Intro to Measure Theory

Andrea Aveni

August 11, 2021

Abstract

This is **not** a mandatory reading. You just have some definitions and facts that could be useful during this course or in your future research.

1 What is Measure Theory?

Is a rigorous (i.e. mathematical) way to associate real numbers to subsets of $\mathbb{R}$, $\mathbb{R}^n$, the Hilbert cube H or any interesting reference space X. These numbers measure "how big" the subset is.
Probabilities are a special case of measures where the "dimension" of the reference space X is chosen to be one.

2 Topological Preliminaries

It is possible to construct measures on any set X, but usually it is more useful to consider measures on particularly nice sets. Here we introduce some fundamental examples.

2.1 Basic definitions

Definition. *A set X, together with a collection of subsets $\tau \subset 2^X$ is a **topological space** if*

- $\emptyset, X \in \tau$
- *τ is closed under finite intersections*
- *τ is closed under arbitrary unions*

*The sets of the collection τ are called **open sets**. The complement of an open set is said to be **closed**.*

If a set A is open, this does NOT imply that it is not closed. For instance, in every topological space, $\emptyset$ and X are always **clopen** (i.e. open and closed at the same time).

Moreover, there are usually many sets that are neither open nor closed, indeed $(0, 1]$ is neither open nor closed in $\mathbb{R}$ with its usual topology.

On the subsets of topological spaces we can define two important operations, the interior and the closure.

Definition. *If $A \subset X$, we define its interior A^o as*

$$A^o := \bigcup_{O \in \tau : O \subseteq A} O$$

And its closure $\bar{A}$ as

$$\bar{A} = \bigcap_{C : C^c \in \tau \,\wedge\, A \subseteq C} C$$

We can immediately see that for any $A \subseteq X$, A^o is open (since it is a union of open sets) and that $\bar{A}$ is closed. Indeed, using the de Morgan laws, we can prove more,

$$\begin{aligned}(\bar{A})^c &= \bigcup_{C : C^c \in \tau \,\wedge\, A \subseteq C} C^c \\ &= \bigcup_{O : O \in \tau \,\wedge\, O \subseteq A^c} O \\ &= (A^c)^o\end{aligned}$$

Finally, if A is open, then $A^o = A$, and if A is closed, $\bar{A} = A$.
If $x \in X$ we define the neighbourhoods of x as the collection of all open sets containing x, that is

$$\mathcal{N}_x := \{U \in \tau : x \in U\}$$

Now, if A is a subset of a metric space X we can classify the points x of X in the following way

- x is an interior point of A if

$$\exists U \in \mathcal{N}_x : U \cap A = U$$

- x is an isolated point of A if

$$\exists U \in \mathcal{N}_x : U \cap A = \{x\}$$

- x is an accumulation point of A if

$$\forall U \in \mathcal{N}_x : U \cap A \setminus \{x\} \neq \emptyset$$

- x is a boundary point of A if

$$\forall U \in \mathcal{N}_x : U \cap A \setminus \{x\} \neq \emptyset \;\wedge\; U \cap A \neq A$$

It is possible to prove that

Theorem. *Let $x \in X$, then*

- *x is an interior point of A iff $x \in A^o$*
- *x is an accumulation point of A iff $x \in \bar{A}$*
- *x is a boundary point ($x \in \partial A$) iff $x \in \bar{A} \cap \overline{A^c} = \bar{A} \cap (A^o)^c$*

Now we introduce a special class of topological spaces, the metric spaces.

Definition. *A set X, together with a function $d : X \times X \to \mathbb{R}_+$ is a* ***metric space*** *if*

- *$d(x, y) = 0$ iff $x = y$ (dropping this we get a pseudo-metric space)*
- *$d(x, y) = d(y, x)$ (dropping this we get a semi-metric space)*
- *$d(x, z) \leq d(x, y) + d(y, z)$*

A metric d naturally induces a topology τ_d on X. In this topology a set A is open iff for every $x \in A$ there is a positive number ε such that $x \in B_\varepsilon(x) := \{y \in X : d(x, y) < \varepsilon\} \subseteq A$.

Example. *Verify that this is a Topology.*

As we would expect, in this topology, each ball $B_\varepsilon(x)$ is open.

Proof. Let $y \in B_\varepsilon(x)$ be different from x (if $y = x$, we can take $B_\varepsilon(x)$ itself), then by the first axiom of metric spaces, $r_y := d(x, y) \in (0, \varepsilon)$. Now, let $z \in B_{\varepsilon - r_y}(y)$, then, using the triangle inequality we get that

$$d(x, z) \leq d(x, y) + d(y, z) < r_y + \varepsilon - r_y = \varepsilon$$

Therefore, we conclude that $z \in B_\varepsilon(x)$, but since z was arbitrary, we obtain that $B_{r_y}(y) \subseteq B_\varepsilon(x)$. □

2.2 Properties of Topological Spaces

Topological spaces can be distinguished on the basis of separation properties. In what follows we will always assume every topological space X to be s.t. every singleton is closed. Then a space X is:

- **Hausdorff** if points can be separated by disjoint open sets
- **Normal** if every point and every closed set can be separated by disjoint open neighborhoods
- **Normal** if closed sets can be separated by disjoint open neighborhoods

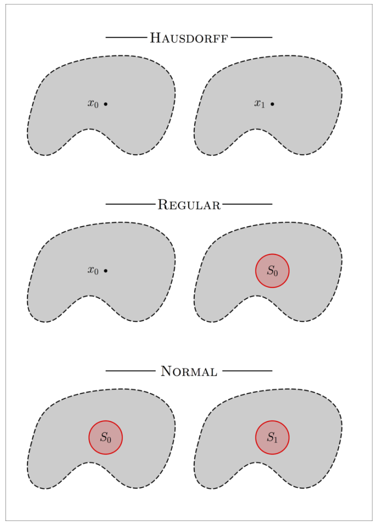

Figure 1: The main separation axioms

Topological spaces can be distinguished on the basis of their countability. A space X is:

- **First-countable** if, for every $x \in X$ we have a countable collection of open sets $\mathcal{B}_x = \{B_n\}_{n\in\mathbb{N}}$ containing x, such that, for every A open containing x, there exists some $n_A \in \mathbb{N}$ such that $B_{n_A} \subseteq A$.
- **Second-countable** if, there is a countable collection of open sets $\mathcal{B} = \{B_n\}_{n\in\mathbb{N}}$ such that, for every open set A, we can find some n_A such that $B_{n_A} \subseteq A$.
- **Separable** if it contains a countable dense subset.

2^{nd}-countability implies 1^{st}-countability and separability. The second property may seem very strict but indeed $\mathbb{R}^n$ is second countable! (We can take take as $\mathcal{B}$ all the boxes with rational endpoints.)

Finally we can look at compactness and connectedness properties.

- A space X is **Lindelöf** if for every open cover of X we can find a countable subcover. (Every 2^{nd}-contable space is Lindelöf!)
- A space X is **compact** if for every open cover of X we can find a finite subcover. (Compactness implies Lindelöfness!)
- A space X is **connected** if the only clopen sets are $\emptyset$ and X. Equivalently, if X is not the union of two disjoint open sets. Equivalently, if X is not the union of two disjoint closed sets.

- A space X is **path-connected** if for every points $x, y \in X$ we can find a continuous path from x to y.

Compactness implies Lindelöfness and path-connectedness implies connectedness.

3 Basic Theorems on compactness

Theorem. *Assuming X is Hausdorff, if A is compact, then it is closed. If A is a closed subset of a compact set X, then, A is compact.*

Proof. 1) Let $y \in A^c$, then for every $x \in A$ we can find disjoint open sets $x \in U_x$ and $y \in V_x$. Now $\{U_x\}_{x \in X}$ is an open cover of A, so we can find a finite subcover $\{U_{x_i}\}_{i=1...n}$. But then $A \subseteq \bigcup_{i=1}^n U_{x_i}$ is open and disjoint from the open set $y \in \bigcap_{i=1}^n V_{x_i}$. But then y is an interior point of A^c and so A is closed.
2) If we have a open cover $\mathcal{O}$ of A, then $\mathcal{O} \cup \{A^c\}$ is an open cover of X, so we can find a finte subcover of X from $\mathcal{O} \cup \{A^c\}$. If this finite subcover does not contain A^c, we are done, otherwise we remove it and we get a finite subcover of A. □

Theorem. *A compact set A in a metric space X is bounded.*

Proof. Let $x \in X$ be any point and consider the family $\mathcal{B} = \{B_n(x)\}$. This is an open cover of A, and we are done! □

Theorem. *HEINE-BOREL If A is a subset of $\mathbb{R}^n$, it is compact iff it is closed and bounded.*

Proof. We already proved one direction since $\mathbb{R}^n$ is Hausdorff and metric.
Now, let A be bounded and closed. Then $A \subset [-a, a]^n =: C_0$ for some $a > 0$. Now we just need to prove that this cube is compact (since the closed subsets of a compact are compact!).
Suppose we have a cover $\mathcal{U}$ of the cube without any finite subcover. We can divide the cube in 2^n subcubes of half size. One of these subcubes C_1 won't admit a finite subcover. We keep dividing obtaining a sequence of nested cubes $C_0 \supseteq C_1 \supseteq C_2 \subseteq ... \subseteq C_k \subseteq ...$ Let now $(x_k)_{k \in \mathbb{N}}$ be a sequence of points such that $x_k \in C_k$. Clearly, $x_k \to x$ for some $x \in \cap_{k=1}^{\infty} C_k \subset C_0$.
Now there must be some set U in the cover contining this x. But since this U is open, there must be some $\varepsilon > 0$ s.t. $x \in B_\varepsilon(x) \subseteq U$. But if $k > 1 + \log_2(a/\varepsilon)$, then $C_k \subseteq B_\varepsilon(x)$. And this is absurd since we assumed that each C_k didn't admit a finite subcover! □

3.1 Properties of Metric Spaces

Theorem. *The following holds*

- *Every metric space is 1^{st} countable, Hausdorff, regular and normal.*

- *A metric space is Lindelöf iff it is 2^{nd}-countable iff it is separable.*
- *(Uryshon Theorem) If X is 2^{nd} and T_3, then X is metrizable.*

Proof. We will only prove the first point.
If X is a metric space and $x \in \mathcal{N}_x$, a countable neighborhood base at x is given by $\mathcal{B}_x := \{B_{1/k}(x) : k \in \mathbb{N}\}$.
Finally, to prove the last three properties, we just need to prove normality, since regularity and Hausdorffness are implied by normality.
First of all, we notice that, if A is a subset in a metric space, then $x \notin \bar{A}$ iff $0 < d(x, A) := \inf\{d(x, y) : y \in A\}$. Indeed, if $x \in \bar{A}$, then, for any k, there is some point $y_k \in A \cap B_{1/k}(x)$, so that $d(x, A) \leq 1/k$. But since k was arbitrary, we get that $d(x, A) = 0$.
On the contrary, if $d(x, A) = 0$, then we can find a sequence $(y_k)_k$ in A such that $d(x, y_k) \leq 1/k$, and this implies that $x \in \bar{A}$.
Now, suppose that A, B are two disjoint closed sets in a metric space, then, for every $a \in A$, $d(a, B) > 0$ and for all $b \in B$, $d(b, A) > 0$. The we consider the following open sets

$$U = \bigcup_{a \in A} B_{d(a,B)/3}(a), \qquad V = \bigcup_{b \in B} B_{d(b,A)/3}(b)$$

Of, course we have that $A \subset U$ and $B \subset V$. Moreover they are disjoint. Indeed, assume $z \in U \cap V$, then $z \in B_{d(a,B)/3}(a)$ for some $a \in A$, so that $r_a := d(z, a) < d(a, B)/3$. Analogously, there is some $b \in B$ such that $r_b := d(b, z) < d(b, A)/3$. If $r_a \leq r_b$, then

$$3r_b = d(b, A) \leq d(b, a) \leq d(b, z) + d(z, a) < r_a + r_s < 2r_b$$

But this would imply that $r_b = 0$ and this is impossible. If instead, $r_b \leq r_a$, we would identically get that $r_a = 0$. $\square$

The best spaces to work with in measure theory are the **Polish** spaces that are simply separable complete metric spaces. The great thing about polish spaces is that if X is Polish, then the set of all probability measures (equipped with the so called weak topology) on X is itself Polish.

Sequences and convergence

Definition. *A sequence $(x_n)_{n \in \mathbb{N}}$ in a topological space X converges to some x if*

$$\forall U \in \mathcal{N}_x,\, \exists N_U \in \mathbb{N} : \forall n \geq N_U,\, x_n \in U$$

And in this case we write $x_n \to x$ or $\lim_{n \to \infty} x_n = x$

This definition still makes sense for metric spaces equipped with the topology induced by the metric, however in this case a simpler condition works

Theorem. *If X is a metric space with metric d and with topology τ_d, then $x_n \to x$ iff*

$$\forall k \in \mathbb{N}, \exists N_k \in \mathbb{N} : \forall n \geq N_k, d(x_n, x) < 1/k$$

For metric spaces, it is possible to consider a weaker notion of convergence

Definition. *A sequence $(x_n)_{n\in\mathbb{N}}$ in a metric space is* ***Cauchy*** *if*

$$\forall k \in \mathbb{N}, \exists N_k \in \mathbb{N} : \forall n, m \geq N_k, d(x_n, x_m) < 1/k$$

Intuitively, in a reasonable metric space any Cauchy sequence would converge, this is indeed the case for **complete** metric spaces.
An example of a non-complete metric space is $\mathbb{Q}$ with its usual topology.
For instance, we can consider the sequence

$$x_n = \frac{(1+\sqrt{2})^{n+1} - (1-\sqrt{2})^{n+1}}{(1+\sqrt{2})^n - (1-\sqrt{2})^n}$$

is all made of rational numbers but its limit would be $1+\sqrt{2}$, an irrational number.[1]

4 Linear Spaces

Definition. *A* ***vector*** *Space on a Field $\mathbb{F}$ (usually the reals or the complex numbers) is a set V with two operations $+ : V \times V \to V$ and $\cdot : \mathbb{F} \times V \to V$ such that:*

- *$(V, +)$ is an abelian group. That is*
  - *(identity)* $\exists \mathbf{0} \in V : \forall v \in V, v + \mathbf{0} = \mathbf{0} + v = v$
  - *(inverse)* $\forall v \in V, \exists w_v \in V : v + w_v = w_v + v = \mathbf{0}$
  - *(associativity)* $\forall x, y, z \in V, (x + y) + z = x + (y + z)$
  - *(commutativity)* $\forall v, w \in V, v + w = w + v$
- *(distributive)*
  - $a(v + w) = av + aw$
  - $(a + b)v = av + bv$
- *(identity)* $1v = v$
- *(compatibility)* $a(bv) = (ab)v$

A ***topological vector space*** *is a vector space $(V, +, \cdot)$ with a topology τ_V such that the two operations are continuous*[2]*.*

[1] See the appendix for further details on this example.
[2] Continuity will be defined later.

Definition. *A function* $f : V \to W$ *from a vector space* V *to another one* W *defined on the same scalar field* $\mathbb{K}$*, is a* ***linear operator*** *if*

$$f(av + bw) = af(v) + bf(w)$$

If the vector space V *is the field itself, the linear operators are called* ***linear functionals****.*

Definition. *A* ***normed space*** *is a real or complex vector space equipped with a map* $\|\cdot\| : V \to \mathbb{R}$ *such that the following properties hold:*

- *Positivity* $\|x\| = 0$ *iff* $x = \mathbf{0}$ *(without this we get a semi-norm)*
- *Homogeneity* $\|av\| = |a|\|v\|$
- *Triangle* $\|x + y\| \leq \|x\| + \|y\|$

A a norm naturally defined a metric on V*:* $d_{\|\|}(x, y) = \|x - y\|$*.*
A normed space where this metric is complete, is called a ***Banach space****.*

Definition. *A linear operator* $f : V \to W$ *between normed spaces is* ***bounded*** *if*

$$\|f\|_{V \to W} := \sup_{v \in V : \|v\|_V = 1} \|f(v)\|_W = \sup_{v \in V : v \neq \mathbf{0}} \frac{\|f(v)\|_W}{\|v\|_V} < \infty$$

Definition. *A* ***complex(real) inner product*** *on a complex(real) vector space is a map* $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{C}(\mathbb{R})$ *such that the following properties hold*

- *Symmetry:* $\langle x, y \rangle = \mathrm{conj}(\langle y, x \rangle)$
- *Linearity:* $\langle ax + by, z \rangle = a\langle x, z \rangle + b\langle y, z \rangle$
- *Positivity:* $\langle x, x \rangle > 0$ *for every* $x \neq \mathbf{0}$

Such an Inner Product defines a norm as follows $\|x\| = \sqrt{\langle x, x \rangle}$*. An inner product space, that induces a norm that induces a complete metric is called an* ***Hilber space****.*

A real inner product also allows to measure angles between vectors

$$\theta_{x,y} = \arccos\left(\frac{\langle x, y \rangle}{\sqrt{\langle x, x \rangle \langle y, y \rangle}}\right)$$

This is well defined thanks to the famous **Cauchy-Schwarz** inequality

$$|\langle x, y \rangle| \leq \|x\|\|y\|$$

That is valid even in complex normed spaces.

Proof. Assuming wlog that $v \neq \mathbf{0}$, we have that, for every $a \in \mathcal{C}$,

$$0 \leq \|x - ay\|^2 = \langle x - ay, x - ay \rangle = \|x\| + |a|^2 \|v\| - a\overline{\langle x, y \rangle} - \overline{a}\langle x, y \rangle$$

Now, if we take $a = \langle x, y \rangle / \|y\|^2$, we get

$$0 \leq \|x\| - |\langle x, y \rangle|^2 / \|y\|^2$$

$\square$

Notice that this also proves that the inequality is binding iff x, y are linearly dependent.

4.1 Continuity

A function $f : X \to Y$ on topological spaces is **continuous** when the preimage of open sets is open that is if

$$\forall B \in \tau_Y,\ f^{-1}(B) \in \tau_X$$

A function $f : X \to Y$ on metric spaces is **continuous** if

$$\forall x \in X, \forall \varepsilon > 0, \exists \delta_{x,\varepsilon} > 0 : f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$$

A function $f : X \to Y$ on metric spaces is **uniformly continuous** if

$$\forall \varepsilon > 0, \exists \delta_\varepsilon > 0, \forall x \in X, f(B_\delta(x)) \subseteq B_\varepsilon(f(x))$$

A function $f : X \to Y$ on metric spaces, is k-**Lipschitz continuous** if

$$\forall x, y \in X, d_Y(f(x), f(y)) \leq k d_X(x, y)$$

A function $f : X \to Y$ on metric spaces, is (α, k)-**Hölder continuous** if

$$\forall x, y \in X, d_Y(f(x), f(y)) \leq k d_X^\alpha(x, y)$$

A function $f : X \to Y$ on normed spaces, is (Fréchet) differentiable if, for every point $x \in X$, we have a bounded linear operator $A_x : X \to Y$ such that

$$\lim_{\|h\|_X \to 0} \frac{\|f(x+h) - f(x) - A_x(h)\|_Y}{\|h\|_X} = 0$$

The operator A_x is called the derivative of f at x.

Theorem.

- *The continuous image of a compact set is compact.*
- *The continuous image of a connected set is connected.*
- *HEINE-CANTOR A continuous function on a compact set is uniformly continuous.*

- *EXTREME VALUE TH. If $f : X \mapsto \mathbb{R}$ is continuous and X is compact, then f attains its maximum and minimum.*
- *If $f : \mathbb{R} \to \mathbb{R}$ is differentaible, then it is continuous.*
- *If $f : A \subseteq \mathbb{R} \to \mathbb{R}$ is differentaible then $\sup_{x \in A} |f'(x)| = k < \infty$, iff f is k-Lipschitz*
- *If f is k-Lipschitz, then f is uniformly continuous.*

Example. *For instance $f(x) = e^x$ is differentiable and therefore continuous but it is not uniformly continuous and therefore it is not Lipschitz.*

4.2 Convexity

- A subset A of a vector space is **convex** if the segment connecting any two points in A lies completely in A.

$$\forall x, y \in A, \forall \alpha \in [0,1], \;\; \alpha x + (1-\alpha) y \in A$$

- A function $f : A \subseteq V \to \mathbb{R}$ from a convex set to the reals, is **convex** if

$$\forall x, y \in A, \forall \alpha \in [0,1] \;\; f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y)$$

- A function $f : A \subseteq V \to \mathbb{R}$ from a convex set to the reals, is **strictly convex** if

$$\forall x, y \in A, \forall \alpha \in (0,1), \;\; f(\alpha x + (1-\alpha)y) < \alpha f(x) + (1-\alpha) f(y)$$

Theorem. *If f is convex on an convex subset A of $\mathbb{R}^n$, then f is continuous on A^o.*

5 Measure Theory

5.1 Algebras

Now we start introducing the basic concepts of measure theory.

Definition. *If X is a set, then an **algebra** on X, is a collection of subsets $\mathcal{A} \subseteq 2^X$ s.t.*

- $X \in \mathcal{A}$
- *$\mathcal{A}$ is closed by complement, i.e.*

$$\forall A \in \mathcal{A}, A^c \in \mathcal{A}$$

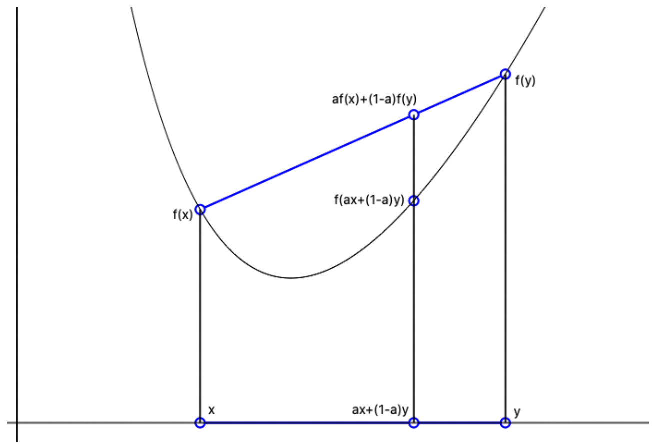

Figure 2: A convex function

- *$\mathcal{A}$ is closed by finite unions, i.e.*

$$\forall A, B \in \mathcal{A}, A \cup B \in \mathcal{A}$$

Why is that $\emptyset \in \mathcal{A}$ in any algebra $\mathcal{A}$?

Definition. *An algebra $\mathcal{A}$ is a σ-algebra, if it is closed under **countable** unions,*

$$(\forall n \in \mathbb{N}, A_n \in \mathcal{A}) \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$$

Example. *These are useful and general algebrae.*

- *Let $\mathcal{A} = \{\emptyset, X\}$ is the smallest σ-algebra on any X.*
- *Let $\mathcal{A} = 2^X$, this is the largest.*
- *Let $\mathcal{A} = \{A \subset \mathbb{R} : |A| \leq \aleph_0 \vee |A^c| \leq \aleph_0\}$ is a σ-algebra. This is the so-called countable-cocountable σ-algebra.*
- *Let $X = \mathbb{R}$, then $\mathcal{A} = \{A \subset \mathbb{R} : |A| < \infty \vee |A^c| < \infty\}$ is an algebra that is not a σ-algebra. This is the so-called finite-cofinite algebra.*

QUIZ: Is there a σ-algebra with precisely 8 elements? And with 7? And with 1?

5.1.1 How to generate (σ-)algebras

Let $\mathcal{C}$ be any collection of subsets. Is there any (σ-)algebra containing $\mathcal{C}$? YEP!
The (σ-)algebra 2^X always works!
If $\mathcal{A}_\alpha$ is a (σ-)algebra for any $\alpha \in I$, is $\bigcap_{\alpha \in I} \mathcal{A}_\alpha$ a (σ-)algebra? YEP!

- X belongs to each of the $\mathcal{A}_\alpha$ and so it belongs to their intersection.
- If A is the intersection, than it belongs to each of them, so A^c belongs to all of them and so A^c belongs to the intersection.
- The same works for finite or countable unions.

Therefore we can define the operator that gives the smallest (σ-)algebra containing any collection $\mathcal{C}$. We use the notation

$$\sigma(\mathcal{C}) = \bigcap_{\mathcal{B} \text{ is algebra containing } \mathcal{C}} \mathcal{B} \qquad \mathcal{A}(\mathcal{S}) = \bigcap_{\mathcal{B} \text{ is } \sigma\text{-algebra containing } \mathcal{C}} \mathcal{B}$$

5.1.2 The Nice σ-Algebras

If X is a topological space, then there is a natural σ-Algebra, the **Borel σ-algebra** generated by the open sets.

$$\mathcal{B} = \sigma(\mathcal{T})$$

This is the most natural σ-algebra on $\mathbb{R}^n$ and its subsets. It contains quite a lot of sets ($|\mathcal{B}_{\mathbb{R}^n}| = 2^{\aleph_0}$) but not all of them ($|2^{\mathbb{R}}| = 2^{2^{\aleph_0}}$). However, it is impossible to "explicitly" construct a non-Borel set. Therefore, don't worry, (probably) you won't ever encounter a non-Borel set in your life.

5.2 Measures

Definition. *A **content** is is a function $\mu : \mathcal{A} \to \overline{\mathbb{R}}_+$ from an algebra to the non-negative numbers that is finitely additive, i.e. If $N \in \mathbb{N}$ and $(A_n)_{n=1}^N$ are disjoint sets in $\mathcal{A}$*

$$\mu\left(\bigcup_{n=1}^N A_n\right) = \sum_{n=1}^N \mu(A_n)$$

Definition. *A **measure** is a content that is σ-additive, i.e. If $(A_n)_{n\in\mathbb{N}}$ are disjoint sets*[3]*, then*

$$\mu\left(\bigcup_{n\in\mathbb{N}} A_n\right) = \sum_{n\in\mathbb{N}} \mu(A_n)$$

In both cases we have that $\mu(\emptyset) = 0$.

Example. *Here some famous measures*

- *We can consider $\mathbb{R}$ with the full σ-algebra $\mathcal{A} = 2^{\mathbb{R}}$. Then a nice measure on this space is the **counting measure** $\mu(A) = |A|$.*
- *If X is any set with any σ-algebra and $x \in X$, then the **Dirac measure** δ_x is defined as*

$$\delta_x(A) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

[3] Such that their union is in $\mathcal{A}$

- *Let $X = \{1, 2, 3\}$ and let $\mathcal{A} = 2^X$, then a unique measure μ can be specified by imposing $\mu(\{1\}) = \mu(\{2\}) = \mu(\{3\}) = 1/3$.*

Here we consider special kinds of measures

Definition. *A measure μ on $\mathcal{A}$ is σ-finite if there are sets $\{A_n\}_{n\in\mathbb{N}}$ in $\mathcal{A}$ s.t.*

$$X = \bigcup_{n\in\mathbb{N}} A_n \ \wedge \ \forall n \in \mathbb{N}, \mu(A_n) < \infty$$

Definition. *A measure μ on $\mathcal{A}$ is finite if $\mu(X) < \infty$*

Definition. *A measure μ on $\mathcal{A}$ is a **probability** if $\mu(X) = 1$*

Notice (and prove) that

- The counting measure on an uncountable set (like $\mathbb{R}$) is not σ-finite.
- The Dirac measure is always a probability measure.

Theorem. *A measure μ is monotone, meaning that if A, B are measurable sets such that $A \subseteq B$, then $\mu(A) \leq \mu(B)$*

Proof. We can write $B = A \cup (B \setminus A)$, and of course A and $B \setminus A$ are disjoint, so we have that, by additivity

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

□

Theorem. *If μ is a σ-finite measure, then it is **continuous**, meaning that*

$$\mu\left(\bigcup_{n\in\mathbb{B}} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

and that

$$\mu\left(\bigcap_{n\in\mathbb{B}} A_n\right) = \lim_{n\to\infty} \mu(A_n)$$

5.3 How to specify a measure

A major practical problem is to identify a small class of measurable sets $\mathcal{S} \subset \mathcal{B}$ such that, if μ is known on $\mathcal{S}$ then μ can be determined on the whole $\mathcal{B}$.
Indeed it would be quite impossible to specify a measure on every Borel subset of the real line.
We will now introduce the theorems that solve this problem.

We start considering semi-algebras

Definition. *A **semi-algebra** on X is a collection of subsets $\mathcal{S} \subset 2^X$ such that*

- $\emptyset \in X$
- *Closed under finite intersections* $\forall A, B \in \mathcal{S}$, $A \cap B \in \mathcal{S}$
- *If* $\emptyset \neq A \in \mathcal{S}$, *then* A^c *is a finite disjoint union of elements in* $\mathcal{S}$.
- If $X = \mathbb{R}$, then $\mathcal{S} = \{(a, b] : a, b \in \mathbb{R} \cup \{-\infty, \infty\} \wedge b > a\}$ is a semi-algebra that generates the Borel σ algebra.

Theorem. *If* μ *is a measure (content) defined on a semi-algebra* $\mathcal{S}$ *then, we have a unique measure (content)* $\tilde{\mu}$ *defined on* $\mathcal{A}(\mathcal{S})$ *such that*

$$\tilde{\mu}|_{\mathcal{S}} = \mu$$

Now we introduce the most fundamental rein measure theory, the **Carathéodory's Extension Theorem**

Theorem. *If* μ *is a measure on an algebra* $\mathcal{A}$*, then there exists a measure* $\tilde{\mu}$ *on* $\sigma(\mathcal{A})$ *such that*

$$\tilde{\mu}|_{\mathcal{A}} = \mu$$

Moreover, if μ *is* σ*-finite, then this* $\tilde{\mu}$ *is unique.*

This is a wonderful result that allows to construct the most important measure of the entire universe.

Definition. *The Lebesgue measure is the unique measure on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *such that*

$$\forall a, b \in \overline{\mathbb{R}} : (b > a \wedge b - a \in \overline{\mathbb{R}}), \mu((a, b]) = b - a$$

The proof is quite involved but you can find each step plainly explained in the wonderful and relaxing video course [**Video**].

5.4 Lebesgue-Stieltjes measures

Measures may seem very complicated at first, but... It is possible to establish a bijection between probability measures on $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ and a nice class of functions.

Definition. *A cdf is a function* $F : \mathbb{R} \to [0, 1]$ *s.t.*

- F *is monotone non-decreasing. (if* $x > y$*, then* $F(x) \geq F(y)$*)*
- F *is right-continuous. (*$\forall x \in \mathbb{R}$, $\lim_{h \downarrow 0} f(x + h) = f(x)$*)*
- F *is s.t.* $\lim_{x \uparrow \infty} F(x) = 1$ *and* $\lim_{x \uparrow \infty} F(x) = 0$.

Theorem. *If* μ *is a probability measure on* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ *then* $F : x \mapsto \mu((-\infty, x])$ *is a cdf.*
Conversely, if F *is a cdf, then there is a unique measure* μ *such that* $\mu((a, b]) = F(b) - F(a)$.

5.5 Completeness

A measure space $(X, \mathcal{A}, \mu)$, where $\mathcal{A}$ is a σ-algebra and μ is a measure on $(X, \mathcal{A})$, is **complete**, if, for every $B \subset X$ such that $B \subset A$ for some A with $\mu(A) = 0$, we have that $B \in \mathcal{A}$. In formulae

$$\forall B \subset X, (\exists A \in \mathcal{A} : B \subset A \wedge \mu(A) = 0) \Rightarrow (B \in \mathcal{A})$$

It is possible to show that $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, \lambda)$ is not complete. It is however possible to add all the subsets of null-sets to $\mathcal{B}_{\mathbb{R}}$, to get the Lebesgue σ-aglebra $\mathcal{L} := \overline{\mathcal{B}_{\mathbb{R}}}$ and to extend μ in a trivial way to get a complete measure space.

The Lebesgue-measurable sets are much more frequent than the Borel-measurable sets, indeed $|\mathcal{L}| = |2^{\mathbb{R}}| = 2^{2^{\aleph_0}} > 2^{\aleph_0} = |\mathcal{B}_{\mathbb{R}}|$. However there are still non-Lebesge-measurable sets, like the famous Vitali set.
For many details on the Vitali sets and other pathological sets, see [**Dude**]

6 Exercises

- Prove that $\mathcal{A}$ is a finite σ-algebra, iff $|\mathcal{A}| = 2^n$ for some $n \in \mathbb{N}$.
- Prove that $\sigma(\{x\} : x \in \mathbb{R}) \neq \mathcal{B}_{\mathbb{R}}$
- Let $\mathcal{A}$ be the finite-cofinite σ-algebra on $\mathbb{Z}$, and let μ be such that $\mu(A) = 1$ if A is infinite and 0 otherwise, prove that this is a content but not a measure.
- Can you find a content on a σ-algebra that is not a measure?
- Let $X = \{1, 2, 3\}$, $\mathcal{A} = \{\emptyset, X, \{1\}, \{2, 3\}\}$, find a probability measure μ on $(X, \mathcal{A})$ that makes $(X, \mathcal{A}, \mu)$ non complete.
- Find a set in $\mathbb{B}_{\mathbb{R}}$ that is not in $\sigma(\{(-x, x) : x > 0\})$
- Show that every continuous cdf is uniformly continuous.
- Show that $\mathbb{B}_{\mathbb{R}} = \sigma(\{[a, b) \cap \mathbb{R} : a, b \in \overline{\mathbb{Q}} \wedge b > a\})$
- Show that, if (X, τ) is 2^{nd}-countable and $\mathcal{C}$ is basis of $\mathcal{T}$, then $\mathcal{B}_{(X,\mathcal{T})} = \sigma(\mathcal{C})$

You should also try the exercises at [**Wo**].
Send your solutions/questions to andrea.aveni@duke.edu

7 Appendix

7.1 Infinite sets

Here we consider the concept of cardinality, that is how big a set is. We denote the cardinality of a set A by $|A|$.

Definition. *Two sets A, B have the same cardinality ($|A| = |B|$) if there exists a bijective function from A to B.*

This is an equivalence relation on the class of all sets (verify it!) and the cardinal numbers are precisely the equivalence classes of this relation.

If A is finite, then $|A| = n$ for some $n \in \mathbb{N}$. If A, is infinite, then there are many possibilities since not all the infinities are the same.

Famous Cardinals

- The smallest infinite cardinal number is denoted $\aleph_0$ and is the cardinality of $\mathbb{N}$, $\mathbb{Z}$, and $\mathbb{Q}$. Sets of this cardinality are called **countable**. Any larger set is said to be **uncountable**.
- The second largest cardinal is denoted $\aleph_1$ and (assuming CH) is equal to $2^{\aleph_0}$, that is, it is the cardinality of the power set of any countable set. This is the cardinality of $\mathbb{R}$, $\mathbb{R}^{27}$, the set of all Borel-measurable sets and the set of all integer sequences. This cardinality is sometimes called **the cardinality of the continuum $\mathfrak{c}$**.
- The next cardinal is (assuming the GCH) $\aleph_2 = 2^{\aleph_1}$ and this is the cardinality of the collection of all Lebesgue-measurable sets and of all real functions.

7.2 Why is that $|\mathbb{N}| = |\mathbb{Z}| = |\mathbb{Q}|$?

To prove that $|\mathbb{N}| = |\mathbb{Z}|$, we need to find a bijection ϕ between them. An example is $\phi(n) = (-1)^{n+1}\lfloor (n+1)/2 \rfloor$

n	0	1	2	3	4	5	6	...	n	...
$\phi(n)$	0	1	-1	2	-2	3	-3	...	$(-1)^{n+1}\lfloor (n+1)/2 \rfloor$	...

Now, to show that $|\mathbb{N}| = |\mathbb{Q}|$, we need to show that there exists some $f : \mathbb{Q} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{Q}$ both surjective. For f we can take a function that associate to a rational number its denominator when reduced in lowest terms. As for g we can take function g in the picture, with the identification in mid, that $(a, b) \in \mathbb{Z}^2$ will correspond to a/b or to (let's say) $7/23$ if $b = 0$.

Indeed when there are surjective maps $f : A \to B$ and $g : B \to A$, the the existence of a bijection is guaranteed. This is a consequence of famous **Cantor–Bernstein–Schröder** theorem. (See [**Top**].)

7.3 Why is that $|\mathbb{N}| \neq |\mathbb{R}|$?

This is the first proof that there are different infinities and was found by Cantor (who ended his life in an asylum after becoming mad studying the infinity) in 1891.
Assume we can list all the real number in (just) $(0, 1]$ written in base two like this

$$
\begin{tikzcd}
(2,-2) \arrow[d] & (2,-1) \arrow[l] & (2,0) \arrow[l] & (2,1) \arrow[l] & (2,2) \arrow[l] \\
\cdots & (1,-1) \arrow[d] & (1,0) \arrow[l] & (1,1) \arrow[l] & (1,2) \arrow[u] \\
 & (0,-1) \arrow[d] & (0,0) \arrow[r] & (1,0) \arrow[u] & (0,2) \arrow[u] \\
 & (-1,-1) \arrow[r] & (-1,0) \arrow[r] & (-1,1) \arrow[r] & (-1,2) \arrow[u]
\end{tikzcd}
$$

Figure 3: A surjective function $g : \mathbb{N} \to \mathbb{Z}^2$

n	x_n
1	0.10101001010111...
2	0.11000100010000...
3	0.10100010101011...
4	0.00101100101001...
$\vdots$	$\vdots$

Then, if we consider the number obtained by "inverting" the n^{th} digit of the n^{th} number (in this example 0.0001...), we would get to conclude that this number is not in the list.

The theory of infinite numbers is a very active topic in Logic and Set theory and you can find a great and moder exposition of this theory in [**Big**].

7.4 The Cantor set

For every $n \geq 1$, consider the set

$$C_n = \bigcup_{k=0}^{3^n-1} [2k/3^n, (2k+1)/3^n]$$

This is a finite union of disjoint intervals. The Cantor set C is defined as

$$C = \bigcap_{n=1}^{\infty} C_n$$

It is easy to see that

$$\lambda \left(\bigcap_{n=1}^{k} C_n \right) = (2/3)^n$$

So that, we immediately get, by continuity of the finite Lebsgue measure on $[0,1]$, that $\lambda(C) = 0$. Therefore, the Cantor set is a null set for the Lebesgue measure. However, it is uncountable! Indeed one can prove that C is the set of all real numbers in $[0,1]$, that one can write in base 3 by using only the digits 0 and 2.
But then you can define the following surjective function $\phi : C \to [0,1]$ that sends a number x in the cantor set to a number whose binary expansion is equal to the ternary expansion of x with all the 2s replaced by 1s. This, once again, thanks to the **Cantor–Bernstein–Schröder** theorem implies that $|C| = |\mathbb{R}|$.
There are many other interesting properties of C that you can find online.

7.5 The Fat Cantor set

The fat cantor set F is a closed subset of $[0,1]$ with Lebesgue measure equal to $1/2$, that does not contain any open interval (and therefore $F^o = \emptyset$). Moreover, its boundary has also measure $1/2$, because indeed we have that $\partial F = F$.

Fore more fun stuff about these two fractals and some other, see [**Fra**].

7.6 Continued fractions and irrational numbers

Here we will reconsider the example showing non-completeness of the rational numbers.
First of all we will prove a classic statement

Theorem. *For any natural number n, its square $\sqrt{n}$ is an irrational number iff n is not the square of a natural number (a perfect square).*

Proof. Indeed, assume that $\sqrt{n} = a/b$ for some a, b coprime natural numbers. Then, we can write $b^2 n = a^2$. Now, since n is not a perfect square, there must be some prime p of odd order k in n.
Since a/b is reduced in lesser terms, at least one of a and b must not be divisible by p. If b is not divisible by p, we have that a must be divisible by p but its order should be $k/2$ and this is impossible since k is odd.
If, on the contrary, a is not divisible by p, we still get a contradiction since then a^2 does not dived p but nb^2 does.
Now we can also prove that any number of the form $c + d\sqrt{n}$ with $c, d \in \mathbb{Z}$ and $n \in \mathbb{N}$ different from a perfect square, is irrational.
Assume that $c + d\sqrt{n} = a/b$, then $\sqrt{n} = (a - cb)/bd$ would itself be irrational, which is impossible. □

Now we will prove that the sequence at page (?) is indeed rational. We want to solve the recurrence equation $a_{n+1} = 2a_n + a_{n-1}$ with $a_0 = 0$ and $a_1 = 1$. This is of course an integer sequence, since each term is a sum of products of integers. We can rewrite the reccurence equation as

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} a_{n-1} \\ a_n \end{bmatrix}$$

Using the initial condition, we can explicitly write that

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}^n \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Using the spectral decomposition, we get

$$\begin{bmatrix} a_n \\ a_{n+1} \end{bmatrix} = \frac{\sqrt{2}}{4} \begin{bmatrix} -1-\sqrt{2} & \sqrt{2}-1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} (1-\sqrt{2})^n & 0 \\ 0 & (1+\sqrt{2})^n \end{bmatrix} \begin{bmatrix} -1 & \sqrt{2}-1 \\ 1 & 1+\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

By isolating the first entry, we immediately find that

$$a_n = \frac{\sqrt{2}}{4} \left((1+\sqrt{2})^n - (1-\sqrt{2})^n \right)$$

Since a_n is made of natural numbers, the sequence

$$x_n = \frac{a_{n+1}}{a_n} = \frac{(1+\sqrt{2})^{n+1} - (1-\sqrt{2})^{n+1}}{(1+\sqrt{2})^n - (1-\sqrt{2})^n}$$

will be made of rational numbers.
The fact that $x_n \to 1+\sqrt{2}$ can be seen using l'Höpital's rule:

$$\begin{aligned} \lim_{n\to\infty} x_n &= \lim_{n\to\infty} \frac{a_{n+1}}{a_n} \\ &= \lim_{n\to\infty} \frac{(1+\sqrt{2})^{n+1} - (1+\sqrt{2})^{-n-1}}{(1+\sqrt{2})^n - (1+\sqrt{2})^{-n}} \\ &= \frac{1}{1+\sqrt{2}} \lim_{n\to\infty} \frac{(1+\sqrt{2})^{2n+2} - 1}{(1+\sqrt{2})^{2n} - 1} \\ &= \frac{1}{1+\sqrt{2}} \lim_{n\to\infty} \frac{2n(1+\sqrt{2})^{2n+2} \ln(1+\sqrt{2})}{2n(1+\sqrt{2})^{2n} \ln(1+\sqrt{2})} = 1+\sqrt{2}. \end{aligned}$$

References

[Wo] Some very useful exercises to be able to solve from the last year's professor of the course. `http://www2.stat.duke.edu/courses/Fall20/sta711/exams/711diag.pdf`

[Video] A great, easy and calm video course on Measure Theory. The professor is slow and very precise. You will find all the details on the extension theorems and some interesting stuff on the Lebsgue measure on the Hilbert cube. `https://www.youtube.com/playlist?list=PLo4jXE-LdDTQq8ZyA8F8reSQHej3F6RFX`

[Dude] This is probably the best book in measure theory from a probabilitic point of view.
Richard Dudley, Real Analysis and Probability, Cambridge.

[Kerkis] This is an advanced book on the study of Borel sets. You will be surprised in seeing how much there is to say about them.
Alexander S. Kechris, Classical Descriptive Set Theory, Springer.

[Bible] This is the final text (5 volumes!) on measure theory. `https://www1.essex.ac.uk/maths/people/fremlin/mt.htm`

[Top] This is the best concise book on point set topology and a bit beyond (the fundamental group is covered but there is no mention of homology or higher homotopy groups)
Munkres, Topology, Springer.

[Big] A big book on big numbers.
Akihiro Kanamori, The Higher Infinite, Springer.

[Fra] DiMartino, R. and Urbina, W. On Cantor-like Sets and Cantor-Lebesgue Singular Functions `https://arxiv.org/pdf/1403.6554.pdf`