Functional Analysis

Daniele Zago

November 22, 2021

Contents

Lecture 1: Introduction to measure theory

2021-10-19

The course contents will be mostly on measure theory with some basic functional analysis notions about Banach and Hilbert spaces.

1.1 Measure spaces

Def. (σ-algebra)
Let X be an arbitrary set and Σ be a collection of subsets of X. Σ is called a ***σ-algebra*** if it is closed w.r.t. to complement and countable unions:

i $A \in \Sigma \implies A^c = X \setminus A \in \Sigma$.

ii $A_1, A_2, \ldots \in \Sigma \implies \bigcup_{n\in\mathbb{N}} A_i \in \Sigma$.

Remark The smallest σ-algebra is $\{\emptyset, X\}$, whereas the biggest possible σ-algebra is the family of all subsets of X, $\mathscr{P}(X) = \{A : A \subseteq X\}$.

Remark In principle we need to consider σ-algebras which are strictly smaller than $\mathscr{P}(X)$, since a coherent definition of measure is very difficult to define on it.

Def. (σ-algebra generated by a set)
If $C \subseteq \mathscr{P}(X)$ is a family of subsets of X, then the ***σ-algebra generated by C*** is the smallest σ-algebra on X which contains C,

$$\sigma(C) = \bigcap_{i:C\subseteq\mathcal{F}} \mathcal{F}, \quad \mathcal{F} \text{ is a } \sigma\text{-algebra of } X.$$

Example (Borel σ-algebra)
If we consider $X = \mathbb{R}$, then the *Borelian σ-algebra* $\mathcal{B}$ is the σ-algebra generated by all open intervals (a, b), $a, b \in \mathbb{R}$. That is, if $C = \{(a, b) : a < b, a, b \in \mathbb{R}\}$, then

$$\mathcal{B} = \sigma(C).$$

In any set X where we can define a *topology*, that is, a notion of open and closed sets, we can also define its associated Borel σ-algebra as the smallest σ-algebra that contains all open sets.

Under the above-defined σ-algebra, all intervals of the form

$$[a, b], [a, b), (a, b], (a, \infty), [a, \infty), (-\infty, b], (-\infty, b)$$

are all contained in $\mathcal{B}$. This is easy to prove by using the property of closeness w.r.t. to countable unions and complements. For instance,

$$(a, +\infty) = \bigcup_{i=1}^{\infty} (a, a+i) \in \mathcal{B}$$

$$\mathbb{R} \setminus (a, +\infty) = (-\infty, a] \in \mathcal{B}.$$

Moreover, $\mathcal{B}$ can be equivalently defined as the σ-algebra generated by the following sets:

$$\sigma\big(\{(a,b) : a < b\}\big)$$

$$\sigma\big(\{(a,b] : a < b\}\big)$$

$$\sigma\big(\{[a,b] : a < b\}\big)$$

$$\sigma\big(\{[a,b) : a < b\}\big)$$

Remark The obvious question is: why would we need a notion of Borel σ-algebra when defining a measure on $\mathbb{R}$ (and, by extension, on $\mathbb{R}^d$)? The reason is that we can construct some subsets A of $\mathbb{R}$ such that $A \notin \mathcal{B}$, for example the Vitali set. On those pathological sets it is not possible to define a function such that it follows the properties that we expect from a *measure*.

Def. (Measure)
Let (X, Σ) be a measurable space, then we say that a function $\mu : \Sigma \to [0, +\infty]$ is a ***measure*** if μ is such that

- $\mu(\emptyset) = 0$
- (σ-*additivity*): $\forall A_i \in \Sigma,\ A_i \cap A_j = \emptyset$, then $\mu\big(\bigcup_{i=1}^{\infty} A_i\big) = \sum_{i=1}^{\infty} \mu(A_i)$.

In this case we say that the triple (X, Σ, μ) is a ***measure space***.

Finiteness If μ is such that it assigns finite mass to X, i.e. $\mu(X) < \infty$, then μ is said to be a *finite measure*. A finite measure with $\mu(X) = 1$ is called a *probability measure*.

$\boldsymbol{\sigma}$-finiteness If X can be written as a countable union of sets, $X = \bigcup_{i=1}^{\infty} A_i$ such that $\mu(A_i) < \infty$ for all i, then μ is σ-*finite*.

General properties of a measure Given a measure space (X, Σ, μ), we can prove some general properties starting from the definition of μ:

a) (*Monotonicity*): if $A, B \in \Sigma$ are such that $A \subseteq B$, then $\mu(A) \leq \mu(B)$.

b) (σ-*subadditivity*): if $(A_i)_{i \in \mathbb{N}}$ is a sequence of elements of Σ, then

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) \leq \sum_{i=1}^{\infty} \mu(A_i).$$

c) (*Continuity*): if $(A_i)_{i\in\mathbb{N}}$ is a monotone increasing sequence of elements of Σ such that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$, then

$$\mu\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} \mu(A_i).$$

d) If $A_{i+1} \subseteq A_i$ at each i we have

$$\mu(\bigcap_i A_i) = \lim_{i\to\infty} \mu(A_i).$$

e) (*Continuity ii*): if $(A_i)_{i\in\mathbb{N}}$ is a monotone decreasing sequence of elements of Σ such that $A_{i+1} \subseteq A_i$ and $\mu(A_{i_0}) < \infty$ for some i_0, then

$$\mu\Big(\bigcap_{i=1}^{\infty} A_i\Big) = \lim_{i\to\infty} \mu(A_i).$$

1.2 Characterization of Borel measures

We try to characterize all measures on the Borel σ-algebra of $\mathbb{R}$, which are called *Borel measures*. To do so, let $F : \mathbb{R} \longrightarrow \mathbb{R}$ be an increasing function which is right-continuous, that is

$$\lim_{x\to a^+} F(x) = F(a).$$

Def. (Measure induced by F)

Let F be a function defined as above, we define the ***measure induced by F*** as the function μ_F such that for a set $(a, b]$ we have

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

For simplicity we define $\mu_F(\emptyset) = 0$.

σ-additivity When restricted to $\mathcal{C} = \{(a, b] : a < b\}$ the measure μ_F is *non-negative*, *additive*, *continuous w.r. to increasing sequences of sets*. Therefore, we have that μ_F is also σ-*additive* on the σ-algebra $\mathcal{B}$.

Example (Measure of some particular sets)

We note that for some particular choices of sets,

$$\mu_F(\mathbb{R}) = \mu_F\big(\bigcup_n (a - n, b + n)\big) = \sup_x F(x) - \inf_x F(x)$$

$$\mu_F(\{a\}) = \mu_F(c, a] - \mu_F(c, a) = F(a) - \lim_{x\to a^-} F(x)$$

To complete the characterization of Borel measures we have the following theorem, which roughly states that defining a measure on $\mathcal{B}$ can be done simply by defining a measure on $\mathcal{C}$.

Thm. 1 (Carathéodory existence theorem)
Let $(X, \mathcal{B})$ be a measurable space with Borel σ-algebra $\mathcal{B}$. Then,

i. *There exists a unique Borel measure $\bar{\mu}_F$ which coincides with μ_F on the intervals $(a, b]$ and is σ-finite $\iff$ $\sup_x F(x) - \inf_x F(x) < \infty$.*

ii. *Given a Borel measure μ on $\mathbb{R}$ there exists a monotone increasing and right-continuous function F defined as*

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

such that μ is equal to the measure induced by F,

$$\mu = \mu_F.$$

Proof.
No.

□

We now have the following relationship, which completely characterizes the Borel measures on $\mathbb{R}$.

$$\boxed{\text{Increasing right-continuous } F \iff \text{Borel measure } \mu_F \text{ on } \mathbb{R}}$$

From this result we can now define the usual notion of measure (i.e. size) of sets in $\mathbb{R}$, which will be the building block for the formalization of the notion of integration over general measure spaces.

Def. (Lebesgue measure)
Let $F(x) = x$ for all $x \in \mathbb{R}$, then if we define $\mu_F = \mu(a, b] = b - a$ $\bar{\mu}_F$ its completion via the Carathéodory is called the ***Lebesgue measure*** on $\mathbb{R}$ and we indicate it by $\mathcal{L}$.

Properties of the Lebesgue measure

i.

ii.

iii.

iv.

v.

1.3 Decomposition of measures

Before introducing the Lebesgue integral, we now describe a useful characterization of the relationship between measures, which is useful to generalize the notion of a random variable from the dichotomy discrete/continuous to a more fundamental description.

Def. (Absolutely continuous measure)
Let ν, μ be measures defined on (X, Σ), then we say that ν is ***absolutely continuous with respect to μ*** and we write $\nu \ll \mu$ if for every $A \in \Sigma$ it holds that

$$\mu(A) = 0 \implies \nu(A) = 0.$$

Prop. 1 (Absolutely continuous measure induced by a density)
Let $f \geq 0$ be a measurable function such that for all $m > 0$, $\int_{-m}^{m} f(x)\,\mathrm{d}x < \infty$. Then, the function ν_f defined as

$$\nu_f(A) = \int_A f(x)\,\mathrm{d}x$$

is a measure on $(\mathbb{R}^n, \mathcal{M})$ which is both σ-finite and absolutely continuous w.r.t. to $\mathcal{L}$. if $f \in L^1(\mathbb{R}^n)$, then the measure is also finite.

Def. (Singular measure)
Let ν, μ be measures defined on (X, Σ), then we say that ν is ***singular with respect to μ*** and we write $\nu \perp \mu$ if there exist $A, B \in \Sigma$ such that

$$\begin{aligned} A \cap B &= \emptyset && \text{(disjoint)} \\ A \cup B &= X && \text{(partition)} \\ \nu(A) = 0 &= \mu(B) && \text{(measures are orthogonal)} \end{aligned}$$

Example (Dirac measure is singular)
Consider the Dirac measure δ_{x_0} centered on a point x_0, then we can write

$$\mathbb{R} = \big(\underbrace{\mathbb{R} \setminus \{x_0\}}_{A} \cup \underbrace{\{x_0\}}_{B}\big),$$

and we have that $\mathcal{L}(B) = 0 = \delta_{x_0}(A)$.

We now state a fundamental theorem which completely characterizes the relationship of continuity and singularity between two measures.

Thm. 2 (Lebesgue decomposition)
*Let ν, μ be two σ-finite measures on a measurable space (X, Σ), then there exist two unique measures η (**absolutely continuous part**) and ρ (**singular part**) such that*

$$\begin{aligned} &\nu = \eta + \rho, \\ &\eta \ll \mu, \\ &\rho \perp \mu. \end{aligned}$$

Lecture 2: Lebesgue integration

2021-10-26

We are now ready to define the notion of Lebesgue integration, which greatly extends integrability beyond the simpler Riemann integral.

2.1 Measurable functions

Def. (Measurable function)
Let (X, Σ) and $(Y, \mathcal{E})$ be two measurable spaces and let $f : X \longrightarrow Y$ be a function. We say that f is ***measurable*** with respect to $\mathcal{E}$ and Σ if for all $E \in \mathcal{E}$ we have that

$$f^{-1}(E) \in \Sigma.$$

Equivalent property For the special case of $f : X \to \mathbb{R}$ we are interested in the following equivalent condition:

$$f^{-1}(t, +\infty) = \{x \in X : f(x) > t\} \text{ is measurable.}$$

Starting from the notion of measurable function, we are now ready to define what it means for a sequence of functions to converge.

Def. (Convergence in measure)
Let $(f_n)_{n\in\mathbb{N}}$ be a sequence of measurable functions and f be a measurable function, all defined on the measure space (X, Σ, μ). Then, ***f_n converge to f in measure*** if for every $\varepsilon > 0$ we have

$$\lim_{n\to\infty} \mu\big(\{x \in X : |f_n(x) - f(x)| \geq \varepsilon\}\big) = 0.$$

Convergence in probability If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, this convergence is called *convergence in probability* since it reads

$$\lim_{n\to\infty} \mathbb{P}\big(\{\omega \in \Omega : |X_n(\omega) - X(\omega)| \geq \varepsilon\}\big) = 0.$$

2.2 Lebesgue integration

Both the definition of Lebesgue integration and the discussion of its properties are always developed in three stages. First, the definition of some property that holds for a simple class of functions; second, the simple functions are used to approximate some positive measurable function f of interest; finally, we can extend the property to any measurable function f by considering it as a sum of its positive and negative parts, f^+ and f^-.

Def. (Simple function)
Let $A_1, A_2, \ldots, A_K$ be a finite family of disjoint sets and $c_1, c_2, \ldots, c_K > 0$ positive constants. We say that φ is a ***simple function*** if

$$\varphi(x) = \sum_{i=1}^{K} x_i \mathbb{1}_{A_i}(x).$$

Measurability of φ A simple function φ is a step function with a finite number of jumps, which can be immediately proven to be measurable: for a single indicator function,

$$t \geq 1 \longrightarrow \{x : \mathbb{1}_A(x) > t\} = \emptyset \in \mathcal{B}$$

$$t < 1 \longrightarrow \{x : \mathbb{1}_A(x) > t\} = A \in \mathcal{B}.$$

Then, for a simple function as defined above we can extend this argument by considering the indicator function of the union $\bigcup_{i=1}^{K} A_i$.

When considering a simple function we can give an intuitive definition of the Lebesgue integral, which represents the total area under the graph of the function.

Def. (Lebesgue integral of a simple function)
Let $\varphi(x) = \sum_{i=1}^{K} c_i \mathbb{1}_{A_i}(x)$, then we define the ***Lebesgue integral of φ*** as the linear functional

$$\int_{\mathbb{R}^N} \varphi(x)\,\mathrm{d}x = \sum_{i=1}^{K} c_i \mathcal{L}(A_i).$$

For a general measurable function f, we can extend the notion of the Lebesgue integral by approximating it from below with simple functions and taking the best approximation over all possible simple functions.

Def. (Lebesgue integral of a positive function)
Let $f : \mathbb{R}^n \longrightarrow R$ be a measurable function such that $f(x) \geq 0$, then we define its ***Lebesgue integral*** as

$$\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = \sup\left\{ \int_{\mathbb{R}^n} \varphi(x)\,\mathrm{d}x : \varphi \text{ simple function and } \varphi < f \right\}.$$

In the general case where f is not positive we define instead the integral in terms of its positive and negative parts, when such an operation is well-defined.

Def. (Lebesgue integral of a function)
Let $f : \mathbb{R}^n \longrightarrow R$ be a measurable function, possibly negative. Let now its ***positive part*** be $f^+(x) = \max\{0, f(x)\}$ and its ***negative part*** be $f^-(x) = \max\{0, -f(x)\}$. Then, we define the ***Lebesgue integral of f*** as

$$\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} f(x)^+\,\mathrm{d}x - \int_{\mathbb{R}^n} f(x)^-\,\mathrm{d}x,$$

whenever this subtraction is well-defined.

Properties of the Lebesgue integral

› If $f = 0$ almost everywhere, then $\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = 0$. Conversely, if $f \geq 0$ is measurable and $\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = 0$ then $f = 0$ almost everywhere.

› If f and g are measurable functions such that $f = g$ almost everywhere, then $\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} g(x)\,\mathrm{d}x$

› If $f, g \in L^1(\mathbb{R}^n)$, i.e. are absolutely integrable, then

$$\int_{\mathbb{R}^n} \big(\alpha f(x) + \beta g(x)\big)\,\mathrm{d}x = \alpha \int_{\mathbb{R}^n} f(x)\,\mathrm{d}x + \beta \int_{\mathbb{R}^n} g(x)\,\mathrm{d}x.$$

› If $f, g \in L^1(\mathbb{R}^n)$ and $f \le g$ almost everywhere, then

$$\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x \le \int_{\mathbb{R}^n} g(x)\,\mathrm{d}x.$$

This definition of the integral is especially useful when dealing with sequences of functions for which we want to establish some sort of convergence of integrals.

Thm. 3 (Monotone convergence)
Let $f_k : \mathbb{R}^n \to \mathbb{R}$ be a sequence of measurable functions such that $f_k \ge 0$ and $f_k(x) \le f_{k+1}(x)$ for all x and for all k, then

$$\lim_{k\to\infty} \int_{\mathbb{R}^n} f_k(x)\,\mathrm{d}x = \int_{\mathbb{R}^n} \lim_{k\to\infty} f_k(x)\,\mathrm{d}x.$$

Proof.
No.

□

Prop. 2 (Repartition function)
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable positive function, and let for every $t > 0$ $F(t)$ be the ***repartition function of f****,*

$$F(t) = \mathcal{L}\{x : f(x) > t\}.$$

Then, it holds that

$$\int_{\mathbb{R}^n} f(x)\,\mathrm{d}x = \int_0^\infty F(t)\,\mathrm{d}t.$$

Proof.
No.

□

Lecture 3: Banach spaces

2021-11-09

We start by defining some useful spaces in functional analysis, namely L^p and M^p spaces.

3.1 L^p spaces

Def. (L^p spaces)
For $p \in [1, \infty)$ we define the following vectorial space

$$L^p(A) = \Big\{ f : \mathbb{R}^n \to \mathbb{R} : f \text{ is measurable and } \int_A |f(x)|^p \, \mathrm{d}x < \infty \Big\}.$$

For $p = \infty$, we define

$$L^\infty(A) = \Big\{ f : R^n \to \mathbb{R} : f \text{ is measurable and } |f(x)| \leq c \text{ for almost every } x \in A \Big\}.$$

Def. (M^p spaces)
When we consider $A = \Omega$, we have the space of random variables with finite p^{th} moment,

$$M^p = \Big\{ X : \Omega \to \mathbb{R} : X \text{ is measurable and } \mathbb{E}[|X|^p] < \infty \Big\},$$

with an analogous definition for $p = \infty$,

$$M^\infty = \Big\{ X : \Omega \to \mathbb{R} : X \text{ is measurable and } |X(\omega)| \leq c \text{ for almost every } \omega \in \Omega \Big\}.$$

Convergence With these definitions, it's immediate to define a notion of convergence in terms of p spaces for general functions (L^p) and for random variables (M^p):

- ***convergence in p-space*** for a sequence of functions $(f_n)_{n \in \mathbb{N}}$,

$$f_n \xrightarrow{L^p} f \iff \int_{\mathbb{R}} |f_n(x) - f(x)|^p \, \mathrm{d}x \xrightarrow{n \to \infty} 0,$$

- ***convergence in p-mean*** for a sequence of random variables $(X_n)_{n \in \mathbb{N}}$

$$X_n \xrightarrow{L^p} X \iff \mathbb{E}[|X_n - X|^p] \xrightarrow{n \to \infty} 0.$$

These spaces are particularly important for functional analysis since they are examples of Banach spaces, whose structure we are going to study more generally.

We now state some inequalities which are useful for studying L^p and M^p spaces in their generality.

Def. (Conjugate exponent)
Let $p > 1$, then the ***conjugate exponent*** of p is $q > 1$ such that

$$\frac{1}{p} + \frac{1}{q} = 1 \iff q = \frac{p}{p-1}.$$

Moreover, if $p = 1$ we say that its conjugate exponent is $q = +\infty$ and vice versa.

Thm. 4 (Young's inequality)
Let p, q be conjugate exponents, then for all $x, y > 0$ we have that

$$xy \leq \frac{x^p}{p} + \frac{y^q}{q}.$$

Remark This property is a generalization of the classic identity $(a-b)^2 > 0 \implies ab < \frac{a^2}{2} + \frac{b^2}{2}$.

Thm. 5 (Hölder's inequality)
Let $p, q \in [1, \infty]$ be conjugate exponents, i.e. $\frac{1}{p} + \frac{1}{q} = 1$, then for all measurable functions f on $O \subseteq \mathbb{R}^n$ we have

$$\|fg\|_1 \leq \|f\|_p \|g\|_q,$$

which if expanded becomes

$$\int_O |f(x)g(x)| \,\mathrm{d}x \leq \left(\int_O |f(x)|^p\right)^{\frac{1}{p}} \left(\int_O |g(x)|^q\right)^{\frac{1}{q}}.$$

Moreover, if $f, g \in L^p(O)$ then $fg \in L^1(O)$ and this becomes an equality $\iff$ $|f|^p$ and $|g|^q$ are linearly dependent in $L^1(O)$.

Corollary 1 (Minkowski's inequality)
Let $f, g \in L^p(O)$, then we have that $\|\cdot\|_p$ satisfies the triangle inequality, i.e.

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p,$$

which if expanded becomes

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

$\|\cdot\|_p$ is a norm This corollary states that $\|\cdot\|_p$ is indeed a norm on L^p, which is therefore a *normed vector space.*

3.2 Banach spaces

Def. (Distance induced by a norm)
Let $(X, \|\cdot\|)$ be a normed vectorial space, then we say that the ***distance induced by the norm*** $\|\cdot\|$ is the function

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

Distance $d(\cdot, \cdot)$ as defined above is indeed a distance between two elements of X.

Topology This distance lets us define an induced topology on X – i.e. a notion of open and closed sets – firstly by defining the ***balls of radius*** r ***centered in*** x_0 as

$$B(x_0, r) = \{x \in X : d(x_0, x) < r\},$$

and then defining a set $A \subseteq X$ as ***open*** if for all $x \in A$ there exists $r > 0$ such that $B(x, r) \subset A$.

Def. (Banach space)
Let $(x_n)_{n\in\mathbb{N}}$ be a Cauchy sequence, i.e. for all $\varepsilon > 0$ there exists a $N \in \mathbb{N}$ such that $\|x_n - x_m\| < \varepsilon$ for all $n, m > N$. If all such sequences have limit in X, then X is called a ***Banach space***.

Completeness This property is called ***completeness***, and the definition states that "a Banach space is a complete normed vector space".

Thm. 6 (L^p are Banach spaces)
The spaces $L^p(\mathbb{R})$, $p \in [1, \infty]$ are Banach spaces w.r. to the distance induced by the norm

$$\|f\|_p = \left(\int_{\mathbb{R}} |f(x)|^p \,\mathrm{d}x\right)^{\frac{1}{p}}, \quad p \in [1, \infty),$$

and

$$\|f\|_\infty = \sup_x |f(x)|.$$

M^p spaces The theorem can be stated in terms of M^p being Banach spaces, when endowed with the norms

$$\|X\|_p = \mathbb{E}[|X|^p]^{\frac{1}{p}}, \quad p \in [1, \infty),$$

$$\|X\|_\infty = \sup_\omega |X(\omega)|, \quad p = \infty.$$

Proof.

□

Lecture 4: L^p spaces and bounded linear operators

2021-11-10

Recall that

$$M^p = \{X : \Omega \to \mathbb{R}, \text{ such that } \mathbb{E}[|X^p|] < \infty\}$$

$$L^p = \{f : \mathbb{R} \to \mathbb{R}, \text{ such that } \int_{\mathbb{R}} |f(x)|^p \, \mathrm{d}x < \infty\}$$

These are actually the same spaces if you consider $M^p = L^p(\Omega)$, indeed you can see that

$$\mathbb{E}[|X|^p] = \int_\Omega |x|^p \, \mathrm{d}\mathbb{P}(\omega)$$

Prop. 3 (L^p spaces inclusion)
$M^1 \supset M^2 \supset \dots$, and in general $M^n \subset M^k$ if $n \geq k$. In particular, if X is a random variable such that $\mathbb{E}[|X|^k] < \infty$, then $\mathbb{E}[|X|^n] < \infty$ for all $n \leq k$.

Proof.
By Jensen's inequality, if $f : \mathbb{R} \longrightarrow \mathbb{R}$ *convex* ($\implies$ meas. and cont.) and X is a random variable, then the random variable $f(X)$ is such that

$$\mathbb{E}[f(X)] \geq f\big(\mathbb{E}[X]\big).$$

We fix $n \geq k \geq 1$ and $X \in M^n \implies \mathbb{E}[|X|^n] < \infty$. We want to prove that also $X \in M^k$: we therefore fix $f(x) = |x|^{\frac{n}{k}}$, and we see that

$$\frac{n}{k} \geq 1 \implies f \text{ is a convex function.}$$

Applying Jensen's inequality,

$$\mathbb{E}[f(Y)] \geq f\big(\mathbb{E}[Y]\big)$$

we use the functions f and r.v. $Y = |X|^k$ to see that

$$\begin{cases} f(Y) = \big(|X|^k\big)^{\frac{n}{k}} = |X|^n \\ \mathbb{E}[f(Y)] = \mathbb{E}[|X|^n] < \infty \quad \text{by assumption} \end{cases}$$

and so we obtain $\infty > \mathbb{E}[|X|^n] \geq \mathbb{E}[|X|^k]^{\frac{n}{k}}$, by taking square roots we have

$$\infty > \mathbb{E}[|X|^n]^{\frac{1}{n}} \geq \mathbb{E}[|X|^k]^{\frac{1}{k}},$$

from which we conclude that $X \in M^n \implies X \in M^k$ for all $1 \leq k \leq n$ and moreover $\|X\|_k \leq \|X\|_n$.

Finally, we observe that this implies a relationship w.r. to convergence for all $1 \leq k \leq n$:

$$X_n \xrightarrow{M_n} X \implies X_n \xrightarrow{M_k} X \quad \text{for all } 1 \leq k \leq n.$$

$\square$

***L^p* spaces inclusion** This proof is analogous when considering more general $L^p(A)$ spaces such that $\mathcal{L}(A) < \infty$, since M^p is a special case of a L^p space when choosing $A = \Omega$. Another proof can be obtained by applying Hölder's inequality:

Proof.
Consider $f \in L^n(A)$, we want to prove that $f \in L^k(A)$ for $1 \leq k \leq n$. Since $f \in L^n(A)$,

$$\int_{\mathbb{R}} |f(x)|^n \mathbb{1}_A(x) \, \mathrm{d}x < \infty,$$

which means that $f(x)\mathbb{1}_A(x) \in L^n(\mathbb{R})$. Moreover, its' trivial to see that $\mathbb{1}_A(x) \in L^p(\mathbb{R})$ for all p if $\mathcal{L}(A) < \infty$, since

$$\int_{\mathbb{R}} |\mathbb{1}_A(x)|^p \, \mathrm{d}x = \mathcal{L}(A).$$

Now, let $q = \frac{n}{n-1}$ be the conjugate exponent of n, we have that

$$\begin{cases} f\mathbb{1}_A \in L^n(\mathbb{R}) \\ \mathbb{1}_A \in L^{\frac{n}{n-1}}(\mathbb{R}) \end{cases}$$

then by Hölder's inequality, $(f\mathbb{1}_A) \cdot \mathbb{1}_A \in L^1(\mathbb{R})$ and we see that

$$\begin{aligned} \|f\|_{L^1(A)} = \int_{\mathbb{R}} |f(x)|\mathbb{1}_A(x) \, \mathrm{d}x &\leq \|f\|_{L^n(A)} \Big(\int_{\mathbb{R}} \mathbb{1}_A(x)^{\frac{n}{n-1}} \, \mathrm{d}x \Big)^{\frac{n-1}{n}} \\ &= \|f\|_{L^n(A)} \cdot \big(\mathcal{L}(A)\big)^{\frac{n-1}{n}} \end{aligned}$$

Therefore, we conclude that $L^n(A) \subseteq L^1(A)$ for any $n \geq 1$. With a similar argument, by accurately choosing the exponent k we also have that $f \in L^k(A)$ for all $1 \leq k \leq n$,

$$f \in L^k(A) \iff |f|^k \in L^1(A).$$

We have that

$$|f|^k \in L^{\frac{n}{k}}(A) \quad \frac{n}{k} > 1$$

since $\big(|f|^k\big)^{\frac{n}{k}} = |f|^n$ and by choosing the conjugate exponent of $\frac{n}{k}$,

$$q = \frac{\frac{n}{k}}{\frac{n}{k} - 1}$$

$\square$

Remark

- Recall that if $L(A) = \mathbb{R}$ then $\mathcal{L}(A) = \infty$ and the inclusion relationship is not true, since we proved before that $L^2(\mathbb{R}) \not\subseteq L^1(\mathbb{R})$.
- This holds in general for random variables since they induce a bounded measure, $\mathbb{P}(\Omega) = 1$.
- We have that $L^\infty(A) \subseteq \bigcap_{k \geq 1} L^k(A)$ but it is not strictly equal to it.

We also have a way of computing the L^k norm in terms of the L^n norm:

$$\int_A |f|^k \, \mathrm{d}x \leq \|f\|_n^k \mathcal{L}(A)^{\frac{n-k}{n}} \implies \|f\|_k \leq \|f\|_n \cdot \mathcal{L}(A)^{\frac{n-k}{nk}}.$$

4.1 Operators between Banach spaces

Since we are working with spaces whose elements are functions (or random variables), we need to define a notion of *functional*, i.e. a transformation between two functions. The notion of a *linear functional* generalizes the linear transformation of standard algebra.

Def. (Linear operator)
Let X, Y be Banach spaces, then we say that a function $T : X \longrightarrow Y$ is a ***linear operator*** if T maintains the vectorial structure, i.e.

$$T(\alpha x_1 + \beta x_2) = \alpha T(x_1) + \beta T(x_2) \quad \text{for all } x_1, x_2 \in X \text{ and } \alpha_1, \alpha_2 \in \mathbb{R}.$$

Def. (Continuity)
We say that $T : X \to Y$ is a ***continuous*** operator if for every sequence $x_n \in X$ such that $x_n \to x \in X$, then $Tx_n \to Tx$, i.e. converging sequences are mapped in converging sequences.

Def. (Boundedness)
We say that $T : X \to Y$ is ***bounded*** if there exists a constant $C > 0$ such that

$$\|Tx\|_Y \leq C\|x\|_X \quad \text{for all } x \in X.$$

Remark Differently from standard functions $f : \mathbb{R} \to \mathbb{R}$ we don't require the image to be bounded, since such a condition turns out to be too restrictive for linear operators between Banach spaces.

Prop. 4 (Continuity implies boundedness)
If X, Y are Banach spaces and $T : X \longrightarrow Y$ is a linear operator, then T is continuous if and only if T is bounded.

Equivalence Because of this theorem, we will always talk about bounded operators instead of continuous operators, since they are the same.

Proof.
$\boxed{\text{Bounded} \implies \text{continuous}}$: Let $x_n \in X$ such that $x_n \xrightarrow{n\to\infty} x \in X$, which by definition means that $\|x_n - x\|_X \xrightarrow{n\to\infty} 0$. From this, we want to prove that $Tx_n \xrightarrow{n\to\infty} Tx$.

Since the operator is bounded, there exists a C such that for all $x \in X$

$$\|Tx\|_Y \leq C\|x\|_X,$$

and applying this to $y = x_n - x \in X$ we have

$$0 \leq \|T(x_n - x)\|_Y \leq C\|x_n - x\|_X \xrightarrow{n\to\infty} 0,$$

and by linearity we have that $\|Tx_n - Tx\|_Y \to 0$.

$\boxed{\text{Continuous} \implies \text{bounded}}$: We actually prove another fact which implies this property, i.e. T not bounded $\implies$ T not continuous, by constructing a sequence which is not converging under the map T. Assume that T is not bounded, then for any $C > 0$ we can always find at least a point $x \in X$ such that

$$\|Tx\|_Y > C\|x\|_X,$$

In particular, let $C = n$, then we can always find a $x_n \in X$ such that

$$\|Tx_n\|_Y > n\|x_n\|_X \quad \text{for } \|x_n\|_X \neq 0$$

Consider now the sequence of points $y_n = \frac{x_n}{n\|x_n\|_X}$, for which we have that $y_n \in X$ since X is a vectorial space. Then,

$$\|y_n\|_X = \left\| \frac{x_n}{n\|x_n\|_X} \right\| \overset{\text{norm.}}{=} \frac{1}{n}.$$

In particular, this means that $\|y_n - 0\|_X = \frac{1}{n} \xrightarrow{n\to\infty} 0$ and if T is continuous, then we should see that $Ty_n \to T0 = 0$ in Y since it is a linear operator. Now,

$$Ty_n = T\left(\frac{x_n}{n\|x_n\|_X}\right) \overset{\text{lin.}}{=} \frac{1}{n\|x_n\|_X} T(x_n),$$

and

$$\|Ty_n\|_Y = \left\| \frac{1}{n\|x_n\|_X} T(x_n) \right\|_Y \overset{\text{lin.}}{=} \frac{1}{n\|x_n\|_X} \|Tx_n\|_Y \overset{\text{Hp.}}{>} \frac{1}{n\|x_n\|_X} n\|x_n\|_X = 1,$$

therefore Ty_n does not converge to 0 in Y and therefore we prove the desired property.

□

Thm. 7 (Space of linear operators)
Let X, Y be Banach spaces, then we have that the set of linear operators between X and Y,

$$\mathcal{B}(X,Y) = \{T : X \to Y \text{ linear bounded operators}\},$$

is a Banach space with norm given by

$$\|T\| = \sup_{\|X\|_X \leq 1} \|Tx\| \overset{lin.}{=} \sup_{x \neq 0} \frac{\|Tx\|_Y}{\|x\|_X}.$$

Proof.

› *Vectorial space*: $\mathcal{B}(X,Y)$ is naturally a vectorial space, since $\alpha T + \beta S$ is a linear operator defined by

$$(\alpha T + \beta S)(x) := \alpha T(x) + \beta S(x).$$

Moreover, if T, S are bounded then $\alpha T + \beta S$ is also bounded by

$$\|\alpha Tx + \beta Sx\|_Y \leq \alpha \|Tx\|_Y + \beta \|Sx\|_Y \leq \underbrace{|\alpha| \cdot C\|x\|_X}_{T \text{ bounded}} + \underbrace{|\beta| \cdot D\|x\|_X}_{S \text{ bounded}}$$
$$= \big(|\alpha|C + |\beta|D\big)\|x\|_X.$$

› *Norm*: $\|T\|$ as defined in the theorem above is indeed a norm since it satisfies the three properties,

1. $\|T\| \geq 0$ for all T.
2. $\|\alpha T\| = \sup\limits_{\|X\|_X \leq 1} \|\alpha Tx\| = |\alpha| \cdot \sup\limits_{\|x\| \leq 1} \|Tx\| = |\alpha| \cdot \|T\|$.
3. $\|T + S\| = \sup\limits_{\|x\| \leq 1} \|Tx + Sx\|_Y \leq \sup\limits_{\|x\| \leq 1} \|Tx\|_Y + \sup\limits_{\|x\| \leq 1} \|Sx\|_Y$.

› *Completeness*: We are not going to prove that $\mathcal{B}(X, Y)$ is a complete space w.r. to convergence induced by the norm, since it is a bit complex.

□

Another important result for linear bounded operators is the following, which is a theorem that can be used as an intermediate step in the proof of the completeness of $\mathcal{B}(X, Y)$.

Thm. 8 (Banach-Steinhaus)

If T_n is a sequence of bounded linear operators from X to Y and for every $x \in X$ there exists $\lim\limits_{n\to\infty} T_n x$ in Y then the operator defined by

$$Tx := \lim_{n\to\infty} T_n x$$

is a bounded linear operator.

Example (Linear bounded operators)

Consider for a fixed $X \in M^p$ and $q = \frac{p}{p-1}$ the operator defined by

$$\begin{aligned} T : M^q &\longrightarrow \mathbb{R} \\ Y &\longmapsto \mathbb{E}[X \cdot Y]. \end{aligned}$$

We prove that for a fixed X, T is a linear bounded operator.

Linearity is immediate because of linearity of the $\mathbb{E}[\cdot]$ operator. As for boundedness we have to show that $\exists C > 0$ such that $\|Tx\|_{\mathcal{Y}} \leq C\|x\|_{\mathcal{X}}$, where $\mathcal{X} = M^q$ and $\mathcal{Y} = \mathbb{R}$. For all $Y \in M^q$ we therefore want to check that

$$|\mathbb{E}[X \cdot Y]| \overset{?}{\leq} C \cdot \mathbb{E}[\|Y\|^q]^{\frac{1}{q}}.$$

We can do so by applying Hölder's inequality, which allows us to write

$$|\mathbb{E}[X \cdot Y]| \overset{\text{Jens.}}{\leq} \mathbb{E}[|X \cdot Y|] \overset{\text{Höld}}{\leq} \underbrace{\mathbb{E}[|X|^p]^{\frac{1}{p}}}_{C} \cdot \mathbb{E}[|Y|^q]^{\frac{1}{q}},$$

and the operator is bounded the constant $C = \mathbb{E}[|X|^p]^{\frac{1}{p}}$ (recall that X is fixed).

Moreover, we can actually prove (exercise) that the norm of T is equal to

$$\|T\| = \mathbb{E}[|X|^p]^{\frac{1}{p}}$$

Example (Set of matrix operators)
If we consider $T : \mathbb{R}^n \to \mathbb{R}^m$, then the set of bounded linear operators is the set of operators defined by the $M_{m\times n}(\mathbb{R})$ matrices,

$$x \mapsto Tx = Ax, \quad A \in M_{m\times n}(\mathbb{R}).$$

Moreover, the norm $\|T\|$ is connected to the norm of the matrix A.

Linear algebra From the example above, we can interpret the space $\mathcal{B}(X, Y)$ as the infinite-dimensional generalization of the space of matrices,

$$\text{Space of } m \times n \text{ matrices} \quad \xrightarrow{\text{Infinite dim.}} \quad \mathcal{B}(X, Y),$$

and the results we are going to prove for infinite-dimensional linear operators are similar to those of standard finite-dimensional vector spaces.