Notes on

Probability

from the lectures of
Federico Bassetti and Matteo Gregoratti,
Politecnico di Milano

S. Licciardi[1], PoliMI Undergraduate Student

Academic Year 2023-2024

[1]simone.licciardi@mail.polimi.it

Contents

Complements to Chapter 1

0.1 Probability construction

Operatively, there are two ways of furnishing a (measure of) probability on a measurable space $(\Omega, \mathcal{F})$: it may be claimed or inferred. In the first case, its law is given *a priori*, while in the second it is deducted *a posteriori*, from the knowledge of the probability values on some elementary events, that is.

The first is common in applications such as Bayesian Statistics, where you make an hypothesis on the distribution and then test it, while the second is also characteristic of measure theory. The key difference lies in the fact that we either have complete or minimal information about the probability.

In both cases some check are in order, as we need to ensure that the probability is *coherent*. Note that in the second case, we also need to ensure that the one generated is unique, the minimal information is *sufficient*, that is.

0.1.1 Discrete setting

Some results of this kind were produced already: we have dealt with the case of discrete partitions of the sample space Ω. We quote such theorems.

Theorem 0.1.1 (Existence and Uniqueness)**.** *Let $\mathcal{E} = \{E_k\}_{k \in I}$ be a discrete partition of Ω and let the function $\mathrm{p} : E_k \in \mathcal{E} \to p_k \in \mathbb{R}$ be such that*

$$\begin{cases} \sum_{k \in I} p_k = 1 & \text{(1a)} \\ p_k \geq 0 \quad \text{for all } k \in I. & \text{(1b)} \end{cases}$$

Then, there exist a unique probability $\mathbb{P}$ on $\sigma(\mathcal{E})$ such that $\mathbb{P}$ and p agree on $\mathcal{E}$. A function p with the properties (1a) and (1b) of is called a discrete probability density.

Morally, the discrete probability density on a partition describes a unique and consistent probability on the generated σ-algebra. Moreover, we have an explicit description of each event $E \in \sigma(\mathcal{E})$ and its probability $\mathbb{P}(E)$.

Lemma 0.1.2. *Let $\mathcal{E} = \{E_k\}_{k \in I}$ be a discrete partition of Ω. Then*

$$\sigma(\mathcal{E}) = \left\{ \bigcup_{k \in J} E_k \ \text{ for } J \subset I \right\}.$$

Now suppose that $\mathrm{p} : E_k \in \mathcal{E} \to p_k \in \mathbb{R}$ is a discrete probability density. Then, if $\mathbb{P}$ is the probability defined in Theorem 0.1.1,

$$\mathbb{P}\left(\bigcup_{k \in J} E_k \right) = \sum_{k \in J} \mathrm{p}_k \quad \text{for all } J \subset I.$$

Remark 0.1.3. A special case worth mentioning is that of the atomic partition on a discrete Ω. Here, the σ-algebragenerated is $\mathcal{P}(\Omega)$ and the discrete probability density is commonly referred to as $p(\{\omega\}) = p_\omega$. By the precedent results, p defines a unique agreeing probability $\mathbb{P}$ such that

$$\mathbb{P}(E) = \sum_{\omega \in E} p_\omega \quad \text{for all } E \subset \Omega. \tag{2}$$

0.1.2 Carathéodory Theorem

For more general settings the short message is that Theorem 0.1.1 holds, provided some conditions, while no explicit characterization like that of Lemma 0.1.2 is possible.

We present a powerful theorem of measure theory, that does just that. It will allow us to extend a *pre-probability*, a function with some coherence that is defined on a collection smaller than a σ-algebra, to a probability, in a unique fashion.

Definizione 0.1.4. Let $\mathcal{A}$ be an algebra defined on Ω. Then, $\tilde{\mathbb{P}} : \mathcal{A} \to \mathbb{R}$ is a *pre-probability* if

1. $\tilde{\mathbb{P}}(\Omega) = 1$ (Normalization)
2. $\tilde{\mathbb{P}}\left(\bigcup_{i=1}^n A_i\right) = \sum_i^n \tilde{\mathbb{P}}(A_i)$ for $A_i \in \mathcal{A}$ (Additivity)
3. if $\bigcup_{i\in\mathbb{N}} A_i \in \mathcal{A}$, then $\tilde{\mathbb{P}}\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i\in\mathbb{N}} \tilde{\mathbb{P}}(A_i)$.

Remark 0.1.5. The latter is the condition that ensure coherence with respect to the probability, and can be read as a *need-based σ-additivity.*

Theorem 0.1.6 (Carathéodory's Theorem)**.** *Let $\mathcal{A}$ be an algebra defined on Ω, and suppose that $\tilde{\mathbb{P}} : \mathcal{A} \to \mathbb{R}$ is a pre-probability. Then, there exists a unique probability $\mathbb{P} : \sigma(\mathcal{A}) \to \mathbb{R}$ such that $\tilde{\mathbb{P}}$ and $\mathbb{P}$ agree on $\mathcal{A}$.*

Remark 0.1.7. This version of the Carathéodory's Theorem is of theoretical interest and provides two results. First, it shows that **existence** of a consistent extension is guaranteed just by requiring the coherence of $\tilde{\mathbb{P}}$ with the conditions that the define a probability: these are encoded in Definition 0.1.4. Moreover, the theorem quantifies the idea that if $\tilde{\mathbb{P}}$, the information provided about $\mathbb{P}$ that is, is defined on a large enough collection, then its extension is **unique**. In particular, we require $\mathbb{P}$ to be given on an algebra, a much smaller collection than a σ-algebra.

We can refine the result for practical purposes by exploiting the equivalence between σ-additivity and continuity[1].

Lemma 0.1.8 (Carathéodory's Theorem, continuity characterization)**.** *Let $\mathcal{A}$ be an algebra defined on Ω, and suppose that $\tilde{\mathbb{P}} : \mathcal{A} \to \mathbb{R}$ satisfies*

1. $\tilde{\mathbb{P}}(\Omega) = 1,$
2. $\tilde{\mathbb{P}}\left(\bigcup_{i=1}^n A_i\right) = \sum_i^n \tilde{\mathbb{P}}(A_i),$
3. *if* $A_i \downarrow \emptyset$, *then* $\tilde{\mathbb{P}}(A_i) \downarrow 0,$

where $A_n \in \mathcal{A}$. Then, there exists a unique probability $\mathbb{P} : \sigma(\mathcal{A}) \to \mathbb{R}$ such that $\tilde{\mathbb{P}}$ and $\mathbb{P}$ agree on $\mathcal{A}$.

Remark 0.1.9. The usefulness of this is that monotone continuity, that is showing a limit is 0, is generally much easier than working on countable unions to arbitrary sets. It is of theoretical interest that this only holds if we use an algebra: as we will see, if it wasn't for this characterization weaker conditions on the collection (namely, that it is a π-system instead of an algebra) could be used.

0.1.3 Practical construction, set structure

An algebra is still a sizeable class, making it hard to define $\mathbb{P}$. This is what Dynkin's $\lambda - \pi$ Theorem was designed to simplify.

Definizione 0.1.10 (π-system)**.** Given a set Ω, a collection of subsets $\mathcal{C}$ is a π-system if stable it is under finite intersection. Explicitly, if $A_1, \ldots, A_n \in \mathcal{C}$ implies that $\bigcap_{i=1}^n A_i \in \mathcal{C}$.

Remark 0.1.11. Operatively, it suffices to show that $A \cap B \in \mathcal{C}$ whenever $A, B \in \mathcal{C}$, for $\mathcal{C}$ to be a π-system by inductive argument.

[1]This result from measure theory is taken to be known and the details are out of the scope of these notes.

Definizione 0.1.12 (λ-system)**.** Given a set Ω, a collection of subsets $\mathcal{C}$ containing Ω is a λ-system if it is stable under (pairwise) disjoint countable union and proper difference. Explicitly, if $A_1, A_2, \dots \in \mathcal{C}$ and $A_i \cap A_j = \emptyset$ for all i,j implies that $\bigcup_{i=1} A_i \in \mathcal{C}$ and if $A, B \in \mathcal{C}$ implies that $A \setminus B \in \mathcal{C}$.

Remark 0.1.13. Proper difference stability is more intelligible under the light of its equivalence with complementation. Recall that set difference can be charachterized as follows:

$$A \setminus B = A \cap B^{\mathsf{C}},$$

where complementation is taken with respect to a space containing both A and B. In particular, if $B \subset A$ the right hand side boils down to the complementation of B with respect to A. Thus, that a collection is closed under proper difference is the same as complementation stability with respect to all $A \in \mathcal{C}$. Formally, for all A in $\mathcal{C}$ it holds that $\mathcal{G}_A = \{B \in \mathcal{C} \text{ such that } B \subset A\}$ is stable under complementation.

Importantly, proper difference stability implies complementation stability with respect to Ω, since $\Omega \in \mathcal{C}$,.

Remark 0.1.14. Remark 0.1.13 highlight the difference between a λ-system and a σ-algebra: in the former the countale unions need to be *disjoint*, while in the latter not so. An example of collection that is closed under disjoint union, but not union, is

$$\mathcal{C} = \{\emptyset, \{a, b\}, \{c, d\}, \{a, c\}, \{b, d\}, \{a, b, c, d\}\}$$

Let us investigate the relationship between a σ-algebra and these simpler systems.

Lemma 0.1.15. *A collection is a σ-algebra if and onl if it is both a π-system and a λ-system.*

This is a much stronger relation than it looks, as the following theorem shows.

Theorem 0.1.16 ($\lambda - \pi$ Theorem)**.** *If $\mathcal{C}$ is a π-system and $\mathcal{D}$ is a λ-system containing it, then $\sigma(\mathcal{C}) \in \mathcal{D}$.*

Remark 0.1.17. An equivalent statement would be that $\sigma(\mathcal{C}) = \lambda(\mathcal{C})$, where the latter is the smallest λ-system generated by $\mathcal{C}_\pi$. That this exists unique follows from the intersection of λ-systems being a λ-system.

A similar result to 0.1.16 is the following, where we weaken the requirements on $\mathcal{D}$ so that it is only required to be a monotone class, at the expense of stronger stability for $\mathcal{C}$.

Theorem 0.1.18 (Monotone Class Theorem, set version)**.** *If $\mathcal{A}$ is an algebra and $\mathcal{M}$ is a monotone class containing it, then $\sigma(\mathcal{A}) \in \mathcal{M}$.*

Remark 0.1.19. Here, an equivalent statement would be that $\sigma(\mathcal{A}) = \mathrm{M}(\mathcal{A})$, where the latter is the smallest monotone system generated by $\mathcal{A}$ and existence and uniquess are as above.

Remark 0.1.20. One of the main applications of the Monotone Class Theorem is that of showing that certain property is satisfied by all sets in an σ-algebra, generally starting by the fact that the field generating the σ-algebra satisfies such property and that the sets that satisfies it constitutes a monotone class.

The following is an immediate application of Theorem 0.1.6 or 0.1.18 equivalently.

Lemma 0.1.21. *Let $\mathbb{P}_1$ and $\mathbb{P}_2$ be probabilities on $\sigma(\mathcal{C})$ and $\mathcal{C}$ be a π-system. If $\mathbb{P}_1$ and $\mathbb{P}_2$ agree on $\mathcal{C}$, then they agree on $\sigma(\mathcal{C})$.*

Remark 0.1.22. The lemma is of practical interest. Evidently, it allows easier comparison between probabilities, as it suffices to check equality on a much smaller collection than the domain. Importantly, the lemma also furnishes sharp[2] conditions for uniqueness in Theorem 0.1.6: it allows us to define the pre-probability on just a π-system containing Ω, and deduce from that its unique extension.

[2]Mathematical gibberish for “minimal”

0.1.4 Practical construction of probability

This continues the precedent section, where we have shown how to divide the construction of the sigma algebra into its components. Now we set as objective that of showing the passages of the most common practice in probability construction: that of defining the pre-probability on a π-system, extending it to a semi-algebra, and then to an algebra in order to apply Carathéodory. That of going from the π-system to the algebra operatively.

Borel's σ-algebra

σ-algebra on Bernoulli Space

Bassetti Unbound

Exercise 1. *Let Ω be a set and let $A \subset \Omega$ a subset from it. Then, show that $\{A, A^{\mathsf{c}}, \emptyset, \Omega\}$ is a σ-algebra.*

Notes. This is the easiest nontrivial σ-algebra. It models a bet: the event may either happen or not (or nor could happen, that is the same as all the outcomes being realized). (??)

Exercise 2. *Let $\{\mathcal{F}_\alpha\}_{\alpha \in I}$ be a colletion of σ-algebras. Is $\bigcap_{\alpha \in I} \mathcal{F}_\alpha$ a σ-algebra. What about $\bigcup_{\alpha \in I} \mathcal{F}_\alpha$?*

Notes. This[3] justifies minimality arguments on the function $\sigma(\cdot)$. Read this masterpiece, this essay and this very general and technical site .

Exercise 3. *Prove the well-definiteness of $\sigma(\mathcal{E})$ as the minimal σ-algebra containing $\mathcal{E}$.*

Notes (Sketch of proof). Let $\Sigma(\mathcal{E})$ be the collection of all the σ-algebras containing the collection $\mathcal{E}$ of subsets of Ω. (Prove that) $\Sigma(\mathcal{E})$ is not empty, and the family intersects to $\bigcap_{S \in \mathcal{E}} S = \sigma(\mathcal{E})$.

Exercise 4. *Let E_1,E_2 be events of Ω. Think of a sample space and construct a measure of probability $\mathbb{P} : \mathcal{P}(\Omega) \to \mathbb{R}$ such that the two events are independent and $\mathbb{P}(E_1) = \mathbb{P}(E_2) = \frac{1}{2}$.*

Notes. (??)

Exercise 5.

Notes.

[3] The answer is that the first is, in fact, a σ-algebra, while the second not so, as it does not contain crossed unions and intersections.