} g(x)dx$.*
- *If $f, g \in L^1(\mathbb{R}^n)$, and $f \leqslant g$ then $\int_{\mathbb{R}^n} f(x)dx \leqslant \int_{\mathbb{R}^n} g(x)dx$.*

Proof. The proof is obtained by exploiting definitions, see [2, Section 2..2] □

Remark 1.25. [On the definition of L^1] Note that due to the previous proposition, in particular the fact that if f, g are measurable functions such that $f = g$ almost everywhere, then $\int_{\mathbb{R}^n} f(x)dx = \int_{\mathbb{R}^n} g(x)dx$, **we identify functions in $L^1(\mathbb{R}^n)$ which coincide almost everywhere.** So a function f in $L^1(\mathbb{R}^n)$ is actually a class of equivalence of functions, we do not distinguish functions which are different on sets of measure zero.

Theorem 1.26 (Monotone convergence)**.** *Let $f_k : \mathbb{R}^n \to \mathbb{R}$ measurables , positive, i.e. $f_k \geqslant 0$ for all k, and such that $f_k(x) \leqslant f_{k+1}(x)$ for all x and for all k. Then*

$$\lim_k \int_{\mathbb{R}^n} f_k(x)dx = \int_{\mathbb{R}^n} \lim_k f_k(x)dx.$$

Proof. See [2, Theorem 2.14]. □

Proposition 1.27. *An equivalent definition of the Lebesgue integral (which can be very useful) is the following. Let $f : \mathbb{R}^n \to \mathbb{R}$ measurable and positive. Let for every $t > 0$ $F(t) = \mathcal{L}(A(t)) = \mathcal{L}\{x \mid f(x) > t\}$. F is called the* **repartition function of** *f. Then*

$$\int_{\mathbb{R}^n} f(x)dx = \int_0^{+\infty} F(t)dt.$$

Proof. See [2, Proposition 6.24] □

1.6 Decomposition of measures

Definition 1.28. *Let ν, ρ be measures defined on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.*

ν is **absolutely continuous** *with respect to $\mathcal{L}$, and we write $\nu << \mathcal{L}$ if $\nu(A) = 0$ for all $A \in \mathcal{B}$ such that $\mathcal{L}(A) = 0$.*

ρ is **singular** *with respect to $\mathcal{L}$, and we write $\rho \perp \mathcal{L}$, if there exist $A, B \in \mathcal{B}$, $A \cap B = \emptyset$, $A \cup B = \mathbb{R}^n$, such that $\mathcal{L}(A) = 0$ and $\rho(B) = 0$.*

Example 1.29. Let $x_0 \in \mathbb{R}$ and consider the Dirac measure δ_{x_0} centered at x_0. Then it is singular with respect to $\mathcal{L}$. Indeed fix $A = \mathbb{R}\backslash\{x_0\}$, $B = \{x_0\}$, and observe that $\mathcal{L}(B) = 0$ and $\delta_{x_0}(A) = 0$.

Proposition 1.30. *Let $f \geqslant 0$, measurable and such that $\int_{-M}^{M} f(x)dx < +\infty$ for all $M > 0$. Define the function*

$$\nu_f : \mathcal{M} \to [0, +\infty] \quad as \quad \nu_f(A) = \int_A f(x)dx.$$

Then ν_f is a measure on $(\mathbb{R}^n, \mathcal{M})$, which is σ-finite and which is absolutely continuous with respect to $\mathcal{L}$. If $f \in L^1(\mathbb{R}^n)$ the measure is finite.

Proof. First of all we show that it is a measure. Observe that $f(x)\chi_\emptyset(x) = 0$ almost everywhere, then $\nu_f(\emptyset) = 0$. Let $A_i \in \mathcal{M}$ which are pairwise disjoint. Define the simple function $\phi_k(x) = \sum_{i=1}^k \chi_{A_i}(x)$. Note that $\lim_k \phi_k(x) = \chi_{\cup_i A_i}(x)$. Moreover $0 \leqslant f(x)\phi_k(x) \leqslant f(x)\phi_{k+1}(x)$ and so by the monotone convergence theorem we get

$$\lim_k \int_{\mathbb{R}^n} \phi_k(x)f(x)dx = \int_{\mathbb{R}^n} \lim_k \phi_k(x)f(x)dx.$$

Observe that

$$\lim_k \int_{\mathbb{R}^n} \phi_k(x)f(x)dx = \lim_k \int_{\mathbb{R}^n} \sum_{i=1}^k \phi_i(x)f(x)dx = \lim_k \sum_{i=1}^k \int_{\mathbb{R}^n} \phi_i(x)f(x)dx$$

$$= \lim_k \sum_{i=1}^k \int_{A_i} f(x)dx = \lim_k \sum_{i=1}^k \nu_f(A_i) = \sum_{i=1}^{+\infty} \nu_f(A_i)$$

and

$$\int_{\mathbb{R}^n} \lim_k \phi_k(x)f(x)dx = \int_{\mathbb{R}^n} \chi_{\cup_i A_i}(x)f(x)dx = \nu_f(\cup_i A_i).$$

Therefore we get that ν_f is a measure.

Since $\nu_f(B(0,k)) = \int_{B(0,k)} f(x)dx < +\infty$ by assumption, then ν_f is σ-finite.

Finally, note that if $A \in \mathcal{M}$ and $\mathcal{L}(A) = 0$, this implies that $\chi_A(x) = 0$ almost everywhere. Therefore also $f(x)\chi_A(x) = 0$ almost everywhere, which implies $\nu_f(A) = 0$.

□

Example 1.31. Let $f(x) = e^{-|x|^2}$. Then $f \in L^1(\mathbb{R}^n)$ and the measure ν_f is called the Gaussian measure. Note that it is a finite measure, and $\int_{\mathbb{R}^n} e^{-|x|^2}dx = \pi^{n/2}$, see [2, Prop. 2.53].

Theorem 1.32 (Lebesgue-Radon-Nikodym theorem)**.** *Let μ a Borelian measure on $\mathbb{R}^n$ which is $\sigma-$finite. Then there exist a unique $\nu << \mathcal{L}$ (absolutely continuous part) and a unique $\rho \perp \mathcal{L}$ (singular part) such that $\mu = \nu + \rho$.*

Moreover there exists $f \geqslant 0$, measurable and such that $\int_{B_R} f(x)dx < +\infty$ for all $R > 0$, for which $\nu = \nu_f$.
f is called the **density** *of ν, or the Radon-Nikodym derivative of ν and can be obtained (if the measure ν is regular) as $f(x) = \lim_{r\to 0} \frac{\nu(B(x,r))}{\mathcal{L}(B(x,r))}$.*

Proof. For the proof we refer to [2, Section 3.2]. □

1.7 Distributions of random variables

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $X : \Omega \to \mathbb{R}$ be a random variable (see Section 2.4). Then **the distribution** $\mathbb{P}_X$ of X is the Borel measure induced on $\mathbb{R}$ by X, defined as follows: for every $A \in \mathcal{B}(\mathbb{R})$,

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

The cumulative distribution function associated to such Borel measure is defined as

$$F_X(x) = \mathbb{P}(\{\omega \mid X(\omega) \leqslant x\}).$$

The distribution identifies the random variable, and often the random variables are described just in terms of their distributions.

Remark 1.33 (The cumulative distribution function)**.** If X is an (absolutely) continuous random variable, $\mathbb{P}_X$ is an absolutely continuous measure and F_X is an absolutely continuous function. The density of P_X with respect to the Lebesgue measure is

$$f_X(x) = F_X'(x) = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h} \qquad \text{for a.e.} x \in \mathbb{R}.$$

If X is a discrete random variable, $\mathbb{P}_X$ is a singular measure with respect to the Lebesgue measure and F_X is a monotone piecewise constant function.

More generally if F_X is the cumulative distribution function associated to a random variable, then F a right continuous, monotone increasing function, which we normalize to have $\inf F_X = 0$ (and obviously $\sup F = 1$). F_X has at most countably many discontinuity points, that are those for which $F(a) > \lim_{x\to a^-} F(x)$, or equivalently for which

$$\mathbb{P}(\{\omega \mid X(\omega) = a\}) > 0.$$

We define

$$F_X^d(x) := \sum_{y \leqslant x} \mathbb{P}(\{\omega \mid X(\omega) = a\}).$$

Note that F_d is a monotone increasing function, which is a.e. constant and has jumps only at discontinuity points of F_X.

So the function $F_X - F_X^d$ is a continuous function, and it is easy to check it is still monotone increasing. A deep result in mathematical analysis (see [2, Thm 3.23]) states that monotone increasing functions F are differentiable a.e.- that is for a.e. $a \in \mathbb{R}$ there exists $F'(a) = \lim_{h\to 0} \frac{F(a+h)-F(a)}{h}$ and moreover $F'(a) \geqslant 0$ a.e. So we define the absolutely continuous part of F_X as

$$F_X^{ac}(x) = \int_{-\infty}^{x} F_X'(y)dy = \int_{-\infty}^{x} (F_X - F_X^d)'(y)dy.$$

So, $F_X'(x)$ is the density of the absolutely continuous measure $\mu_{F_X^{ac}}$.

It is possible to prove that in general

$$F_X(x) = F_X^d(x) + F_X^{ac}(x) + F_X^s(x)$$

where F_X^s is a continuous and increasing function, whose derivative is zero in almost all x, but it can be not identically zero (a typical example is the devil's staircase function, or the Cantor function).

The three functions F_X^d, F_X^{ac}, F_X^s are all increasing, but are of very different nature:

- F_X^d can only increase by jumps and it is constants between two consecutive jumps,
- F_X^{ac} is a "nice" function with the property of being the integral of its derivative, which coincide with the distribution density,
- F_X^s is quite weird function, indeed quite hard to imagine (continuous, increasing with zero derivative a.e.).

We typically deal with real random variables such that the singular part F_X^s of their distribution function is identically zero.

Moreover, we see that a real random variable is discrete if and only if $F_X = F_X^d$ and it is absolutely continuous if and only if $F_X = F_X^{ac}$ and in this case $f_X(x) = F'_X(x)$.

Remark 1.34 (Joint distribution)**.** If X, Y are random variables on the same probability space, that is $X, Y : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$, we may define the joint cumulative distribution function as

$$F_{X,Y}(x,y) = \mathbb{P}(\{\omega \mid X(\omega) \leqslant x\} \cap \{\omega \mid Y(\omega) \leqslant y\}).$$

If X, Y are independent then $F_{X,Y}(x,y) = F_X(x)F_Y(y)$. Two random variables X and Y are jointly continuous if there exists a nonnegative function $f_{X,Y} : \mathbb{R}^2 \to \mathbb{R}$ such that for any measurable set $A \subseteq \mathbb{R}^2$ there holds

$$\mathbb{P}(\{\omega \mid (X(\omega), Y(\omega)) \in A\}) = \int_A f_{a,y}(x,y)dxdy.$$

The function $f_{X,Y}(x,y)$ is called the joint probability density function and is obtained as

$$f_{X,Y}(x,y) = \frac{d^2}{dxdy} F_{X,Y}(x,y) \qquad \text{a.e..}$$

Given the joint probability density function it is possible to recover the density functions of X and Y as the marginals:

$$f_X(x) = \int_{\infty}^{+\infty} f_{X,Y}(x,y)dy \qquad f_Y(y) = \int_{\infty}^{+\infty} f_{X,Y}(x,y)dx.$$

On the other hand, given the marginals f_X, f_Y, there is not a unique associated joint probability density function (apart from the case in which X, Y are independent, in which case $f_{X,Y}(x,y) = f_X(x)f_Y(y)$).

Remark 1.35. Some examples of widely used random variables/distributions:

- the Dirac measure δ_c centered at c is the distribution associated to the constant random variable c (so the random variable X such that $X(\omega) = c$ almost surely).
- the **gamma distribution** with parameters a, b is an absolutely continuous measure with density $f(x) = \Gamma(a)^{-1} b^a x^{a-1} e^{-bx} \chi_{(0,+\infty)}(x)$
- the **chi-square distribution** is a gamma distribution with parameters $n/2, 1/2$,
- the **normal or Gaussian distribution** with parameters μ, σ is an absolutely continuous random variable, with density $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma}}$,
- the **standard normal distribution** is a normal distribution with parameters $0, 1$, that is an absolutely continuous measure with density $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$,

- the **binomial distribution** of parameters n, p is a singular measure, and it is given by $\sum_{k=0}^{n} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \delta_k$ where δ_k is the Dirac measure centered at k,
- the **Poisson distribution** of parameter λ is a singular measure, and it is given by $e^{-\lambda} \sum_{k=0}^{+\infty} \frac{\lambda^k}{k!} \delta_k$ where δ_k is the Dirac measure centered at k.

Definition 1.36. *The nth-moment of a random variable X is given by $\mathbb{E}(X^n)$, more precisely*

- *if X is a (asbsolutely) continuous random variable (whose associated distribution has density f) then*
$$\mathbb{E}(X^n) = \int_{\mathbb{R}} x^n f(x) dx.$$
- *if X is a discrete random variable (taking values on $\mathbb{Z}$),*
$$\mathbb{E}(X^n) = \sum_{k \in \mathbb{Z}} k^n P(\omega \mid X(\omega) = k).$$

Note that $\mathbb{E}(X^n) < +\infty$ if and only if $\mathbb{E}(|X|^n) < +\infty$.

We recall that the moment for $n = 1$, that is $\mathbb{E}(X)$, is called the **mean**, *whereas $\mathbb{E}(X^2) - (\mathbb{E}(X))^2$ is called the* **variance**.

1.8 Problems

(i) Let $f : \mathbb{R} \to \mathbb{R}$ be a monotone function. Show that f is Lebesgue measurable.

(ii) Consider the right continuous increasing function on $\mathbb{R}$
$$F(x) = \begin{cases} x & x < 0 \\ x + 1 & x \geqslant 0. \end{cases}$$
Which is the Borel measure associated to this function?

2 L^p spaces and spaces of random variables with finite p-moment.

2.1 Banach spaces

Let X be a vectorial space on $\mathbb{R}$ (this means that it is closed by summation and by multiplication by scalars, that is if $x, y \in X$, $\lambda, \mu \in \mathbb{R}$, then $\lambda x + \mu y \in X$).

Definition 2.1. *A norm $\| \cdot \| : X \to [0, +\infty)$ is a function such that*

- *$\|x\| \geqslant 0$ for all $x \in X$ and $\|x\| = 0$ iff $x = 0$ (positivity);*
- *$\|\lambda x\| = |\lambda| \|x\|$ for all $x \in X, \lambda \in \mathbb{R}$ (homogeneity);*
- *$\|x + y\| \leqslant \|x\| + \|y\|$ (triangular inequality).*

$(X, \| \cdot \|)$ is a normed space.

Example 2.2. On $\mathbb{R}^n$ we may define the euclidean norm $|x| = \sqrt{x_1^2 + \cdots + |x_n|^2}$.

A norm induces on X a metric structure on X in the following way.

Definition 2.3 (Metric structure and notion of convergence). *Let $(X, \|\cdot\|)$ be a normed space. We define a distance between elements in X as*

$$d(x,y) = \|x-y\|.$$

Note that this is a good definition, since it is positive, zero only if $x = y$, and satisfies the triangular inequality, that is $d(x,z) \leqslant d(x,y) + d(y,z)$ for all x, y, z.

We define the balls associated to this distance as follows: we fix a center $x_0 \in X$ and a radius $r > 0$ and we set

$$B(x_0, r) = \{x \in X \mid \|x - x_0\| < r\}.$$

A set $A \subseteq X$ is **open** *if for all $x \in A$ there exists $r > 0$ such that $B(x,r) \subseteq A$. A set C is* **closed** *is $X \backslash C$ is open.*

Let $(x_n)_n$ a sequence of element in X and $x \in X$. Then

$$\lim_n x_n = x \qquad \textit{iff} \lim_{n\to+\infty} \|x_n - x\| = 0.$$

Proposition 2.4. *The following are equivalent:*

i) C is closed

ii) for every sequence (x_n) of elements in C such that there exists $x \in X$ with $\lim_n x_n = x$, there holds that $x \in C$.

Proof. Assume that C is closed and ii) is false. Then there exists (x_n) of elements in C such that $\lim_n x_n = x \notin C$. This implies that there exists $r > 0$ such that $B(x,r) \subseteq X \backslash C$. Therefore $x_n \notin B(x,r)$ for all n, which means that $\|x_n - x\| \geqslant r$ for all n, in contradiction with the fact that $\lim_n x_n = x$.

Assume that ii) holds and assume that C is not closed. So there exists $x \notin C$ such that for all $r > 0$ there holds that $B(x,r) \cap C \neq \varnothing$. Let $x_n \in C$ such that $x_n \in B(x, \frac{1}{n}) \cap C$. So $\|x_n - x\| < \frac{1}{n}$ and then $\lim_n x_n = x$. But this would imply $x \in C$. □

Definition 2.5 (Banach space).
A sequence $(x_n)_n$ in X is a Cauchy sequence if $\lim_{n,m} \|x_n - x_m\| = 0$.
A normed space is called a Banach space if all the Cauchy sequences have limit in X.

Remark 2.6. Note that if $(x_n)_n$ is a sequence which converge to $x \in X$, then it is also a Cauchy sequence, since by triangular inequality $\|x_n - x_m\| \leqslant \|x_n - x\| + \|x - x_m\|$ and then $0 \leqslant \lim_{n,m\to+\infty} \|x_n - x_m\| \leqslant \lim_{m,n\to+\infty} \|x_n - x\| + \|x - x_m\| = 0$.

The viceversa is not always true. Let's think e.g. of the case $X = \mathbb{Q}$ and the euclidean norm. Define (x_n) as follows: $x_0 = 1$, $x_1 = 1,01$, $x_2 = 1,01001$, $x_3 = 1,010010001$, $x_4 = 1,01001000100001$ and so on, that is $x_n = 1,1010010001\dots1\ 0\overset{n}{\dots}0\ 1$. It is easy to check that $x_n \in \mathbb{Q}$ for all n, that $x_n \to x$ (so $(x_n)_n$ is a Cauchy sequence, but this can also be checked directly) and that $x \notin \mathbb{Q}$. This implies that $(\mathbb{Q}, |\cdot|)$ is not a Banach space.

An important theorem in Banach spaces (more generally in complete metric spaces) is the **contraction lemma,** or **Banach-Caccioppoli theorem**:

Theorem 2.7. *Let $(X, \|\cdot\|)$ a Banach space and $F : X \to X$ such that there exists $0 < a < 1$ for which*

$$\|F(x) - F(y)\| \leqslant a\|x - y\| \qquad \forall x, y \in X.$$

(F is a contraction) Then the map F admits a unique fixed point, that is a point such that $\bar{x} = F(\bar{x})$.

Proof. See problem 1 at the end of the chapter. □

2.2 Bounded linear operators

Definition 2.8. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be two Banach space.*

A **linear operator** *is a map $T : X \to Y$ such that $T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$ for all $\alpha, \beta \in \mathbb{R}$, $x, y \in X$.*

A **bounded operator** *is a map $T : X \to Y$ such that*

$$\|T\| = \sup_{\{x \in X \|x\| \leqslant 1\}} \|Tx\| < +\infty.$$

If this quantity if finite, it is called the norm of T.

A **continuous operator** *is a map $T : X \to Y$ such that*

$$\lim Tx_n = Tx \qquad \text{for all sequences } x_n \text{ such that } \lim_n x_n = x.$$

Proposition 2.9. *A linear operator $T : X \to Y$ is continuous if and only if it is bounded.*

Proof. Assume that T is bounded, then

$$\|Tx_n - Tx\| = \|T(x_n - x)\| = \|x_n - x\| T\left(\frac{x_n - x}{\|x_n - x\|}\right) \leqslant \|x_n - x\|\|T\|.$$

Therefore if $\|x_n - x\| \to 0$, then also $\|Tx_n - Tx\| \to 0$.

Assume that T is continuous, and we want to prove that T is bounded. Assume by contradiction that it is not true. So for every $n \in \mathbb{N}$ there exists $x_n \in X$ such that $\|x_n\| = 1$ and $\|Tx_n\| \geqslant n$. Define $y_n = \frac{x_n}{n}$. Then $\|y_n\| = \frac{\|x_n\|}{n} = \frac{1}{n} \to 0$. This implies that $y_n \to 0$. Observe that by linearity $Ty_n = \frac{1}{n}Tx_n$ and then $\|Ty_n\| = \frac{1}{n}\|Tx_n\| \geqslant \frac{n}{n} = 1$. Therefore $y_n \to 0$ but $Ty_n \not\to 0$, in contradiction with continuity. □

Theorem 2.10. *The set of all bounded linear operators between two Banach spaces X, Y, is a Banach space $\mathcal{B}(X, Y)$, with norm $\|T\|$ as defined above.*

Proof. See [1, Theorem 2.12]. □

Example 2.11. Let $X = \mathbb{R}^n$ and $Y = \mathbb{R}^m$ both with the euclidean norm. Let $\mathbf{A} \in M_{m\times n}(\mathbb{R})$ be a $n \times m$ matrix. Then

$$Tx = \mathbf{A}x = (\sum_{j=1}^{n} a_{ij}x_j)_{i=1,\ldots,m}$$

is a bounded linear operator from $\mathbb{R}^n$ to $\mathbb{R}^m$.

Theorem 2.12 (Uniform boundedness principle, or Banach-Steinhaus theorem)**.** *Let T_n be a sequence of bounded linear operators between the Banach spaces X and Y, that is $T_n \in \mathcal{B}(X, Y)$ for all n. Assume that for all $x \in X$ there exists $C_x \in \mathbb{R}$ such that $\sup_n \|T_n x\| \leqslant C_x$.*

Then there exists $C \in \mathbb{R}$ such that $\|T_n\| \leqslant C$ for all n.

In particular this implies that if the sequence $T_n x$ is convergent for every $x \in X$, then $Tx := \lim_n T_n x$ defines a bounded linear operator.

Proof. See [1, Theorem 4.1]. □

2.3 L^p spaces

We consider the L^p spaces defined as follows

Definition 2.13 (L^p spaces)**.** *We define for $p \geqslant 1$,*

$$L^p(\mathbb{R}^n) = \left\{ f : \mathbb{R}^n \to \mathbb{R} \; |f \text{ is measurable and } \int_{\mathbb{R}^n} |f(x)|^p dx < +\infty \right\}.$$

Note that also functions in L^p which differ on sets of measure zero are identified.

If $A \in \mathcal{M}$, then we define

$$L^p(A) = \left\{ f : \mathbb{R}^n \to \mathbb{R} \mid f \text{ is measurable and } \int_{\mathbb{R}^n} |f(x)|^p \chi_A(x) = \int_A |f(x)|^p dx < +\infty \right\}.$$

We define

$$L^\infty(\mathbb{R}^n) = \{ f : \mathbb{R}^n \to \mathbb{R} \mid f \text{ is measurable and there exists } c > 0 \text{ such that } |f(x)| \leqslant c \text{ for almost every } x \}$$

and analogously

$$L^\infty(A) = \{ f : \mathbb{R}^n \to \mathbb{R} \mid f \text{ is measurable and there exists } c > 0 \text{ such that } |f(x)| \leqslant c \text{ for almost every } x \in A \}.$$

Definition 2.14. *Let $p > 1$. Then the conjugate exponent of p is the number $q > 1$ such that $1/p + 1/q = 1$. In particular the conjugate exponent of 2 is 2.*

We say that the conjugate exponent of 1 is $+\infty$.

Lemma 2.15 (Young inequality)**.** *Let p, q be conjugate exponents. Then $xy \leqslant x^p/p + y^q/q$ for all $x, y \geqslant 0$.*

Proof. Fix $x > 0$ and consider $\sup_{y \geqslant 0}(xy - y^q/q)$. First of all observe that the supremum is actually a maximum, since $\lim_{y \to +\infty} xy - y^q/q = -\infty$. Differentiating in y, we get that the unique point where the derivative is 0 is given by $y = x^{1/(q-1)}$. This is the maximum. Therefore for all $y \geqslant 0$, $xy - y^q/q \leqslant x^{1+1/(q-1)} - x^{q/(q-1)}/q = x^p/p$, since $p = q/(q-1)$. □

Theorem 2.16 (Holder inequality)**.** *Let $O \subseteq \mathbb{R}^n$ be an open set (it can also be $O = \mathbb{R}^n$), $p \in [1, +\infty]$ and q its conjugate exponent. Assume that $f \in L^p(O), g \in L^q(O)$. Then $f(x)g(x) \in L^1(O)$ and*

$$\int_O |f(x)g(x)| dx \leqslant \left(\int_O |f(x)|^p dx \right)^{1/p} \left(\int_O |g(x)|^q dx \right)^{1/q}.$$

Proof. Let $\tilde{f}(y) = |f(y)| \left(\int_O |f(x)|^p dx \right)^{-1/p}$ and $\tilde{g}(y) = |g(y)| \left(\int_O |g(x)|^q dx \right)^{-1/q}$. We apply the Young inequality to $\tilde{f}(y)$ and $\tilde{g}(y)$ and we get

$$|f(y)g(y)| \left(\int_O |f(x))^p| dx \right)^{-1/p} \left(\int_O |g(x))|^q dx \right)^{-1/q} \leqslant \frac{1}{p} \frac{|f(y)|^p}{\int_O |f(x)|^p dx} + \frac{1}{q} \frac{|g(y)|^q}{\int_O |g(x)|^q dx}$$

Integrating in O both sides we conclude

$$\frac{\int_O |f(x)g(x)| dx}{\left(\int_O |f(x))^p| dx \right)^{1/p} \left(\int_O |g(x))|^q dx \right)^{1/q}} \leqslant \frac{1}{p} + \frac{1}{q} = 1.$$

□

Corollary 2.17 (Minkowski inequality)**.** *Let $f, g \in L^p(O)$, then*

$$\left(\int_O |f(x) + g(x)|^p dx \right)^{\frac{1}{p}} \leqslant \left(\int_O |f(x)|^p dx \right)^{\frac{1}{p}} + \left(\int_O |g(x)|^p dx \right)^{\frac{1}{p}}.$$

Proof. For $p = 1, \infty$, the inequality is straightforward. We consider the case $p \in (1, +\infty)$. First of all, we observe that if $f, g \in L^p$ the $f + g \in L^p$. This is due to the fact that

$$\frac{|f(x) + g(x)|^p}{2^p} = \left| \frac{f(x)}{2} + \frac{g(x)}{2} \right|^p \leqslant \frac{|f(x)|^p}{2} + \frac{|g(x)|^p}{2}$$

by the convexity of the function $r \mapsto r^p$ on $[0, +\infty)$ when $p \geqslant 1$. Now we observe that

$$|f(x) + g(x)|^p = |f(x) + g(x)||f(x) + g(x)|^{p-1} \leqslant |f(x)||f(x) + g(x)|^{p-1} + |f(x)||f(x) + g(x)|^{p-1}$$

and that $|f(x)+g(x)|^{p-1} \in L^q$ where $q = \frac{p}{p-1}$ is the conjugate exponent of p. Moreover

$$\int_O \left(|f(x)+g(x)|^{p-1}\right)^q dx = \int_O |f(x)+g(x)|^p dx. \tag{2.1}$$

So by Holder inequality applied to f and $|f+g|^{p-1}$ we get

$$\int_O |f(x)||f(x)+g(x)|^{p-1} dx \leqslant \left(\int_O |f(x)|^p dx\right)^{\frac{1}{p}} \left(\int_O |f(x)+g(x)|^p dx\right)^{\frac{p-1}{p}}$$

and analogously by Holder inequality applied to f and $|f+g|^{p-1}$ we get

$$\int_O |g(x)||f(x)+g(x)|^{p-1} dx \leqslant \left(\int_O |g(x)|^p dx\right)^{\frac{1}{p}} \left(\int_O |f(x)+g(x)|^p dx\right)^{\frac{p-1}{p}}.$$

Integrating (2.1) and using the previous inequalities we get

$$\begin{aligned}\int_O |f(x)+g(x)|^p dx &\leqslant \left(\int_O |f(x)|^p dx\right)^{\frac{1}{p}} \left(\int_O |f(x)+g(x)|^p dx\right)^{\frac{p-1}{p}} \\ &\quad + \left(\int_O |g(x)|^p dx\right)^{\frac{1}{p}} \left(\int_O |f(x)+g(x)|^p dx\right)^{\frac{p-1}{p}} \\ &= \left(\int_O |f(x)+g(x)|^p dx\right)^{\frac{p-1}{p}} \left[\left(\int_O |f(x)|^p dx\right)^{\frac{1}{p}} + \left(\int_O |g(x)|^p dx\right)^{\frac{1}{p}}\right]\end{aligned}$$

from which we deduce the statement by dividing both sides by $\left(\int_O |f(x)+g(x)|^p dx\right)^{\frac{p-1}{p}}$. □

Theorem 2.18. *Let $p \geqslant 1$.*

The spaces $L^p(O)$ are Banach spaces, with norm given by $\|f\|_p = \left(\int_O |f(x))^p| dx\right)^{1/p}$.
The space $L^\infty(O)$ is a Banach space with norm given by $\|f\|_\infty = \inf\{c > 0 \; |\mathcal{L}\{x \mid f(x) \geqslant c\} = 0\}$.

Proof. Proving that L^p is a vectorial space is an easy task only if $p = 1, +\infty$, otherwise it is a consequence of Minkowski inequality (which also gives that $\|\cdot\|_p$ satisfies the triangular inequality), which is a consequence of the Holder inequality. For the proof see [2, Section 6.1]. □

A direct consequence of the Holder inequality is the following interpolation inequality.

Corollary 2.19 (Interpolation inequality)**.** *Let $O \subseteq \mathbb{R}^n$ be an open set (it can also be $O = \mathbb{R}^n$), $p, r \in [1, +\infty]$ such that $p < r$. Assume that $f \in L^p(O) \cap L^r(O)$. Then $f \in L^s(O)$ for every $s \in [p, r]$ and moreover*

$$\|f\|_s = \|f\|_p^\alpha \|f\|_r^{1-\alpha} \qquad \text{where } \alpha \in [0,1] \text{ is such that} \qquad \frac{1}{s} = \frac{\alpha}{p} + \frac{1-\alpha}{r}.$$

Proof. Take $s \in (p, r)$ and $\alpha \in (0,1)$ such that $\frac{1}{s} = \frac{\alpha}{p} + \frac{1-\alpha}{r}$. Since $1 = \frac{\alpha s}{p} + \frac{(1-\alpha)s}{r}$ we deduce that $\frac{p}{\alpha s} > 1$ and $\frac{r}{(1-\alpha)s} > 1$ are conjugate exponents.

Since $f \in L^p$, we get that $|f|^{\alpha s} \in L^{\frac{p}{\alpha s}}$ and moreover since $f \in L^r$ then $f^{(1-\alpha)s} \in L^{\frac{r}{(1-\alpha)s}}$. Therefore by the Holder inequality we get that $|f|^{\alpha s}|f|^{(1-\alpha)s} = |f|^s \in L^1$, which implies that $f \in L^s(O)$ and moreover

$$\|f\|_s^s = \int_O |f|^s dx \leqslant \left(\int_{\mathbb{R}} (|f|^{\alpha s})^{\frac{p}{\alpha s}} dx\right)^{\frac{\alpha s}{p}} \left(\int_{\mathbb{R}} (|f|^{(1-\alpha)s})^{\frac{r}{(1-\alpha)s}} dx\right)^{\frac{(1-\alpha)s}{r}} = \|f\|_p^{\alpha s} \|f\|_r^{(1-\alpha)s}.$$

□

Another consequence of the Holder inequality is the following:

Corollary 2.20. *Let O be an open set with $\mathcal{L}(O) < +\infty$. Then $L^p(O) \subseteq L^r(O)$ for every $1 \leqslant r \leqslant p$, and moreover $\|f\|_r \leqslant \|f\|_p \mathcal{L}(O)^{\frac{p-r}{pr}}$.*

Proof. Fix $p > 1$ and $f \in L^p(O)$. We want to prove that $f \in L^1(O)$. Note that since $\mathcal{L}(O) < +\infty$, then $\chi_O \in L^q(\mathbb{R}^n)$ for every q, so in particular it is in $L^q(O)$ for q conjugate exponent of p. By Holder inequality we get

$$\|f\|_1 \leqslant \|f\|_p \mathcal{L}(O)^{1/q} = \|f\|_p \mathcal{L}(O)^{\frac{p-1}{p}}$$

which give the conclusion of the theorem for $r = 1$. The case $r \in (1, p)$ is obtained just using the interpolation inequality, proved in the previous corollary: indeed $\frac{1}{r} = \frac{\alpha}{p} + 1 - \alpha$, with $\alpha = \frac{p(r-1)}{r(p-1)}$ and then

$$\|f\|_r \leqslant \|f\|_p^\alpha \|f\|_1^{1-\alpha} \leqslant \|f\|_p^\alpha \|f\|_p^{1-\alpha} \mathcal{L}(O)^{\frac{p-1}{p}(1-\alpha)} = \|f\|_p \mathcal{L}(O)^{\frac{p-r}{pr}}.$$

□

Finally we present an important example of linear bounded operators from L^p to $\mathbb{R}$.

Example 2.21. Let $g \in L^q(\mathbb{R}^n)$ with $q \geqslant 1$. Consider the following operator $T : L^p(\mathbb{R}^n) \to \mathbb{R}$, where p is the conjugate exponent of q, defined as

$$Tf = \int_{\mathbb{R}^n} f(x)g(x)dx.$$

It is immediate to check that it is linear. Moreover, by Holder inequality we get, for all $f \in L^p(\mathbb{R}^n)$ with $\|f\|_p \leqslant 1$,

$$|Tf| = \left| \int_{\mathbb{R}^n} f(x)g(x)dx \right| \leqslant \int_{\mathbb{R}^n} |f(x)g(x)|dx \leqslant \|f\|_p \|g\|_q \leqslant \|g\|_q.$$

Therefore T is a bounded operator, with norm $\|T\| \leqslant \|g\|_q$.

Define now $f_g(x) = |g(x)|^{q/p} \|g\|_q^{-q/p} \frac{g(x)}{|g(x)|}$. Then $f_g \in L^p(\mathbb{R}^n)$ and $\|f_g\|_p = \|g\|_q^{q/p} \|g\|_q^{-q/p} = 1$. We compute, recalling that $q/p + 1 = q$

$$Tf_g = \|g\|_q^{q/p} \int_{\mathbb{R}^n} |g(x)|^{q/p+1} dx = \|g\|_q^{q/p} \int_{\mathbb{R}^n} |g(x)|^q dx = \|g\|_q^{q-q/p} = \|g\|_q.$$

Therefore $\|T\| = \|g\|_q$.

2.4 Convergence in L^p spaces

Note that $f_n \to f$ in $L^p(O)$ means that $\lim_n \int_O |f_n(x) - f(x)|^p dx = 0$. Moreover $f_n \to f$ in L^∞ if $\lim_n \sup_O |f_n - f| = 0$.

Definition 2.22. *$f_n \to f$ almost everywhere if $\lim_n f_n(x) = f(x)$ for almost every x.*

It is not always true that convergence almost everywhere is sufficient for convergence in L^p as the following example shows.

Example 2.23. Let

$$f_n(x) = \begin{cases} 0 & x \geqslant \frac{1}{n} \\ n - nx & 0 \leqslant x \leqslant \frac{1}{n} \\ n + nx & -\frac{!}{n} \leqslant x \leqslant 0 \\ 0 & x \leqslant -\frac{1}{n}. \end{cases}$$

Then $\lim_n f_n(x) = 0$ for almost every x, but $\int_{\mathbb{R}} f_n(x)dx = 2 \neq 0$.

Theorem 2.24. *Let $p \geqslant 1$, $(f_n)_n, f \in L^p(O)$ and assume that $f_n \to f$ almost everywhere.*

If there exists $g \in L^p(O)$ such that $|f_n(x)| \leqslant g(x)$ for almost every x and every n then $\lim_n f_n = f$ in $L^p(O)$.

If $\lim_n f_n = f$ in $L^p(O)$, then up to passing to a subsequence $f_n \to f$ almost everywhere.

If $\lim_n f_n = f$ in $L^p(O)$, then $f_n \to f$ in measure.

Proof. The first part is the Lebesgue dominated convergence theorem, see [2, Theorem 2.24].

The second part is proven in [2, Corollary 2.32].

The third part is a consequence of the Chebycheff inequality (see Problem ii). Indeed

$$\forall \varepsilon > 0 \quad \mathcal{L}\left(\{x \in \mathbb{R}^n \mid |f_n(x) - f(x)| > \varepsilon\}\right)^{\frac{1}{p}} \leqslant \frac{1}{\varepsilon}\|f_n - f\|_p.$$

□

By the Holder inequality we can multiply functions in L^p by functions in L^q. This gives another notion of convergence.

Definition 2.25 (Weak convergence)**.** *Let $1 < p < +\infty$. Given $(f_n)_n, f \in L^p(O)$, we say that $f_n \rightharpoonup f$ (weakly) in $L^p(O)$ if for all $g \in L^q(O)$, with q the conjugate exponent of p, ($q = +\infty$ if $p = 1$) there holds*

$$\lim_n \int_O f_n(x)g(x)dx = \int_O f(x)g(x)dx.$$

Given $(f_n)_n, f \in L^\infty(O)$, we say that $f_n \rightharpoonup^\star f$ (weakly star) in $L^\infty(O)$ if for all function $g \in L^1(O)$, there holds

$$\lim_n \int_O f_n(x)g(x)dx = \int_O f(x)g(x)dx.$$

Given $(f_n)_n, f \in L^1(O)$, we say that $f_n \rightharpoonup f$ (weakly) in $L^1(O)$ if for all continuous and bounded functions $g : O \to \mathbb{R}$, there holds

$$\lim_n \int_O f_n(x)g(x)dx = \int_O f(x)g(x)dx.$$

Given $(f_n)_n, f \in L^1(O)$, we say that $f_n \rightharpoonup f$ (vaguely) in $L^1(O)$ if for all continuous functions $g : O \to \mathbb{R}$ such that $\lim_{x\to\partial O} g(x) = 0$, there holds

$$\lim_n \int_O f_n(x)g(x)dx = \int_O f(x)g(x)dx.$$

If f_n are densities of continuous random variables X_n, this convergence is also called **convergence in distribution** *of X_n.*

Proposition 2.26. *If f_n converge to f in L^p then it also converge weakly in L^p, whereas the viceversa is not true.*

Proof. The statement is a consequence of Holder inequality: let $f_n, f \in L^p$ and $g \in L^q$ (for $p > 1$) or g continuous and bounded (if $p = 1$), then

$$\left|\int_O (f_n(x) - f(x))g(x)dx\right| \leqslant \int_O |f_n(x) - f(x||g(x)|dx \leqslant \|f_n - f\|_p\|g\|_q.$$

Therefore if $\|f_n - f\|_p \to 0$, then $\int_O (f_n(x) - f(x))g(x)dx \to 0$. □

The main examples of sequence of functions which are converging weakly but not strongly are rapidly oscillating functions.

Example 2.27. [Weak convergence of periodic functions] Let $f(x)$ be a continuous periodic function (e.g, $f(x) = \sin x$) in $\mathbb{R}$ with period T.

Define $f_n(x) = f(nx)$ (note that this is a periodic function with period T/n, so as $n \to +\infty$ this is more and more oscillating).

Then for every $O \subseteq \mathbb{R}^n$ borelian bounded set

$$f_n \rightharpoonup \frac{1}{T}\int_0^T f(x)dx \qquad \text{in } L^p(O) \text{ for all } 1 \leqslant p < +\infty$$

and moreover

$$f_n \rightharpoonup^\star \frac{1}{T}\int_0^T f(x)dx \qquad \text{in } L^\infty(O).$$

For the proof see [1, Example 5.16].

Intuitively weak convergence is convergence of mean values.

2.5 Spaces of random variables with finite moments

We fix a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we consider the random variables $X : \Omega \to \mathbb{R}$. We introduce the spaces of random variables with finite p-moment (see definition in Section 2.7)

$$M^p = \{X \text{ random variable } \mathbb{E}(|X|^p) < +\infty\}$$

ans we $\|X\|_p = (\mathbb{E}(|X|^p))^{1/p}$.

First of all we have the following Holder inequality and Minkowski inequality

Proposition 2.28. *Let $X \in M^p$ and $Y \in M^q$, with q conjugate exponent of p, then*

$$\mathbb{E}(|XY|) \leqslant \mathbb{E}(|X|^p))^{1/p}(\mathbb{E}(|Y|^q))^{1/q}.$$

Moreover if $X, Z \in M^p$, then

$$\mathbb{E}(|X + Z|^p))^{1/p} \leqslant \mathbb{E}(|X|^p))^{1/p} + \mathbb{E}(|Z|^p))^{1/p}$$

Proof. It is sufficient to apply the Young inequality to $|X|\mathbb{E}(|X|^p))^{-1/p}$ and to $|Y|\mathbb{E}(|Y|^q))^{-1/q}$ and proceed as for L^p spaces (instead of integrating in $\mathcal{O}$ one needs to take the average of both side of the inequality). The proof of Minkowski also follows again as for the case of L^p spaces. □

Theorem 2.29. *The space M^p with the norm $\|X\|_p$ for $p \in [1, +\infty)$ is a Banach space.*

Similarly as for $L^p(O)$ spaces, where $\mathcal{L}(O) < +\infty$ (see Corollary 2.20), the spaces M^p are decreasing as we will show. First of all we recall the Jensen inequality:

Lemma 2.30 (Jensen's inequality)**.** *Let $g : \mathbb{R} \to \mathbb{R}$ be a convex function, then for every random variable X*

$$\mathbb{E}(g(X)) \geqslant g(\mathbb{E}(X)).$$

Theorem 2.31. *There holds that $M^k \subseteq M^n$ for every $1 \leqslant n \leqslant k$. Moreover if $X \in M^k$ then $(\mathbb{E}(|X|^n))^{\frac{1}{n}} \leqslant (\mathbb{E}(|X|^k))^{\frac{1}{k}}$ for all $n \leqslant k$.*

Proof. Let $1 \leqslant n \leqslant k$, $g(x) = |x|^{\frac{k}{n}}$. Since $\frac{k}{n} \geqslant 1$, the function g is convex. Let $X \in M^k$ and we apply Jensen's inequality to the random variable $|X|^n$, observing that $g(|X|^n) = |X|^k$,

$$\mathbb{E}(|X|^k) = \mathbb{E}(g(|X|^n)) \geqslant g(\mathbb{E}(|X|^n)) = (\mathbb{E}(|X|^n))^{\frac{k}{n}}.$$

□

Example 2.32. $T : M^k \to \mathbb{R}$ such that $T(X) = \mathbb{E}(X)$ is a bounded linear operator.

If we consider $X \in M^2$, then $T_X : M^2 \to \mathbb{R}$ defined as $T_X(y) = \mathbb{E}(XY)$ is again a bounded linear operator.

2.6 Modes of convergence for random variables

Analogously to the case of measurable functions we have several notion of convergence in the space of random variables (and in the associated space of distributions).

Definition 2.33. *Let X_n be a sequence of real randos variables.*

- $X_n \to X$ *in probability if for every* $\varepsilon > 0$, $\lim_n \mathbb{P}(|X_n - X| > \varepsilon) = 0$.
- $X_n \to X$ *in* M^1 *if* $\mathbb{E}(|X_n - X|) \to 0$, *that is* **convergence is mean** *and* $X_n \to X$ *in* M^2 M^2 *if* $\mathbb{E}((X_n - X)^2) \to 0$, *that is* **convergence is the mean square convergence**.
- $X_n \to X$ *in distribution if* $\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X))$ *for every bounded continuous function* g. *Note that if* X_n, X *are absolutely continous random variables with associated densities* f_n, f, *then this is equivalent to say that* f_n *converges vaguely in* L^1 *to* f.

Theorem 2.34 (Prokhorov's theorem)**.** *Let X_n be a sequence of random variables which are* **tight** *in the following sense: for every $\varepsilon > 0$ there exist $n_\varepsilon > 0$ and a compact set K_ε (so a bounded closed set) such that $\mathbb{P}\{\omega, X_n(\omega) \in K_\varepsilon\} \geqslant 1 - \varepsilon$ for all $n \geqslant n_\varepsilon$. Then, there exists a random variable X such that, up to a subsequence, $X_n \to X$ in distribution.*

2.7 Problems

(i) Let $(X, \|\cdot\|)$ a Banach space and $F : X \to X$ such that there exists $0 < a < 1$ for which

$$\|F(x) - F(y)\| \leqslant a\|x - y\| \qquad \forall x, y \in X.$$

(F is a contraction)

(a) Show that the map F is continuous.

(b) Let $x_0 \in X$. Define $x_1 = F(x_0)$, $x_2 = F(x_1)$ and so on $x_n = F(x_{n-1})$. Prove that

$$\|x_n - x_{n+1}\| \leqslant a^n \|x_0 - x_1\|.$$

Deduce that $(x_n)_n$ is a Cauchy sequence.

(c) Let $\bar{x} = \lim_n x_n$, where (x_n) has been defined in the previous step. Show that $F(\bar{x}) = \bar{x}$. So, $\bar{x}$ is a fixed point of F.

(d) Show that the map F admits a unique fixed point, that is a point such that $\bar{x} = F(\bar{x})$.

This is called **Banach-Caccioppoli theorem**.

(ii) Let $f \in L^p(\mathbb{R}^n)$ and $\alpha > 0$. Prove that

$$\mathcal{L}\left(\{x \in \mathbb{R}^n \mid |f(x)| > \alpha\}\right)^{\frac{1}{p}} \leqslant \frac{1}{\alpha}\|f\|_p.$$

This is called **Chebycheff inequality**.

(iii) Prove that if $f \in L^2(-1,1)$ then $f \in L^1(-1,1)$ and moreover

$$\|f\|_1 \leqslant \sqrt{2}\|f\|_2.$$

Provide an example of a function $f \in L^1(-1,1)$ such that $f \notin L^2(-1,1)$.

(iv) Consider the following operator $T : L^2(0,2) \to L^2(0,2)$ defined as

$$Tf(x) = \int_0^x f(y)dy.$$

Show that this is a bounded continuous operator.

Hint Recall the **Jensen inequality**:

$$\left(\frac{1}{b-a}\int_a^b f(x)dx\right)^2 \leqslant \frac{1}{b-a}\int_a^b f(x)^2 dx.$$

3 Hilbert spaces

3.1 Hilbert spaces

Hilbert spaces are spaces where it is possible to define the notions of length and orthogonality, which allow to work with the elements geometrically, as if they were vectors in Euclidean space. First of all we recalls some basic definitions.

Definition 3.1. *A set X is a vector space on $\mathbb{R}$ (a real vector space) if it is a set equipped with two operations, vector addition (which allows to add two vectors $x, y \in X$ to obtain another vector $x + y \in X$) and scalar multiplication (which allows us to "scale" a vector $x \in X$ by a real number c to obtain a vector $cx \in X$). Moreover we require that X contains a neutral element for the vector addiction, that is an element $0 \in X$ such that $0 + x = x$ for every $x \in X$ and $x - x = 0$.*

A **scalar product** *on X is a function $(\cdot, \cdot) : X \times X \to \mathbb{R}$ such that*

- *$(x, x) \geqslant 0$ for all x and $(x, x) = 0$ iff $x = 0$;*
- *it is symmetric $(x, y) = (y, x)$ for all $x, y \in X$;*
- *it is linear, that is $(\alpha x + \beta y, z) = \alpha(x, z) + \beta(y, z)$ for all $x, y, z \in X, \alpha, \beta \in \mathbb{R}$.*

We associate to a scalar product a norm in this way $\|x\| = \sqrt{(x, x)}$.

Proposition 3.2. *The function $\|\cdot\| : X \to [0, +\infty)$ defined as $\|x\| = \sqrt{(x, x)}$ is a norm. Moreover the scalar product is continuous, that is if $x_n \to x$ in X and $y \in X$, then $(x_n, y) \to (x, y)$ in $\mathbb{R}$.*

Proof. Positivity and homogeneity are obvious. To prove the triangle inequality one first need to to prove the Cauchy Schwartz inequality $|(x, y)| \leqslant \|x\|\|y\|$. See [1, Theorem 5.1].

The continuity is an easy consequence of the Cauchy Schwartz inequality:

$$|(x_n - x, y)| \leqslant \|x_n - x\|\|y\|.$$

□

Definition 3.3 (Hilbert space)**.** *A space X with a scalar product which induces on X a norm such that X is a Banach space is called Hilbert space.*

Proposition 3.4 (Parallelogram identity)**.** *For every $x, y \in H$, there holds*

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

Proof. By definition and by linearity and symmetry of the scalar product $\|x+y\|^2 = (x+y, x+y) = (x, x) + 2(x, y) + (y, y) = \|x\|^2 + 2(x, y) + \|y\|^2$, and $\|x - y\|^2 = (x + y, x + y) = \|x\|^2 - 2(x, y) + \|y\|^2$. It is sufficient to sum. □

Example 3.5. In $\mathbb{R}^n$ we define the scalar product $(x, y) = x_1y_1 + x_2y_2 + \cdots + x_ny_n$. The euclidean norm is the norm associated to this scalar product. So $\mathbb{R}^n$ with this scalar product is a Hilbert space. This is the basic example of Hilbert space of finite dimension.

3.2 Orthogonality and projections in Hilbert spaces

Definition 3.6 (Orthogonal space)**.** *We say that $x, y \in X$ are orthogonal if $(x, y) = 0$.*

If $S \subseteq X$ is a subset of X, we define the orthogonal subspace

$$S^\perp = \{x \in X \mid (x, s) = 0 \ \forall s \in S\}.$$

This a vectorial subspace of X.

Example 3.7. *If we consider $S \subset M^2$ the subspace of constant random variables, then $S^\perp = \{X \in M^2 \ |\mathbb{E}(X) = 0\}$.*

Theorem 3.8 (Orthogonal projection)**.** *Let $V \subseteq H$ be a closed subspace of a Hilbert space, $V \neq \{0\}$ and let $h \in H$.*

Then there exists a unique element $v \in V$ at minimal distance from h, that is such that $\|h - v\| = \min_{w\in V} \|h - w\|$. Moreover there exists a unique element $s \in V^{\perp}$ such that $h = v + s$.

The map $Pr_V : H \to V$ which associate $h \to v$ is called the orthogonal projection of H in V and it is a bounded linear operator of norm 1.

Proof. We consider the minimization problem $\min_{w\in V} \|h - w\|$ and we show that it admits a solution which is unique. Since $\|h - w\| \geqslant 0$ we get that $\inf_{w\in V} \|h - w\| = \delta \geqslant 0$. Let $v_n \in V$ such that $\delta \leqslant \|v_n - h\| \leqslant \delta + 1/n$. Then $(v_n)_n$ is a Cauchy sequence, since by parallelogram identity and linearity

$$\|v_n-v_m\|^2 = 2\|v_n-h\|^2+2\|v_m-h\|^2-\|(v_n+v_m)-2h\|^2 \leqslant 2(\delta+1/n)^2+2(\delta+1/m)^2-4\|h-(v_n+v_m)/2\|^2.$$

We conclude by recalling that since $(v_n + v_m)/2 \in V$ then $\|h - (v_n + v_m)/2\| \geqslant \delta$,

$$\|v_n-v_m\|^2 \leqslant 2(\delta+1/n)^2+2(\delta+1/m)^2-4\delta^2 = 4\delta/n+4\delta/m+1/n^2+1/m^2 \to 0 \qquad \text{as } n, m \to +\infty.$$

Since H is a Banach space there exists $v \in H$ such that $\lim_n v_n = v$ and since V is closed then $v \in V$. By continuity, we conclude that $\|v - h\| = \delta = \inf_{w\in V} \|h - w\|$. v is the unique minimizer. Indeed if it were not the case, there would exists $v' \in V$ with $\|v - h\| = \|v' - h\| = \delta$. By parallelogram identity

$$\|v - v'\|^2 = 2\|v - h\|^2 + 2\|v' - h\|^2 - 4\|(v + v')/2 - h\|^2 \leqslant 2\delta^2 + 2\delta^2 - 4\delta^2 = 0$$

which implies $\|v - v'\| = 0$.

Let $w \in V$. We claim that $(h - v, w) = 0$. Since v is the point at minimum distance, then the function $\lambda \to \|h - v + \lambda w\|^2$ has minimum in $\lambda = 0$. Differentiating the function in λ it should be that the derivative in 0 is 0. $\frac{\|h-v+\lambda w\|^2}{d\lambda} = \frac{(h-v+\lambda w,h-v+\lambda w)}{d\lambda} = 2(h - v, w)$. Therefore $(h - v, w) = 0$. This means that $h - v \in V^{\perp}$.

Let $v = Pr_V(h), v' = Pr_V(h')$ and let $\alpha, \beta \in \mathbb{R}$. Then $\alpha v + \beta v' \in V$ and $\alpha v + \beta v' - \alpha h - \beta h' \in V^{\perp}$. Therefore by uniqueness $Pr_V(\alpha h + \beta h') = \alpha v + \beta v'$. Then Pr_V is linear. Moreover since $(Pr_V h - h, Pr_V h) = 0$,

$$\|h\|^2 = \|h - Pr_V h + Pr_V h\|^2 = (h - Pr_V h + Pr_V h, h - Pr_V h + Pr_V h) = \|h - Pr_V h\|^2 + \|Pr_V h\|^2.$$

This implies that for all h with $\|h\| \leqslant 1$, $\|Pr_V h\|^2 = \|h\|^2 - \|h - Pr_V h\|^2 \leqslant 1$. So Pr_V is bounded. Moreover if $h \in V$, then $Pr_V h = h$. Therefore $\|Pr_V\| = 1$.

□

Definition 3.9 (Orthonormal set)**.** *A set $\{u_i, i \in I\}$ of elements in H is an orthonormal set if $\|u_i\| = 1$ for all i and $(u_i, u_j) = 0$ for all $i \neq j$.*

Proposition 3.10. *Let $\{u_i, i \in I\}$ be a orthonormal set. Then the following are equivalent*

- *if $(x, u_i) = 0$ for all i, then $x = 0$*
- *$\|x\|^2 = \sum_i |(x, u_i)|^2$ for all $x \in H$,*
- *for all $x \in H$, $x = \sum_i (x, u_i)u_i$, (where the convergence is with respect to the norm of H).*

An orthonormal set for which one of the previous conditions hold is called an orthonormal basis. Every Hilbert space admits a orthonormal basis.

Proof. See [2, Proposition 5.28]. □

Definition 3.11 (Separable space)**.** *H is separable if it admits a countable orthonormal basis.*

Theorem 3.12 (Computation of the orthogonal projection)**.** *Let V be a closed subspace of H and let $\{v_i, i \in I\}$ be an orthonormal basis of V. Then for all $h \in H$,*

$$Pr_V(h) = \sum_{i \in I} (h, v_i) v_i.$$

Proof. See [1, Theorem 5.10]. □

Theorem 3.13 (Parseval theorem)**.** *Let $\{u_i, i \in I\}$ be a countable orthonormal set in H. The following are equivalent*

- *if $(h, u_i) = 0$ for all i then $h = 0$,*
- *for each $h \in H$ there holds $h = \sum_i (h, u_i) u_i$, which means that $\lim_n \|h - \sum_{i=1}^n (h, u_i) u_i\| = 0$,*
- *for each $h \in H$, $\|h\|^2 = \sum_i |(h, u_i)|^2$.*

In particular $\{u_i, i \in I\}$ is an orthonormal basis of H.

3.3 Hilbert space of random variables and conditional expectation

We fix a probability space $(\Omega, \mathbb{P}, \mathcal{F})$ and we define the space

$$M^2 = \{X : (\Omega, \mathbb{P}, \mathcal{F}) \to \mathbb{R} \mid \ X \text{ random variable with } \mathbb{E}(X^2) < +\infty\}.$$

Recall that X is a random variable if $X^{-1}(A) \in \mathcal{F}$ for every $A \in \mathcal{B}$ (so for every A in the $\sigma-$algebra of Borel sets. Given X random variable, we define $\sigma(X) \subseteq \mathcal{F}$, that is **the σ-algebra generated by** X, as the minimal σ- algebra contained in $\mathcal{F}$ which contains all the elements $X^{-1}(A) = \{\omega \in \Omega \mid X(\omega) \in A\}$ for every $A \in \mathcal{B}$. So it is the minimal σ-algebra which assures that X is measurable.

Note that if X is a constant random variable, so $X(\omega) = c$ for all $\omega \in \Omega$, then $X^{-1}(A) = \Omega$ if $c \in A$, and $X^{-1}(A) = \emptyset$ if $c \notin A$. So in this case $\sigma(X) = \{\emptyset, \Omega\}$, which is the minimal possible σ-algebra.

We define on M^2 the scalar product

$$(X, Y) = \mathbb{E}(XY)$$

and the induced norm is

$$\|X\| = \sqrt{\mathbb{E}(X^2)}.$$

It is possible to prove that M^2 with this norm and this scalar product is a Hilbert space. Observe that, as we did for L^p spaces, we are actually considering class of equivalence of random variables, since we are identifying two random variables X, Y such that $\mathbb{P}(\omega \mid X(\omega) = Y(\omega)) = 1$.

We consider a σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, and consider the probability space $(\Omega, \mathbb{P}, \mathcal{G})$. On this space we may define the space

$$M^2_{\mathcal{G}} = \{X : (\Omega, \mathbb{P}, \mathcal{G}) \to \mathbb{R} \mid \ X \text{ random variable with } \mathbb{E}(X^2) < +\infty\}.$$

Note that $M^2_{\mathcal{G}}$ is a closed subspace of M^2.

Definition 3.14 (Conditional expectation)**.** *We define the* **conditional expectation of X given** *$\mathcal{G}$ as the orthogonal projection of $X \in M^2$ in the space $M^2_{\mathcal{G}}$ as defined and characterized in Theorem 3.8 that is*

$$\mathbb{E}(X|\mathcal{G}) = Pr_{M^2_{\mathcal{G}}}(X),$$

or equivalently $\mathbb{E}(X|\mathcal{G})$ is the unique random variable in $M^2_{\mathcal{G}}$ such that

$$\mathbb{E}(X - \mathbb{E}(X|\mathcal{G}))^2 = \min_{Z \in M^2_{\mathcal{G}}} \mathbb{E}(X - Z)^2.$$

In particular $\mathbb{E}(X|\mathcal{G})$ is the minimum mean squared predictor of X based on the information contained in $\mathcal{G}$.

Note that $X - \mathbb{E}(X|\mathcal{G})$ is orthogonal to every element of $M^2_{\mathcal{G}}$ that is

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(X|\mathcal{G})Y) \qquad \forall\ Y \in M^2_{\mathcal{G}}.$$

In particular, since constant random variables are in $M^2_{\mathcal{G}}$ for every $\mathcal{G}$, we get $\mathbb{E}(X) = \mathbb{E}(\mathbb{E}(X|\mathcal{G}))$.

Remark 3.15 (Conditioning with respect to a random variable X)**.** A particular case of the previous definition is the following. Let us consider a random variable $X \in M^2$, and let $\mathcal{G} = \sigma(X)$ as before. It is possible to show that in this case every $\mathcal{G}$ measurable random variable is a Borel function of X, which means that

$$M^2_{\mathcal{G}} := \{h(X),\ \text{for } h : \mathbb{R} \to \mathbb{R}, \text{borelian function}\}.$$

$h : \mathbb{R} \to \mathbb{R}$ is a Borel function if for all borelian set $B \subseteq \mathcal{B}(\mathbb{R})$, the set $h^{-1}(B) := \{x \in \mathbb{R}\ h(x) \in B\}$ is in the Borel σ-algebra (Note that this condition is slightly stronger than asking that h is measurable, since measurable functions satisfies $h^{-1}(B) := \{x \in \mathbb{R}\ h(x) \in B\} \in \mathcal{M}$, that is are elements of the σ-algebra of measurable sets (given by sets which differs from Borel sets by subsets of sets of zero Lebesgue measure).

In this case $\mathbb{E}(Y|\sigma(X)) = \mathbb{E}(Y|X)$ is the best predictor of Y given X. In particular $\mathbb{E}(Y|X)$ the unique Borel function $h(X)$ which minimizes $\mathbb{E}(Y - h(X))^2$:

$$\mathbb{E}[(Y - \mathbb{E}(Y|X))^2] = \mathbb{E}[(Y - h(X))^2] = \min_{f:\mathbb{R}\to\mathbb{R},\text{borelian}} \mathbb{E}[(Y - f(X))^2]$$

and moreover

$$\mathbb{E}(Yf(X)) = \mathbb{E}(h(X)f(X)) \qquad \forall f : \mathbb{R} \to \mathbb{R}.\ \text{borelian}.$$

Note that solving this minimization problem can be very difficult, so in general we consider a reduced problem, adding some conditions on the functions f on which we are minimizing.

The simplest case is the case in which we consider the minimization problem among *linear functions*: that is

$$\min_{f:\mathbb{R}\to\mathbb{R},\text{linear}} \mathbb{E}[(Y - f(X))^2].$$

$h : \mathbb{R} \to \mathbb{R}$ is linear if and only if there exists $a, b \in \mathbb{R}$ such that $h(r) = ar + b$. So the problem reduced to a finite dimensional problem: given $X \in M^2$ we want to find for all Y, $a, b \in \mathbb{R}$ for which it is minimal $\mathbb{E}((Y - a - bX)^2)$. So, the **linear least square estimator** is given by

$$L(Y|X) = a + bX,$$

where a, b are the optimal values which minimize $\mathbb{E}((Y - a - bX)^2)$. This problem can be restated exactly as a projection problem: we define S as the space generated by $X, 1$ in M^2, that is $S = \{Z = aX + b \in M^2, a \in \mathbb{R}, b \in \mathbb{R}\}$ and we want to find $Pr_S(Y)$.

In order to solve the problem, first of all we choose an orthonormal basis of S. A basis of S is given by $\{1, X\}$. Observe that if $\mathbb{E}(X) = (X, 1) \neq 0$, we have that X and 1 are not orthogonal, so we substitute X with the element $X - \mathbb{E}(X)$ which is orthogonal to 1. Moreover we have to normalize this element by choosing $c \in \mathbb{R}$ such that $c^2\mathbb{E}(X - \mathbb{E}(X))^2 = 1$. Since $\mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = Var(X)$, it is sufficient to choose $c = \sqrt{VarX}$. Therefore an orthonormal basis of S is given by $1, \frac{X-\mathbb{E}(X)}{\sqrt{Var(X)}}$. Recalling Theorem 3.12, we get

$$Pr_S(Y) = (Y, 1)1 + \left(Y, \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}}\right) \frac{X - \mathbb{E}(X)}{\sqrt{Var(X)}}.$$

So the linear least square estimator coincides with

$$L(Y|X) = \mathbb{E}(Y) + \frac{Cov(X, Y)}{Var(X)}(X - \mathbb{E}(X)).$$

Finally we compute the average error

$$\mathbb{E}(Y - L(Y|X))^2 = Var(Y) + \frac{Cov^2(X,Y)}{Var^2(X)} VarX - 2\frac{Cov(X,Y)}{Var(X)} Cov(X,Y)$$
$$= VarY - \frac{Cov^2(X,Y)}{Var(X)} = \frac{Var(Y)Var(X) - Cov^2(XY)}{Var(X)}.$$

In general the best linear predictor is different from the general minimum mean squared predictor. Let $Y = X^2 + Z$ with X, Z independent and both normals with mean 0 and variance 1. Then $\mathbb{E}(Y|X) = X^2$, whereas $L(Y|X) = 1$ (check it!).

Remark 3.16 (Conditioning with respect to a constant random variable)**.** A very simple case to compute $\mathbb{E}(Y|\sigma(X)) = \mathbb{E}(Y|X)$ is the case in which $X \equiv k$ (that is X is constant). In this case $\sigma(X) = \{\emptyset, \Omega)$ and the space

$$M^2_{\mathcal{G}} := \{\text{constant random variables}\}.$$

So, $\mathbb{E}(Y|X)$ is the unique constant c such that

$$\mathbb{E}[(Y-c)^2] = \min_{\lambda \in \mathbb{R}} \mathbb{E}[(Y-\lambda)^2]$$

and moreover

$$\lambda \mathbb{E}(Y) = \mathbb{E}(Y\lambda) = \mathbb{E}(c\lambda) = c\lambda \qquad \forall \lambda \in \mathbb{R}.$$

It is immediate to verify that $c = \mathbb{E}(Y|\mathcal{G}) = \mathbb{E}(Y)$. Another simple case is the case in which $X = \chi_A$, for some $A \in \mathcal{F}$ which means that $\chi_A(\omega) = 1$ if $\omega \in A$ and $\chi_A(\omega) = 0$ if $\omega \notin A$. It is simple to see that in this case $\sigma(\chi_A) = \{\emptyset, \Omega, A, \Omega\backslash A\}$. In this case

$$M^2_{\mathcal{G}} := \{a\chi_A + b\chi_{\Omega\backslash A} = (a-b)\chi_A + b \qquad a, b \in \mathbb{R}\}.$$

So, $\mathbb{E}(Y|A)$ is obtained by solving the finite dimensional minimization problem

$$\min_{a,b \in \mathbb{R}} \mathbb{E}[(Y - a\chi_A - b)^2].$$

Since $M^2_{\mathcal{G}}$ is a finite dimensional space (of dimension 2), we compute a orthonormal basis of it. We start from the basis given by $\{1, \chi_A\}$ and we orthonormalize it by Gram-Schmidt procedure. Let $X_1 = 1$ and $X_2 = \frac{\chi_A - \mathbb{P}(A)}{\sqrt{\mathbb{P}(A)(1-\mathbb{P}(A))}}$. Note that $\mathbb{E}|X_1|^2 = 1 = \mathbb{E}|X_2|^2$ and moreover $\mathbb{E}(X_1X_2) = 0$. Therefore by Theorem 3.12 we deduce that

$$\mathbb{E}(Y|A) = \mathbb{E}(YX_1)X_1 + \mathbb{E}(YX_2)X_2 = \mathbb{E}(Y) + \frac{\mathbb{E}(Y\chi_A)}{\mathbb{P}(A)(1-\mathbb{P}(A))}\chi_A - \mathbb{E}(Y)\frac{\mathbb{P}(A)}{\mathbb{P}(A)(1-\mathbb{P}(A))} =$$

$$= \frac{\mathbb{E}(Y\chi_A)}{\mathbb{P}(A)(1-\mathbb{P}(A))}\chi_A - \frac{\mathbb{P}(A)}{1-\mathbb{P}(A)}\mathbb{E}(Y).$$

3.4 Bounded linear operators in Hilbert spaces

Let H be a Hilbert space. We consider linear bounded operators $T : H \to H$.

Definition 3.17 (Adjoint of an operator)**.** *Let $T : H \to H$ be a bounded linear operator. The adjoint of T is the operator $T^* : H \to H$ such that $(Th, k) = (h, T^*k)$ for all $h, k \in H$. T is symmetric if $T = T^*$.*

Proposition 3.18. *Let T be a linear bounded symmetric operator. Then $\|T\| = \sup_{\|x\|=1} |(Tx, x)|$.*

Proof. By Cauchy Schwartz inequality we get

$$|(Tx, x)| \leqslant \|Tx\|\|x\| \leqslant \|T\|\|x\|^2 = \|T\|.$$

On the other hand take $x \in H$ with $\|x\| = 1$ and $Tx \neq 0$ and define $y = Tx/\|Tx\|$. Then $\|y\| = 1$ and by symmetry and linearity of the operator

$$(Ty, x) = \frac{1}{\|Tx\|}(T(Tx), x) = \frac{1}{\|Tx\|}(Tx, Tx) = \|Tx\|.$$

A simple computation gives that $4(Ty, x) = (T(x + y), x + y) - (T(x - y), x - y)$, and then we get

$$4\|Tx\| = 4|(Ty, x) = (T(x + y), x + y) - (T(x - y), x - y) \leqslant \sup_{\|z\|=1} |(Tz, z)|(\|x + y\|^2 + \|x - y\|^2)$$

$$= \sup_{\|z\|=1} |(Tz, z)|(2\|x\|^2 + 2\|y\|^2)$$

where at the end we used the parallelogram identity. So we deduce, recalling that $\|x\| = 1 = \|y\|$ that $\|Tx\| \leqslant \sup_{\|z\|=1} |(Tz, z)|$. This gives the conclusion taking the supremum with respect to x. □

Definition 3.19 (Compact operators)**.** *Let $T : H \to H$ be a linear bounded operator. T is compact if for every bounded sequence $(h_n)_n$, there exists a subsequence such that $(Th_n)_n$ has a limit, that is $\lim_n Th_n = v$.*

Equivalently (it has to be proved though), an operator is compact if for every sequence $h_n \rightharpoonup h$ (h_n is weakly converging to h), there holds that $\lim_n \|Th_n - Th\| = 0$, so Th_n converge strongly to Th.

Definition 3.20 (Point spectrum (eigenvalues) of an operator)**.** *The* **point spectrum** *$\sigma_p(T)$ of a operator is given by the* **eigenvalues** *of T, that is by the elements $\lambda \in \mathbb{R}$ such that there exists $v \in H$ (called eigenvector) for which $Tv = \lambda v$:*

$$\sigma_p(T) := \{\lambda \in \mathbb{R} \mid \exists v \in H, Tv = \lambda v\}.$$

Given $\lambda \in \sigma_p(T)$, every element $v \in H$ such that $Tv = \lambda v$ is called **eigenvector** *relative to the eigenvalue λ.*

The **kernel** *of an operator is the subspace N of H composed by vectors $h \in H$ such that $Th = 0$ (N is the space of eigenvectors relative to the eigenvalue 0).*

Theorem 3.21 (Spectral theorem for compact symmetric operators)**.** *Let T be a symmetric compact operator on H, separable Hilbert space.*

If H is infinite dimensional, then T is not invertible (even if the kernel of T can be 0).

There exist $x_k \in H$, $\lambda_k \in \mathbb{R}$, such that $\{x_k\}$ is an orthonormal basis of H and $Tx_k = \lambda_k x_k$ (that is λ_k are eigenvalues of T with associated eigenvectors x_k) and the space $\{x \in H \; Tx = \lambda_k x\}$ for every $\lambda_k \neq 0$ has finite dimension (so the multiplicity of every non zero eigenvalue is finite). Moreover the set of all the eigenvalues $\{\lambda_k\}$ of T is either finite or countable, and in this case $\lim_k \lambda_k = 0$:

$$\text{either } \sigma_p(T) = \{\lambda_1, \dots, \lambda_N\} \qquad \text{or } \sigma_p(T) = \{\lambda_k,\ k \in \mathbb{N}\} \text{ and } \lim_{k\to+\infty} \lambda_k = 0.$$

Finally, let

$$m = \inf_{\{h\in H\|h\|=1\}} (Th, h) \qquad M = \sup_{\{h\in H\|h\|=1\}} (Th, h).$$

Then $m, M \in \sigma_p(T)$ and $\sigma_p(T) \subseteq [m, M]$.

Proof. For the proof we refer to [1, Theorem 6.3, Lemma 6.5]. □

Remark 3.22. *Let T be a symmetric compact operator and $\{\lambda_k\}$ the set of all eigenvalues of T. Let $V_k = \{x \in H, Tx = \lambda_k x\}$ for every k, and P_k be the projection on V_k. Then*

$$T = \sum_k \lambda_k P_k.$$

Definition 3.23 (Hilbert-Schmidt operator)**.** *Let H be a separable. Hilbert space and T a compact symmetric operator. We say that T is an Hilbert Schmidt operator if*

$$\sum_k \lambda_k^2 = \sum_k \|Tv_k\|^2 < +\infty$$

where λ_k are the eigenvalues of T.

Actually it can be proved that if T is a Hilbert Schmidt operator, the value of the sum $\sum_k \|Tv_k\|^2$ does not depend on the choice of the orthonormal basis v_k.

Proposition 3.24. *Let H be a separable metric space with orthonormal basis u_i and $\mathcal{H}$ be the space of all Hilbert Schmidt operators. Then this space with the norm*

$$\|T\| = \sum_i \|Tu_i\|$$

is a Hilbert space with scalar product given by

$$(S, T) = \sum_i (Su_i, Tu_i).$$

Example 3.25. [Finite dimensional case] Let $H = \mathbb{R}^n$ and $Tx = \mathbf{A}x$ for some $n \times n$ matrix $\mathbf{A}$ with values in $\mathbb{R}$. The adjoint of T is $T^*x = \mathbf{A}^T x$, where $\mathbf{A}^T$ is the traspose of the matrix $\mathbf{A}$. T is symmetric if and only if $\mathbf{A}$ is symmetric. Moreover the eigenvalues of T are the eigenvalues of the matrix $\mathbf{A}$.

Finally the spectral theorem for compact symmetric operators says that if $\mathbf{A}$ is a symmetric matrix, then it can be reduced to diagonal form by a orthogonal transformation.

Remark 3.26. Let T be a Hilbert-Schmidt operator, such that 1 is not an eigenvalue of T. Then for all $f \in H$, the equation

$$h - Th = f$$

admits a unique solution $h \in H$. Indeed, consider v_k an orthonormal basis of H composed by eigenvectors of T. Then we rewrite the equation as

$$h - Th = \sum_k (h, v_k)v_k - \sum_k (h, v_k)\lambda_k v_k = \sum_k (1 - \lambda_k)(h, v_k)v_k = f = \sum_k (f, v_k)v_k.$$

Therefore the equation is satisfied if

$$(1 - \lambda_k)(h, v_k) = (f, v_k) \qquad \text{that is } (h, v_k) = \frac{(f, v_k)}{1 - \lambda_k}.$$

Then the solution h to the equation is given by

$$h := \sum_k \frac{(f, v_k)}{1 - \lambda_k} v_k.$$

3.5 Problems

(i) Let $X_n, Y_n \in \mathcal{H}$ such that $X_n \to X$ and $Y_n \to Y$. Show that

  – $\mathbb{E}(X_n) \to \mathbb{E}(X)$,

- $(X_n, Y_n) = \mathbb{E}(X_nY_n) \to \mathbb{E}(XY) = (X, Y)$,
- $Cov(X_n, Y_n) = \mathbb{E}(X_nY_n) - \mathbb{E}(X_n)\mathbb{E}(Y_n) \to Cov(XY) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
- $Var(X_n) = Cov(X_n, X_n) \to Var(X) = Cov(X, X)$.

(ii) Consider $H = L^2(-1, 1)$.

(a) Let $V_1 = \{a + bx \mid a, b \in \mathbb{R}, x \in (-1, 1)\}$ (the subspace of polynomials of degree less than 1.) Find the orthogonal projection of x^2 on V_1.

(b) Let $V_2 = \{a + bx + cx^2 \mid a, b, c \in \mathbb{R}, x \in (-1, 1)\}$ (the subspace of polynomials of degree less than 2). Find the orthogonal projection of x^3 on V_2.

(iii) Consider $X, Y, Z \in \mathcal{H}$ and assume X, Z are not constant. Compute the least linear quadratic estimator $L(Y|X, Z)$. Show that $L(Y|X, Z) = L(Y|X) + L(Y|Z - L(Z|X)) - \mathbb{E}(Y)$. (Hint: look at Remark 3.15).

(iv) Let $T : L^2(0.1) \to L^2(0, 1)$ defined as $Tf(x) = \int_0^x f(y)dy$.

Show that this is a compact operator and compute its adjoint.

4 Elements of Fourier Analysis

Fourier Analysis has several important applications in mathematics and statistics, in particular in data analysis and estimation. Loosely speaking, Fourier analysis refers to the tool used to compress complex data into exponential functions (or trigonometric functions). So, it permits to analyze data in terms of their frequency components. Two of the central ingredients of Fourier Analysis are the convolution operator and the Fourier transform.

In this last chapter we will consider also functions taking complex values, that is $f : \mathbb{R} \to \mathcal{C}$. In this case f can be written in terms of 2 real functions f_1, f_2 which correspond to the real and imaginary part of f, that is $f(x) = f_1(x) + if_2(x)$.

We recall also the formula for the complex exponential

$$e^{ix} = \cos x + i \sin x.$$

4.1 Convolution operator

Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be measurable functions and we define the convolution between f and g as the function

$$f * g(x) := \int_{\mathbb{R}} f(x - y)g(y)dy \qquad (\text{ or equivalently} = \int_{\mathbb{R}} f(y)g(x - y)dy)$$

for all x such that the integral exists finite. Note that $f * g$ is a function of x!
Intuitively: let $x \in \mathbb{R}^n$ and consider the function $y \to f(x-y)$. This is the same as the function f, but we have to shift the graph of f by x and then flip it around the axis $y = x$. Assume that f is a smooth function which is positive only in a neighborhood of 0 and null elsewhere, with integral 1. Computing $f * g(x)$ we are taking a sort of weighted average of the values of g near the point x (weighted by the values of g)..

Basic properties of the convolution are the following. For the proof we refer to the Section 8.2 in [2].

(i) $f * g(x) = g * f(x)$ and $(f * g) * h(x) = f * (g * h)(x)$,

(ii) The support of a function h is the closure of the set of points where $h \neq 0$. The support of $f * g$ is contained in the closure of the sum of the support of f and the support of g.

(iii) **Young inequality for convolutions**. If $f \in L^p(\mathbb{R}^n)$ for some $p \in [1, +\infty]$ and $g \in L^1(\mathbb{R}^n)$ then $f * g \in L^p(\mathbb{R}^n)$ and moreover $\|f * g\|_p \leqslant \|f\|_p \|g\|_1$.

One of the main important features of the convolution operator is that it has *regularizing properties.*

Proposition 4.1. *Let $f \in L^2(\mathbb{R}^n)$ and $g \in L^2(\mathbb{R}^n)$, then $f * g(x)$ is a continuous function such that $\lim_{|x|\to+\infty} f * g(x) = 0$ (so it is also bounded). Moreover $\|f * g\|_\infty \leqslant \|f\|_2\|g\|_2$.*

If $f \in L^1(\mathbb{R}^n)$ and $g \in C^k(\mathbb{R}^n)$ bounded and with bounded derivatives up to order k, then $f * g \in C^k(\mathbb{R}^n)$ and for every $i \in \{1, \dots, n\}$ and $h \in \{1, \dots, k\}$, $\partial^h_{x_i}(f * g)(x) = f * (\partial^h_{x_i} g)(x)$.

Let

$$g(x) = \begin{cases} ce^{\frac{1}{|x|^2-1}} & |x| \leqslant 1 \\ 0 & \text{elsewhere} \end{cases}$$

where $c > 0$ is chosen such that $\int_\mathbb{R} g(x)dx = 1$. Note that $g \in C^\infty(\mathbb{R})$ and $g(x) = 0$ for $|x| \geqslant 1$.

Let $t > 0$ and consider $g_t(x) = tg\left(\frac{x}{t}\right)$. Then $\int_\mathbb{R} g_t(x)dx = 1$ (by change of variable formula!) and $g_t(x) = 0$ if $|x| \geqslant t$.

As $t \to 0$ g_t becomes more and more concentrated at $x = 0$. Observe that by its properties, g_t **is the density function of a continuous random variable X_t.**

Proposition 4.2 (Approximation of the Dirac measure and regularization by convolution)**.** *Let X_t be the continuous random variable with density given by g_t as defined before. Then X_t converges in distribution as $t \to 0^+$ to the* **discrete** *random variable X_0 with associated distribution the Dirac measure δ_0 (that is $X \equiv 0$ almost surely).*
Let $f \in L^p(\mathbb{R})$ for $p \in [1, +\infty)$. Then $g_t * f(x)$ is smooth (that is, it is in C^∞) and moreover $g_t * f(x) \to f$ in L^p.

Proof. To prove the convergence in distribution we need to show that for every f which is continuous and bounded there holds

$$\lim_{t\to 0^+} \int_\mathbb{R} f(x)g_t(x)dx = \delta_0(f) = f(0).$$

By definition and changing the variable posing $y = \frac{x}{t}$

$$\int_\mathbb{R} f(x)g_t(x)dx = \int_{-t}^{t} f(x)g_t(x)dx = c\int_{-1}^{1} f(ty)e^{-\frac{1}{|y|^2-1}}dy.$$

Sending $t \to 0$ and applying the dominated convergence theorem we conclude.

The second part of the theorem is a consequence of the properties of convolutions. We refer to [2, Chapt. 8.2]. ☐

The convolution is also useful to compute density functions of the sum of independent random variables.

Theorem 4.3. *Let X and Y be independent continuous random variables and let f, g the associated density functions. So $Z = X + Y$ is a continuous random variable with density function given by $f * g$.*

Remark 4.4. The same statement holds also with discrete random variables, substituting the integral with sum and convolution with a discrete convolution. That is if X, Y are discrete independent random variables, then $X + Y = Z$ is discrete random variable and the following holds: for every $n \in \mathbb{Z}$,

$$\mathbb{P}(Z = n) = \sum_{-\infty}^{+\infty} \mathbb{P}(X = k)\mathbb{P}(Y = n - k).$$

The proof of this formula can be checked easily in the case of random variables taking a finite number of values.

Proof. Observe that for every a, b, by independence

$$\mathbb{P}(X \leqslant a, Y \leqslant b) = \mathbb{P}(X \leqslant a)\mathbb{P}(Y \leqslant b) = \int_{-\infty}^{a} f(x)dy \int_{-\infty}^{b} g(y)dy.$$

So in particular we get

$$\mathbb{P}(X + Y \leqslant t) = \mathbb{P}(X \leqslant x, Y \leqslant y, x + y \leqslant t) = \int_{(x,y)\in\mathbb{R}^2, x+y\leqslant t} f(x)g(y)dxdy$$

where the integral is an integral computed in $\mathbb{R}^2$. We change variables to (z, w) where $x = z$ and $w = x + y$ (so $y = w - z$). So we get that $z \in \mathbb{R}$ and $w \leqslant t$:

$$\mathbb{P}(X + Y \leqslant t) = \int_{(x,y)\in\mathbb{R}^2, x+y\leqslant t} f(x)g(y)dxdy = \int_{-\infty}^{t} \int_{\mathbb{R}} f(z)g(w - z)dwdz = \int_{-\infty}^{t} f * g(z)dz$$

where in the last equality we use the definition of convolution. ☐

4.2 Fourier series

Assume that f is 2π periodic and bounded. Then we may reduce to consider it a function in $L^2(-\pi, \pi)$. This is a Hilbert space and we may apply the theory discussed in Chapter 4.

Let $L^2(-\pi, \pi)$. It is possible to show that an orthonormal basis of this space is given by $\{\frac{1}{\sqrt{2\pi}}, \frac{\cos(nx)}{\sqrt{\pi}}, \frac{\sin(nx)}{\sqrt{\pi}}, n \in \mathbb{N}\}$.

By Parseval theorem every function $f \in L^2(-\pi, \pi)$ can be written as

$$f(x) = a_0 + \sum_{n=1}^{+\infty} a_n \cos nx + \sum_{n=1}^{+\infty} b_n \sin nx,$$

where

$$a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)dx \quad a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nxdx, \qquad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nxdx.$$

This is called the Fourier serie of f. The equality holds in the sense that

$$\lim_{N} \left\| f - a_0 - \sum_{n=1}^{N} a_n \cos nx - \sum_{n=1}^{N} b_n \sin nx \right\|_2 = 0. \tag{4.1}$$

It is possibile to prove that if f is more regular than in L^2, then the convergence holds also in pointwise sense. We refer for the proof to [1].

Proposition 4.5. *If f is differentiable at a point $\bar{x}$, then the convergence in* (4.1) *holds also pointwise in $\bar{x}$, that is*

$$f(\bar{x}) = a_0 + \sum_{n=1}^{+\infty} a_n \cos n\bar{x} + b_n \sin n\bar{x}.$$

Some control on the derivatives at a point is necessary to assure the pointwise convergence (there exists continuous functions such that the Fourier serie does not converge pointwise).

The Parseval identity gives that

$$\|f\|_2^2 = 2\pi a_0^2 + \pi \sum_n (a_n^2 + b_n^2).$$

If we consider functions with complex variables the orthonormal basis is given by $\{\frac{e^{inx}}{\sqrt{2\pi}}\}$ and the coefficients of the Fourier serie are

$$c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-inx}dx.$$

For a generic function of period T, the Fourier serie has the form

$$\sum_{n=-\infty}^{n=+\infty} c_n e^{\frac{2\pi i n x}{T}},$$

where $c_n = \frac{1}{T}\int_0^T f(x)e^{-i2\pi nx/T}dx$.

4.3 Fourier transform

The Fourier transform is an isometry among Hilbert spaces as we will see (so a bijection which maintains the distance) and in some sense it can be interpreted as a generalization of the Fourier serie in non periodic context.

Let $f \in L^1(\mathbb{R})$. We define the Fourier transform of f as the complex valued function

$$\hat{f}(x) = \int_{\mathbb{R}} f(y)e^{ixy}dy.$$

It can be generalized to several dimension: if $f \in L^1(\mathbb{R}^n)$ then

$$\hat{f}(x) = \int_{\mathbb{R}^n} f(y)e^{ix\cdot y}dy.$$

Observe that since $|e^{ixy}| = 1$ for all $x, y \in \mathbb{R}$, $|\hat{f}(x)| \leqslant \int_{\mathbb{R}} |f(y)|e^{ixy}dy \leqslant \int_{\mathbb{R}} |f(y)|dy = \|f\|_{L^1}$. More precisely we get the following result (see for the proof [2], Section 8.3), stating that the Fourier transform sends integrable functions in bounded continuous functions.

Proposition 4.6 (Riemann Lebesgue lemma). *Let $f \in L^1(\mathbb{R})$. Then $\hat{f} \in C(\mathbb{R})$ and moreover $\lim_{|x|\to+\infty} \hat{f}(x) = 0$, $\|\hat{f}\|_\infty \leqslant \|f\|_{L^1}$.*

Other important properties of the Fourier transform are stated in the following proposition.

Proposition 4.7. *Let $f, g \in L^1(\mathbb{R})$. Then*

(i) $\widehat{(f * g)} = \hat{f}\hat{g}$. *So the Fourier transform of a convolution is the product of the Fourier transform.*

(ii) *If $|x|^k f \in L^1(\mathbb{R})$, then $\hat{f} \in C^k(\mathbb{R})$ and $d_x^k \hat{f}(x) = [\widehat{(iy)^k f}]$.*

(iii) *If $f \in C^k(\mathbb{R})$, $d_x^k f(x) \in L^1$, $\lim_{|x|\to+\infty} d_x^n f(x) = 0$ for $n \leqslant k$, then $\widehat{(d_x^n f)}(x) = (-ix)^n \hat{f}(x)$ for all $n \leqslant k$.*

Proof. (i) By definition, properties of the exponential and changing at the end variables (from (y, t) to (s, t) where $s = y - t$)

$$\begin{aligned}\widehat{(f * g)}(x) &= \int_{\mathbb{R}} f * g(y)e^{ixy}dy = \int_{\mathbb{R}}\int_{\mathbb{R}} f(t)g(y-t)e^{ixy}dtdy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(t)g(y-t)e^{ix(y-t)}e^{ixt}dtdy \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} f(t)g(s)e^{ixs}e^{ixt}dtds = \hat{f}(x)\hat{g}(x).\end{aligned}$$

(ii) We get that

$$d_x\hat{f}(x) = d_x\int_{\mathbb{R}} f(y)e^{ixy}dy = \int_{\mathbb{R}} d_x f(y)e^{ixy}dy = \int_{\mathbb{R}} f(y)(iy)e^{ixy}dy = \widehat{(iyf)}(x).$$

Repeat the argument we conclude with the result for every $k \in \mathbb{N}$.

(iii) We integrate by parts and we have that

$$\widehat{d_y f}(x) = \int_{\mathbb{R}} d_y f(y) e^{ixy} dy = \left[f(y)e^{ixy}\right]_{-\infty}^{+\infty} - \int_{\mathbb{R}} f(y)(ix)e^{ixy} dy = -ix\hat{f}(x).$$

Iterating the procedure we conclude.

□

The previous proposition has a very important consequence:

$$\text{let } a > 0 \text{ and } f_a(x) = e^{-a|x|^2}, \text{ then } \widehat{f_a}(x) = \sqrt{\frac{\pi}{a}} e^{-\frac{|x|^2}{4a}}. \tag{4.2}$$

More generally in $\mathbb{R}^n$, if $f_a(x) = e^{-a|x|^2}$, for $x \in \mathbb{R}^n$, then $\widehat{f_a}(x) = \sqrt{\frac{\pi^n}{a^n}} e^{-\frac{|x|^2}{4a}}$.

We prove (4.2). Observe that by the previous proposition, items (ii) and (iii) we get that

$$d_x \widehat{f_a}(x) = \int_{\mathbb{R}} e^{-a|y|^2}(iy)e^{ixy} dy = \int_{\mathbb{R}} \frac{-i}{2a} d_y(e^{-a|y|^2}) e^{ixy} dy = -\frac{i}{2a}\widehat{d_y f_a}(x) = -\frac{x}{2a}\hat{f}_a(x).$$

So the function $\hat{f}_a = \phi$ satisfies $\phi'(x) = -\frac{x}{2a}\phi(x)$, integrating we get that $(\log\phi(x))' = -\frac{x^2}{4a} + c$ and then $\phi(x) = ke^{-\frac{1}{4a}x^2}$. Finally to compute k we need to compute $\phi(0) = \hat{f}_a(0)$.

$$\hat{f}_a(0) = \int_{\mathbb{R}} e^{-a|y|^2} e^0 dy = \sqrt{\frac{\pi}{a}}.$$

Proposition 4.8. *Let $f, g \in L^1(\mathbb{R})$, then*

$$\int_{\mathbb{R}} \hat{f}(x) g(x) dx = \int_{\mathbb{R}} f(x)\hat{g}(x) dx.$$

Proof. By definition and by changing the order of integration (thanks to Fubini Tonelli theorem)

$$\int_{\mathbb{R}} \hat{f}(x)g(x)dx = \int_{\mathbb{R}}\int_{\mathbb{R}} f(y)g(x)e^{ixy} dy dx = \int_{\mathbb{R}} f(x)\hat{g}(x) dx.$$

□

For $f \in L^1(\mathbb{R})$ we may define also the anti transform of f as follows:

$$\check{f}(x) = \frac{1}{2\pi}\int_{\mathbb{R}} f(y)e^{-ixy} dy = \frac{1}{2\pi}\hat{f}(-x).$$

Obviously, this operator satisfies the same properties as the Fourier transform.

Theorem 4.9 (Fourier inversion theorem)**.** *Let $f \in L^1(\mathbb{R})$ such that also $\hat{f} \in L^1(\mathbb{R})$. Then f is continuous and bounded (that is, it coincides almost everywhere with a continuous function) and $\check{\hat{f}} = f = \hat{\check{f}}$.*

In particular if $f, g \in L^1(\mathbb{R})$ with $\hat{f} = \hat{g}$, then $f = g$ almost everywhere.

Proof. We give a sketch of the proof, for the rigorous demonstration we refer to [2], Theorem 8.26.

We have that

$$\int_{\mathbb{R}} \hat{f}(y)e^{-ixy} dy = \lim_{\varepsilon\to 0}\int_{\mathbb{R}} \hat{f}(y)e^{-ixy}e^{-\varepsilon y^2} dy = \lim_{\varepsilon\to 0}\int_{\mathbb{R}}\int_{\mathbb{R}} f(z)e^{iyz} dz e^{-ixy} e^{-\varepsilon y^2} dy =$$

by changing the order of integration

$$= \lim_{\varepsilon\to 0}\int_{\mathbb{R}} f(z)\int_{\mathbb{R}} e^{iy(z-x)}e^{-\varepsilon y^2} dy dz.$$

Now we observe that

$$\int_{\mathbb{R}} e^{iy(z-x)} e^{-\varepsilon y^2} dy = \widehat{e^{-\varepsilon y^2}}(z-x)$$

and then by (4.2) we conclude

$$\int_{\mathbb{R}} e^{iy(z-x)} e^{-\varepsilon y^2} dy = \widehat{e^{-\varepsilon y^2}}(z-x) = \frac{\sqrt{\pi}}{\sqrt{\varepsilon}} e^{-(x-z)^2/4\varepsilon}.$$

We substitute in the previous integral and we get

$$2\pi \check{\hat{f}}(x) = \int_{\mathbb{R}} \hat{f}(y) e^{-ixy} dy = \lim_{\varepsilon \to 0} \int_{\mathbb{R}} f(z) \frac{\sqrt{\pi}}{\sqrt{\varepsilon}} e^{-(x-z)^2/4\varepsilon} dz = 2\sqrt{\pi} \lim_{\varepsilon \to 0} \int_{\mathbb{R}} f\left(x - \frac{\xi}{2\sqrt{\varepsilon}}\right) e^{-\xi^2} d\xi$$

$$= 2\sqrt{\pi} f(x) \int_{\mathbb{R}} e^{-\xi^2} d\xi = 2\pi f(x).$$

The last conclusion comes from the fact that $\widehat{(f-g)} = \hat{f} - \hat{g} = 0$. Therefore $f - g \in L^1(\mathbb{R})$ is such that $\widehat{(f-g)} = 0 \in L^1(\mathbb{R})$, which implies by the inversion theorem that $f - g = \check{\widehat{(f-g)}} = 0$. $\square$

Using the inversion theorem, we get also the following result:

Corollary 4.10. *Let*

$$\mathcal{S} = \{g : \mathbb{R} \to \mathbb{R},\ g \in C^\infty,\ \forall k\ \sup_x ||x|^k g(x)| \leqslant C_k,\ |x|^k g(x) \in L^1(\mathbb{R})\}.$$

Then the Fourier transform is a bijection of $\mathcal{S}$ into itself.

Note that for all $a > 0$, $e^{-ax^2} \in \mathcal{S}$.

Proof. By Proposition 4.7, we get that if $g \in \mathcal{S}$ then $\hat{g} \in C^\infty$ and moreover $x^k \hat{g}$ is bounded continuous and integrable, so in particular $\hat{g} \in \mathcal{S}$. The conclusion comes from the inversion theorem. $\square$

Lemma 4.11. *The set $\mathcal{S}$ is dense in the space $C_0(\mathbb{R}) = \{g \in \mathcal{C}(\mathbb{R}) \mid \lim_{|x| \to +\infty} g(x) = 0\}$ (with respect to $\|\cdot\|_\infty$ norm) and in the space $L^p(\mathbb{R})$ for all $p \in [1, +\infty)$ (with respect to $\|\cdot\|_p$ norm).*

For this lemma we refer to [2, Proposition 8.17].

Theorem 4.12 (Plancherel theorem)**.** *If $f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ then $\hat{f} \in L^2(\mathbb{R})$, and the Fourier transform extends in a unique way to a isomorphism (so a linear bijection)*

$$\mathcal{F} : L^2(\mathbb{R}) \to L^2(\mathbb{R}), \qquad \mathcal{F}(f) = \hat{f}$$

with $2\pi \|f\|_2^2 = \|\hat{f}\|_2^2$.

Finally we conclude with the following theorem.

Theorem 4.13. *Let $f_n, f \in L^1(\mathbb{R})$. Assume that $\hat{f}_n \to \hat{f}$ pointwise and that there exists $C > 0$ such that $\|f_n\|_{L^1} \leqslant C$ for all n. Then $f_n \rightharpoonup f$ vaguely in $L^1(\mathbb{R})$, that is for all $g \in C_0(\mathbb{R})$, there holds $\lim_n \int_{\mathbb{R}} f_n(x) g(x) dx = \int_{\mathbb{R}} f(x) g(x) dx$.*

Proof. Let $g \in C_0(\mathbb{R})$. Then by Lemma 4.11 there exists $g_k \in \mathcal{S}$ such that $\sup_{x \in \mathbb{R}} |g_k(x) - g(x)| \leqslant \frac{1}{k}$.

Since $g_k \in \mathcal{S}$ then $g_k = \hat{\check{g}}_k$ by Corollary 4.10. Therefore we get

$$\int_{\mathbb{R}} (f_n - f)(x) g_k(x) dx = \int_{\mathbb{R}} (f_n - f)(x) \hat{\check{g}}_k(x) dx = \int_{\mathbb{R}} \int_{\mathbb{R}} (f_n - f)(x) \check{g}_k(y) e^{ixy} dy dx$$

exchanging the order of integration

$$= \int_{\mathbb{R}} \int_{\mathbb{R}} (f_n - f)(x) \check{g}_k(y) e^{ixy} dx dy = \int_{\mathbb{R}} (\hat{f}_n - \hat{f})(y) \check{g}_k(y) dy.$$

Since $\sup_{y\in\mathbb{R}} |\hat{f}_n(y) - \hat{f}(y)| \leqslant \|f_n - f\|_L^1 \leqslant \|f_n\|_L^1 + \|f\|_L^1 \leqslant C + \|f\|_L^1$ and $g_k \in L^1$, we get that $|(\hat{f}_n - \hat{f})(y)\check{g_k}(y)| \leqslant C + \|f\|_{L^1}|g_k| \in L^1$. Moreover $\hat{f}_n(y) - \hat{f}(y) \to 0$ as $n \to +\infty$ by assumption, then by the Lebesgue dominated convergence we conclude that

$$\lim_n \int_{\mathbb{R}} (f_n - f)(x) g_k(x) dx = 0$$

for all $k > 0$. Using the fact that $\sup_{x\in\mathbb{R}} |g_k(x) - g(x)| \leqslant \frac{1}{k}$ we get

$$\begin{aligned}
\left|\int_{\mathbb{R}} (f_n - f)(x) g(x) dx\right| &\leqslant \left|\int_{\mathbb{R}} (f_n - f)(x)(g_k - g)(x) dx\right| + \left|\int_{\mathbb{R}} (f_n - f)(x) g_k(x) dx\right| \\
&\leqslant \int_{\mathbb{R}} |f_n(x) - f(x)||g_k(x) - g(x)| dx + \left|\int_{\mathbb{R}} (f_n - f)(x) g_k(x) dx\right| \\
&\leqslant \frac{1}{k}\|f_n - f\|_{L^1} + \left|\int_{\mathbb{R}} (f_n - f)(x) g_k(x) dx\right| \leqslant \frac{1}{k}(C + \|f\|_{L^1}) + \left|\int_{\mathbb{R}} (f_n - f)(x) g_k(x) dx\right|.
\end{aligned}$$

Therefore we conclude that for all $k \in \mathbb{N}$,

$$\lim_n \left|\int_{\mathbb{R}} (f_n - f)(x) g(x) dx\right| \leqslant \frac{1}{k}(C + \|f\|_{L^1})$$

which gives the conclusion sending $k \to +\infty$. □

4.4 Characteristic functions of random variables and the Central Limit theorem

Let X be a random variable, with associated $\mathbb{P}_X$ probability distribution. The characteristic function of X is defined as the (complex valued) function

$$\phi_X(t) = \mathbb{E}(e^{itX}).$$

More precisely

- if X is a (asbsolutely) continuous random variable (with density f) then

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} f(x) dx = \hat{f}(t).$$

  So in this case the characteristic function of X is the Fourier transform of the density function f associated to X.

- if X is a discrete random variable (taking values on $\mathbb{Z}$),

$$\phi_X(t) = \sum_{k\in\mathbb{Z}} e^{ikt} P(\omega \mid X(\omega) = k).$$

Note that ϕ_X is a continuous function such that $\phi(0) = 1$.

Proposition 4.14. *If X_1, X_2 are independent random variables, then the characteristic function of $X_1 + X_2$ satisfies*

$$\phi_{X_1+X_2}(t) = \phi_{X_1}(t)\phi_{X_2}(t).$$

Proof. We consider only the case in which X_1, X_2 are absolutely continuous random variables (for the other case the argument is similar). The probability density of the sum of X_1 and X_2 is given by the convolution between the density of X_1 and the density of X_2 by Theorem 4.3. Then the Fourier transform of a convolution is the product of the Fourier transforms, see Proposition 4.7. □

The characteristic function associated to a random variable characterizes completely the random variable, and moreover the functional from the spaces of random variables with the convergence in distribution to the space of characteristic functions with the pointwise convergence is continuous, in the sense that if a sequence of random variables is converging in distribution to a random variable, then the same holds for the characteristic functions (and viceversa).

Theorem 4.15. *Let X_n be a family of random variables.*

(i) *If X_n are converging in distribution to X, then $\phi_{X_n}(t) \to \phi_X(t)$ for every t.*

(ii) *If $\phi_{X_n}(t) \to \phi(t)$ for every t, where ϕ is a function continuous at $t = 0$, then ϕ is the characteristic function of a random variable X and X_n converge in distribution to X.*

Proof. (i) $X_n \to X$ in distribution for every bounded continuous function g it holds

$$\mathbb{E}(g(X_n)) \to \mathbb{E}(g(X)).$$

So, taking for every t, $g_t(y) = e^{ity}$ (which is bounded and continuous), we get $\phi_{X_n}(t) \to \phi_X(t)$.

(ii) We prove this part theorem only in the case of absolutely continuous random variables X_n, with associated densities f_n. The general case can be obtained similarly.

We claim that X_n are tight. If the claim is true, then by Theorem 2.34, up to a subsequence we get that X_{n_k} converge in distribution to a random variable X. By (i), we get that $\phi_{X_{n_k}}(t) \to \phi_X(t)$ for every t and so $\phi(t) = \phi_X(t)$. Since the limit is unique (does not depend on subsequences), we conclude the convergence of the whole sequence of X_n.

So to conclude it is sufficient to show that X_n are tight. Since we are assuming X_n to have a density f_n, we get that $\phi_{X_n}(t) = \hat{f}_n(t)$. Fix $\delta > 0$ and consider

$$\frac{1}{2\delta}\int_{-\delta}^{\delta}(1-\phi_{X_n}(t))dt = \frac{1}{2\delta}\int_{-\delta}^{\delta}(1-\hat{f}_n(t))dt = \frac{1}{2\delta}\int_{-\delta}^{\delta}\int_{\mathbb{R}}(1-e^{iyt})f_n(y)dydt$$

$$= \frac{1}{2\delta}\int_{\mathbb{R}}\int_{-\delta}^{\delta}(1-e^{iyt})dtf_n(y)dy = \frac{1}{2\delta}\int_{\mathbb{R}}\left[2\delta - \frac{2\sin\delta y}{y}\right]f_n(y)dy = \int_{\mathbb{R}}\left[1-\frac{\sin\delta y}{\delta y}\right]f_n(y)dy$$

$$\geqslant \frac{1}{2}\int_{|\delta y|\geqslant 2} f_n(y)dy = \frac{1}{2}\mathbb{P}\left(|X_n| \geqslant \frac{2}{\delta}\right).$$

Hence

$$\mathbb{P}\left(|X_n| \geqslant \frac{2}{\delta}\right) \leqslant \frac{1}{\delta}\int_{-\delta}^{\delta}(1-\phi_{X_n}(t))dt \to \frac{1}{\delta}\int_{-\delta}^{\delta}(1-\phi(t))dt.$$

Since ϕ is continuous and $\phi(0) = 1$, we get that for every $\varepsilon > 0$ there exists δ such that $(1-\phi(t)) \leqslant \varepsilon/4$ for $t \in [-\delta, \delta]$. So

$$\frac{1}{\delta}\int_{-\delta}^{\delta}(1-\phi(t))dt \leqslant \frac{\varepsilon}{2}.$$

We fix $\varepsilon > 0$, we choose δ as above, and $K_\varepsilon = \{|x| \leqslant \frac{2}{\delta}\}$ and then we choose $\bar{n}$ such that $\frac{1}{\delta}\int_{-\delta}^{\delta}(1-\phi_{X_n}(t))dt \leqslant \varepsilon$ for all $n \geqslant \bar{n}$. This gives the desired tightness: $\mathbb{P}\left(|X_n| \in K_\varepsilon\right) \geqslant 1-\varepsilon$ for all $n \geqslant \bar{n}$.

□

We conclude showing that actually the Central Limit theorem can be interpreted as a result in Fourier analysis. The theorem says that if we have a sufficiently large sample of observations-randomly produced in a way that does not depend on the values of the other observations- the probability distribution of the observed averages will closely approximate a normal distribution.

Theorem 4.16 (Central Limit theorem)**.** *Let X_n be independent identically distributed random variables with (common) mean μ and a variance σ. Then $\frac{X_1+\cdots+X_n-\mu}{\sqrt{n}\sigma}$ converges in distribution to the normal random variable with mean 0 and variance 1.*

We are not going to prove in full generality this theorem, but we are just giving an idea of what is going on in the case in which every X_i is an absolutely continuous random variable with density f. Up to a renormalization we may assume that the mean of X_i is 0 and the variance is 1.

Proposition 4.17. *Let $f : \mathbb{R} \to [0, +\infty)$ such that*

$$\int_{\mathbb{R}} f(x)dx = 1, \qquad \int_{\mathbb{R}} xf(x)dx = 0 \qquad \int_{\mathbb{R}} x^2 f(x)dx = 1.$$

Let $f^{*n} := f * \cdots * f$ (the convolution of f by itself n times).

Then $f_n(x) := \sqrt{n} f^{*n}(\sqrt{n}x)$ converges vaguely as $n \to +\infty$ to $\frac{e^{-x^2/2}}{\sqrt{2\pi}}$.

Proof. The first assumption on f implies that $\widehat{f}(0) = 1$. Moreover, recalling Proposition 4.7, item ii, we get that the second and third assumption on f imply that $f \in C^2$. Moreover

$$\widehat{d_x f}(0) = \int_{\mathbb{R}} (iy)f(y)dy = 0 \qquad \widehat{d_x^2 f}(0) = \int_{\mathbb{R}} (-iy)^2 f(y)dy = -1.$$

By Taylor theorem we conclude that for $x \to 0$,

$$\hat{f}(x) = 1 - \frac{1}{2}x^2 + o(x^2).$$

We compute now $\widehat{f_n}(x)$. We have that

$$\widehat{f_n}(x) = \int_{\mathbb{R}} f_n(y)e^{ixy}dy = \int_{\mathbb{R}} \sqrt{n} f^{*n}(\sqrt{n}y)e^{ixy}dy =$$

changing variable $z = \sqrt{n}y$

$$= \int_{\mathbb{R}} f^{*n}(z)e^{ix\frac{z}{\sqrt{n}}}dz = \int_{\mathbb{R}} f^{*n}(z)e^{i\frac{x}{\sqrt{n}}z}dz = \widehat{f^{*n}}\left(\frac{x}{\sqrt{n}}\right)$$

and recalling by Proposition 4.7, item i, that $\widehat{f^{*n}}(x) = (\widehat{f}(x))^n$ we conclude that

$$\widehat{f_n}(x) = \widehat{f^{*n}}\left(\frac{x}{\sqrt{n}}\right) = \left(\widehat{f}\left(\frac{x}{\sqrt{n}}\right)\right)^n.$$

So, we get for x fixed and $n \to +\infty$

$$\widehat{f_n}(x) = \left(\widehat{f}\left(\frac{x}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{x^2}{2n} + o\left(\frac{1}{n}\right)\right)^n = e^{n\log\left(1-\frac{x^2}{2n}+o\left(\frac{1}{n}\right)\right)}.$$

Recalling that for x fixed and $n \to +\infty$, we get $\log\left(1 - \frac{x^2}{2n} + o\left(\frac{1}{n}\right)\right) = -\frac{x^2}{2n} + o\left(\frac{1}{n}\right)$ we get

$$\widehat{f_n}(x) = e^{-\frac{x^2}{2}+o(1)}$$

and therefore $\lim_n \widehat{f}_n(x) = e^{-\frac{x^2}{2}}$. By (4.2) with $a = \frac{1}{2}$ we have that $\widehat{\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}} = e^{-\frac{x^2}{2}}$. Therefore

$$\lim_n \widehat{f_n}(x) = \widehat{\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}}.$$

Moreover $\|f_n\|_1 = 1$ for all n. So, we may apply Theorem 4.13 to obtain that f_n is converging vaguely to $\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$. □

4.5 Problems

(i) Let $c > 0$, and

$$h_c(x) = \begin{cases} 1 & |x| \leqslant c \\ 0 & \text{elsewhere} \end{cases}.$$

Compute $h_c * h_c$. Then compute $h_c * h_c * h_c$. What we can say about the regularity of these functions?

(ii) Let $X_1, X_2, \ldots X_n$ are n independent continuous random variables with the same distribution (and so with the same density function f). Assume that $\mathbb{E}(X_i) = \mu$ and $\mathbb{E}(X_i - \mu)^2 = \sigma^2$. Show that the density function of $\frac{X_1 + \cdots + X_n - \mu}{\sqrt{n}\sigma}$ is given by $\sqrt{n}\sigma f^{*n}(x\sqrt{n}\sigma + \mu n)$, where $f^{*n}(x)$ is the convolution of f with itself repeated n times.

(iii) Let $\delta < \pi$ and $f : (-\pi, \pi) \to \mathbb{R}$ defined as

$$f(x) = \begin{cases} 1 & -\delta < x < \delta \\ 0 & -\pi < x < -\delta \text{ and } \delta < x < \pi. \end{cases}$$

(a) Compute the Fourier serie of f.

(b) Show that

$$\sum_n \frac{\sin n}{n} = \frac{\pi - 1}{2}.$$

(c) Compute

$$\sum_n \frac{\sin^2 n}{n^2}.$$

(iv) (a) Compute the Fourier transform of $g(x) = e^{-x}\chi_{(0,+\infty)}(x)$.
Recall the following formulas (obtained by integration by parts):

$$\int e^{-y}\sin(xy)dy = -\frac{1}{x^2+1}e^{-y}(x\cos xy + \sin xy) + c$$

$$\int e^{-y}\cos(xy)dy = \frac{1}{x^2+1}e^{-y}(x\sin xy - \cos xy) + c.$$

(b) Compute the Fourier transform of $f(x) = xe^{-x}\chi_{(0,+\infty)}(x)$ (that is the characteristic function of the Gamma distribution).
Use item (a) and Proposition 4.7.

References

[1] A. Bressan *Lecture notes on Functional Analysis, with applications to Linear Partial Differential Equations* Graduate Studies in Mathematics, vol 143, AMS, 2013.

[2] G. Folland *Real Analysis: modern tecniques and their applications.* Wiley 1999 (2nd ed).

A Solutions to problems Section 2

(i) Let $f : \mathbb{R} \to \mathbb{R}$ be a monotone function. Show that f is Lebesgue measurable.

It is sufficient to show that for all $c \in \mathbb{R}$ the set $\{x \in \mathbb{R} \mid f(x) > c\}$ is measurable.

Assume that f is monotone increasing (if it is monotone decreasing the argument is analogous). Let $c \in \mathbb{R}$. If $f(x) \leqslant c$ for all $x \in \mathbb{R}$ then $\{x \in \mathbb{R} \mid f(x) > c\}$ is the empty set and we are done.

Assume now that there exists $\bar{x} \in \mathbb{R}$ such that $f(\bar{x}) > c$. By monotonicity we get that $f(y) > c$ for all $y > \bar{x}$. We consider now the set $A_c = \{x \in \mathbb{R} \,|\, f(x) > c\}$. Our aim is to show that this is a measurable set.

We observed that by monotonicity, if $x \in A_c$, then $[x, +\infty) \subseteq A_c$. So, if A_c is not bounded from below, this implies that $A_c = \mathbb{R}$ and so we are done. Assume now that A_c is bounded from below and define $x_c = \inf A_c$. For all $x > x_c$ we get that $f(x) > c$ and $f(x) \leqslant c$ for all $x < x_c$. This implies that $A_c = (x_c, +\infty)$ if $f(x_c) \leqslant c$, and $A_c = [x_c, +\infty)$ if $f(x_c) > c$. In both cases, $A_c \in \mathcal{M}$.

Note that actually we get something more: for all c, we get that A_c is a Borel set, so the function f is Borel measurable.

(ii) Consider the right continuous increasing function on $\mathbb{R}$

$$F(x) = \begin{cases} x & x < 0 \\ x+1 & x \geqslant 0. \end{cases}$$

Which is the Borel measure associated to this function?

We define $\mu_F(a,b] = F(b) - F(a)$, and then we extend it to a measure on the Borel σ-algebra. Given F as in the statement, we get that $\mu_F(a,b] = b - a$ if $a < b < 0$, $\mu_F(a,b] = b+1-(a+1) = b-a$ if $0 \leqslant a < b$, whereas if $a < 0 \geqslant b$, then $\mu_F(a,b] = b+1-a = b-a+1$. Therefore $\mu_F = \mathcal{L} + \delta_0$.

B Solutions to problems Section 3

(i) Let $(X, \|\cdot\|)$ a Banach space and $F : X \to X$ such that there exists $0 < a < 1$ for which

$$\|F(x) - F(y)\| \leqslant a\|x - y\| \qquad \forall x, y \in X.$$

(F is a contraction)

(a) Show that the map F is continuous.

(b) Let $x_0 \in X$. Define $x_1 = F(x_0)$, $x_2 = F(x_1)$ and so on $x_n = F(x_{n-1})$. Prove that

$$\|x_n - x_{n+1}\| \leqslant a^n \|x_0 - x_1\|.$$

Deduce that $(x_n)_n$ is a Cauchy sequence.

(c) Let $\bar{x} = \lim_n x_n$, where (x_n) has been defined in the previous step. Show that $F(\bar{x}) = \bar{x}$. So, $\bar{x}$ is a fixed point of F.

(d) Show that the map F admits a unique fixed point, that is a point such that $\bar{x} = F(\bar{x})$.

This is called **Banach-Caccioppoli theorem**.

(a) Let (x_n) be a sequence in X which is converging to x. Then $0 \leqslant \|F(x_n) - F(x)\| \leqslant a\|x_n - x\|$, and so $\lim_{n\to+\infty} F(x_n) = F(x)$ since $\lim_{n\to+\infty} x_n = x$.

(b) By the property of the function F and the definition of the we get that

$$\|x_{n+1} - x_n\| = \|F(x_n) - F(x_{n-1})\| \leqslant a\|x_n - x_{n-1}\| =$$

$$= a\|F(x_{n-1}) - F(x_{n-2})\| \leqslant a^2\|x_{n-1} - x_{n-2}\| \leqslant ... \leqslant a^n\|x_1 - x_0\|.$$

Let $n > m$. Then,by using the triangular inequality, we get

$$\|x_n - x_m\| \leqslant \|x_n - x_{n-1}\| + \|x_{n-1} - x_{n-2}\| + \cdots + \|x_{m+1} - x_m\|.$$

By using the previous inequality and recalling that $\sum_{i=0}^n a^i = \frac{1-a^{n+1}}{1-a}$, we get

$$\|x_n - x_m\| \leqslant (a^n + a^{n-1} + \cdots + a^m)\|x_0 - x_1\| \leqslant = \frac{a^{m+1} - a^{n+1}}{1-a}|x_0 - x_1\|.$$

Since $0 < a < 1$, we get that $a^{n+1}, a^{m+1} \to 0$ as $n, m \to +\infty$. So in particular the previous inequality implies that (x_n) is a Cauchy sequence.

(c) Since (x_n) is a Cauchy sequence, and the space is complete, it is converging to some point x. Using the continuity of F we have that $\lim_n F(x_n) = F(x)$. But we recall that $\lim_n F(x_n) = \lim_n x_{n-1} = x$. So $F(x) = x$.

(d) Let x, z such that $F(x) = x$ and $F(z) = z$. The by the property of F, and recalling that $a < 1$,

$$\|x - z\| = \|F(x) - F(z)\| \leqslant a\|x - z\| < \|x - z\|.$$

This is not possible unless $\|x - z\| = 0$, which implies $z = x$.

(ii) Let $f \in L^p(\mathbb{R}^n)$ and $\alpha > 0$. Prove that

$$\mathcal{L}\left(\{x \in \mathbb{R}^n \mid |f(x)| > \alpha\}\right)^{\frac{1}{p}} \leqslant \frac{1}{\alpha}\|f\|_p.$$

This is called **Chebycheff inequality**.

Let $A_\alpha = \{x \in \mathbb{R}^n \mid |f(x)| > \alpha\}$. Then $\mathbb{R}^n = A_\alpha \cap (\mathbb{R}^n \backslash A_\alpha)$. So we compute, recalling definitions,

$$\|f\|_p^p = \int_{\mathbb{R}^n} |f(x)|^p dx = \int_{A_\alpha} |f(x)|^p dx + \int_{\mathbb{R}^n \backslash A_\alpha} |f(x)|^p dx \geqslant \int_{A_\alpha} |f(x)|^p dx,$$

since $|f|^p \geqslant 0$. Now if $x \in A_\alpha$, then $|f(x)|^p \geqslant \alpha^p$. Therefore in the previous inequality we get

$$\|f\|_p^p \geqslant \int_{A_\alpha} |f(x)|^p dx \geqslant \alpha^p \int_{A_\alpha} dx = \alpha^p \ \mathcal{L}(A_\alpha).$$

This gives the desired inequality,

$$\mathcal{L}(A_\alpha) \leqslant \left(\frac{\|f\|_p}{\alpha}\right)^p$$

after extracting the p rooth.

(iii) Prove that if $f \in L^2(-1, 1)$ then $f \in L^1(-1, 1)$ and moreover

$$\|f\|_1 \leqslant \sqrt{2}\|f\|_2.$$

Provide an example of a function $f \in L^1(-1, 1)$ such that $f \notin L^2(-1, 1)$.

Since $\chi_{(-1,1)} \in L^2(-1, 1)$, then by Holder inequality $f = f\chi_{(-1,1)} \in L^1(-1, 1)$. Moreover

$$\|f\|_1 \leqslant \|f\|_2 \left(\int_{-1}^1 dx\right)^{\frac{1}{2}} = \sqrt{2}\|f\|_2.$$

The function $f(x) = \frac{1}{\sqrt{|x|}}$ is in $L^1(-1, 1)$ but not in $L^2(-1, 1)$.

(iv) Consider the following operator $T : L^2(0, 2) \to L^2(0, 2)$ defined as

$$Tf(x) = \int_0^x f(y)dy.$$

Show that this is a bounded continuous operator.

First of all T is linear by linearity of the integral.

Note that, by Jensen, and changing the order of integration (observe that $0 < y < x < 2$) we get

$$\|Tf\|_2^2 = \int_0^2 \left(\int_0^2 \chi_{(0,x)}(y) f(y) dy \right)^2 dx \leqslant \int_0^2 2 \int_0^2 \chi_{(0,x)}(y) |f(y)|^2 dy = 2 \int_0^2 \int_0^x |f(y)|^2 dy$$

$$= 2 \int_0^2 \int_y^2 |f(y)|^2 dx dy = 2 \int_0^2 (2-y)|f(y)|^2 dy \leqslant 4\|f\|_2^2$$

where the last inequality comes from the fact that $(2-y) \leqslant 2$ for $y \in (0,2)$. Therefore

$$\sup_{\{\|f\| \leqslant 1\}} \|Tf\|_2 \leqslant \sup_{\{\|f\| \leqslant 1\}} 2\|f\|_2 \leqslant 2.$$

(a) We have to prove that $\int_O |f_k(x)g_k(x) - f(x)g(x)|dx \to 0$ as $k \to +\infty$.
We observe that, by Holder inequality,

$$\int_O |f_k(x)g_k(x) - f(x)g(x)|dx = \int_O |f(x)(g_k(x) - g(x)) + (f_k(x) - f(x))g_k(x)|dx$$
$$\leqslant \int_O |f(x)||g_k(x) - g(x)|dx + \int_O |f_k(x) - f(x)||g_k(x)|dx$$
$$\leqslant \|g_k - g\|_q \|f\|_p + \|f_k - f\|_p \|g_k\|_q. \quad \text{(B.1)}$$

Sending $k \to +\infty$, and recalling that $\lim_k \|g_k\|_q = \|g\|_q$, we get the conclusion.

(b) Arguing as in (B.1), we get

$$\int_O |f_k(x)g_k(x) - f(x)g(x)|dx \leqslant \int_O |f(x)||g_k(x) - g(x)|dx + \int_O |f_k(x) - f(x)||g_k(x)|dx$$
$$\leqslant \int_O |f(x)||g_k(x) - g(x)|dx + \|f_k - f\|_p \|g_k\|_q.$$

Note that $\int_O |f(x)||g_k(x) - g(x)|dx \to 0$ by weak convergence of g_k to g, whereas $\|f_k - f\|_p \to 0$ by strong convergence of f_k to f. Moreover recall that there exists $M > 0$ such that $\|g_k\|_q \leqslant M$. So, we conclude, sending $k \to +\infty$ that $\int_O |f_k(x)g_k(x) - f(x)g(x)|dx \to 0$.

C Solutions to problems Section 4

(i)
- $\mathbb{E}(X_n) = (X_n, 1) \to (X, 1) = \mathbb{E}(X)$, by continuity of the scalar product (as a consequence of Cauchy Schwartz inequality).
- $(X_n, Y_n) = (X_n - X, Y_n - Y) + (X, Y_n) + (X_n, Y) - (X, Y)$. We conclude observing that $(X_n - X, Y_n - Y) \to 0$, $(X_n, Y) \to (X, Y)$ and $(X, Y_n) \to (X, Y)$.
- the convergence of covariance and variance are immediate consequences of the first two items.

(ii) Consider $H = L^2(-1, 1)$.

(a) Let $V_1 = \{a + bx \; |a, b \in \mathbb{R}, x \in (-1, 1)\}$ (the subspace of polynomials of degree less than 1.) Find the orthogonal projection of x^2 on V_1.

(b) Let $V_2 = \{a + bx + cx^2 \; |a, b, c \in \mathbb{R}, x \in (-1, 1)\}$ (the subspace of polynomials of degree less than 2). Find the orthogonal projection of x^3 on V_2.

(a) We look for an orthonormal basis of V_1. A basis of V_1 is given by $1, x$. Note that $(x, 1) = \int_{-1}^{1} x dx = 0$, so 1 and x are orthogonal. We have to normalize them. We compute $\int_{-1}^{1} dx = 2$ and $\int_{-1}^{1} |x|^2 dx = \frac{2}{3}$. Therefore an orthonormal basis of V_1 is given by $\left(\frac{1}{\sqrt{2}}, \frac{\sqrt{3}x}{\sqrt{2}}\right)$. By the theorem on orthogonal projection we have that

$$P_{V_1}(x^2) = a_0 \frac{1}{\sqrt{2}} + a_1 \frac{\sqrt{3}x}{\sqrt{2}}$$

where $a_0 = (x^2, \frac{1}{\sqrt{2}})$ and $a_1 = (x^2, \frac{\sqrt{3}x}{\sqrt{2}})$. We compute

$$a_0 = \left(x^2, \frac{1}{\sqrt{2}}\right) = \frac{1}{\sqrt{2}} \int_{-1}^{1} x^2 dx = \frac{2}{3\sqrt{2}} \qquad a_1 = \left(x^2, \frac{\sqrt{3}x}{\sqrt{2}}\right) = \frac{\sqrt{3}}{\sqrt{2}} \int_{-1}^{1} x^3 dx = 0.$$

Therefore

$$P_{V_1}(x^2) = \frac{2}{3\sqrt{2}} \frac{1}{\sqrt{2}} = \frac{1}{3}.$$

Another way to compute the orthogonal projection is just using the definition: we have to find the point in V_1 with minimal distance from x^2. Every point in V_1 is defined as $a + bx$ for some a, b, so we have to solve the minimization problem

$$\min_{a,b \in \mathbb{R}} \|x^2 - a - bx\|^2 = \min_{a,b} \int_{-1}^{1} |x^2 - a - bx|^2 dx$$

$$= \min_{a,b} \int_{-1}^{1} (x^4 + a^2 + b^2x^2 - 2ax^2 - 2bx^3 + 2abx) dx = \min_{a,b} \left(\frac{1}{5} + 2a^2 + \frac{b^2}{3} - \frac{2}{3}a\right).$$

The minimum is 1/5 and there are two minimum couples of points: $a = \frac{1}{3}$, $b = 0$ and $a = 0$, $b = 0$. Therefore the projection is $1/3 + 0x = 1/3$.

(b) The previous point implies that $x^2 - \frac{1}{3} \in V_1^{\perp}$. We have to find a orthonormal basis of V_2. In order to have three generators which are orthogonal, we consider $1, x, x^2 - \frac{1}{3}$. Moreover we normalize them to have norm 1. We compute

$$\int_{-1}^{1} (x^2 - \frac{1}{3})^2 dx = \frac{2}{5} + \frac{2}{9} - \frac{4}{9} = \frac{2}{5} - \frac{2}{9} = \frac{8}{45}.$$

So an orthonormal basis of V_2 is given by $(\frac{1}{\sqrt{2}}, \frac{\sqrt{3}x}{\sqrt{2}}, \frac{\sqrt{45}}{\sqrt{8}}(x^2 - \frac{1}{3})$. Again by theorem on orthogonal projections we have

$$P_{V_2}(x^3) = a_0 \frac{1}{\sqrt{2}} + a_1 \frac{\sqrt{3}x}{\sqrt{2}} + a_2 \frac{\sqrt{45}}{\sqrt{8}}(x^2 - \frac{1}{3}) = (t^3, \frac{\sqrt{3}t}{\sqrt{2}}) \frac{\sqrt{3}x}{\sqrt{2}}$$

where $a_0 = (x^3, \frac{1}{\sqrt{2}})$, $a_1 = (x^3, \frac{\sqrt{3}x}{\sqrt{2}})$, $a_2 = (x^3, \frac{\sqrt{45}}{\sqrt{8}}(x^2 - \frac{1}{3}))$. It is easy to check that $a_0 = 0 = a_2$. We compute

$$a_1 = (x^3, \frac{\sqrt{3}x}{\sqrt{2}}) = \frac{\sqrt{3}}{\sqrt{2}} \int_{-1}^{1} x^4 dt = \frac{\sqrt{3}}{\sqrt{2}} \frac{2}{5}.$$

Therefore

$$P_{V_2}(x^3) = \frac{\sqrt{3}}{\sqrt{2}} \frac{2}{5} \frac{\sqrt{3}x}{\sqrt{2}} = \frac{3}{5}x.$$

(iii) Recalling Remark 3.15 we have that

$$L(Y|X, Z) = Pr_S(Y) = a + bX + cZ$$

where S is the space with basis $1, X, Z$.

Observe that by the same argument $L(Z|X) = Pr_T(Z)$ where T is the space with a basis given by $1, X$. In particular by Theorem 3.8 we have that $Z - L(Z|X) \in T^{\perp}$ and arguing as in Remark 3.15 $L(Z|X) = \mathbb{E}(Z) + \frac{Cov(X,Z)}{Var(X)}(X - \mathbb{E}(X))$.

An orthonormal basis of S can be therefore obtained by considering an orthonormal basis of T, which is given by $1, \frac{X-\mathbb{E}(X)}{\sqrt{Var(X)}}$ as proved in Remark 3.15 and then adding the element $k(Z - L(Z|X))$ where k is such that $\mathbb{E}(k(Z - L(Z|X))^2 = 1$. Since $\mathbb{E}((Z - L(Z|X))^2 = \frac{Var(Z)Var(X)-Cov^2(X,Z)}{Var(X)}$ as proved in Remark 3.15, we get that $k = \frac{\sqrt{VarX}}{\sqrt{Var(Z)Var(X)-Cov^2(X,Z)}}$.

So, as in Remark 3.15,

$$\begin{aligned} L(Y|X,Z) &= \mathbb{E}(Y) + \frac{Cov(X,Y)}{Var(X)}(X - \mathbb{E}(X))) \\ &\quad + \frac{Var(X)Cov(Z,Y) - Cov(X,Z)Cov(X,Y)}{Var(Z)Var(X) - Cov^2(X,Z)}(Z - L(Z|X)) \\ &= \mathbb{E}(Y) \\ &\quad + \frac{Var(Z)Cov(X,Y) - Cov(Z,Y)Cov(X,Z)}{Var(Z)Var(X) - Cov^2(X,Z)}(X - \mathbb{E}(X))) \\ &\quad + \frac{Var(X)Cov(Z,Y) - Cov(X,Z)Cov(X,Y)}{Var(Z)Var(X) - Cov^2(X,Z)}(Z - \mathbb{E}(Z)). \end{aligned}$$

Observe that

$$\mathbb{E}(Y) + \frac{Cov(X,Y)}{Var(X)}(X - \mathbb{E}(X))) = L(Y|X)$$

and moreover

$$\mathbb{E}(Y) + \frac{Var(X)Cov(Z,Y) - Cov(X,Z)Cov(X,Y)}{Var(Z)Var(X) - Cov(X,Z)}(Z - L(Z|X)) = L(Y|Z - L(Z|X)).$$

This conclude the proof.

(iv) Let $T : L^2(0.1) \to L^2(0,1)$ defined as $Tf(x) = \int_0^x f(y)dy$.

Show that this is a compact operator and compute its adjoint.

To prove that it is compact we have to show that if f_n converge weakly to f, that is $\lim_n \int_0^1 f_n(x)g(x)dx = \int_0^1 f(x)g(x)dx$ for all g, then $\|Tf_n - Tf\|_2 \to 0$, We compute, reasoning as above,

$$\|Tf_n - Tf\|_2^2 = \int_0^1 \left(\int_0^1 \chi_{(0,x)}(y)(f_n(y) - f(y))dy\right)^2 dx.$$

Now $F_n(x) = \left(\int_0^1 \chi_{(0,x)}(y)(f_n(y) - f(y))dy\right)^2$ is a function such that $\lim_n F_n(x) = 0$ for all x (by weak convergence, since $\chi_{(0,x)} \in L^2$). Moreover by Jensen inequality $F_n(x) \leqslant \int_0^x |f_n(y) - f(y)|^2 dy \leqslant \|f_n - f\|_2^2 \leqslant (\|f_n\|_2 + \|f\|_2)^2$. Recall that since f_n converge weakly then there exists C such that $\|f_n\|_2 \leqslant C$ (see Problem v)). This implies that $0 \leqslant F_n(x) \leqslant 2C^2$. Since constant functions are element in $L^2(0,1)$, we conclude by Lebsegue dominated convergence that $\lim_n \|Tf_n - Tf\|_2^2 = 0$.

To compute the adjoint we recall that $(Tf, g) = (f, T^*g)$ and so we compute, changing the order of integration

$$(Tf, g) = \int_0^1 \int_0^x f(y)g(x)dydx = \int_0^1 \int_y^1 g(x)dxf(y)dy.$$

Therefore $T^*g(x) = \int_x^1 g(y)dy$.

D Solutions to problems Section 5

(i) Let $c > 0$, and $h_c(x) = \begin{cases} 1 & |x| \leqslant c \\ 0 & \text{elsewhere} \end{cases}$. Compute $h_c * h_c(x)$. Then compute $h_c * h_c * h_c$. What we can say about the regularity of these functions?

By definition of h_c

$$h_c * h_c(x) = \int_{\mathbb{R}} h_c(x-y)h_c(y)dy = \int_{-c}^{c} h_c(x-y)dy = |[-c,c] \cap [x-c, x+c]|$$

where we indicated with $|[-c,c] \cap [x-c,x+c]|$ the length of the intersection between the two intervals. Since

$$[-c,c] \cap [x-c,x+c] = \begin{cases} \varnothing & x \geqslant 2c \text{ or } x \leqslant -2c \\ [-c, x+c] & -2c < x < 0 \\ [x-c, c] & 0 < x < 2c \end{cases}$$

we conclude that

$$h_c * h_c(x) = \begin{cases} 0 & x \geqslant 2c \text{ or } x \leqslant -2c \\ x+2c & -2c < x < 0 \\ 2c - x & 0 < x < 2c. \end{cases}$$

The graph is a triangular.. Then again by definition

$$h_c * h_c * h_c(x) = \int_{\mathbb{R}} (h_c * h_c)(x-y)h_c(y)dy = \int_{-c}^{c} (h_c * h_c)(x-y)dy$$

$$= \int_{[-c,c]\cap[x-2c,x+2c]} (h_c * h_c)(x-y)dy.$$

We observe that $h_c * h_c * h_c(x) = h_c * h_c * h_c(-x)$ so it is sufficient to compute the function for x positive and then symmetrize it (as an even function). If $x > 3c$ then $h_c * h_c * h_c(x) = 0$. If $x \in (2c, 3c)$ then $[-c,c] \cap [x-2c, x+2c] = [x-2c, c]$ with $x - 2c > 0$ and so

$$h_c * h_c * h_c(x) = \int_{x-2c}^{c} h_c * h_c(y)dy = \frac{(4c-x)^2}{2} - \frac{c^2}{2}.$$

If $x \in (c, 2c)$ then $[-c,c] \cap [x-2c, x+2c] = [x-2c, c]$ with $x - 2c < 0$ and so

$$h_c * h_c * h_c(x) = \int_{x-2c}^{0} h_c * h_c(y)dy + \int_{0}^{c} h_c * h_c(y)dy = \frac{4c^2 - x^2}{2} + \frac{3}{2}c^2.$$

If $x \in (0, c)$ then $[-c,c] \cap [x-2c, x+2c] = [-c, c]$ and so

$$h_c * h_c * h_c(x) = \int_{-c}^{c} h_c * h_c(y)dy = 3c^2.$$

(ii) Let $X_1, X_2, \ldots X_n$ are n independent continuous random variables with the same distribution (and so with the same density function f). Assume that $\mathbb{E}(X_i) = \mu$ and $\mathbb{E}(X_i - \mu)^2 = \sigma^2$. Show that the density function of $\frac{X_1 + \cdots + X_n - \mu}{\sqrt{n}\sigma}$ is given by $\sqrt{n}\sigma f^{*n}(x\sqrt{n}\sigma + \mu n)$, where $f^{*n}(x)$ is the convolution of f with itself repeated n times.

By Theorem 4.3 we get that the density function associated to the sum of X_1, X_2 is $f * f$. Then again by the theorem, the density function associated to the sum of $X_1 + X_2$ with X_3 is $(f * f) * f = f^{*3}$ and so on.

By linearity $\mathbb{E}(X_1+\cdots+X_n)=n\mu$ and by independence we get $\mathbb{E}((X_1+\cdots+X_n-\mu n)^2)=n\sigma^2$. So the sum as $Z=\frac{X_1+\cdots+X_n-\mu n}{\sqrt{n}\sigma}$, we get that Z has $\mathbb{E}(Z)=0$ and $\mathbb{E}(Z^2)=1$ (so it has mean 0 and variance 1).

f^{n*} is the density associated to $X_1+\ldots X_n$, we get that $\sqrt{n}\sigma f^{*n}\left(x\sqrt{n}\sigma+\mu n\right)$ is the density associated to Z. Indeed we compute, changing variable,

$$\int_{\mathbb{R}} x\sqrt{n}\sigma f^{*n}\left(x\sqrt{n}\sigma+\mu n\right)dx=\int_{\mathbb{R}}\frac{y-\mu n}{\sqrt{n}\sigma}f^{*n}(y)dy=\frac{1}{\sqrt{n}\sigma}\mathbb{E}(X_1+\cdots+X_n-n\mu)=0$$

$$\int_{\mathbb{R}} x^2\sqrt{n}\sigma f^{*n}\left(x\sqrt{n}\sigma+\mu n\right)dx=\int_{\mathbb{R}}\frac{(y-\mu n)^2}{n\sigma^2}f^{*n}(y)dy=1.$$

(iii) Let $\delta<\pi$ and $f:(-\pi,\pi)\to\mathbb{R}$ defined as

$$f(x)=\begin{cases}1 & -\delta<x<\delta\\ 0 & -\pi<x<-\delta \text{ and } \delta<x<\pi.\end{cases}$$

(a) Compute the Fourier serie of f.

(b) Show that
$$\sum_n\frac{\sin n}{n}=\frac{\pi-1}{2}.$$

(c) Compute
$$\sum_n\frac{\sin^2 n}{n^2}.$$

(a) $a_0=\frac{1}{2\pi}\int_{-\pi}^{\pi}f(t)dt=\frac{\delta}{\pi}$, $b_n=0$ since f is an even function.
$$a_n=\frac{1}{\pi}\int_{-\pi}^{\pi}f(t)\cos(nt)dt=\frac{1}{\pi}\int_{-\delta}^{\delta}\cos(nt)dt=\frac{2}{n\pi}\sin(n\delta).$$
So the Fourier serie of f is
$$\frac{\delta}{\pi}+\frac{2}{\pi}\sum_n\frac{\sin(n\delta)}{n}\cos(nx).$$

(b) We take $\delta=1$. The Fourier serie of f at $x=0$ is converging to $f(0)$ so
$$\frac{1}{\pi}+\frac{2}{\pi}\sum_n\frac{\sin n}{n}=1.$$

(c) The Parseval identity reads
$$\|f\|_2^2=2\pi a_0^2+\sum_n\pi(a_n^2+b_n^2).$$
We apply Parseval equality for $\delta=1$ and we get
$$\frac{2}{\pi}+\frac{4}{\pi}\sum_n\frac{\sin^2 n}{n^2}=2.$$

(iv) (a) Compute the Fourier transform of $g(x)=e^{-x}\chi_{(0,+\infty)}(x)$.
Recall the following formulas (obtained by integration by parts):
$$\int e^{-y}\sin(xy)dy=-\frac{1}{x^2+1}e^{-y}(x\cos xy+\sin xy)+c$$
$$\int e^{-y}\cos(xy)dy=\frac{1}{x^2+1}e^{-y}(x\sin xy-\cos xy)+c.$$

(b) Compute the Fourier transform of $f(x) = xe^{-x}\chi_{(0,+\infty)}(x)$ (that is the characteristic function of the Gamma distribution).
Use item a. and Proposition 4.7.

(a) By definition and using the primitive of the functions $e^{-y}\cos xy$ and $e^{-y}\sin xy$, we get

$$\hat{g}(x) = \int_{\mathbb{R}} g(y)e^{ixy}dy = \int_0^{+\infty} e^{-y}e^{ixy}dy = \int_0^{+\infty} e^{-y}\cos xy dy + i\int_0^{+\infty} e^{-y}\sin xy dy$$

$$= \frac{1}{x^2+1} + i\frac{x}{x^2+1}.$$

(b) By Proposition 4.7,

$$d_x\hat{g}(x) = \int_{\mathbb{R}} (iy)g(y)e^{ixy}dy = i\int_{\mathbb{R}} f(y)e^{ixy}dy = i\hat{f}(x).$$

Therefore

$$\hat{f}(x) = -i\left(\frac{1}{x^2+1} + i\frac{x}{x^2+1}\right)' = -i\left(\frac{-2x}{(x^2+1)^2} - i\frac{x^2-1}{(x^2+1)^2}\right)$$

$$= \frac{1-x^2}{(x^2+1)^2} + i\frac{2x}{(x^2+1)^2} = \left(\frac{1+ix}{1+x^2}\right)^2 = (1-ix)^{-2}$$

where the last identity is obtained by using the fact that $\frac{1}{1-ix} = \frac{1+ix}{1+x^2}$.