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 $(\Omega, \mathcal{F})$: we provide *a priori* the complete description of it, or we deduct it from the information we have. That is, key difference is how much information about the probability we are starting with. The first is common in applications such as Bayesian Statistics, where you make an hypothesis on the distribution and then update it with experiments, while the second is recurrent in modelisation.

In both cases some check are in order, as we need to ensure that the information we have is *coherent*, that is there is a probability which fits the given description, and *sufficient* to uniquely charachterize the probability.

We start with the complitely resolved case of discrete partitions, then move to theoretical results on general settings, and finally deal with the practical aspects of the matter.

0.1.1 Discrete setting

The case of discrete partitions of the sample space Ω is complitely resolved: that is we know everything about it with the bare minimal information. We state the theorems that allow and describe this kind of reasoning.

Remark 0.1.1. The reason why we care about partitions is that they model experiments where, at the end, only one of some states of the system is true. Think of rolling 4 dices, and being interested in the results of the first only. Then, only (and exactly) one the *states* "the first dice rolled to n" can suceed! Formally, that is equivalent to observing that even though $\omega \in \Omega$ describes the 4 results of the 4 rolls, that is it contains information we don't really care about (the other 3 rolls), we can partition the set according to the first result only. Importantly, the remaining structure of Ω is *not* garbage: for example, partitioning for the first roll could just be the first part of answering "how likely is it that the second dice rolls to 3, given the first rolled to 1?".

Theorem 0.1.2 (Existence and Uniqueness)**.** *Let $\mathcal{E} = \{E_k\}_{k\in I}$ be a discrete*[1] *partition of Ω and suppose that the function* $\mathrm{p} : E_k \in \mathcal{E} \to p_k \in \mathbb{R}$ *satisfies*

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

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, if the probability p_k of the events of a discrete partition is known, we are guaranteed that there is a unique probability extending them with consistency to the generated σ-algebra.

In this setting, some very operative results are also available: we can describe explicitly each event $E \in \sigma(\mathcal{E})$ and its probability $\mathbb{P}(E)$.

[1] That is, it is at most countable: $I \subset \mathbb{N}$

Lemma 0.1.3. *Let $\mathcal{E} = \{E_k\}_{k\in I}$ be a discrete partition of Ω. Then,*

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

Lemma 0.1.4. *Let $\mathrm{p} : E_k \in \mathcal{E} \to p_k \in \mathbb{R}$ and $\mathbb{P} : \sigma(\mathcal{E}) \to \mathbb{R}$ be as is Theorem 0.1.2. Then,*

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

Remark 0.1.5. A special case worth mentioning is that of the atomic partition on a discrete Ω. Here, the σ-algebra generated 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 probability $\mathbb{P}$ consistent over $\mathcal{E}$ such that

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

Now, we want to show by examples that $\mathbb{P}$ need not be given over just a single partition. Actually, since a probability is a very special kind of measure, it is not even necessary that the information is about its values! We will explain the matter thorugh examples, but first we state an handy result.

Corollary 0.1.6. *Let $\mathcal{E} = \{E_k\}_{k\in I}$ be a partition of Ω and suppose that the function $\mathrm{p} : E_k \in \mathcal{E} \to p_k \in \mathbb{R}$ satisfies $\sum_{k\in I} p_k = 1$ and $p_k \geq 0$ for all $k \in I$. Moreover, let $\mathcal{F} = \{F_h\}_{h\in J}$ be another partition and suppose for all k such that $p_k > 0$ the function $q^{(k)} : F_h \in \mathcal{F} \to q_{h\,|k} \in \mathbb{R}$ satisfies $\sum_{h\in J} q_{h\,|k} = 1$ and $q_{h\,|k} \geq 0$ for all $h \in J$.*

Then, there exist a unique probability $\mathbb{P}$ on $\sigma(\mathcal{E} \cup \mathcal{F})$ such that $\mathbb{P}$ and p agree on $\mathcal{E}$ and

$$\mathbb{P}(F_h|E_k) = q_{h|k} \quad \textit{for all } k \in I \textit{ and } h \in J.$$

The result benefits some explaining.

Remark 0.1.7. The difference with Theorem 0.1.2 is that the σ-algebra is *quite bigger*. To understands *how much*, it is sufficient to observe that it can also be charachterized as the σ-algebra generated by the partition $\mathcal{Q} = \{E_k \cap F_h\}_{k\in I, h\in J}$ of Ω.

Remark 0.1.8. Recall Remark 0.1.1. In proability, it is quite common to study the relationships between different states and this result guarantees that if we were given not only the probability of a state, but also the probability of another state *in relation with* the first, then we can furnish consistently and uniquely the probability over combinations of all sorts of these states.

A useful representation is that of a tree of events. Let us present it with an example.

Example 1. (??)(??)

Remark 0.1.9. By the precedent Remark, you understand that a special relationship correlating states is that of independence. In that case, we can do one of two things depending on the request of the problem. First, we can use the definition: that is we put $\mathbb{P}(F_h \cap E_k) = \mathbb{P}(F_h)\mathbb{P}(E_k)$, find the probabilities on the partition in Remark 0.1.7 and then, apply Theorem 0.1.2 and Lemmas 0.1.3,0.1.4 to find $\mathbb{P}$. Second, we make use of the fact that for $\mathbb{P}(E_k) > 0$ independence is equivalent to $\mathbb{P}(F_h|E_k) = \mathbb{P}(F_h)$. We establish all the conditional probabilities, and then apply Theorem 0.1.6 to show the existence of $\mathbb{P}$. Finding the values of $\mathbb{P}$ is then a matter of manipulating conditional probabilities.

Example 2. (??) (??)

Moreover, this theorem allows us to easy the question of probability modelisation for repeated experiments. (??)

And now a final Remark on *why* the discrete partition setting is solvable.

Remark 0.1.10. For the "partition" part, it all boils down to the fact that we only consider unions: the complementation is just the union of the other elements, and the intersection is the union of partition elements that are in all the sets. If now you add the "discrete" part, we not only describe everything as unions, but as at most countable unions: that is, complementation will be always well defined. Technically, if we considered an uncountable partition, even if S was a countable union of some of its elements, then the complementary S^{c} would be an uncountable union. That is because an uncountable collection without countably many elements is still uncountable. That would make S^{c} untractable under 0.1.4!

0.1.2 Carathéodory Theorem

For more general settings the short message is that Theorem 0.1.2 holds, provided some conditions, while no explicit characterization like that of Lemmas 0.1.3,0.1.4 is possible.

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

Definition 0.1.11. Let $\mathcal{A}$ be an algebra defined on Ω. Then, $\tilde{\mathbb{P}} : \mathcal{A} \to \mathbb{R}$ is a *pre-measure* 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 0.1.12. The latter is the condition that ensure coherence with respect to the probability, and can be read as a *need-based σ-additivity*.

Theorem 0.1.13 (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-measure. 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 0.1.14. 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 0.1.11. 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[2]. Better: we aim at simplifing the checks on $\tilde{\mathbb{P}}$ rather than changing the statement of 0.1.13 and we do so by furnishing an equivalent set of conditions.

Lemma 0.1.15 (Pre-measure, Continuity characterization)**.** *Let $\mathcal{A}$ be an algebra defined on Ω. Then $\tilde{\mathbb{P}} : \mathcal{A} \to \mathbb{R}$ is a pre-measure if and only if*

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 $A_n \in \mathcal{A}$.

Remark 0.1.16. 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.

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

0.1.3 Practical construction, set structure

The following two resources cover the topic well (and probably better than I can).

- π-λ Theorem and Monotone Class Theorem: link
- Practical construction: Read the "extension.pdf" file in "assets" directory. In particular, recall continuity charchterization.

I will leave some operative remarks that integrate the resources, in the rest of the chapter.

Definition 0.1.17 (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 $A_1, \dots, A_n \in \mathcal{C}$ implies that $\bigcap_{i=1}^{n} A_i \in \mathcal{C}$.

Remark 0.1.18. 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.

Definition 0.1.19 (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 complementation. 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 .

Remark 0.1.20. Some textbooks require proper difference stability, that is $A, B \in \mathcal{C}$ implies $A \setminus B \in \mathcal{C}$, instead of complementation.

We show that they are equivalent. The question boils down to the identity

$$A \setminus B = A \cap B^{\mathsf{C}}, \tag{3}$$

where complementation is taken with respect to a space containing both A and B.

Since $B \subset A$ the above equation yields that $B \setminus A$ is the complementation of B with respect to A, then $\mathcal{C}$ being closed under proper difference is the same as $\mathcal{G}_A = \{B \in \mathcal{C} : B \subset A\}$, the collection of subsets of A in $\mathcal{C}$, being closed under complementation for all $A \in \mathcal{C}$. Importantly, this implies complementation stability with respect to Ω.

The converse, that complementation implies proper difference, holds as well. Suppose $B \subset A$. Then, $A \setminus B = A \cap B^{\mathsf{C}} = (A^{\mathsf{C}} \cup B)^{\mathsf{C}} \in \mathcal{C}$ by (3) and countable disjoint union.

Remark 0.1.21. The Monotone Class Theorem furnishes a tool to show that a certain property is satisfied by all sets in a σ-algebra. It is sufficient to show that the π-system generating the σ-algebra satisfies it, and that the set satysfing it is a λ-system. We can show the same if the property is satisfied over an algebra and the class satisfying the property constitues a monotone class. This patter of reasoning is termed a *monotone class argument*.

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. *Let $\Omega = \mathbb{N}$. Show that $\mathrm{p}(\{n\}) = \theta^n(1-\theta)$ for all $n \in \mathbb{N}$ is a discrete probability density. That is, show that it is coherent enough for a probability extending it to $\mathcal{P}(\mathbb{N})$ to exist.*

Notes. (??)

Exercise 6. *Let E_1,E_2 be independent events on Ω such that $\mathrm{p}(E_1) = \mathrm{p}(E_2) = \frac{1}{2}$. Determine the sigma-algebra, and find the probability $\mathbb{P}$ on this space consistent with the two values of* p.

Exercise 7.

Remark 0.1.22. These are some examples of probability modelization Note that Ω is basically irrelevant. Here independence, a property *of the probability*[4], furnishes the necessary information for $\mathbb{P}$ to be defined uniquely. It allows us to work in a context of minimal information, by relating different partitions (that is, states).

Exercise 8 (Jacod Protter, 7.1)**.**

Remark 0.1.23. The idea is that only finitely many disjoint events can have probability $\mathbb{P}(E) \leq \alpha$. That is all infinite sequences (convergent or divergent doesn't really matter, as we can restrict ourselves to the lim sup) need to tend to zero.

Exercise 9 (Jacod Protter, 7.2)**.**

Remark 0.1.24. Same idea as in 8, but the fact that here we also apply results about cardinality. That is we group events by having a probability larger than $\frac{1}{n}$ and then use 7.1 to show that their cardinality need be discrete, as the whole collection is countable union of finite collections.

[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.

[4]Independence and conditioal probabilities are the charachteristic that really distinguish a probability from a measure.

Exercise 10 (Jacod Protter, 7.10)**.**

Remark 0.1.25. This is an analytical result, but shows that the definition of discrete random variables is coherent with its charachterization in terms of cumulative density function.

To prove the result, one could create a bijection between $\mathbb{N}$ and the set of jump discontinuities $\mathbb{D}$, by using monotonicity and order on reals. A more instructive approach, though, is that of 0.1.24. We consider for each n the set

$$D_n = \left\{ x_0 \in [0,1] : \text{ in } x_0 \text{ is located a jump discontinuity larger than } \frac{1}{n} \right\}.$$

By boundedness of $[0,1]$, D_n need be finite. Then, $\bigcup_{n\in\mathbb{N}} D_n$ is discrete.

Analitically, you could also show that removable discontinuities need be discrete. See here for further considerations. This does not have direct applications in probability.

Exercise 11. *Let*

$$F(x) = \begin{cases} 0 & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0. \end{cases}$$

Show that this is a CDF. Does there exist, and if so, is it unique, the distribution P that generates F? If it exists and is unique, find it.

This distribution is denoted by $\mathcal{E}(\lambda)$ and is called the negative exponential distribution.

Remark 0.1.26. For $0 < a < b < +\infty$,

$$P((a,b]) = e^{-a} - e^{-b}.$$

Exercise 12. *Find an example of X and Y random variables from a measurable space such that $P_X = P_Y$ but $P(X = Y) \neq 1$.*

Remark 0.1.27. Start by setting $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1) = (\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. Then, observe that

$$\mathbb{P}\{X = Y\} = \mathbb{P}\{\omega : X(\omega) = Y(\omega)\} = \mathbb{P}\{\omega : X(\omega) = t, Y(\omega) = t, \text{ for } t \in \mathbb{R}\}$$

and comparatively,

$$P_X = P_Y \implies \mathbb{P}\{X \in A\} = \mathbb{P}\{Y \in A\},$$

for every $A \in \mathcal{B}$.

In other words, the first condition requires that the random variables have indistinguishable (probabilistically) preimages, while the second condition only requires that the probability of preimage sets be equal: the identity in the first case is element-wise, in the second, it's about the law.

Idea: The symmetry of the law of X and Y produces structures indistinguishable by the law but with different relations.

Solution: For $\Omega = \{a, b, c\}$ with $\mathbb{P}(\{a\}) = \mathbb{P}(\{c\}) = \frac{1}{4}$ and $\mathbb{P}(\{b\}) = \frac{1}{2}$, and $X : (a, b, c) \to (1, 0, -1)$, $Y : (a, b, c) \to (-1, 0, 1)$, the two random variables will be symmetric, fulfilling the required condition. Moreover, if Ω contains only 2 elements, then $P\{X = Y\} = 0$.

Exercise 13. *Let $U(0,1) : (\Omega, \mathcal{F}, \mathbb{P}) \to (\mathbb{R}, \mathcal{B})$ be the random variable with a uniform distribution between 0 and 1. Explicitly,*

$$F_U(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases}$$

Find the distribution of

$$X : \omega \in (\Omega, \mathcal{F}, \mathbb{P}) \to -\log(U(\omega))\mathbb{1}\{U(\omega) > 0\} \in (\mathbb{R}, \mathcal{B}).$$

Remark 0.1.28. The problem of finding the distribution of a random variable Y can be approached in two ways:

1. If Y is discrete, it suffices to find the the PMF, i.e. determine the support S first and then the value $\mathbb{P}(X = s)$ for each $s \in S$. P_X is the image law.

2. If Y is not discrete, it is not possible to describe P_X by finding its value on atoms. It is necessary to find its value on a π-system. Often, this means finding the CDF F_X.

In this case, we obviously adopt the second approach.

Idea: See the function as the transformation of U through a function t.

Idea: Work with the monotonically increasing and invertible composite function to determine the interval for which you want to find the preimage with respect to U, and then solve using F_U.

Exercise 14 (JP, 9.5)**.**

Remark 0.1.29. This shows that the role of e.v. on a class of functions is similar to that of the probability on a σ-algebra, a case which can be found for $X = 1$.

Exercise 15. *Consider the function $Q : \mathcal{F} \to \mathbb{R}$ such that*

$$Q(A) = \int_A f \, \mathrm{d}m,$$

where f is a PMF (that is, f is measurable, $f \leq 0$, $\int_{\mathbb{R}} f \, \mathrm{d}m$). Show that it is a probability.

Remark 0.1.30. The result is a particular case of the precedent exercise. Since we consider a PMF f, it is equivalent to restrict the X of exercise [JP, 9.5] to absolutely continuous r.v.'s.

Exercise 16. *Let $(\Omega, \mathcal{F}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}, \mathcal{E}(\lambda))$ for $\lambda > 0$, where we recall that $f_{\mathcal{E}}(x) = \lambda e^{-\lambda x}$.*

1. *Consider $X : \omega \in \Omega \to \lfloor \omega \rfloor \in \mathbb{R}$. Show that $X \sim \mathcal{G}(1 - e^{-\lambda})$, that is that the image law is a geometric distribution.*

2. *Compute $\mathbb{E}[X]$ directly, that is without making use of the expectation rule.*

3. *Compute $\mathbb{E}[X]$ using the expectation rule, and compare the proceedings with the above.*

Remark 0.1.31. First, observe that since $X(\Omega) = \mathbb{Z}$, X is discrete since $P_X(\mathbb{Z}) = 1$. Moreover, the support is just $\mathbb{N}$, since the exponential distribution is null for negatives and so we can consider instead $\tilde{X}$, a.s. agreeing, that is null for $x < 0$. Then, it suffices to find the PMF:

$$p_x(n) = \mathbb{P}\{X = n\} = \mathbb{P}\{X = n\} = \mathbb{P}((n, n+1]) = \int_{(n,n+1]} f \, \mathrm{d}m = \int_n^{n+1} \lambda e^{-\lambda t} \, \mathrm{d}t = \left(1 - e^{-\lambda}\right) e^{-\lambda n}.$$

This is, in fact, a geometric distribution. For the second point, we first verify that the expected value exists: by the above discussion, since $\tilde{X} = X$ a.s. and since $\tilde{X} \leq 0$, it follows that X admitts e.v. We can find it with the following equialities:

$$\mathbb{E}[X] = \int_\Omega \lfloor \omega \rfloor \, \mathrm{d}\mathbb{P}(\omega) = \int_\Omega \sum^{\infty} k \mathbb{1}_{(n,n+1]} \, \mathrm{d}\mathbb{P}(\omega) = \sum^{\infty} k \int_\Omega \mathbb{1}_{(n,n+1]} \, \mathrm{d}\mathbb{P}(\omega) =$$

$$= \sum^{\infty} k \mathbb{P}((n, n+1]) = \frac{1}{1 - e^{-\lambda}}.$$

We have used the corollary for series to the MCT in order to commute integration and summation (note that the r.v. is positive and so this is allowed). The last passage is motivated by the known result

$$\sum^{\infty} k \, t^{k-1} = \frac{1}{(1-t)^2}.$$

With the expectation rule, this gets notably shortened.

$$\mathbb{E}[X] = \int_{\mathbb{R}} \mathrm{id} \, \mathrm{d}P_X = \int_{\mathbb{R}} \lfloor x \rfloor \, \mathrm{d}P_X(x) = \sum_{k \in \mathbb{N}} k p_X(k) = \frac{1}{1 - e^{-\lambda}}.$$

Thus, the expectation rule, known as e.r., can make computations much easier. Not only that, but if we were just given the image law of X and not the probabiltiy on $(\Omega, \mathcal{F})$, we would have still been able to compute the expected value.

Exercise 17. *Find the e.v. and var. of the discrete uniform distribution.*

Remark 0.1.32. Let $X \sim \mathrm{d}\mathcal{U}(\{1,\dots,n\})$. Then, the e.v. exists because the X is positive, and $\mathbb{E}[X] = \frac{n+1}{2}$. Moreover, $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \frac{n^2-1}{12}$. Where the computation of the second moment comes from the closed form for the sum

$$\sum_{k=0}^{n} k^2 = \frac{n(2n+1)(n+1)}{6}.$$

Exercise 18. *By computing the e.v. of the distribution $p(k) = \frac{6}{\pi^2}\frac{1}{k^2}$ of $\mathbb{N}$, show that the expected value need not be finite. Otherwise said, show that there are r.v. that admitt the e.v., but not variance.*

Remark 0.1.33. As the distribution is positive, one has that the e.v. exists. Moreover, X is discrete, and so

$$\mathbb{E}[X] = \sum^{\infty} kp(k) = \frac{6}{\pi^2}\sum^{\infty}\frac{1}{k} = \infty.$$

Obviusly, $X \notin L^1$ and so Var is not defined.

Chapter 1

Esperimenti congiunti

1.1 Prodotto di spazi di probabilità

Supponiamo di svolgere due esperimenti aleatori diversi, rappresentati dai due spazi di probabilità $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ e $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$. Per fare previsioni, è del tutto lecito cercare di modellizzarli complessivamente come un solo esperimento, detto *congiunto*. Per produrre uno spazio di probabilità, prima costruiremo uno spazio strutturato, e poi introdurremo una misura.

1.1.1 Prodotto di spazi misurabili

Partiamo dall'insieme degli esiti Ω: che nel primo esperimento si verifichi un certo esito e nel secondo un altro è del tutto equivalente al verificarsi di un esito "bidimensionale" che contenga come prima coordinata l'esito del primo esperimento e come seconda coordinata l'esito del secondo. In altri termini, Ω è in bigezione con $\Omega_1 \times \Omega_2$: quindi, scegliamo proprio il prodotto per rappresentare l'esperimento congiunto:

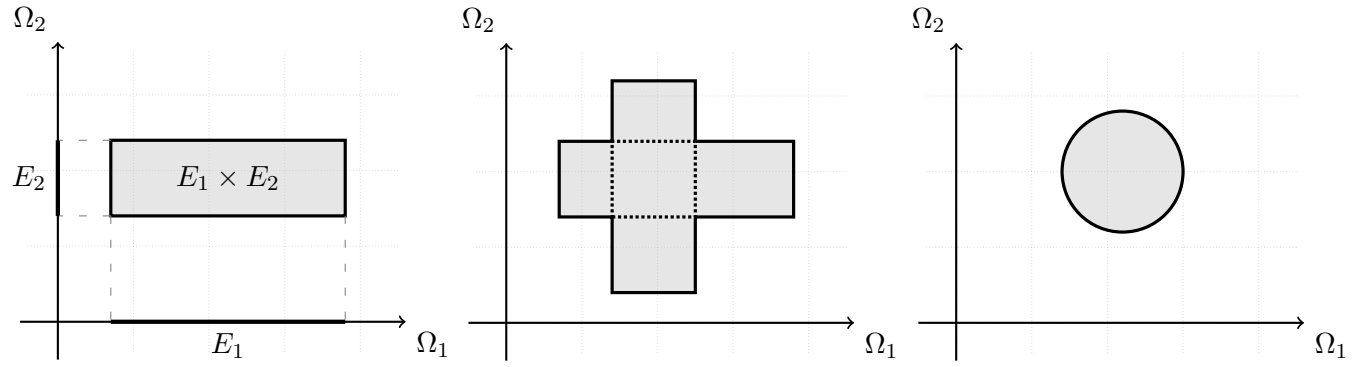
$$\Omega = \Omega_1 \times \Omega_2.$$

A questo punto, stabiliamo una struttura σ-additiva che dia concretezza ai predicati sugli esiti: imponiamo su $\Omega_1 \times \Omega_2$ una σ-algebra $\mathcal{A}$. Ovviamente, vogliamo poterci ridurre a studiare i due esperimenti singolarmente, e quindi è necessario che per $E \in \mathcal{F}_1$ si abbia che $E \times \Omega_2 \in \mathcal{A}$. Non solo: stiamo considerando l'esperimento congiunto proprio per studiare le relazioni tra i due esperimenti, quindi un altro requisito è che se $E_1 \in \mathcal{F}_1$ e $E_2 \in \mathcal{F}_2$ allora $E_1 \times E_2 \in \mathcal{A}$. Poichè la collezione $\{E_1 \times E_2 : E_i \in \mathcal{F}_i \text{ for } i = 1, 2\}$, detta "collezione dei rettangoli", non è una σ-algebra, considereremo come struttura dell'esperimento congiunto

$$\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma\left(\{E_1 \times E_2 : E_i \in \mathcal{F}_i \text{ for } i = 1, 2\}\right),$$

e cioè la più piccola[1] σ-algebra che contiene gli esiti di interesse. Diamo un'intuizione considerando il caso $(\Omega_i, \mathcal{F}_i) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ illustrato in Figura 1.1.

[1] Per quelli che non hanno ancora bevuto il caffè mattutino, questo è desiderabile perchè più spazzatura aggiungiamo, più si restringe la classe di probabilità che potremo definirci sopra. Un esempio: la probabilità uniforme può essere definita su $\mathcal{B}([0,1])$, ma non su $\mathcal{P}([0,1])$.



(a) Un insieme della collezione dei rettangoli.

(b) L'unione di due rettangoli che non sia un rettangolo.

(c) Un evento che non sia un rettangolo.

Figure 1.1: Esempi di eventi in $\mathcal{F}_1 \otimes \mathcal{F}_2$, per $\Omega_i = \mathbb{R}$ e $\mathcal{F}_i = \mathcal{B}(\mathbb{R})$.

Il generico insieme della collezione dei rettangoli $E_1 \times E_2$ può essere rappresentato come in Figura 1.1a, in accordo con l'idea di rettangolo[2]. Pertanto, i due insiemi in Figura 1.1b sono entrambi appartenenti alla collezione, ma la loro unione evidentemente no: questo esemplifica come la collezione dei rettangoli può fallire ad essere una σ-algebra. Ma $\mathcal{F}_1 \otimes \mathcal{F}_2$ contiene anche insiemi più complessi: la Figura 1.1c mostra un evento di questa σ-algebra.

Infine, alcune considerazioni. In primo luogo, quanto visto può essere facilmente esteso a più di due spazi procedendo iterativamente.

Inoltre, un risultato notevole è che definendo la σ-algebra di Borel di $\mathbb{R}^n$ come quella generata dalla topologia, e cioè $\mathcal{B}(\mathbb{R}^n) \stackrel{\text{def}}{=} \sigma(\mathcal{T}_{\mathbb{R}^n})$, si ottiene che $\otimes^n \mathcal{B}(\mathbb{R}) = \mathcal{B}(\mathbb{R}^n)$. Questo conferma la buona definizione dello spazio prodotto e, per le prossime sezioni, motiva il fatto che comunemente i vettori di variabili aleatorie siano definiti su $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

1.1.2 Prodotto di misure di probabilità

Finora, le scelte sono state fatte in maniera obbligata: se avessimo considerato oggetti diversi da $\times\Omega_i$ e $\otimes\mathcal{F}_i$ la descrizione dello spazio misurabile sarebbe stata logicamente equivalente[3], o avrebbe perso di flessibilità o di informazione.

Per la misura di probabilità $\mathbb{P}$ su questo spazio, non avremo lo stesso lusso: infatti, l'unico requisito che possiamo imporre è quello di consistenza con gli spazi di partenza. Formalmente, che per $E_1 \in \mathcal{F}_1$ si abbia

$$\mathbb{P}_1(E_1) = \mathbb{P}(E_1 \times \Omega_2),$$

e analogamente per $E_2 \in \mathcal{F}_2$. Si può dimostrare che esistono più misure di probabilità con questa caratteristica: possiamo garantire l'unicità solo con ulteriori vincoli modellistici e, in particolare, è sufficiente che gli esperimenti marginali siano indipendenti[4], e cioè valga la fattorizzazione

$$\mathbb{P}(E_1 \times \Omega_2 \cap \Omega_1 \times E_2) = \mathbb{P}(E_1 \times \Omega_2)\mathbb{P}(\Omega_1 \times E_2).$$

Poichè $E_1 \times \Omega_2 \cap \Omega_1 \times E_2 = E_1 \times E_2$, possiamo imporre il requisito come segue.

Theorem 1.1.1. *Siano $(\Omega_1, \mathcal{F}_1, \mathbb{P}_1)$ e $(\Omega_2, \mathcal{F}_2, \mathbb{P}_2)$ due spazi di probabilità. Allora esiste ed è unica la probabilità $\mathbb{P}$ su $(\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2)$ tale che*

$$\mathbb{P}(E_1 \times E_2) = \mathbb{P}_1(E_1)\mathbb{P}_2(E_2)$$

per $E_1 \in \mathcal{F}_1$ e $E_2 \in \mathcal{F}_2$. Chiamiamo $\mathbb{P}$ la "probabilità prodotto" e la indichiamo con $\mathbb{P}_1 \otimes \mathbb{P}_2$.

[2] Achtung: anche l'insieme $\mathbb{Q}^2$, per esempio, appartiene a questa collezione, ma la sua rappresentazione geometricamente non è un rettangolo.

[3] E cioè tutto sarebbe stato identico ai fini modellistici, "modulo" una bigezione.

[4] Ricordiamo che $A \perp B$ se e solo se $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Una proprietà desiderabile di $\mathbb{P}_1 \otimes \mathbb{P}_2$ è la computabilità. Sia $C \in \mathcal{F}_1 \otimes \mathcal{F}_2$, e siano le "sezioni" di C definite come segue:

$$C_1(\omega_2) = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in C\} \qquad C_2(\omega_1) = \{\omega_2 \in \Omega_2 : (\omega_1, \omega_2) \in C\}$$

Allora, le funzioni "probabilità della sezione"

$$\omega_1 \in \Omega_1 \longmapsto \mathbb{P}_2(C_2(\omega_1)) \in \mathbb{R}$$
$$\omega_2 \in \Omega_2 \longmapsto \mathbb{P}_1(C_1(\omega_2)) \in \mathbb{R}$$

sono misurabili e limitate, e quindi integrabili. Questo conferma la buona posizione della regola

$$(\mathbb{P}_1 \otimes \mathbb{P}_2)(C) = \int_{\Omega_1 \times \Omega_2} \mathbb{1}_C(\boldsymbol{\omega})\, \mathbb{P}_1 \otimes \mathbb{P}_2(\mathrm{d}\boldsymbol{\omega}) = \\ = \int_{\Omega_1} \mathbb{P}_2(C_2(\omega_1))\, \mathbb{P}_1(\mathrm{d}\omega_1).$$

Proof. Proveremo un solo caso. Sia $C = E_1 \times E_2$ con $E_i \in \mathcal{F}_i$. Si dimostra che $\pi_2 : (\omega_1, \omega_2) \in \Omega_1 \times \Omega_2 \to \omega_2 \in \Omega_2$ è misurabile ed integrabile. Allora,

$$\begin{aligned} \mathbb{P}_1 \otimes \mathbb{P}_2(E_1 \times E_2) &= \mathbb{P}_1(E_1)\mathbb{P}_2(E_2) = \\ &= \mathbb{P}_1(E_1)\mathbb{E}_{\mathbb{P}_2}[\mathbb{1}_{E_2}] = \mathbb{P}_1(E_1)\mathbb{E}_{\otimes\mathbb{P}}[\mathbb{1}_{E_2} \circ \pi_2] = \\ &= \int_{\Omega_2} \mathbb{P}_1(E_1)\mathbb{1}_{E_2}(\omega_2)\mathbb{P}_2(\mathrm{d}\omega_2). \end{aligned}$$

Poichè se $\omega_2 \in E_2$ allora $C_1(\omega_2) = \{\omega_1 \in \Omega_1 : (\omega_1, \omega_2) \in C = E_1 \times E_2\} = E_1$, si ha che $t_2(\omega_2) = \mathbb{P}(E_1)$. Similmente, se $\omega_2 \notin E_2$, allora $C_1(\omega_2) = \emptyset$ e $t_2(\omega_2) = 0$. Quindi, $t_2 = \mathbb{P}_1(E_1)\mathbb{1}_{E_2}$, da cui segue la tesi. Per linearità, segue il caso di $C = \bigcup^n E_{1,i} \times E_{2,i}$ con $E_{2,i}$ disgiunti (se non lo sono, si può recastare il set in una sommatoria dove lo sono). Infine, per C generico, è sufficiente trovare un'approssimante $C^{(n)} = \bigcup^n E_{1,i}^{(n)} \times E_{2,i}^{(n)}$ dal basso (?? esplicitarla ??), e osservare che per continuità monotona di $\otimes\mathbb{P}$ e MCT applicato alla sequenza $t_2^{(n)} \uparrow t_2$ segue la tesi in generale. □

Ovviamente, anche in questo caso tutto si può estendere a più di due spazi iterativamente.

1.2 Vettori aleatori e prodotto di spazi

Rimangono da introdurre le variabili aleatorie su o da spazi prodotto. Per farlo, in primo luogo stabiliamo dei criteri di misurabilità. Poi, discuteremo un criterio di indipendenza e infine troveremo un regola computazionale per il valore atteso.

1.2.1 Misurabilità

La situazione più semplice, quella in cui date delle variabili aleatorie definite sullo stesso spazio ci chiediamo come si comporta il loro vettore, è stata già incontrata nel caso di vettori aleatori reali: il vettore era misurabile se e solo lo erano le componenti. A conferma[5] della buona definizione dello spazio prodotto, questa proprietà continua a valere.

Lemma 1.2.1 (Misurabilità delle componenti)**.** *Per $X : \Omega \to E$, $Y : \Omega \to E$, si ha che $X : (\Omega, \mathcal{A}) \to (E, \mathcal{E})$ e $Y : (\Omega, \mathcal{A}) \to (E, \mathcal{F})$ sono misurabili se e solo se lo è $(X, Y) : (\Omega, \mathcal{A}) \to (E \times F, \mathcal{E} \otimes \mathcal{F})$.*

La seconda casistica, quella in cui il dominio è uno spazio prodotto, è più delicata e l'implicazione vale solo in un verso. Per fissare le idee, consideriamo solo un caso.

[5] Che il criterio valesse per i vettori reali lo rende *desiderabile* in generale

Lemma 1.2.2 (Misurabilità delle restrizioni)**.** *Sia* $X : (\Omega_1 \times \Omega_2, \mathcal{F}_1 \otimes \mathcal{F}_2) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *un vettore aleatorio. Allora, la restrizione*

$$X(\cdot, \omega_2) : \Omega_1 \to \mathbb{R}$$

è misurabile per ogni $\omega_2 \in \Omega_2$ *rispetto a* $\mathcal{F}_1$*.*

Questa è solo una condizione necessaria e non un criterio per la misurabilità: non vale l'inverso. Esplicitamente, la misurabilità di $X(\cdot, \omega_2)$ e $X(\omega_1, \cdot)$ rispetto ad ogni $\omega_1 \in \Omega_1$ e $\omega_2 \in \Omega_2$ non implica quella del vettore (come ci si potrebbe aspettare...).

Ciò detto il risultato caratterizza gli insiemi misurabili dello spazio prodotto: se $C \in \otimes\mathcal{F}_i$ allora $X = \mathbb{1}_C$ è misurabile, la restrizione $\mathbb{1}_C(\cdot, \omega_2)$ è misurabile, e la sezione $C_1(\omega_2)$ è un evento[6]. In altri termini, tutte le "sezioni" di un insieme misurabile nello spazio prodotto sono misurabili rispetto agli spazi marginali.

1.2.2 Criterio di Indipendenza

Supponiamo di considerare gli esperimenti $(E_i, \mathcal{E}_i, \mathbb{P}_i)$ congiuntamente, ed imponiamo che siano indipendenti: per quanto visto, consideriamo cioè la misura prodotto $\otimes\mathbb{P}_i$ su $(\times E_i, \otimes\mathcal{E}_i)$. Ora *accendiamo* le variabili aleatorie: se ognuno di questi spazi fosse lo spazio immagine di un solo esperimento $(\Omega, \mathcal{F}, \mathbb{P})$ attraverso $X_i : \Omega \to E_i$, allora avremmo che

$$X = (X_1, \dots, X_n) : \Omega \to E_1 \times \cdots \times E_n.$$

Facciamo un ragionamento modellistico: che gli spazi iniziali considerati congiuntamente fossero indipendenti significa che la conoscenza di un evento su uno non influenza le predizioni di un evento sull'altro. Introducendo le variabili aleatorie, lasciamo la struttura delle dipendenze inalterata, quindi sarebbe desiderabile che la conoscenza di un evento su uno spazio immagine non influenzasse le previsioni degli eventi sugli altri spazi immagine.

In altri termini, se le definizioni fossero ben poste dovremmo avere l'indipendenza delle variabili aleatorie, e cioè (nel caso $n = 2$)

$$P_{(X_1,X_2)}(E_1 \times E_2) = \mathbb{P}\{X_1 \in E_1, X_2 \in E_2\} = \mathbb{P}\{X_1 \in E_1\}\mathbb{P}\{X_2 \in E_2\} = P_{X_1}(E_1)P_{X_2}(E_2).$$

Questo è confermato dal seguente risultato, per cui vale anche l'inverso: se le variabili aleatorie sono indipendenti, allora il loro vettore è definito sullo spazio prodotto.

Lemma 1.2.3. *Le variabili aleatorie* $X : \Omega \to E$ *e* $Y : \Omega \to F$ *sono indipendenti se e solo se*

$$P_{(X,Y)} = P_X \otimes P_Y. \tag{1.1}$$

A livello operativo, questo risultato ha due conseguenze. In primo luogo, è un criterio per stabilire se due variabili aleatorie sono indipendenti: è necessario e sufficiente che la probabilità congiunta fattorizzi nelle probabilità marginali per ogni scelta di eventi negli spazi di arrivo.

La seconda è che permette di provare il seguente lemma, che a sua volta conferma che è sufficiente informare sulla legge di ogni variabile X_i e dire che sono indipendenti perchè esista un'unica variabile aleatoria nello spazio prodotto con queste caratteristiche.

Lemma 1.2.4. *Siano* $X_i : \Omega \to (E_i, \mathcal{E}_i, P_i)$ *variabili aleatorie. Allora, il vettore* $X = (X_1, \dots, X_n) : \Omega \to (\times E_i, \otimes\mathcal{E}_i, \otimes P_i)$ *è tale che*

- *La sua i-esima componente è distribuita come* X_i*: in simboli,* $(X)_i \sim X_i$*,*
- *Le sue componenti sono una famiglia di variabili aleatorie indipendenti.*

Lasciamo in appendice al capitolo la dimostrazione di 1.2.3.

?? da sistemare. Osserviamo che 1.1 vale se e solo se $P_{(X,Y)}(A \times B) = P_X(A)P_Y(B)$ per ogni $A \in \mathcal{E}$ e $B \in \mathcal{F}$. Consideriamo $P_{(X,Y)}(A \times B) = \mathbb{P}(\{X \in A\}, \{Y \in B\}) = \mathbb{P}\{X \in A\}\mathbb{P}\{Y \in B\}$. Se le r.v. sono indipendenti, percorrerla da sinistra dimostra la tesi; se vale 1.1 è sufficiente osservare che si ottine la def di indipendendenza stocastica. □

[6] Questo prova quanto visto nella regola per computare $\otimes\mathbb{P}$

1.2.3 Teorema di Fubini-Tonelli

Tornando al caso di variabili aleatorie definite sul prodotto di spazi, il calcolo del valore atteso diventa un'estensione dell'integrale multivariabile. Operativametne, potremmo ragionare in termini di integrale di Riemann multivariabile ma, come vedremo, ha più senso ricondursi all'integrale astratto in una sola variabile, e solo a quel punto impiegare l'integrale di Riemann.

In particolare, sotto positività o integrabilità, è possibile calcolare l'integrale sullo spazio prodotto come un integrale iterato su ognuno degli spazi di partenza (e senza considerare l'ordine di integrazione).

Theorem 1.2.5 (Fubini)**.** *Sia* $X : (\times\Omega_i, \otimes\mathcal{F}_i, \otimes\mathbb{P}_i) \to E$ *una variabile aleatoria. Allora se* $X \in L^1(otimes\mathbb{P}_i)$ *vale che ??*

In particolare, è sufficiente che valga la condizione ??

Theorem 1.2.6 (Tonelli)**.** *Sia* $X : (\times\Omega_i, \otimes\mathcal{F}_i, \otimes\mathbb{P}_i) \to E$ *una variabile aleatoria. Allora se* $X \leq 0$ *vale che ??*

Il teorema di Tonelli ha ipotesi più blande e dimostra Fubini scomponendo la funzione in parte positiva e negativa.

Chapter 2

Funzione Caratteristica

Un presupposto[1]: questa sezione non può essere capita a fondo finchè non si sarà fatta analisi complessa ad AM3. Quindi, la scelta del corso è quella di fornire solo gli strumenti operativi, e ignorare l'aspetto teoretico.

Operativamente, forniremo una caratterizzazione di P su $(\mathbb{R}^n, \mathcal{B}^n)$ che sia alternativa alla funzione di ripartizione F (CDF) e, dove definite, alle densità di probabilità f/p (PDF/PMF) che permette di agevolare certe operazioni.

2.1 Fundamentals

2.1.1 Prerequisiti

Dello spazio vettoriale $\mathbb{C}$ assumeremo come nota la bigezione con $\mathbb{R}^2$, e cioè dimestichezza con la loro forma algebrica.

Inoltre, daremo per nota - *come definizione* della funzione esponenziale complessa - l'espressione

$$e^{i\theta} = \cos(\theta) + i\sin(\theta).$$

Meno nota, ma comunque ben definita (in corsi futuri), è l'integrazione di una funzione a valori complessi. Senza entrare nei dettagli, si estende la teoria vista nel caso reale: sia $h : \mathbb{R}^n \to (\mathbb{C}, \mathcal{B}^2)$ misurabile, allora per garantire la linearità *definiremo* il suo integrale come

$$\int_{\mathbb{R}^n} h(x)\,dx = \int_{\mathbb{R}^n} \mathbb{Re}(h(x))\,dx + i\int_{\mathbb{R}^n} \mathbb{Im}(h(x))\,dx,$$

e parleremo di funzioni integrabili (ma non positive, visto che il concetto stesso è *ill-defined*). Vale anche una stima di *boundedness* (che può essere estesa dal caso reale):

$$\left|\int_{\mathbb{R}^n} h(x)\,dx\right| \leq \int_{\mathbb{R}^n} |h(x)|\,dx.$$

2.1.2 Definizione

Non ci perderemo nella comprensione dei concetti: ci importano i risultati e i limiti dello strumento. Partiamo da una definizione astratta.

Definition 2.1.1 (Funzione Caratteristica di P)**.** Sia P una probabilità su $\mathbb{R}^n$. Allora, la sua funzione caratteristica $\hat{P} : \mathbb{R}^n \to \mathbb{C}$ è data da

$$\hat{P}(\mathbf{u}) = \varphi(\mathbf{u}) = \int_{\mathbb{R}^n} e^{i\mathbf{u}\cdot\mathbf{s}}\,P(d\mathbf{s}).$$

[1] Più tipo una *supposta*, per chi deve studiare Probabilità sigh...

Dove osserviamo che aver utilizzato i numeri complessi garantisce la buona definizione: che l'intgranda sia limitata in quanto $\left|e^{i\mathbf{u}\cdot\mathbf{s}}\right| \leq 1$, implica che sia integrabile.

Diamo ora la definizione operativamente più utile.

Definition 2.1.2 (Funzione Caratteristica di X)**.** Sia $X : \Omega \to \mathbb{R}^n$ una variabile aleatoria. Allora, la sua funzione caratteristica $\hat{P}_X : \mathbb{R}^n \to \mathbb{C}$ è data da

$$\hat{P}_X(\mathbf{u}) = \varphi_X(\mathbf{u}) = \int_{\mathbb{R}^n} e^{i\mathbf{u}\cdot\mathbf{s}}\, P_X(d\mathbf{s}) \overset{\text{e.r.}}{=} \int_{\Omega} e^{i\mathbf{u}\cdot X(\omega)}\, \mathbb{P}(d\omega) = \mathbb{E}\left[e^{i\mathbf{u}\cdot X(\omega)}\right].$$

Quindi, la funzione caratteristica di una variabile aleatoria è la funzione caratteristica della legge immagine P_X.

2.1.3 Caratterizzazione di P

Ecco i risultati teoretici cruciali per usare questi risultati.

Theorem 2.1.3 (Di Caratterizzazione)**.** *Siano P e Q probabilità su $\mathbb{R}^n$. Allora,*

$$\hat{P} = \hat{Q} \quad \Longleftrightarrow \quad P = Q.$$

Corollary 2.1.4. *Siano X e Y vettori aleatori in $\mathbb{R}^n$. Allora,*

$$\hat{P}_X = \hat{P}_Y \quad \Longleftrightarrow \quad P_X = P_Y,$$

e cioè hanno la stessa distribuzione.

Operativamente, se per verificare che $P = Q$, con la caratterizzazione f tutti i controlli erano del tipo q.o., qua è sufficiente un controllo puntuale: se il criterio fallisce in un solo punto, allora sono probabilità diverse.

2.1.4 Computazione

Vale la pena notare che la definizione non è neanche lontanamente opeartiva. A meno che uno non voglia lavorare che integrali di funzioni miste trigonometriche e algebriche, bisogna computare integrali complessi, per i quali non abbiamo gli strumenti. Quindi, il nostro approccio sarà quello di determinare (*alla carlona*) gli integrali delle distribuzioni note e cercare di ricondurre i risultati a queste, o a loro combinazioni (attraverso certe regole di compatibilità).

Riportiamo qua solo una tabella sintetica. ??

Alla luce del teorema di Caratterizzazione sarebbe bello avere anche un risultato che permetta di *invertire* questo processo: partendo dalla funzione caratteristica, ricavare la probabilità. Questo è possibile (circa), ma visto il taglio *operativo* di questa sezione, non ci interessa.

2.1.5 Flop: Caratterizzazione analitica

Avere un oggetto che caratterizzi le distribuzioni è comodo, ma fornisce solo un criterio di unicità. Negli altri casi - CDF, PDF e PMF - è sempre stato possibile trovare delle *proprietà analitiche* che caratterizzavano tutte (e solo) le funzioni di quel tipo. Questo, per le funzioni caratteristiche, non sarà possibile. Quindi, per utilizzarle si deve avere a disposizione un ricco database di risultati, che ne permettano l'invertibilità.

Al di là dei risultati noti abbiamo delle condizioni necessarie, per curare parzialmente il problema.

Lemma 2.1.5. *Sia $\varphi : \mathbb{R}^n \to \mathbb{C}$ una funzione caratteristica. Allora,*

- $\varphi(0) = 1$
- $|\varphi| \leq 1$
- *φ è (uniformemente) continua.*

Operativamente, φ non è una funzione caratteristica se fallirà uno di questi test, mentre lo sarà se e solo se troveremo una P associata.

2.1.6 Top: Momenti

Con le funzioni caratteristiche diventa immediato calcolare i momenti: riduciamo l'operazione da un'integrazione a una derivazione.

Theorem 2.1.6 (Funzione Caratteristica e Momenti)**.** *Siano $X_k : \Omega \to \mathbb{R}$ per $k = 1, \dots, n$ variabili aleatorie tali che $X_k \in L^m(\mathbb{P})$ per ogni $k \leq n$. Allora, per $X = (X_k)$,*

- $\sigma_X \in \mathcal{C}^m$
- *Vale la formula*
$$\frac{\partial^m}{\partial u_{k_1} \dots \partial u_{k_m}} \varphi(\mathbf{u}) = i^m \mathbb{E}\left[X_{k_1} \dots X_{k_m} e^{i\mathbf{u}\cdot X}\right]$$

Concettualmente, tanto è più integrabile X, tanto è più derivabile φ_X. Inoltre, vale la commutatività di integrale e derivata complessi:

$$\frac{\partial^m}{\partial u_{k_1} \dots \partial u_{k_m}} \mathbb{E}\left[e^{i\mathbf{u}\cdot X}\right] = i^m \mathbb{E}\left[\frac{\partial^m}{\partial u_{k_1} \dots \partial u_{k_m}} e^{i\mathbf{u}\cdot X}\right].$$

La conseguenza operativa più importante, comunque, è che ora trovare i momenti, misti o non, è facilissimo.

Corollary 2.1.7. *Se $X_1, \dots, X_m$ sono variabili in $L^m(\mathbb{P})$, allora vale*

$$\frac{\partial^m}{\partial u_{k_1} \dots \partial u_{k_m}} \varphi(\mathbf{0}) = i^m \mathbb{E}\left[X_{k_1} \dots X_{k_m}\right],$$

e in particolare si hanno queste regole

- $\mathbb{E}[X_k] = \frac{1}{i} \frac{\partial}{\partial u_k} \varphi(0)$
- $\mathbb{E}[X_k^2] = -\frac{\partial^2}{\partial {u_k}^2} \varphi(0)$
- $\mathbb{E}[X_k X_j] = -\frac{\partial^2}{\partial u_k \partial u_j} \varphi(0)$

Da questo è facile trovare anche al varianza e la covarianza.

Una nota importante è che l'integrabilità è richiesta in ipotesi: non è sufficiente che φ sia $\mathcal{C}^k$ per applicarlo, ma si deve verificare che le componenti siano L_k. Operativamente, questo può essere fatto per esempio con quale (brutale) approssimazione, e applicando il DCT.

2.1.7 Top: Trasformazioni Affini

Il seguente risultato ci permette di trovare molto facilmente la funzione caratteristica della trasformazione affine di una variabile aleatoria nota.

Lemma 2.1.8. *Sia X un vettore aleatorio e $Y = AX + b$. Allora,*

$$\varphi_Y(\mathbf{u}) = e^{i\mathbf{u}\cdot b} \varphi_X(A^T \mathbf{u}).$$

In particolare, questo fornisce un metodo efficace per gestire le proiezioni e le somme. Infatti, attraverso banali trasformazioni lineari di tipo riga, ricaviamo

- $\varphi_{X_k}(u) = \varphi_X(0, \dots, 0, u, 0, \dots, 0)$
- $\varphi_{\sum X_k} = \varphi_X(u, \dots, u)$

Un esempio pratico di applicazione è determinare la funzione caratteristica di una normale generica. (??)

2.1.8 Top: Indipendenza (e somme)

Chapter 3

Convergenza Debole

Le convergenze viste finora lavorano sulla distanza puntuale $|X_n(\omega) - X(\omega)|$, considerata in diverse maniere (su un evento quasi certo, come distanza media, come distanza media a code appiattite...). Quindi, è sempre necessario rifarsi a un dominio comune e valutare questo oggetto per ogni ω, anche nel caso in cui si applichi la expectation rule, visto che comunque $X_n - X$ deve essere **una** variabile aleatoria, e cioè avere un solo dominio. Ci proponiamo di definire un tipo di convergenza che lavori esclusivamente sul codominio, e cioè sugli spazi di probabilità immagine (o meglio, le leggi immagine, visto che $\mathbb{R}$ e $\mathcal{B}$ saranno un codominio fisso).

3.0.1 Convergenza sul codominio

Operativamente, il nostro piano per costruire una convergenza è partire imponendo che le successioni costanti convergano, e poi richiedendo che questa definizione sia consistente con quelle più forti.

Remark 3.0.1 (Caratterizzazzione dell'identità di probabilità). L'unico risultato noto a riguardo è il Criterio di Carathèordory: siano P e Q probabilità su $(\mathbb{R}, \mathcal{B})$, e sia $\mathcal{A} \subset \mathcal{B}$ una π-classe. Se $P|_{\mathcal{A}} = Q|_{\mathcal{A}}$ allora $P = Q$. In particolare questo vale se $\mathcal{A} = \tau$, la topologia di $\mathbb{R}$. Quindi, sarà sufficiente richiedere che $P(E) = Q(E)$ per ogni $E \in \tau$.

Questo è già più maneggevole, ma possiamo migliorarlo: raffiniamo questo risultato spostando il problema da un setting insiemistico a uno funzionale. Da $\mathbb{E}_P(\mathbb{1}_{\mathbb{E}}) = P(E)$, ricaviamo affinchè $P = Q$ è sufficiente *testare* P e Q sull'insieme delle variabili aleatorie indicatrici, e cioè verificare che $\mathbb{E}_P(\mathbb{1}_{\mathbb{E}}) = \mathbb{E}_Q(\mathbb{1}_{\mathbb{E}})$ per ogni $E \in \mathcal{B}$.

Sfruttando la definizione topologica di continuità[1] possiamo recastare ulteriormente questo risultato: è sufficiente testare su $\mathcal{C}_b^0$[2].

Lemma 3.0.2 (Identità su $\mathcal{C}_b^0$)**.** *Siano P e Q probabilità su $(\mathbb{R}, \mathcal{B})$. Allora, $P = Q$ se e solo se*

$$\int_R h(x)P(\mathrm{d}\,x) = \int_R h(x)Q(\mathrm{d}\,x)$$

per ogni $h \in \mathcal{C}_b^0$.

Questa è una caratterizzazione dell'identità ragionevolmente maneggevole e quindi possiamo estenderla a convergenza. Cioè, consideriamo due misure di probabilità *vicine* se la distanza euclidea tra gli integrali è piccola.

Definition 3.0.3 (Convergenza debole)**.** $\mathbb{P}_n \implies \mathbb{P}$ se e solo se per ogni $h \in \mathcal{C}_b^0$ si ha

$$\int h(s)\mathbb{P}_n(\mathrm{d}s) \to \int h(s)\mathbb{P}(\mathrm{d}s)$$

E ovviamente per le variabili aleatorie abbiamo la definizione associata.

[1] ??

[2] $\mathcal{C}_b^0$ è l'insieme delle funzioni continue e limitate su $\mathbb{R}$

Definition 3.0.4 (Convergenza in legge)**.** Sia (X_n) una successione di variabili aleatorie e siano $X_n : \Omega_n \to \mathbb{R}$,$X : \Omega \to \mathbb{R}$. Allora,$X_n \xrightarrow{d} X$ se e solo se $P_{X_n} \implies P_X$.

Lemma 3.0.5. *$X_n \xrightarrow{d} X$ se e solo se per ogni $h \in \mathcal{C}_b^0$ si ha $\mathbb{E}[h(X_n)] = \mathbb{E}[h(X)]$.*

Remark 3.0.6 (La definizione è ben posta)*.* Assumendo la misurabilità, affinchè h sia integrabile rispetto a ogni probabilità è, fondamentalmente, necessario che sia limitata: se non lo fosse, sarebbe sufficiente concentrare in quel punto la massa. Ci si potrebbe chiedere perchè non usare solo la misurabilità di h: il fatto è che anche se questo non dà problemi nell'identità 3.0.2, imponendo di testare la convergenza su una classe più ampia di funzioni si restringe la classe di convergenze che valgono. In questo caso, perdiamo la classe delle convergenze numeriche[3]. Per esempio, per $X_n = \frac{1}{n}$ q.c. e $X = 0$ q.c., testando rispetto alla funzione $h = \delta_0$ (misurabile, non continua), otteniamo $\mathbb{E}[h(X_n)] = h\left(\frac{1}{n}\right) = 0 \nrightarrow 1 = h(0) = \mathbb{E}[h(X)]$. Per $h \in C_b^0$, non abbiamo questo problema, visto che le funzioni continue commutano con il limite.

Lavorare sul codominio vuole dire che gli spazi su cui ambientiamo X_n e X possono essere diversi. Consideriamo per esempio la dimostrazione del Lemma 3.0.5.

Proof. La dimostrazione si riduce a

$$\mathbb{E}[h(X_n)] = \int_{\Omega_n} h(X_n(\omega))\mathbb{P}_n(\mathrm{d}\omega) \overset{e.r.}{=} \int_{\mathbb{R}} h(s)P_{X_n}(\mathrm{d}s) \xrightarrow{Hp}$$

$$\xrightarrow{Hp} \int_{\mathbb{R}} h(s)P_X(\mathrm{d}s) \overset{e.r.}{=} \int_{\Omega} h(X(\omega))\mathbb{P}(\mathrm{d}\omega) = \mathbb{E}[h(X)]. \quad (3.1)$$

□

Crucialmente, è necessario computare un limite rispetto all'intero spazio. La convergenza debole permette di ridurre questo problema (attraverso l'expectation rule) ad un limite sulla misura nello stesso spazio, rendendolo computabile.

Remark 3.0.7 (Considerazioni Modellistiche)*.* Modellisticamente, lo sperimentatore che osserva le distribuzioni di X_n, individua la legge del limite, di X, senza dedurla dal confronto con le altre che ha davanti, ma semplicemente accorgendosi che assomiglia a una certa distribuzione (eventualmente definita su un altro spazio di probabilità). Per concretizzare questo, pensiamo al TCL: siano le $X_n \sim \mathcal{B}\rceil\nabla(p)$, $\Omega_n = \{0, 1\}$, allora $\bar{X}_n \xrightarrow{d} N$. Cioè lo sperimentatore sta dicendo che anche se la normale non appare da nessuna parte (cioè non c'è un *confronto*), tutto è discreto e addirittura lo spazio di probabilità è binario, per come distribuisce la media campionaria è indistinguibile da una normale.

Remark 3.0.8 (Considerazioni Teoretiche)*.* Si fa presente che spesso si riesce ad ambientare tutte le variabili aleatorie su uno stesso spazio di probabilità. Anche se questo sembrerebbe rendere inutile il discorso fatto finora, in realtà stiamo solo nascondendo la polvere sotto al tappeto: non possiamo imporre infatti l'indipendenza. In altri termini, sopravvive il fatto che mentre per computare tutte le precedenti convergenze serviva conoscere la legge congiunta di (X_n, X) per ogni $n \in \mathbb{N}$, qua è sufficiente conoscere quella marginale di X_n. Questo diventerà evidente alla luce del Teorema di Levy.

Remark 3.0.9 (Unicità del limite)*.* Il Tradeoff è che le precedenti convegenze erano drasticamente più forti, perchè erano tra variabili aleatorie mentre questa è tra leggi. Basti pensare che il limite non è unico: è unica la sua legge. Formalmente, vale che se $X_n \xrightarrow{d} X$ e $X_n \xrightarrow{d} \tilde{X}$, allora $P_X = P_{\tilde{X}}$.

[3]Che sono ovviamente desiderabili: infatti, quando si dà una definizione più ampia della precednete, è buona norma che quella precedente ne sia un sottocaso, e cioè continui a valere. Poichè siamo partiti della convergenza certa, che a sua volta estende la convergenza puntuale, che estende quella numerica, è essenziale che la convergenza numerica valga.

3.0.2 Criteri di convergenza specializzati

Partiamo da alcuni criteri specializzati, ma solo sufficienti.

Theorem 3.0.10 (PMF Criterion)**.** *Suppose X_n and X are discrete, and $p_{X_n}(k) \to p_X(k)$ for all $k \in S = S_X \cup \bigcup S_{X_n}$. Then, $X_n \xrightarrow{d} X$.*

Crucialmente, questo criterio può fallire se S ha punti di accumulazione. Basti considerare il caso di $\delta_{\frac{1}{n}} \xrightarrow{d} \delta_0$, per cui però $p_{X_n}(0) = 0 \nrightarrow 1 = p_X(0)$.

Theorem 3.0.11 (PDF Criterion)**.** *Suppose X_n and X are discrete, and $p_{X_n}(k) \to p_X(k)$ for all $k \in S = S_X \cup \bigcup S_{X_n}$. Then, $X_n \xrightarrow{d} X$.*

Alcune applicazioni notevoli di questi criteri sono le seguenti.
??
??

3.0.3 Criteri di convergenza generici

Il principale motivo per cui servono criteri generali è che non solo ci sono distribuzioni che non sono nè discrete, nè continue, ma in maniera fondamentalmente più contorta il limite di discrete può essere continuo e viceversa.

Ricordiamo preliminarmente che abbiamo 2 caratterizzazioni di tipo generale a cui possiamo appoggiarci: la CDF e la CF. Per entrambe esiste un criterio (necessario e sufficiente) che le relaziona con la convergenza in legge.

Cominciamo dalla funzione di ripartizione.

Theorem 3.0.12 (CDF Criterion)**.** *$X_n \xrightarrow{d} X$ se e solo se $F_{X_n}(x) \to F_X(x)$ per ogni $x \in \mathbb{R}$ punto di continuità di F_X.*

Si osservi che per ricavare il valore di F_X nei punti di discontinuità è sufficiente applicare la continutià da destra (una proprietà delle funzioni di ripartizione). La classica situazione di fallimento è quella di $\delta_{\frac{1}{n}}$: il limite di $F_{X_n}(0)$ è 0, ma per continuità da destra $F_X(0) = 1$. Operativamente, se il limite puntuale delle F_{X_n} è una CDF ed è continuo, allora è verificata la convegenza.

Passiamo al caso della funzione caratteristica.

Theorem 3.0.13 (Levy)**.** *Per X_n,X variabili aleatorie reali,*

1. *Se $X_n \xrightarrow{d} X$ allora $\varphi_n(u) \to \varphi(u)$ per ogni $u \in \mathbb{R}$.*
2. *Se $\varphi_n(u) \to \psi(u)$ e ψ è continua in 0, allora esiste una variabile aleatoria X tale che $\varphi_X = \psi$ e $X_n \xrightarrow{d} X$.*

Da cui ricaviamo la caratterizzazione della convergenza in legge attraverso la funzione caratteristica.

Corollary 3.0.14 (CF Criterion)**.** *$X_n \xrightarrow{d} X$ se e solo se $\varphi_n(u) \to \varphi(u)$ per ogni $u \in \mathbb{R}$.*

Non Solo una caratterizzazione: Il Teorema di Levy dice di più. Tutti i criteri visti finora alla base avevano qualche forma di claim su come fosse fatto il limite[4]. Questo teorema ci dota invece di un criterio per stabilire addirittura **se** la successione converge[5], senza conoscere il limite (che, anzi, ci viene fornito come corollario).

Corollary 3.0.15 (Convergenza)**.** *Siano X_n variabili aleatorie. Allora, X_n converge in legge se e solo se $\varphi_n(u) \to \psi(u)$ e ψ è continua in 0.*

[4] Come dice Greg, capire see X_n converge ad X è un mestiere da boys.
[5] E, sempre secondo Greg, stabilire se X_n converge è una questione per men.

Perchè è devastante? La ragione per cui le altre convergenze non hanno questo genere di criterio è, di nuovo, che si tratta di una convergenza sul codominio. Infatti, per staibilire le altre convergenze è necessario conoscere la distribuzione congiunta di (X_n, X) per ogni $n \in \mathbb{N}$: per esempio,

$$\mathbb{E}\,|X_n - X| = \int_\Omega |x - y| P_{(X_n,X)}(\mathrm{d}\,x\,\mathrm{d}\,y).$$

Nel caso della convergenza in legge, non è necessario. Quindi, solo in questo caso abbiamo speranza di avere un criterio che funzioni a prescindere dalla legge congiunta, e quinda da X.

Questo conclude anche il discorso precedente: qua si vede che assumere lo stesso spazio, benchè teoricamente possibile, non elimina la flessibilità questa convergenza, ma semplicemnete la nasconde.

??

3.0.4 Proprietà

3.0.5 Forza ed inversione

Chapter 4

Martingales

Discrete-time deterministic processes are an idealization: in reality, they are affected by uncertainty and this might be important for our analysis. In particular, even if we could model the process, determining its limiting behaviour it's a whole different task altogether.

Usually we can determine (statistically or modellistically) the law of X_n, but cannot study its limit as that would require to evaluate infinitely many X_n's, enough times. Thus, we will assume that the law of the random variables in the process is known, and we will focus on determining the limiting law.

We will study a kind of very general kind of stochastic process: martingales. While general, this structure will allow some powerful results on convergence and large deviations. Here is a formalization of what we have just described.

Definition 4.0.1. A discrete time stochastic process is a sequence of random variables $\{X_n\}_{n\in\mathbb{N}}$ defined on $(\Omega, \mathcal{F}, P)$.

Example 3. A classic example of a discrete time stochastic process is a simple random walk, where:

$$X_n = \sum_{i=1}^{n} Y_i$$

and $\{Y_i\}$ are independent, identically distributed random variables taking values $+1$ or -1 with equal probability.

4.1 Well Posedness

4.1.1 Independence

We first recall the idea behind stochastic independence.

4.1.2 Borel-Cantelli Lemma

Let's first fix the ideas with a use case: The Devious Bet (blog post).

This result is pretty sleek and thus mathematicians like it; but to an application minded scientist, it doesn't really light on any synapsis. To me, its formulation is not intrepretable and I was clueless about why we had to introduce it.

Here is the actual Lemma.

Theorem 4.1.1 (First Borel-Cantelli Lemma)**.** *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of events in a probability space $(\Omega, \mathcal{F}, P)$. If*

$$\sum_{n=1}^{\infty} P(A_n) < \infty,$$

then

$$P(A_n,\ i.o.) = 0.$$

Theorem 4.1.2 (Second Borel-Cantelli Lemma)**.** *Let $\{A_n\}_{n=1}^{\infty}$ be a sequence of independent events in a probability space $(\Omega, \mathcal{F}, P)$. If*

$$\sum_{n=1}^{\infty} P(A_n) = \infty,$$

then

$$P(A_n,\ i.o.) = 1.$$

We are then interested furnishing an interpretation of lemma (what is it saying?) and understanding why it matters.

First Borel-Cantelli Lemma

Lemma 4.1.1 is actually a very unintelligible way to say something very stupid; namely that *if* $\mathbb{E}[X] < \infty$ *implies* $X < \infty$ *almost surely.*

Indeed, we can restate condition 4.1.1 as

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \mathbb{E}[\mathbb{1}_{A_n}] = \mathbb{E}\left[\sum_{n=1}^{\infty} \mathbb{1}_{A_n}\right] < \infty,$$

where the first equality comes from the expectation of indicators and the second from the Monotone Convergence Theorem. So, it is basically saying *if we expect X_n to be finite...*.

We can also restate the conclusione 4.1.1 by observing that

$$P(A_n,\ \text{i.o.}) = P(\sum_{n=1}^{\infty} \mathbb{1}_{A_n} < \infty),$$

where one simply checks that the two set share the same ω's. Then, the result is equivalent to

$$\sum_{n=1}^{\infty} \mathbb{1}_{A_n} < \infty \text{ a.s.}$$

Then, if $S = \sum_{n=1}^{\infty} \mathbb{1}_{A_n}$, the First Borel-Cantelli Lemma says just that $\mathbb{E}[S] < \infty$ implies $S < \infty$ almost surely.

So the interpretation is that when we collect the status of a light at each second n (has A_n happened or not?), if we expect finitely many lights to be on, that will happen almost surely. Physically, if we think of $\mathbb{E}$ as a centroid of the mass distribution, if the centroid is finite the sequence cannot have an unbalanceable mass at $+\infty$.

The above discussion furnishes trivially a proof. Another idea would be that of assuming that the probability of the i.o. set A is nonnull, and show that every n sets A is repeated, then lower bounding $\sum \mathbb{1}_{A_n}$ with an (divergent) series of costant terms. This won't work: we could spread A so that its first element was in each set A_k, the second in each set A_{2k} and so on. To save this approach, one could regroup only a fraction $0 < \alpha < 1$ of the mass, removing the spread, and then lower bounding.

Here is the standard proof. The tools are the tail necessary condition on series and Boole's inequality[1].

Proof. Since $\sum_{m \geq n} P(A_m)$ by the tail necessary condition on series we have $\sum_{m \geq n} P(A_m) \to 0$.

Then, by applying Boole's Inequality one finds

$$P(A_n,\ \text{i.o.}) = P(\bigcap_n \bigcup_{m \geq n} A_m) = \lim_n P(\bigcup_{m \geq n} A_m),$$

which proves the thesis by monotone continuity of probabilty measures. □

[1] Which is just disjointification and sigma additivity.

Second Borel-Cantelli Lemma

This one is interesting. Here the idea is that if the events in the process are independent, it doesn't really matter in which order I am taking them or if we consider them all: the asymptotic result shouldn't be affected by this. That is, (this is nontrivial, but intuitive) it should either have null or unit probability.

Interpretation-wise, it happens almost surely that either infinitely many lights come up every time or only finitely do: the independence of events makes it impossible to have chances.

This is very hand-wavy, but an intuition of the *why* is all we really need: the final form of this result will be the Zero-One Law, where we will get a better feel about the underlying asymptotic information structure (or tail sigma algebra).

Proof-wise, this is more demanding. Under independence, we have two equivalent formulations:

$$\sum_{n=1}^{\infty} P(A_n) = \infty \implies P(A_n, \text{ i.o.}) = 1, \tag{4.1}$$

$$P(A_n, \text{ i.o.}) = 0 \implies \sum_{n=1}^{\infty} P(A_n) < \infty. \tag{4.2}$$

For the proof, our choice will be 4.1, as to show 4.2 we would need to find a convergent series upper bounding ours: this cannot be done easily, as we could spread the i.o. set over the sequence.

Proof. Both Jacod Protter and Karr present this proof, basically leveraging the same tools:

1. That an infinite product is 1 iff each of its members is,
2. Converting unions into intersections formula,
3. Independent sets definition,
4. (Karr) Upper bounding $1 + x$,
5. (JP) Transforming product into sum (2ith log) and upper bounding $\log(1 + x)$.

□

Mathematically, this theorem furnishes equivalence conditions between the moment being finite and the series being finite almost surely.

Proof. □

4.1.3 0-1 Kolgomorov Law

Morally, we introduced Borel-Cantelli Lemma because a sequence of sets is, really, the sequence of indicator functions of these sets. That is, it is a naive type of stochastic process, which can be studied without delving into measure-theoretic issues. We now want to do exactly that, in order to extend the second Lemma to generic (independent) processes.

Tail Sigma Algebra