1 Background

1.1 Classical dynamical systems

Theorem 1.1: Implicit function theorem

Send help...

1.2 Linear stability of space-time PDEs

1.2.1 Dimensionality reduction via decomposition methods

1.3 Stochastic processes

1.3.1 Measure-theoretic probability

Let us start with the basic ingredients needed for the rigorous definition of probabilities associated to a set of events. To do that we will need the basic notions of measurability associated to arbitrary sets and in particular we want to review the minimal concepts that allow us to define the probability over a sample space to be the measure of such space. The idea is that while defining a measure of a simple set (interpreted as a numerical quantity of the amount of *"stuff"* inside such set) may be straightforward, things get less clear when sets contain non-canonical objects s.a. events. For example, if we consider intervals $I = (a, b) \subset \mathbb{R}$ then the mos intuitive measure that we may define is the length $L := |a-b|$; in this way we are able to associate bigger measures to larger sets in a consistent way. This generalises trivially in $\mathbb{R}^n$ as well. However what measure can we define for subsets s.a. $Q = I \cap \mathbb{Q}$ i.e. the discrete sets of (countably-many) rational numbers in $[a, b]$? Is the number of elements a good candidate for measuring the subsets? What if we want to measure the sets of irrational numbers in $[a, b]$? Does it make sense to count them as opposed to the length $L := |a - b|$ that we defined for the previous case? In principle given a *"whole space"* Ω (which we can think of the largest possible space, in some sense, containing all the possible elements for a particular case), we want to be able to define a measure that can be arbitrarly used for any subset $A \subset \Omega$ which quantifies the *size* of the subsets in a consistent way. So the natural question that arises after these two examples is how do we formally generalise the notion of a measure to any arbitrary set in order to retain a consistent property of the larger the set the bigger the measure?

Definition 1.1 ($\sigma-$algebra)**.** Let Ω be a non-empty set and $\mathcal{P}(\Omega)$ be its power set, then the set $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is called a $\sigma-$algebra if it satisfies the following properties

1. $\varnothing$, $\Omega \in \mathcal{A}$;
2. if $A \in \mathcal{A}$ then $A^{\mathrm{c}} \in \mathcal{A}$ (i.e. $\mathcal{A}$ is closed under complement);
3. if A_1, A_2, $\cdots \in \mathcal{A}$ then $\cup_{j=1}^{\infty} A_j \in \mathcal{A}$ (closed under countable unions) and $\cap_{j=1}^{\infty} A_j \in \mathcal{A}$ (closed under countable intersection).

Observation 1.1. Notice that closure under countable intersections is made redundant by the closure under complement and closure under countable unions since $A \cap B = (A^c \cup B^c)^c$.

The $\sigma-$algebra is the centerpiece of measure-theoretic probability. It is defined as a collection of subsets of a non-empty set Ω. This is obvious from our statement $\mathcal{A} \subseteq \mathcal{P}(\mathcal{A})$ of the $\sigma-$algebra $\mathcal{A}$ being a subset of the power set of Ω. It follows immediately that a $\sigma-$algebra of Ω is not unique as we can choose any subset of the power set as long as it satisfies the properties 1. – 3. of Definition 1.1. In fact for any set Ω the collection $\{\varnothing, \Omega\}$ is the trivial $\sigma-$algebra while the power set $\mathcal{P}(\Omega)$ is the full/largest $\sigma-$algebra.

Example 1.1

Let $\Omega = \{a, b, c, d\}$; its power set contains $2^4 = 16$ elements, specifically

$$\begin{aligned}\mathcal{P}(\Omega) = \{\varnothing, &\{a\}, \{b\}, \{c\}, \{d\}, \{a,b\}, \{a,c\}, \{a,d\}, \{b,c\}, \{b,d\}, \{c,d\}, \\ &\{a,b,c\}, \{a,b,d\}, \{a,c,d\}, \{b,c,d\}, \Omega\}\,.\end{aligned}$$

Consider $\mathcal{A} := \{\varnothing, \{a,b\}, \{c,d\}, \Omega\} \subset \mathcal{P}(\Omega)$ and let us check whether $\mathcal{A}$ is a $\sigma-$algebra. Both the empty set $\varnothing$ and the whole space Ω are included in $\mathcal{A}$ hence property 1. is satisfied. Regarding property 2. it is easy to check that

- $A_1 := \varnothing \in \mathcal{A} \implies A_1^c = \Omega \in \mathcal{A}$;
- $A_2 := \{a,b\} \in \mathcal{A} \implies A_2^c = \{c,d\} \in \mathcal{A}$;
- $A_3 := \{c,d\} \in \mathcal{A} \implies A_3^c = \{a,b\} \in \mathcal{A}$;
- $A_4 := \Omega \in \mathcal{A} \implies A_4^c = \varnothing \in \mathcal{A}$;

so $\mathcal{A}$ is closed under complement. Finally we observe that $\mathcal{A}$ is closed under arbitrarly countable unions

$$\begin{aligned}
&A_1 \cup A_2 = A_2 \in \mathcal{A}\,,\\
&A_1 \cup A_3 = A_3 \in \mathcal{A}\,,\\
&A_1 \cup A_4 = A_4 \in \mathcal{A}\,,\\
&A_2 \cup A_3 = \Omega \in \mathcal{A}\,,\\
&A_2 \cup A_4 = \Omega \in \mathcal{A}\,,\\
&A_3 \cup A_4 = \Omega \in \mathcal{A}\,,\\
&A_1 \cup A_2 \cup A_3 = \Omega \in \mathcal{A}\,,\\
&A_1 \cup A_2 \cup A_4 = \Omega \in \mathcal{A}\,,\\
&A_2 \cup A_3 \cup A_4 = \Omega \in \mathcal{A}\,,\\
&A_1 \cup A_2 \cup A_3 \cup A_4 = \Omega \in \mathcal{A}\,,
\end{aligned}$$

and closed under arbitrarly countable intersections

$$\begin{aligned}
&A_1 \cap A_2 = \varnothing \in \mathcal{A}\,,\\
&A_1 \cap A_3 = \varnothing \in \mathcal{A}\,,\\
&A_1 \cap A_4 = \varnothing \in \mathcal{A}\,,\\
&A_2 \cap A_3 = \varnothing \in \mathcal{A}\,,\\
&A_2 \cap A_4 = A_2 \in \mathcal{A}\,,\\
&A_3 \cap A_4 = A_3 \in \mathcal{A}\,,\\
&A_1 \cap A_2 \cap A_3 = \varnothing \in \mathcal{A}\,,\\
&A_1 \cap A_2 \cap A_4 = \varnothing \in \mathcal{A}\,,\\
&A_2 \cap A_3 \cap A_4 = \varnothing \in \mathcal{A}\,,\\
&A_1 \cap A_2 \cap A_3 \cap A_4 = \varnothing \in \mathcal{A}\,.
\end{aligned}$$

Since all the 3 properties hold we conclude that $\mathcal{A}$ is a $\sigma-$algebra on Ω. Similarly to the above, we can easily show that the family $\mathcal{C} := \{\varnothing, \{a\}, \{b, c, d\}, \Omega\}$ is also a $\sigma-$algebra on Ω. Conversely $\mathcal{D} := \{\varnothing, \{a, b\}, \{b, c\}, \Omega\}$ is not a $\sigma-$algebra on Ω as it violates property 3..

There are 2 key reasons of why we need to introduce $\sigma-$algebras in measure theory:

- we can use $\sigma-$algebras to keep track of any collection of subsets of the whole space Ω whose size we are *permitted* to measure;
- out of all the possible subsets in the power set of Ω, we are interested only in those whose measure tell us all the information about the whole space.

Highlight

The first reason is the most intuitive. With a $\sigma-$algebra we want to be able to immediately identify a restricted family of subsets of Ω that we want to be able to measure. Obviously we want to be able to measure the size of Ω itself, and also the empty set $\varnothing$, which we intuitevly give a measure of 0. Hence the reason of property 1. in Definition 1.1. Also given a subset of Ω we would like to measure both the size of the stuff in such subset but also the amount of stuff that it is not in the subset, which would be the complement. Hence why property 2.. Finally given any 2 (or any number) of subsets of Ω we want to be able to measure their union and their intersection. Hence why property 3.. A $\sigma-$algebra allows us to identify which subsets of the whole space can have nice properties in terms of measurability. In particular, in the context of probabilities, $\sigma-$algebras allow us to identify the families of subsets of the event space whose probability is defined.

This notion of measurability is formalised for any subset in the collection of subsets identified by a $\sigma-$algebra.

Definition 1.2 (Measureable sets)**.** Let Ω be a non-empty set and $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ be a $\sigma-$algebra of Ω, then any subset $A \in \mathcal{A}$ is called $\mathcal{A}-$measurable.

This is further extended in a single object without the need to define the actual measure yet.

Definition 1.3 (Measurable space)**.** The double $(\Omega, \mathcal{A})$ is called a measurable space.

Succintly we can now specify what a measure needs to be in order to have meaning in the context of estimating the amount of stuff inside a set.

Definition 1.4 (Unsigned measure)**.** Let $(\Omega, \mathcal{A})$ be a measurable space, then the function $\mu : \mathcal{A} \to \mathbb{R}^+$ is called a measure if it satisfies the following properties:

1. $\mu(\varnothing) = 0$;
2. given a pairwise disjoint collection $\{A_j\}_{j=1}^{\infty} \subseteq \mathcal{A}$ then $\mu(\cup_{j=1}^{\infty} A_j) = \sum_{j=1}^{\infty} \mu(A_j)$ (i.e. μ is countably additive).

Once we specify a measure on a measurable space $(\Omega, \mathcal{A})$ we can construct a more specialised object for our purposes.

Definition 1.5 (Measure space)**.** Let $(\Omega, \mathcal{A})$ be a measurable space and $\mu : \mathcal{A} \to \mathbb{R}^+$ be a measure on such space, then the triple $(\Omega, \mathcal{A}, \mu)$ is called a measure space.

All the Definitions from 1.2 to 1.5, with the exception of 1.4, are straightforward and do not require further explanations. One key observation however from Definition 1.5 is that whenever we are dealing with a measure space s.t. $\mu(\Omega) = 1$ then we say that $(\Omega, \mathcal{A}, \mu)$ is a probability space and in particular that μ is a probability measure. Regarding Definition 1.4 we clarify things by looking at an example.

Example 1.2

Continuing from Example 1.1, let us try to define a suitable measure for the $\sigma-$algebra $\mathcal{A} = \{\varnothing, \{a,b\}, \{c,d\}, \Omega\}$. Arguably the most intuitive form of measure on a finite and countable space $(\Omega, \mathcal{A})$ is the *counting measure*

$$\begin{aligned} \mu :\ \mathcal{A} &\longrightarrow \mathbb{N} \\ A_j &\longrightarrow \text{card}(A_j) \end{aligned}$$

Let us verify whether it is indeed a measure as per Definition 1.4. Property 1. is trivial. We then specify the measure of the remaining 3 subsets in $\mathcal{A}$

$$\mu(A_2) = 2\,, \quad \mu(A_3) = 2\,, \quad \mu(A_4) = 4\,.$$

We can now verify property 2.. Notice that the only pairwise disjoint sub-collection in $\mathcal{A}$ is $\{A_2, A_3\}$. Trivially then $\mu(A_2 \cup A_3 = \Omega) = 4$ and $\sum_{j=2}^{3} \mu(A_j) = \mu(A_2) + \mu(A_3) = 2 + 2 = 4$ which lead us to conclude that the counting measure is a measure for $(\Omega, \mathcal{A})$. We can also define a probability measure

$$\begin{aligned} \nu :\ \mathcal{A} &\longrightarrow [0,1] \\ A_j &\longrightarrow \frac{\mu(A_j)}{\mu(A_4)} \end{aligned}$$

and easily verify that it satisfies both properties 1. and 2. as we did for the counting measure. As a final remark note that

$$\nu(A_2) = 0.5\,, \quad \nu(A_3) = 0.5\,, \quad \nu(A_4) = 1\,.$$

is consistent with the characteristic of the counting measure of assigning bigger quantities to larger sets.

At this point we might be tempted to think that, since the power set $\mathcal{P}(\Omega)$ is itself a $\sigma-$algebra (and in particular it is the full $\sigma-$algebra associated to Ω as we mentioned earlier), then we don't we always select $\mathcal{P}(\Omega)$ to build any probability space? This is a legitimate question and it can be linked to a subsequent one: are there some $\sigma-$algebras that are more efficient than others? Here the concept of efficiency in constructing $\sigma-$algebras can be thought as the steps required for their construction. It is easy to see in fact that as Ω gets more complicated (e.g. it becomes now a set of countably infinite elements) the construction of $\mathcal{P}(\Omega)$ gets more impractical (and in the limit impossible) and too big to be useful. We would like ideally a relatively small subset of the power set to be our $\sigma-$algebra in a way that it still retains all of the information in the whole space Ω. Well, for starters, we know that a $\sigma-$algebra is not unique to Ω; therefore the immediate intuition is that in order to build a smaller $\sigma-$algebra we can take the intersections of two distinct ones.

Proposition. *Let $\{\mathcal{A}_j\}$ be $\sigma-$algebras on Ω and $j \in J$ with J being an index set, then the set $\cap_{j \in J} \mathcal{A}_j$ is also a $\sigma-$algebra on Ω.*

Proof. TBD... $\square$

This initial step can be further developed by noticing that we can use any subset $\mathcal{M}$ of the power set (even those that are not $\sigma-$algebras) to generate smaller $\sigma-$algebras by relying on countable intersections.

Proposition. *Let $\mathcal{M} \subseteq \mathcal{P}(\Omega)$, $\{\mathcal{A}_j\}$ be all the $\sigma-$algebras on Ω that contain $\mathcal{M}$ with $j \in J$ being in some index set J, then there exist the smallest $\sigma-$algebra that contains $\mathcal{M}$ given by*

$$\sigma(\mathcal{M}) := \bigcap_{j \in J} \mathcal{A}_j \,, \quad \mathcal{M} \subseteq \mathcal{A}_j \;\; \forall j \in J \,,$$

called the $\mathcal{M}-$generated $\sigma-$algebra.

Proof. TBD... □

We remark that it is not difficult to find a generic $\sigma-$algebra that contains $\mathcal{M}$ (the trivial one would be indeed the power set) but rather which one of the $\sigma-$algebras on Ω that contain $\mathcal{M}$ is the most efficient one, i.e. the $\sigma-$algebra with the least amount of subsets of Ω.

Example 1.3

We continue with Example 1.1 and let us choose subset $\mathcal{M} = \{\{a\}, \{b\}\}$. It is straightforward to see that $\mathcal{M}$ is not a $\sigma-$algebra on $\Omega = \{a, b, c, d\}$. Now we ask ourselves, what is the smallest possible $\sigma-$algebra that we can construct that still contains $\mathcal{M}$? According to the Proposition we will need to find all the $\sigma-$algebras that contain $\mathcal{M}$ and proceed with the countable intersections. Clearly we can already exclude the two $\sigma-$algebras, $\mathcal{A} = \{\varnothing, \{a, b\}, \{c, d\}, \Omega\}$ and $\mathcal{C} = \{\varnothing, \{a\}, \{b, c, d\}, \Omega\}$, that we constructed in Example 1.1, since neither of them contain $\mathcal{M}$. So our effort would be to find all the remaining, countable $\sigma-$algebras on Ω that do contain $\mathcal{M}$. This would be hard to do in practice, especially if Ω was infinite or uncountable. Instead we will consider another, more practical way to construct the $\sigma-$algebra generated by $\mathcal{M}$. We start by including $\mathcal{M}$ with the empty set and the whole space in a collection $\mathcal{B} = \{\varnothing, \mathcal{M}, \Omega\}$. We can now proceed with taking the complement and the union of each subset in $\mathcal{B}$. Notice that these operations correspond to property 2. and 3. in 1.1. Everytime we get a new element we perform the same operations again until we cannot generate any new subset by performing unions and complements of the elements that are already in $\mathcal{B}$. This leads us sistematically to

$$\mathcal{B} =: \sigma(\mathcal{M}) = \{\varnothing, \mathcal{M}, \{a, b\}, \{c, d\}, \{b, c, d\}, \{a, c, d\}, \Omega\} \,.$$

Attention must be paid in order to make sure that we are not taking the union of 2 distinct $\sigma-$algebras since, contrary to the intersection as precribed by the above Proposition, the resulting collection of the union of $\sigma-$algebras might not be a $\sigma-$algebra[a].

[a] https://proofwiki.org/wiki/Union_of_Sigma-Algebras_may_not_be_Sigma-Algebra

We now have all the ingredients to generalise these steps of getting to the smallest $\sigma-$algebra generated by a subset $\mathcal{M}$ to the formal case of infinite sample spaces.

Definition 1.6 (Borel $\sigma-$algebra)**.** Let Ω be a topological or metric space, or again a subset of $\mathbb{R}^n$, and let $\mathcal{O} = \{\mathcal{M} \subseteq \mathcal{P}(\Omega) \;:\; \mathcal{M} \text{ is open}\}$, then

$$\mathcal{B}(\Omega) := \sigma(\mathcal{O}) \,,$$

is called the Borel $\sigma-$algebra on Ω.

In other words the Borel $\sigma-$algebra is the $\sigma-$algebra on Ω generated by the collection $\mathcal{O}$ of all open subsets of Ω. Better yet, the Borel $\sigma-$algebra is the smallest $\sigma-$algebra on Ω that contains all its open sets. The introduction of the Borel $\sigma-$algebra is the last building block that we require for the introduction of measure theoretic random variables acting on sample spaces. The objects that we introduce in the following are necessary for dealing with the most abstrac probability settings in a consistent and rigorous framework. As such, from this moment onward we will abandon the more abstract setting of measure theory and focus specifically to the case of probability measures defined on sets of outcomes as clarified in the next example.

Example 1.4

Adapted from [1, pp. 14–15]. Consider a fair die with 6 faces and we are interested in framing the outcomes of 1 single throw of such die in the language of measure theoretic probability. We immediately define the sample space to be

$$\Omega = \{1, 2, 3, 4, 5, 6\}\,.$$

The power set of Ω is made by the $2^6 = 64$ possible combinations of its elements. Notice however that the elements of Ω are not themselves events, but rather the subsets in $\mathcal{P}(\Omega)$ are. Therefore, in probabilistic terms, we may think of Ω as the set of samples, or outcomes, of a given experiment, while we think of $\mathcal{P}(\Omega)$ as the set of events that we are interested to measure. While the set of outcomes Ω is fixed by the nature of what we are observing, the set of events $\mathcal{P}(\Omega)$ is subject to our own desire and interpretation. The events associated to the single throw of a die need to be defined arbitrarly based on what we want to measure. So for instance say that we are interested in measuring the probability of the event of an even number as the outcome of a single throw. We therefore specify the following $\sigma-$algebra

$$\mathcal{P}(\Omega) \supset \mathcal{A} := \{\varnothing, A_1, A_1^c, \Omega\}\,, \quad A_1 = \{2, 4, 6\}.$$

On the $\sigma-$algebra we define a probability measure as follows (see Example 1.2 to check whether the probability measure is in fact a measure, i.e. it satisfies the properties 1. and 2. of Definition 1.4)

$$\begin{aligned} p : \mathcal{A} &\longrightarrow [0, 1] \\ A_j &\longrightarrow \frac{\text{card}(A_j)}{\text{card}(\Omega)}\,. \end{aligned}$$

From the above we easily conclude that the probability associated to the event of getting an even number from the single throw of a fair die is $p(A_1) = \frac{3}{6} = \frac{1}{2}$. Similarly we quantify the event of getting an odd number by choosing the complement of A_1

$$\mathcal{A} \ni A_2 := \{1, 3, 5\} = A_1^c\,,$$

and measuring it according to the same probability measure defined on the $\sigma-$algebra

$$p(A_2) = \frac{3}{6} = \frac{1}{2} = 1 - p(A_1)\,.$$

To conclude we can define the event of getting a prime number or a number less than or equal to 5 seemingly easy by repeating the same process

$$A_2 \subset A_3 = \{1, 2, 3, 5\}\,, \quad A_2 \subset A_3 \subset A_4 = \{1, 2, 3, 4, 5\}\,,$$

and computing their respective probability measures

$$p(A_3) = \frac{2}{3} > p(A_2)\,, \quad p(A_4) = \frac{4}{5} > p(A_3) > p(A_2)\,.$$

There are 2 important observations that we can take from the example above.

Observation 1.2. Let $(\Omega, \mathcal{A}, p)$ be a measure space under probability measure p and $A_j \in \mathcal{A}$, then

$$p(A_j^c) = 1 - p(A_j)\,,$$

for each event A_j. Furthermore given any two events $A, B \in \mathcal{A}$ it follows that

$$A \subseteq B \quad \text{implies} \quad p(A) \leq p(B)\,.$$

Highlight

In general we refer to the triple $(\Omega, \mathcal{A}, p)$ as a probability space for the special case of a measure space upon which the measure is the probability measure. The three ingredients of any probability space have the following interpretations:

- Ω (sample space): the set of all possible outcomes (all the things that can happen);
- $\mathcal{A}$ (event space): the set of all specific events (all possible *sets* of things that can happen);
- $p : \mathcal{A} \to [0, 1]$ (probability measure): function that assigns numbers to events based on their *"likelihood"*.

One clear object that we already introduced and yet it is still missing from our example above is the Borel $\sigma-$algebra. To motivate its adoption in probability problems we need to further introduce another important object.

Definition 1.7 (Random variable)**.** Let $(\Omega, \mathcal{A}, p)$ be a probability space, then the function $X : \Omega \to \mathbb{R}^n$ is called a ($n-$dimensional) random variable (r.v.) if

$$X^{-1}(B) \in \mathcal{A}\,, \quad \forall B \in \mathcal{B}(\mathbb{R}^n)\,.$$

The above statement is equivalent in demanding that the mapping X is $\mathcal{A}-$measurable. The adoption of a r.v. in probability problems is necessary whenever the sample space Ω and the $\sigma-$algebra we need to work with have complicated structures that makes them very hard to work with. Whenever that is the case we specify a r.v. X that brings us to the Borel $\sigma-$algebra of $\mathbb{R}^n$ which is has a much nicer structure to work with. This trick however is allowed as long as we make sure that each subset of $\mathcal{B}(\mathbb{R}^n)$ is the image of an event in our initial $\sigma-$algebra. The following example clarifies what r.v. we are allowed to specify given a particular choice of $\sigma-$algebra.

Example 1.5

Adapted from [1, pp. 16–17]. Let us again consider the throw of a fair die. From Example 1.4 we constructed the $\sigma-$algebra $\mathcal{A} = \{\varnothing, A_1, A_2 := A_1^c, \Omega\}$ associated to the event of getting an even number ($A_1 = \{2, 4, 6\}$). Now let us assign a positive score $+1$ to the event A_1 and a negative score to its complement A_2. This introduces a r.v. defined as follows

$$X : \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \longrightarrow \begin{bmatrix} -1 & +1 & -1 & +1 & -1 & +1 \end{bmatrix}\,,$$

which we need to verify whether it satisfies the condition of measurability imposed by Definition 1.7. This means that we need to show that for all subsets $B \in \mathcal{B}(\mathbb{R})$ an event in $\mathcal{A}$ is defined as its pre-image. Let us set

$$\mathcal{B}(\mathbb{R}) = \{\varnothing, \{-1\}, \{+1\}, \{-1, +1\}, (-\infty, -1) \cup (-1, +\infty), (-\infty, +1) \cup (+1, +\infty), \mathbb{R}\}\,.$$

Now we compute the pre-images of each subset in $\mathcal{B}(\mathbb{R})$

$$
\begin{aligned}
X^{-1}(\varnothing) &= \varnothing \in \mathcal{A}\,,\\
X^{-1}(\{-1\}) &= A_2 \in \mathcal{A}\,,\\
X^{-1}(\{+1\}) &= A_1 \in \mathcal{A}\,,\\
X^{-1}(\{-1,+1\}) &= \Omega \in \mathcal{A}\,,\\
X^{-1}((-\infty,-1)\cup(-1,+\infty)) &= \varnothing \in \mathcal{A}\,,\\
X^{-1}((-\infty,+1)\cup(+1,+\infty)) &= \varnothing \in \mathcal{A}\,,\\
X^{-1}(\mathbb{R}) &= \varnothing \in \mathcal{A}\,,
\end{aligned}
$$

which thus entails that X is $\mathcal{A}-$measurable. Suppose now we construct a different mapping Y which retains the same score properties as X but it assigns a value of 0 to the outcome $\{5\}$, i.e.

$$
Y : \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \end{bmatrix} \longrightarrow \begin{bmatrix} -1 & +1 & -1 & +1 & 0 & +1 \end{bmatrix}\,.
$$

This function is clearly not a r.v. for the even/odd event (i.e. it is not $\mathcal{A}-$measurable) since

$$
Y^{-1}(\{0\}) = \{5\} \notin \mathcal{A}\,.
$$

Notice however that Y is $\mathcal{P}(\Omega)-$measurable, which is however uninteresting since it is associated to any possible event.

From the above we observe that the choice of a $\sigma-$algebra in the starting probability space somewhat constrains the type of r.v.s we are allowed to construct. We can flip this point of view as in practice one starts with a r.v. X instead and wants to generate a $\sigma-$algebra containing all the events that are of interest for X.

Lemma 1.1. *Let $X : \Omega \to \mathbb{R}^n$ be a r.v. on Ω, then*

$$
\mathcal{G}(X) := \sigma(X^{-1}(B)) = \{A \in \mathcal{P}(\Omega) \;:\; A = X^{-1}(B)\,,\;\; \forall B \in \mathcal{B}(\mathbb{R}^n)\}\,,
$$

is a $\sigma-$algebra on Ω.

In particular we refer to $\mathcal{G}(X)$ as the $\sigma-$algebra generated by X.

Proof. TBD... □

Highlight

Given this interpretation, we can see the $\sigma-$algebra generated by a r.v. X as the set of events containing the minimal amount of information that is sufficient to fully describe X.

As we mentioned above, we introduce the r.v.s as mapping from a probability space $(\Omega, \mathcal{A}, P)$ to $\mathbb{R}^n$ out of the necessity of being able to work with tuples of numbers rather than abstract events modelled as $\sigma-$algebras. Nevertheless the probability measure is still defined on the starting space and as such, the probability of an event $A \in \mathcal{A}$ is quantified indirectly as the measure associated to the pre-image of a subset $B \in \mathcal{B}(\mathbb{R}^n)$. That is we assign probability measures to the events of the $\sigma-$algebra generated by the r.v. in the following way

$$P(A \in \mathcal{G}(X)) = P(X^{-1}(B)) = P(X \in B)\,, \quad \forall B \in \mathcal{B}(\mathbb{R}^n)\,. \tag{1}$$

The interpretation of Equation 1 is as follow:

- the leftmost quantity is the *traditional* interpretation of probability of event A in a $\sigma-$algebra, however it is impractical (or impossible) to compute given the abstractness of the probability space;
- the quantity in the middle is the result of $A = X^{-1}(B)$, i.e. the event A is an element of not just any $\sigma-$algebra but specifically the one that is generated by a r.v. X;
- the rightmost quantity is the *novel* interpretation of probability measure on the probability space, which in this case it is the probability of the event A associated to the r.v. X taking **numerical** values in $B \in \mathcal{B}(\mathbb{R}^n)$. This quantity is easy to compute since B will be a set of $n-$dimensional tuples of real numbers, hence it is directly observable.

These 3 points are reppresented in Figure 1 below.

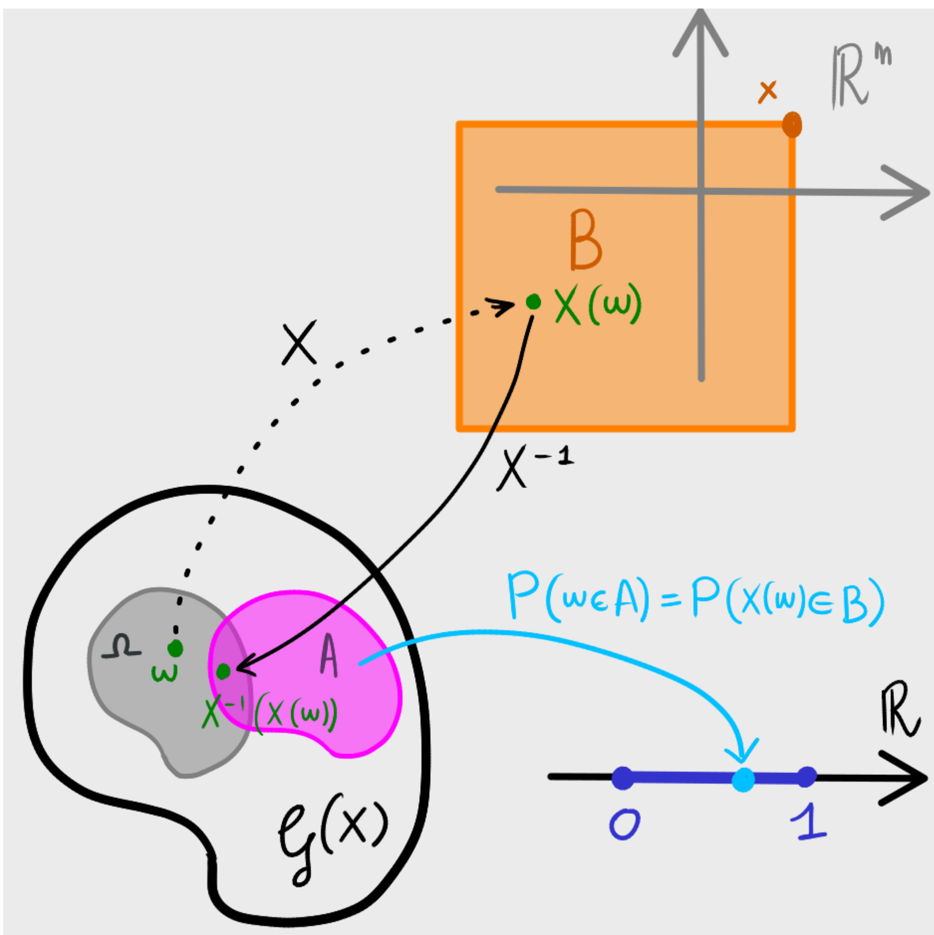

Figure 1: Visual depiction of the action of the r.v. X acting as a mapping from the sample space Ω to the **observable space** $\mathbb{R}^n$. Note that the probability measure P is defined on the $\sigma-$algebra generated by X and it is an indirect measure of the X taking values in $B \in \mathbb{R}^n$ being the pre-image of $A \in \mathcal{G}(X)$.

It would be ideal to be able to describe the properties of a r.v. X not by looking at all its values but rather on a small set of quantities that still describe its global behaviour.

Definition 1.8 (Expectation operator)**.** Let $X : \Omega \to \mathbb{R}^n$ be a r.v. on a probability space $(\Omega, \mathcal{G}(X), P)$, then the quantity

$$E(X) := \int_\Omega X \, dP \,,$$

is called the expected value, or mean, of X.

The expectation of X in Definition 1.8 above is defined as an integral w.r.t. the probability measure which we remark is defined over Ω which is often *unobservable.* So we want to take advantage of X being a map from the sample space to an observable space to make this quantity easier to compute.

Definition 1.9 (Distribution function)**.** Let $X : \Omega \to \mathbb{R}^n$ be a r.v. on a probability space $(\Omega, \mathcal{G}(X), P)$, then the mapping $F_X : \mathbb{R}^n \to [0, 1]$ is called a distribution function of X if it is defined by the following rule

$$F_X(x) := P(X \leq x) \,, \quad \forall x \in \mathbb{R}^n \,.$$

So now what we are constructing is a function F whose values are the probabilities of the r.v. taking numerical values bounded by some element in $\mathbb{R}^n$. The inputs of such distribution functions are all the possible numerical values that can bound the r.v. while the output will obviously be monotonically increasing with x (the larger the bound put on X the larger the associated pre-image event will be). We are now getting closer in being able to quantify the expected value of our r.v. since, as per Definition 1.8, what we need to do now is to introduce the integral in $\mathbb{R}^n$ through the distribution function.

Definition 1.10 (Probability density function (pdf))**.** Let $X : \Omega \to \mathbb{R}^n$ be a r.v. on a probability space $(\Omega, \mathcal{G}(X), P)$ and $F_X : \mathbb{R}^n \to [0, 1]$ be its distribution function, then the non-negative, integrable function $f : \mathbb{R}^n \to \mathbb{R}$ is called the density of X if, when exists, it is defined by the following rule

$$F_X(x) = \int_{-\infty}^{x} f(y) dy \,.$$

In essence, we are assigning an integrand function f, defined in the observable space $\mathbb{R}^n$, which characterises the distribution of the probabilities of X taking values in $\mathbb{R}^n$. This is exactly what we want because with a trivial equality of the distribution function F_X as provided in Definition 1.9 with the one provided in Definition 1.10 we can succintly write, remembering the interprepation provided for Equation 1, the hallmark of measure-theoretic probability theory

$$P(X \in B) = \int_B f(x) dx \,, \quad \forall B \in \mathcal{B}(\mathbb{R}^n) \,. \tag{2}$$

The importance of Equation 2 cannot be overstated. It provides a direct, computable quantity for measuring the events of a sample space via its probability measure by merely relying on the numerical realizations of the r.v. X in an observable space. Not only this, it actually provides a simple way of computing the expectation value of the r.v. in a nicer fashion compared to the abstract setting in Definition 1.8.

Lemma 1.2. *Let $X : \Omega \to \mathbb{R}^n$ be a r.v. on a probability space $(\Omega, \mathcal{G}(X), P)$ and $f : \mathbb{R}^n \to \mathbb{R}$ be its pdf, then for every integrable function $g : \mathbb{R}^n \to \mathbb{R}$ the expectation operator of $g(X)$ is*

$$E(g(X)) = \int_{\mathbb{R}^n} g(x) f(x) dx \,,$$

and in particular, for $g(X) = x$, we get

$$E(X) = \int_{\mathbb{R}^n} x f(x) dx \,.$$

Proof. TBD... □

Example 1.6

We want to put meaning into the interpretation of the quantities defined from Definition 1.8 to 1.10 and how they come together in the providing us with the identities in Equation 2 and the above Lemma. Let us consider...

Of particular importance in characteristing the pdf of a r.v. is not only the expected value but also higher-order *moments*,

Definition 1.11 (Central moments)**.** Let $X : \Omega \to \mathbb{R}^n$ be a r.v. on a probability space $(\Omega, \mathcal{G}(X), P)$, $f : \mathbb{R}^n \to \mathbb{R}$ be its pdf and $\langle X \rangle := E(X)$ being the mean of the pdf, then the quantity

$$\langle X \rangle_m = \int_{\mathbb{R}^n} (x - \langle X \rangle)^m f(x) dx \,,$$

is called the $m-$th central moment of X.

Arguably the most famous central moment in probability is the second moment about the mean, or variance

$$V(X) := \langle X \rangle_2 = \int_{\mathbb{R}^n} (x - \langle X \rangle)^2 f(x) dx = \langle X^2 \rangle - \langle X \rangle^2 \,, \tag{3}$$

which quantifies the average, squared departures of the realizations of a r.v. X w.r.t. its mean $E(X)$.

1.3.2 Methods of probability theory

Multivariate random variables

Joint, conditional and marginal probabilities Having established the connection between the abstract probability space and the observable space where the distribution and density functions live, we now wish to etsablish the same framework for a multivariate setting. That is to say we want to formaly characterise the probabilities of different events in the $\sigma-$algebra generated by X in terms of probability density functions of X and their central moments. To start let us assume that we have a probability space $(\Omega, \mathcal{A}, P)$ and we have two events $A, B \in \mathcal{A}$. Given a sample point $\omega \in \Omega$ we are told that it belongs to event $B \in \mathcal{A}$. Based on this information we want to assign a probability measure to the event A having previously established that B has occurred, i.e. we want to quantify the probability that $\omega \in A$ given that $\omega \in B$. One can naively interpret this as the probability assigned to the intersection of A with B, however this would be wrong as it disregard the fact that we are not accounting for the probability of $\omega \in B$. To do that let us define a new probability subspace as $(B, \mathcal{C}, \tilde{P})$, that is event $B \in \mathcal{A}$ is now a sample space and its $\sigma-$algebra is readily given by $\mathcal{C} = \{C \cap B \,, \ \forall C \in \mathcal{A}\}$. The new probability measure will be simply the rescaled w.r.t. the measure of B, i.e. $\tilde{P}(C) := \frac{P(C)}{P(B)}$. So now the probability of the sample point ω being in both B and A can be immediately quantified by measuring with $\tilde{P}$ the *parts* of $A \in \mathcal{A}$ that lie in the intersection with B being interpreted as a probability space

$$\tilde{P}(A \cap B) = \frac{P(A \cap B)}{P(B)} \,. \tag{4}$$

In other words Equation 4 motivates the adoption of two distinct types of probability measures: one is $\tilde{P}(A \cap B) =: P(A|B)$ which is the (**conditional**) probability of $\omega \in A$ given that $\omega \in B$; the other is $P(A \cap B) =: P(A, B)$ which is the (**joint**) probability of $\omega \in A \cap B$.

Example 1.7

Adapted from Wikipedia[a].

[a] https://en.wikipedia.org/wiki/Conditional_probability_distribution

Example 1.8

Adapted from Wikipedia[a].

[a] https://en.wikipedia.org/wiki/Marginal_distribution

Example 1.9

Adapted from a blog post[a].

[a] https://jeffreyfreeman.me/blog/conditional-probabilities-and-bayes-theorem/

Bayesian inference

Theorem 1.2: Bayes' theorem

$$\underbrace{p(X=x|Y=y)}_{\text{posterior}} = \frac{\overbrace{p(X=x,Y=y)}^{\text{joint}}}{\underbrace{p(Y=y)}_{\text{marginal}}} = \frac{\overbrace{p(Y=y|X=x)}^{\text{likelihood}}\overbrace{p(X=x)}^{\text{prior/bias}}}{\underbrace{\sum_{y'\in\mathcal{Y}} p(X=x|y')p(y')}_{\text{evidence}}}$$

A visual intuition[1] of Bayes' theorem 1.2 helps with its understanding.

Example 1.10

Taken from [2, pp. 46–47]. $H=1$ positive COVID-19 infection, $H=0$ negative COVID-19 infection. $Y=1$ positive test, $Y=0$ negative test. We thing of H being an unkown class label and Y being a feature vector; we are interested in estimating $p(H|y)$ in a form of binary classification. The sensitivity (true positive rate) is $p(Y=1|H=1)=0.875$ while the specificity (true negative rate) is $p(Y=0|H=0)=0.975$. The prior $p(H=1)=0.1$ is quantified as the prevalence of the disease in a given area. Conversely $p(H=0)=1-p(H=1)=0.9$ quantifies the absence rate of the disease in the same area. Suppose you test positive, what is the chance of you actually being infected by COVID-19?

$$p(H=1|Y=1) = \frac{\overbrace{p(Y=1|H=1)}^{\text{sensitivity}}\overbrace{p(H=1)}^{\text{prevalence}}}{\underbrace{P(Y=1|H=1)}_{\text{sensitivity}}\underbrace{p(H=1)}_{\text{prevalence}}+\underbrace{P(Y=1|H=0)}_{\text{specificity}}\Big(1-\underbrace{p(H=1)}_{\text{prevalence}}\Big)} = 0.795\,.$$

[1] https://setosa.io/ev/conditional-probability/

If instead you tested negative the chances of having contracted the infection are

$$p(H=1|Y=0) = \frac{\Big(1-p(Y=1|H=1)\Big)p(H=1)}{\Big(1-p(Y=1|H=1)\Big)p(H=1)+p(Y=0|H=0)\Big(1-p(H=1)\Big)} = 0.014\,.$$

Bayesian interpretation of probability[2] differs from the frequentist point of view. The inference requires 4 ingredients which we will address, individually, in the following.

What is the posterior? The posterior probability distribution[3] describes...

What is the likelihood? The likelihood[4] is a conditional probability that...

What is the prior? The prior[5] or bias quantifies our assumptions regarding the probability distribution of the phenomena of interest.

What is the evidence? The evidence[6] or marginal likelihood is a statistical estimation of the observed truth of the conditions upon which our inference is based.

Maximum likelihood estimation

The maximum likelihood estimation[7] (MLE) serves as a tool to build the likelihood part of Bayesian inference.

1.3.3 Discrete-time processes

Transition probability

Random walk

Markov chain

See here[8].

[2] `https://en.wikipedia.org/wiki/Bayesian_inference`
[3] `https://en.wikipedia.org/wiki/Posterior_probability`
[4] `https://en.wikipedia.org/wiki/Likelihood_function`
[5] `https://en.wikipedia.org/wiki/Prior_probability`
[6] `https://en.wikipedia.org/wiki/Marginal_likelihood`
[7] `https://www.probabilitycourse.com/chapter8/8_2_3_max_likelihood_estimation.php`
[8] `https://en.wikipedia.org/wiki/Markov_chain`

Chapman-Kolmogorov equation (CKE) See here[9].

1.3.4 Stochastic calculus

Brownian motion

Ornstein-Uhlenbeck process (OUP)

See here[10].

Fokker-Plank equation (FPE)

Kramers' escape problem

See here[11].

[9] https://en.wikipedia.org/wiki/Chapman\OT1\textendashKolmogorov_equation
[10] https://en.wikipedia.org/wiki/Ornstein-Uhlenbeck_process
[11] https://en.wikipedia.org/wiki/First-hitting-time_model