Probability theory notes

Kotatsu

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
$$\mathbb{P}\left(\bigcup_n H_n\right) = \sum_n \mathbb{P}(H_n)$$

- **Boole inequality (sub-additivity)**: if we have a collection $(H_n)_{n \in \mathbb{N}} \subset \mathscr{H}$ (not necessarily disjoint) then
$$\mathbb{P}\left(\bigcup_n H_n\right) \leqslant \sum_n \mathbb{P}(H_n)$$

---

[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.

- **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 $(E, \mathscr{E})$.

> **Definition 1.1**
>
> A mapping $X : \Omega \to E$ is called **random variable taking values in E** if X is measurable relative to $\mathscr{H}$ and $\mathscr{E}$.

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

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

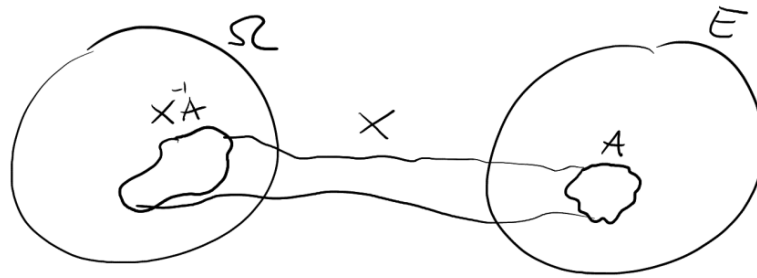so that $X^{-1}A$ is an event for all A in $\mathscr{E}$.

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 $\mathscr{H}$: otherwise it is not in $\mathscr{H}$. 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 $\mathscr{H}$[3]). If something is not in $\mathscr{H}$ 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 ▉▉▉▉▉ by me. But most of the times, we have $E = \mathbb{R}$ or $E = \mathbb{R}^d$ with respectively $\mathscr{E} = \mathscr{B}$[5]$(\mathbb{R}) = \mathscr{B}_{\mathbb{R}}$ and $\mathscr{B}_{\mathbb{R}^d}$.

> **Remark**
>
> The simplest random variables are indicator functions of events. Example: take $H \in \mathscr{H}$.

---

[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

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.

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

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

by the monotone class theorem. So we have now specified the measure on the p-system. The part $\mathbb{P}(X \le x)$ 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 have been 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 $(F, \mathcal{F})$ which comes from the fact that measurable functions of measurable functions are still measurable.

Definition 1.3

Consider two random variables X, Y taking values in $(E, \mathcal{E})$ and $(F, \mathcal{F})$ respectively. Consider the pair

$$Z = (X, Y) : \Omega \to E \times F.$$

Why would we want to call it Z? It's because, beside being a random vector, it is in turn a random variable:

$$Z(\omega) = (X(\omega), Y(\omega)).$$

Since $E \times F$ is a product space, we should attach it the product σ-algebra. So Z is a random variable taking values in $E \times F$.

Note that the product space $E \times F$ is endowed with the σ-algebra $\mathcal{E} \otimes \mathcal{F}$, that is the product σ-algebra generated by the collection of all possible rectangles between E and F. We frequently have to look to special cases like random vectors that must take values in measurable spaces for them to make sense. This measurable space is naturally generated by the product σ-algebra (but it may be generated by other σ-algebras[7]!).

Definition 1.4

We call **joint distribution** of X and Y the distribution of Z.

This is interesting, since we know that this variable has the specific structure of a random vector: we identify the distribution of this vector as the joint distribution of its two coordinates[8].

Remark

The product σ-algebra $\mathcal{E} \otimes \mathcal{F}$ is generated by the p-system of measurable rectangles.

On the product space, it is enough to only specify it on this p-system.
Let denote with π the joint distribution of X, y. It is sufficient to specify

$$\pi(A \times B) = \mathbb{P}(X \in A, Y \in B) \qquad \forall A \in \mathcal{E}, B \in \mathcal{F}.$$

We exploited the measurability of X and Y

Definition 1.5

Given the joint distribution π, consider sets $A \in \mathcal{E}, B \in \mathcal{F}$. Then we call **marginal distribution of X**

$$\mathbb{P}(X \in A) = \pi(A \times F) \qquad \forall A \in \mathcal{E}$$

and we call **marginal distribution of Y**

$$\mathbb{P}(Y \in B) = \pi(E \times B) \qquad \forall B \in \mathcal{F}.$$

We call it distribution because it is actually a measure! So we can call it with the notation of measure

$$\mu(A) = \mathbb{P}(X \in A) = \pi(A \times F) \qquad \forall A \in \mathcal{E}$$

---

[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.

[7]Repeatedly inflicting painful kicks on my gonads.

[8]it eludes me how anyone could find this interesting. We have to think about the whole vector as being distributed like its components separately

and

$$\nu(B) = \mathbb{P}(Y \in B) = \pi(E \times B) \qquad \forall B \in \mathscr{F}.$$

This actually means that the second coordinate is fixed in being the <u>whole space</u> F. Think about integrating the second coordinate along the real line when doing <u>marginal distributions</u>... this is the same thing here.

Now that we have joint and marginal distributions, what is the next step[9]?

> **Definition 1.6**
>
> Let X, Y be random variables taking values in $(E, \mathscr{E})$ and $(F, \mathscr{F})$ respectively and let μ and ν be their respective distributions. Then X and Y are said to be **independent** if their joint distribution is the product measure formed by their marginals.
>
> $$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B) \qquad \forall A \in \mathscr{E}, B \in \mathscr{F}.$$
>
> This also means that
>
> $$\pi = \mu\nu$$

Here the marginals do not interact with each other. This is true for two random variables and we need[10] something more general.

> **Definition 1.7**
>
> Let $(X_1, X_2, \ldots, X_n)$ be a finite collection of random variables. The collection is said to be an **independency** if the distribution of $(X_1, X_2, \ldots, X_n)$ is the product of $\mu_1, \mu_2, \ldots, \mu_n$ where μ_i is the distribution of X_i, for $i = 1, \ldots, n$.

Cynlar is stupid I wish him dead to be frank for this independecy shit. Independecy is not even an english word. What the fuck? Anyway, what about infinite collections?

> **Definition 1.8**
>
> Let $(X_n)_n$ be an infinite collection of random variables. It is said to be an **independency** if every finite sub-collection of it is an independecy.

We now turn to stochastic processes[11]! But first...

1.3 Infinite product spaces

Let T be an arbitrary (countable or uncountable) set. We will think about this set as an "index" set. For each $t \in T$ consider the measurable $(E_t, \mathscr{E}_t)$. So we have a space for each index (plenty of measurable spaces hanging around). Consider a point x_t in E_t for each $t \in T$. The collection[12] $(x_t)_{t \in T}$. If $(E_t, \mathscr{E}_t) = (E, \mathscr{E})$ then $(x_t)_{t \in T}$ is actually a function of T taking values on $(E, \mathscr{E})$.

The set F of all possible functions $x = (x_t)_{t \in T}$ is called the **product space**$((E_t \mathscr{E}_t))_{t \in T}$.

This is the natural generalization of what we do when we construct product spaces, albeit with a different notation. Usually F is denoted by $X_{t \in T} E_t$. But we know we also need a σ-algebra ...

A **rectangle** in F is a subset of the form

$$\{x \in F : x_t \in A_t \ \forall t \in T\}$$

Where A_t differs from E_t for only a finite number of t.

1.4 Stochastic processes

---

[9]Committing the deed, methinks.

[10]No.

[11]Please no.

[12]We could consider it a function of t but that wouldn't be exactly correct since each t has a different measurable space. We may have the same space but it's not true in general... I am thrilled to say the least.