Contents

1 Basics of probability

We start with the probability triplet: $(\Omega, \mathscr{H}, \mathbb{P})$ Here Ω is the set of sample space, $\mathscr{H}$ is the σ-algebra built upon Ω and $\mathbb{P}$ is the probability measure. Since $\mathbb{P}$ is a measure, it will take values in $\mathbb{R}$.
We are interested in probability measure, which means:

- $\mathbb{P}$ is a **finite measure** and $\mathbb{P}(\Omega) = 1$;
- $\omega \in \Omega$ will be called **outcomes**.

So consider the example of the roll of the die. If we roll it,

$$\Omega = \underbrace{\{1, 2, 3, 4, 5, 6\}}_{\text{outcomes}}$$

And if we considere the elements $A \in \mathscr{H}$ (which will be subsets of Ω) will be called **events**.
We want to quantify the possibility that the event A occurs: we want to measure, through $\mathbb{P}$, the set A: from a measure theory point of view, it's only sets in the σ-algebra.
The probability measure has the following properties:

- $\mathbb{P}(\Omega) = 1, \quad \mathbb{P}(\varnothing) = 0$
- **monotonicity of** $\mathbb{P}$: take 2 events $H, K \in \mathscr{H}$ such that $H \subset K$. Then $\mathbb{P}(H) \leqslant \mathbb{P}(K)$[1].
- **finite additivity**: take $H, K \in \mathscr{H}$ such that $H \cap K = \varnothing$. The $\mathbb{P}(H \cup K) = \mathbb{P}(H) + \mathbb{P}(K)$;
- **countable additivity**: this requires that we consider collection of events. We denote them in this way:
$$(H_n)_{n\in\mathbb{N}} \subset \mathscr{H}$$
with $\mathbb{N} = \{0, 1, 2, 3, \ldots\}$ and $\mathbb{N}^\star = \{1, 2, 3, 4, \ldots\}$ such that they are disjoint pairwise (except identical pairs). Then
$$\mathbb{P}\left(\bigcup_n H_n\right) = \sum_n \mathbb{P}\left(H_n\right)$$
- **Boole inequality (sub-additivity)**: if we have a collection $(H_n)_{n\in\mathbb{N}} \subset \mathscr{H}$ (not necessarily disjoint) then
$$\mathbb{P}\left(\bigcup_n H_n\right) \leqslant \sum_n \mathbb{P}\left(H_n\right)$$
- **sequential continuity**: consider the sequence $(H_n)_{n\in\mathbb{N}} \subset \mathscr{H}$ such that $H_n \nearrow H \in \mathscr{H}$ (H_n is an increasing sequence of numbers that has H as limit) then $\mathbb{P}(H_n) \nearrow \mathbb{P}(H)$. Moreover, if $(F_n)_{n\in\mathbb{N}} \subset \mathscr{H}$ such that $F_n \searrow F \in \mathscr{H}$ then $\mathbb{P}(F_n) \searrow \mathbb{P}(F)$. The second property is actually true because $\mathbb{P}$ is finite (it is not true for infinite measures).

In measure theory we encounter the concept of **negligible sets**: these are sets of measure zero or non measurable sets included in measure zero sets. In probability theory, sets are **events**: so we have negligible events (events with probability 0 or non measurable events included in events with probability 0). Analogously, in measure theory a property which holds **almost everywhere** is allowed not to hold on negligible sets. In probability theory a property which holds **almost surely** is allowed not to hold on negligible events. We also have, in measure theory, *measurable functions* that in probability theory are **random variables**.

1.1 Random variables

Consider a measurable space $(E, \mathscr{E})$.

[1] note that the notation is loose since we have proper subset on one side and leq on the other side. But this is not much of a problem, since i will kill myself very soon.

Definition 1.1

A mapping $X : \Omega \to E$ is called **random variable taking values in E** if X is measurable relative to $\mathcal{H}$ and $\mathcal{E}$.

What does it mean[2]? The inverse image of the set A through X ($X^{-1}A$) with $A \in \mathcal{E}$ is actually the set of the ωs such that $X(\omega)$ arrives to A. So

$$X^{-1}A = \{\omega \in \Omega : X(\omega) \in A\} = \{X \in A\}$$

so that $X^{-1}A$ is an event for all A in $\mathcal{E}$.

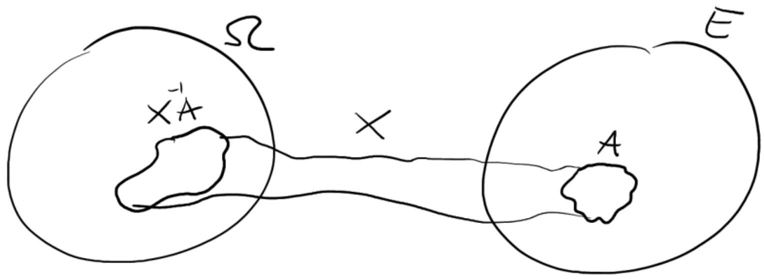

Figure 1: this is an early reminder of the fact that I will take my own life very soon.

So if $X^{-1}A$ is measurable by $\mathbb{P}$ then it is in $\mathcal{H}$: otherwise it is not in $\mathcal{H}$. So

$$\mathbb{P}(X^{-1}A) = \mathbb{P}\left(\{\omega \in \Omega : X(\omega) \in A\}\right).$$

The message is that I am interested/able to evaluate $\mathbb{P}$ over the set only if what I am evaluating is indeed an event (which means: it belongs to $\mathcal{H}$[3]). If something is not in $\mathcal{H}$ get it off my fucking face man and kill yourself NOW[4]. This is the only restriction for a random variable. E can be whatever we need it to be: a graph, a tree, your mom being absolutely [redacted] by me. But most of the times, we have $E = \mathbb{R}$ or $E = \mathbb{R}^d$ with respectively $\mathcal{E} = \mathcal{B}^5(\mathbb{R}) = \mathcal{B}_{\mathbb{R}}$ and $\mathcal{B}_{\mathbb{R}^d}$.

Remark

The simplest random variables are indicator functions of events. Example: take $H \in \mathcal{H}$. Define the function

$$\mathbb{1}_H : \Omega \to \mathbb{R}$$
$$\mathbb{1}_H(\omega) = \begin{cases} 0 & \omega \notin H \\ 1 & \omega \in H \end{cases}$$

Remark

A random variable is said to be **simple** if it takes only finitely many values in $\mathbb{R}^d$.

Remark

A random variable is said to be **discrete** if it takes only countably many values.

[2] who asked

[3] il lettore più arguto avrà notato che, a questo punto, il dio è ormai irrimediabilmente cane.

[4] 

[5] Borel σ-algebra

Definition 1.2

Distribution of a random variable. Let X be a random variable taking values in $(E, \mathcal{E})$ and let μ be the image of $\mathbb{P}$ under X, that is,

$$\mu(A) = \mathbb{P}(X^{-1}A) = \mathbb{P}(X \in A) = \mathbb{P} \circ X^{-1}(A)^{a}, \ A \in \mathcal{E}.$$

Then μ is a probability measyre on $(E, \mathcal{E})$ and it is called **distribution of X**.

[a] you would know this if you knew fucking measure theory I guess

So we map, by means of X, sets belonging to $\mathcal{E}$ into $\mathcal{H}$ and then evaluates this sets by means of the measure $\mathbb{P}$. This is what we mean when we say that distributions are ultimately built with the probability measure and the random variable. Distribution is itself a measure!

Remark

You should remember (LOL) that when we want to specify a measure on a σ-algebra, it's enough to do it on a *p-system*[a] generating that σ algebra: by means of the monotone class theorem we are then able to extend the measure to the σ-algebra.
This means that to specity μ it is enough to specify it on a *p-system* which generates $\mathcal{E}$. For example, consider $E = \overline{\mathbb{R}}, \mathcal{E} = \mathcal{B}_{\overline{\mathbb{R}}}$. Consider the collection of sets $[-\infty, x]$, $x \in \mathbb{R}$ which is of course a p-system because it is closed under intersection. Moreover, this shit generates the Borel sigma algebra on $\overline{\mathbb{R}}$.
If we want to define a distribution, that is a measure, it is enough to define it on this p-system. Imagine that we apply our distribution measure to one set of this p-system

$$c(x)^{b} = \mu\left([-\infty, x]\right) = \mathbb{P}(X \leq x), \qquad x \in \mathbb{R}$$

by the monotone class theorem. So we have now specified the measure on the p-system. The part $\mathbb{P}(X \leq x)$ reminds us of the undergraduate times[c]: it is a distribution function! This is what our professor did implicitly to avoid using measure theory[d]

[a] a p-system is a simpler object than a σ-algebra: it is simply a collection of sets closed under intersection
[b] because it is a function of x
[c] I already wanted to kill myself at that time.
[d] I have noticed that my life has not benefited in ANY form since I was introduced to measure theory.

1.2 Functions of random variables

Consider X, a random variable taking values in $(E, \mathcal{E})$ and consider further a measurable space $(F, \mathcal{F})$. Let $f : E \to F$ be a measurable function relative to $\mathcal{E}$ and $\mathcal{F}$[6]. This function should me measurable by means of $\mathbb{P}$, otherwise we couldn't do anything useful with it. Consider the composition

$$Y = f \circ X \qquad \text{such that } Y(\omega) = f \circ X(\omega) = f\big(X(\omega)\big), \ \omega \in \Omega.$$

This composition is a random variable taking values in $(F, \mathcal{F})$

[6] This basically means that this bitch won't do anything evil. The whole point of measure theory, σ algebras and all other shit is to ensure everything behaves.