

Chapter 1

Basics

1.1 Probability spaces.

Definition 1.1. Let Ω be a set. We say that $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is a σ -algebra on Ω if:

1. $\emptyset \in \mathcal{A}$.
2. $A \in \mathcal{A} \Rightarrow \Omega \setminus A \in \mathcal{A}$.
3. \mathcal{A} is closed under countable union, i.e.

$$\forall n \in \mathbb{N} (A_n \in \mathcal{A}) \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}.$$

Definition 1.2. Let Ω be a set and \mathcal{A} a σ -algebra on Ω . We say that $\mathbb{P} : \mathcal{A} \rightarrow [0, 1]$ is a *probability measure* on \mathcal{A} if:

1. $\mathbb{P}(\emptyset) = 0$ and $\mathbb{P}(\Omega) = 1$.
2. \mathbb{P} is σ -additive, i.e.

$$\left[\forall n \in \mathbb{N} (A_n \in \mathcal{A}) \wedge \forall i \neq j (A_i \cap A_j = \emptyset) \right] \Rightarrow \mathbb{P} \left[\bigcup_{i=1}^{\infty} A_n \right] = \sum_{i=1}^{\infty} \mathbb{P}[A_i].$$

Definition 1.3. A *probability space* is a triple $(\Omega, \mathcal{A}, \mathbb{P})$, where Ω is a set (called also *sample space*), \mathcal{A} is a σ -algebra of subsets of Ω and \mathbb{P} is a probability measure on \mathcal{A} . The elements of \mathcal{A} are called *events*.

Remark. If Ω is at most countable, then we usually set $\mathcal{A} := \mathcal{P}(\Omega)$. If $\Omega = \mathbb{R}^d$ (or, more generally, Ω is a topological space), we often set $\mathcal{A} := \mathcal{B}$, where \mathcal{B} is the Borel σ -algebra (i.e. the smallest σ -algebra containing all open sets).

We now recall some properties of probability measures. The reader should know them already, therefore we omit a proof.

Proposition 1.4. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. Let $A, B, A_n \in \mathcal{A}$, for all $n \in \mathbb{N}$. The following holds:

- *Monotony:* $A, B \in \mathcal{A}, A \subseteq B \Rightarrow \mathbb{P}[A] \leq \mathbb{P}[B]$.
- $\mathbb{P}[A^c] = 1 - \mathbb{P}[A]$.
- $\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$.
- *Continuity:*
 - $\forall n [A_n \subseteq A_{n+1}] \Rightarrow \mathbb{P}(A_n) \nearrow \mathbb{P}[\bigcup_{n=1}^{\infty} A_n]$.
 - $\forall n [A_{n+1} \subseteq A_n] \Rightarrow \mathbb{P}(A_n) \searrow \mathbb{P}[\bigcap_{n=1}^{\infty} A_n]$.

- Let

$$\limsup A_n := \bigcap_{k=1}^{\infty} \bigcup_{n \geq k} A_n \quad \text{and} \quad \liminf A_n := \bigcup_{k=1}^{\infty} \bigcap_{n \geq k} A_n.$$

Then

INSERT CHAIN OF INEQUALITIES.

- $\mathbb{P} \left[\bigcup_{n=1}^{\infty} A_n \right] \leq \sum_{n=1}^{\infty} \mathbb{P}[A_n]$.
- **Borel-Cantelli lemma:** If $\sum_{n=1}^{\infty} \mathbb{P}[A_n] < \infty$, then $\mathbb{P}[\limsup A_n] = 0$. (NOTA: inserire versione più potente)

1.2 Measurable functions and random variables.

Definition 1.5. Let $(\Omega_1, \mathcal{A}_1)$ and $(\Omega_2, \mathcal{A}_2)$ be couples of sets and σ -algebras. We say that a function $f : \Omega_1 \rightarrow \Omega_2$ is *measurable* if

$$\forall A_2 \in \mathcal{A}_2 \left(f^{-1}[A_2] \in \mathcal{A}_1 \right).$$

We call *algebra induced by f* the set

$$f^{-1}(\mathcal{A}_2) := \{f^{-1}[A_2] \mid A_2 \in \mathcal{A}_2\}.$$

Observe that $f^{-1}(\mathcal{A}_2)$ is the smallest σ -algebra on Ω such that $f : (\Omega_1, \mathcal{A}_1) \rightarrow (\Omega_2, \mathcal{A}_2)$ is measurable. We denote $f^{-1}(\mathcal{A}_2)$ by $\sigma(f)$ and we call it *σ -algebra generated by f* .

If \mathbb{P}_1 is a probability measure on $(\Omega_1, \mathcal{A}_1)$, then $\mathbb{P}_2 : \mathcal{A}_2 \rightarrow [0, 1]$ given by $\mathbb{P}_2[A_2] := \mathbb{P}_1[f^{-1}[A_2]]$ is a probability measure on $(\Omega_2, \mathcal{A}_2)$, and it's called *probability measure induced by f* .

Notation. Pensare a un modo decente per spiegare la notazione $[\varphi(x)]$ e la nozione di “almost surely”.

Definition 1.6. Let Ω be a set and $\mathcal{F} \subseteq \mathcal{P}(\Omega)$. We define $\sigma(\mathcal{F})$ as the smallest σ -algebra on Ω which contains \mathcal{F} .¹

Definition 1.7. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A *random variable* is a measurable function $X : (\Omega, \mathcal{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and $\overline{\mathcal{B}} = \sigma(\{I \subseteq \overline{\mathbb{R}} \mid I \text{ is an interval}\})$.

Definition 1.8. Let X be a random variable on a probability space. Then $P^X : \overline{\mathcal{B}} \rightarrow [0, 1]$ given by $P^X[B] := \mathbb{P}[X \in B]$ is a probability measure on $(\overline{\mathbb{R}}, \overline{\mathcal{B}})$. P^X is called *distribution of X* .

Remark. We usually work with random variables X such that $\mathbb{P}[|X| = \infty] = 0$.

Examples. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. The following X 's are random variables:

1. *Discrete equidistribution on $\{x_1, \dots, x_N\}$.* We have that $\text{ran } X = \{x_1, \dots, x_N\} \subseteq \overline{\mathbb{R}}$ and $\mathbb{P}[X = x_k] = \frac{1}{N}$. So, by additivity, $\mathbb{P}[B] = \frac{|B \cap \{x_1, \dots, x_N\}|}{N}$.
2. *Bernoulli distribution.* $X = \mathbb{1}_A$, where $A \in \mathcal{A}$ and $\mathbb{1}_A$ is the indicator function of A , i.e. $\mathbb{1}_A(\omega) = 1$ if $\omega \in A$ and 0 elsewhere. Let $\theta := \mathbb{P}(A)$. We write $X \sim B(1, \theta)$, because $\mathbb{P}[X = 1] = \theta$ and $\mathbb{P}[X = 0] = 1 - \theta$. We obtain immediately that

$$P^X[B] = \begin{cases} 1, & \text{if } 0, 1 \in B \\ \theta, & \text{if } 1 \in B \text{ and } 0 \notin B \\ 1 - \theta, & \text{if } 1 \notin B \text{ and } 0 \in B \\ 1, & \text{if } 0, 1 \notin B. \end{cases}$$

3. *Binomial distribution.* We write $X \sim B(n, \theta)$ and we have that $\text{ran } X = \{0, \dots, n\}$. The distribution of X is given by

$$P^X(\{k\}) = \binom{n}{k} \theta^k (1 - \theta)^{n-k}.$$

¹Observe that $\sigma(\mathcal{F})$ is always well-defined, since $\mathcal{P}(\Omega)$ is a σ -algebra, and the intersection of σ -algebras is again a σ -algebra.

4. *Hypergeometric distribution*. NOTA: inserire.
5. *Poisson distribution*. NOTA: inserire.
6. *Continuous uniform distribution on $[a, b]$* . We have $\text{ran } X = [a, b]$ with $a, b \in \mathbb{R}$. The distribution of X is given by $P^X(B) = \frac{\lambda(B)}{b-a}$, where λ is the Lebesgue measure on \mathbb{R} .
7. *Exponential distribution*. NOTA: inserire.
8. *Normal distribution*. NOTA: inserire.

Remark. Let X, X' be random variables on the same probability space. Write $X \sim X'$ iff $\mathbb{P}[X = X'] = 1$. This is trivially an equivalence relation. From now on, we look at random variables which lay in the same equivalence class as they are the same random variable.

Of course, all of our objects will be well-defined with respect to this convention.²

Definition 1.9. Let X be a random variable on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The *expectation* (or *expected value*, *mean*) of X is

$$\mathbb{E}[X] = \int_{\Omega} X \, d\mathbb{P},$$

if the above Lebesgue integral exists.

We now recall the construction of the Lebesgue integral. For simplicity we focus only on random variables.

Construction of the Lebesgue integral.

1. A random variable is called *simple* if it is of the form $X = \sum_{i=1}^n c_i \mathbb{1}_{A_i}$, with $c_i \in \mathbb{R}^+$ and $A_i \in \mathcal{A}$. For such a random variable, we define

$$\mathbb{E}[X] := \sum_{i=1}^n c_i \mathbb{P}[A_i].$$

2. Let X be a non-negative random variable. Then we define

$$\mathbb{E}[X] := \sup\{\mathbb{E}[Y] \mid 0 \leq Y \leq X, \text{ with } Y \text{ simple r.v.}\}.$$

²This convention is motivated by the following fact: all the properties of random variables we are interested in, are always related to the Lebesgue integral. But the Lebesgue integral “can’t see” sets with zero measure.

3. Prove *Monotone Convergence Theorem*: If $(X_n)_{n \in \mathbb{N}}$ is a sequence of random variables such that $0 \leq X_n \leq X_{n+1}$ and $X_n \nearrow X$, then $\mathbb{E}[X_n] \nearrow \mathbb{E}[X]$.
4. Show that for any non-negative random variable X there is a sequence $(X_n)_{n \in \mathbb{N}}$ of simple random variables such that $0 \leq X_n \leq X_{n+1}$ almost surely. The sequence can be constructed as follows:

$$X_n := \underbrace{\sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{1}_{\left[\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right]}}_{=: Y_n} + n \mathbb{1}_{[X \geq n]}.$$

Observe that $\left[\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right]$ and $[X \geq n]$ are both in $\sigma(X)$.

Corollary.

- (a) By monotone convergence we immediately have

$$\mathbb{E}[X] = \lim_{n \rightarrow \infty} \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{P}\left[\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right] + n \mathbb{P}[X \geq n].$$

- (b) If $\mathbb{P}[X = +\infty] > 0$, then $\mathbb{E}[X] = +\infty$. This is because $\mathbb{P}[X \geq n] \nearrow \mathbb{P}[X = \infty]$.
- (c) If $\mathbb{E}[X] < +\infty$, then $n \mathbb{P}[X \geq n] \rightarrow 0$.

Proof. First observe that $Y_n \leq Y_{n+1}$ for all $n \in \mathbb{N}$. Now, if $X(\omega) < \infty$, then $\exists n_0 \in \mathbb{N}$ s.t. $X(\omega) \geq n_0$. So $n \mathbb{1}_{[X \geq n]}(\omega) = 0$ for all $n > n_0$, i.e. $X(\omega) = \lim_n Y_n(\omega)$. Thus $Y_n \nearrow X \cdot \mathbb{1}_{[X < +\infty]}$ almost surely, because by (b) we have $\mathbb{E}[X] < +\infty \Rightarrow \mathbb{P}[X < +\infty] = 1$.

Therefore $Y_n \leq X_n \leq X$, and hence by monotone convergence we obtain $\mathbb{E}[Y_n] \rightarrow \mathbb{E}[X]$ and $\mathbb{E}[X_n] \rightarrow \mathbb{E}[X]$. Since $\mathbb{E}[X] < +\infty$, we can finally conclude that $\mathbb{E}[X_n] - \mathbb{E}[Y_n] \rightarrow 0$, i.e. $n \mathbb{P}[X \geq n] \rightarrow 0$. \square

5. Let X be an arbitrary random variable. We can write $X = X^+ - X^-$. So the expected value of X is defined as

$$\mathbb{E}[X] := \mathbb{E}[X^+] - \mathbb{E}[X^-],$$

unless $\mathbb{E}[X^+]$ and $\mathbb{E}[X^-]$ are both $+\infty$.

Definition 1.10. A random variable X is called *integrable* if $\mathbb{E}[X]$ exists and is finite (i.e. $\mathbb{E}[|X|] < \infty$).

Theorem 1.11. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $(\Omega^*, \mathcal{A}^*)$ a measurable space, $Y : (\Omega, \mathcal{A}) \rightarrow (\Omega^*, \mathcal{A}^*)$ a measurable function. Consider the following probability measure on $(\Omega^*, \mathcal{A}^*)$: for any $A^* \in \mathcal{A}^*$, define

$$\mathbb{P}^*[A^*] := \mathbb{P}[Y^{-1}[A^*]] = \mathbb{P}[Y \in A^*].$$

Let $f : (\Omega^*, \mathcal{A}^*) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$. The following holds:

If $\int_{\Omega^*} f \, d\mathbb{P}^*$ exists, then $\int_{\Omega} f(Y) \, d\mathbb{P}$ exists³ and

$$\int_{\Omega} f(Y) \, d\mathbb{P} = \int_{\Omega^*} f \, d\mathbb{P}^*.$$

Proof. In order to lighten the notation, for all $\omega \in \Omega$ we define $\omega^* := Y(\omega)$. The proof follows almost completely the argument for the construction of the expected value. We proceed by steps:

1. Suppose f is simple, i.e. $f = \sum_{i=1}^n c_i \mathbb{1}_{A_i^*}$. Then $f \circ Y = \sum_{i=1}^n c_i \mathbb{1}_{Y^{-1}[A_i^*]}$. So

$$\int_{\Omega} f(Y) \, d\mathbb{P} = \sum_{i=1}^n c_i \mathbb{P}[Y^{-1}[A_i^*]] = \sum_{i=1}^n c_i \mathbb{P}^*[A_i^*] = \int_{\Omega^*} f \, d\mathbb{P}^*.$$

2. If $f \geq 0$, we already know that we can find a sequence $(f_n)_{n \in \mathbb{N}}$ of simple functions such that $0 \leq f_n \leq f_{n+1} \nearrow f$. So, by monotone convergence we obtain

$$\int_{\Omega} f(Y) \, d\mathbb{P} = \lim_{n \rightarrow \infty} \int_{\Omega} f_n(Y) \, d\mathbb{P} = \lim_{n \rightarrow \infty} \int_{\Omega^*} f_n \, d\mathbb{P}^* = \int_{\Omega^*} f \, d\mathbb{P}^*,$$

where the second equality holds by first point.

3. Finally, for any general f , it is sufficient to write $f = f^+ - f^-$.

□

As a special case of the last theorem, we obtain the following:

Corollary 1.12. Let X be a random variable on $(\Omega, \mathcal{A}, \mathbb{P})$ a.s. finite. Let $f : (\mathbb{R}, \mathcal{B}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ measurable. If $\int_{\mathbb{R}} f \, dP^X$ exists, then

$$\mathbb{E}[f(X)] = \int_{\mathbb{R}} f \, dP^X.$$

³Of course, $f(Y) : \Omega \rightarrow \overline{\mathbb{R}}$ means just $f \circ Y$.

Definition 1.13. Let X be a random variable, a.s. finite. The *distribution function* of X is $F_X : \mathbb{R} \rightarrow [0, 1]$ given by

$$F_X(t) = \mathbb{P}[X \leq t] = P^X[(-a, t]].$$

We now recall some properties of distribution functions. The reader should know them already, therefore we omit a proof.

Proposition 1.14. Let F_X be a distribution function. The following hold:

1. F is monotone increasing.
2. F is continuous from the right⁴, but $\lim_{t \rightarrow t_0^-} F(t) = \mathbb{P}[X < t]$ is not necessarily equal to $\mathbb{P}[X \leq t]$.
3. $\lim_{t \rightarrow +\infty} F(t) = 1$.
4. $\lim_{t \rightarrow -\infty} F(t) = 0$.

Theorem 1.15 (Measure theory). If $F : \mathbb{R} \rightarrow [0, 1]$ is a function with the properties (1)–(4) of Proposition 1.14, then there exists a unique probability measure \mathbb{P} on $(\mathbb{R}, \mathcal{B})$ such that, for any interval $(a, b]$, $\mathbb{P}[(a, b]] = F(b) - F(a)$.

Notation. If \mathbb{P} is a probability measure on \mathbb{R} and $f : \mathbb{R} \rightarrow \mathbb{R}$ measurable, we write

$$\int_{\mathbb{R}} f \, dF := \int_{\mathbb{R}} f(t) \, dF(t) := \int_{\mathbb{R}} f \, d\mathbb{P},$$

where the latter is the Lebesgue integral.

Lemma 1.16. Let F be a distribution function. Suppose F is continuous on \mathbb{R} , differentiable (???) except for finitely many point and continuous integrable (???). Then $\int_{\mathbb{R}} f \, d\mathbb{P} = \int_{-\infty}^{+\infty} f(t) F'(t) \, dt$.

1.3 Independence of random variables.

In this section we always work in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$.

⁴Roughly speaking, a function is right-continuous if no jump occurs when the limit point is approached from the right. Formally, F is said to be right-continuous at the point c if the following holds: $\forall \varepsilon > 0 \, \exists \delta > 0 \, \forall x [c < x < c + \delta \Rightarrow |f(x) - f(c)| < \varepsilon]$.

Definition 1.17. Two sub- σ -algebras $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathcal{A}$ are called *independent* if

$$\forall A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2 \quad \left(\mathbb{P}[A_1 \cap A_2] = \mathbb{P}[A_1]\mathbb{P}[A_2] \right).$$

Sub- σ -algebras $\mathcal{A}_1, \dots, \mathcal{A}_n \subseteq \mathcal{A}$ are called *independent* if $\forall A_i \in \mathcal{A}_i$ with $i = 1, \dots, n$, A_1, \dots, A_n are mutually independent, i.e. if for all choices of indices $\{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$ we have

$$\mathbb{P}[A_{i_1} \cap \dots \cap A_{i_k}] = \mathbb{P}[A_{i_1}] \cdot \dots \cdot \mathbb{P}[A_{i_k}].^5$$

Arbitrarily many sub- σ -algebras $\mathcal{A}_i \subseteq \mathcal{A}$, $i \in I$ are called *independent* if they are independent for any finite subset of I .

Definition 1.18. Let $(\Omega_i, \mathcal{A}_i)$ with $i \in I$ be measurable spaces. Let $f_i : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\Omega_i, \mathcal{A}_i)$ be a measurable function for any $i \in I$. The f_i 's are called *independent* if the sub- σ -algebras $\sigma(f_i) \subseteq \mathcal{A}$ generated by them are independent (cfr. Definition 1.5).

Examples. Let X_1, \dots, X_n be discrete random variables (NOTA: AGGIUNGERE DEFINIZIONE). Define $p_i(x) := \mathbb{P}[X_i = x]$ and suppose $\sum_{x \in \mathbb{R}} p_i(x) = 1$. Consider the random vector

$$(X_1, \dots, X_n) : (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}),$$

which of course is discrete as well. Define

$$P_{X_1, \dots, X_n}(x_1, \dots, x_n) := \mathbb{P}[X_1 = x_1, \dots, X_n = x_n] = \mathbb{P} \left[\bigcap_{i=1}^n X_i^{-1}[\{x_i\}] \right].$$

Then:

Lemma 1.19. X_1, \dots, X_n are independent if and only if $P_{X_1, \dots, X_n}(x_1, \dots, x_n) = p(x_1) \cdot \dots \cdot p(x_n)$, for all x_1, \dots, x_n .

Proposition 1.20. Let X_1, \dots, X_n be continuous random variables. Let f_1, \dots, f_n be their density functions w.r.t. Lebesgue measure. Now consider the random vector $(X_1, \dots, X_n) : (\Omega, \mathcal{A}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$. We say that $f(x_1, \dots, x_n) : \mathbb{R} \rightarrow \mathbb{R}$ is the *density function of (X_1, \dots, X_n)* (w.r.t. Lebesgue measure on \mathbb{R}^n) if $\mathbb{P}[(X_1, \dots, X_n) \in B] = \int_B f \, d\lambda_n$ for all $B \in \mathcal{B}_{\mathbb{R}^n}$, where λ_n is the Lebesgue measure on \mathbb{R}^n . Such a function might not exist. Nevertheless, the following are equivalent:

1. X_1, \dots, X_n are independent.
2. The density function $f(x_1, \dots, x_n)$ exists and $f(x_1, \dots, x_n) = f_1(x_1) \cdot \dots \cdot f_n(x_n)$ a.s.

⁵By Exercise 7, this condition is equivalent to the following: for all choices of $\varepsilon_1, \dots, \varepsilon_n \in \{1, -1\}$, $\mathbb{P}[A_1^{\varepsilon_1} \cap \dots \cap A_n^{\varepsilon_n}] = \mathbb{P}[A_1^{\varepsilon_1}] \cdot \dots \cdot \mathbb{P}[A_n^{\varepsilon_n}]$, where $A_i^1 = A_i$ and $A_i^{-1} = A_i^c$.

Proof.

(b) \Rightarrow (a): By hypothesis

$$\mathbb{P}[(X_1, \dots, X_n) \in B] = \int_B f_1(x_1) \dots f_n(x_n) \underbrace{d\lambda_n(x_1, \dots, x_n)}_{dx_1 \dots dx_n},$$

for all $B \in \mathcal{B}_{\mathbb{R}^n}$. Now take $B_i \in \mathcal{B}_{\mathbb{R}}$. So we have

$$\begin{aligned} \mathbb{P}[X_1 \in B_1, \dots, X_n \in B_n] &= \int_{B_1 \times \dots \times B_n} f_1(x_1) \dots f_n(x_n) d\lambda_n(x_1, \dots, x_n) \\ &= \int_{B_1} f_1 d\lambda \int_{B_2} f_2 d\lambda \dots \int_{B_n} f_n d\lambda = \mathbb{P}[x_1 \in B_1] \dots \mathbb{P}[x_n \in B_n]. \end{aligned}$$

(a) \Rightarrow (b): Copiare se si ha tempo.

□

Chapter 2

The law of large numbers and Birkhoff's Ergodic theorem

Throughout the whole chapter, all the random variables live in a probability space $(\Omega, \mathcal{A}, \mathbb{P})$, unless stated otherwise.

Theorem 2.1. Let (X_n) be a sequence of i.i.d. real random variables with $\mathbb{E}[X_n] = \mu$ and $\text{Var}[X_n] = \sigma^2 < \infty$. Define

$$\bar{X}_n := \frac{1}{n}(X_1 + \dots + X_n).$$

Then $\bar{X}_n \rightarrow \mu$ in probability, i.e.

$$\mathbb{P}[|\bar{X}_n - \mu| \geq a] \leq \frac{\sigma^2}{na^2} \rightarrow 0.$$

Proof. We split the proof in four steps.

1. Markov's inequality:

If X is a non-negative random variable, then $\mathbb{P}[X \geq a \mathbb{E}[X]] \leq \frac{1}{a}$ for all $a > 0$.

In order to prove this, first observe that if $\mathbb{E}[X] = 0$, then $X = 0$ a.s., and thus the statement holds (??? a me non sembra che funzioni). Suppose now $\mathbb{E}[X] > 0$. Of course $X \geq X \mathbb{1}_{[X \geq a\mu]}$, therefore

$$\mathbb{E}[X] \geq \mathbb{E}[X \mathbb{1}_{[X \geq a\mu]}] \geq \mathbb{E}[a\mu \mathbb{1}_{[X \geq a\mu]}] = a \mathbb{E}[X] \mathbb{P}[X \geq a \mathbb{E}[X]].$$

2. Chebyshev's inequality.

If $\text{Var}[X] < \infty$, then $\mathbb{P}[|X - E[X]| \geq a] \leq \frac{\text{Var}[X]}{a^2}$ for all $a > 0$.

The proof follows immediately by Markov's inequality, since

$$\mathbb{P}[|X - \mu| \geq a] = \mathbb{P}\left[\underbrace{(X - \mu)^2}_{=: Y \geq 0} \geq \frac{a^2}{\sigma^2} \underbrace{\sigma^2}_{=: \mathbb{E}[Y] = \text{Var}[X]}\right] \leq \frac{\sigma^2}{a^2}.$$

3. The following holds:

If X_1 and X_2 are independent then $\text{Var}[X_1 + X_2] = \text{Var}[X_1] + \text{Var}[X_2]$.

Indeed, observe that

$$\begin{aligned} \text{Var}[X_1 + X_2] &= \mathbb{E}[(X_1 - \mu_1) + (X_2 - \mu_2)]^2 \\ &= \mathbb{E}[(X_1 - \mu_1)^2] + \mathbb{E}[(X_2 - \mu_2)^2] + \underbrace{2\mathbb{E}[(X_1 - \mu_1)(X_2 - \mu_2)]}_{=0} = \text{Var}[X_1] + \text{Var}[X_2]. \end{aligned}$$

4. Finally, observe that

$$\mathbb{E}[\bar{X}_n] = \mu \quad \text{and} \quad \text{Var}[\bar{X}_n] = \text{Var}\left[\frac{1}{n}(X_1 + \dots + X_n)\right] = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}.$$

So the statement immediately follows by Chebyshev's inequality.

□

Definition 2.2. Let (X_n) and X be random variables. We say that

- $X_n \rightarrow X$ *in probability* if $\mathbb{P}[|X_n - X| \geq a] \rightarrow 0$ for all $a > 0$.
- $X_n \rightarrow X$ *a.s.* if $\mathbb{P}[X_n \rightarrow X] = 1$, that is

$$\Omega_0 \in \mathcal{A} \quad \text{and} \quad \mathbb{P}[\Omega_0] = 1,$$

where $\Omega_0 := \{\omega \in \Omega \mid \lim_n X_n(\omega) \text{ exists and is equal to } X(\omega)\}$.

Lemma 2.3. A.s. convergence implies convergence in probability.

Proof. Define $U_k := \sup\{|X_n - X| : n \geq k\}$. Then $X_n \rightarrow X$ a.s. implies $U_k \rightarrow 0$ in probability. □

Theorem 2.4 (Strong law of large numbers). Let (X_n) be iid and suppose $\mathbb{E}[X_n] = \mu < \infty$ for all $n \in \mathbb{N}$. Then $\bar{X}_n \rightarrow \mu$ a.s.

2.1 Products of probability spaces.

Question. Given a probability measure P on $(\mathbb{R}, \mathcal{B})$, is there a probability space $(\Omega, \mathcal{A}, \mathbb{P})$ on which one can define a sequence of iid random variables X_n with distribution P ?

Answer. Yes. Define $(\Omega, \mathcal{A}, \mathbb{P}) := \times_{n \in \mathbb{N}} (\mathbb{R}, \mathcal{B}, P)$. (??? e poi?)

2.1.1 Finitely many spaces

Let $(\Omega_i, \mathcal{A}_i, \mathbb{P}_i)$ be probability spaces, with $i = 1, 2$. Define

- $\Omega := \Omega_1 \times \Omega_2$
- $\mathcal{A}_1 \otimes \mathcal{A}_2 := \sigma(\{A_1 \times A_2 \mid A_i \in \mathcal{A}_i\})$
- $\mathbb{P}_0[A_1 \times A_2] := \mathbb{P}_1[A_1] \mathbb{P}_2[A_2]$.

It's easy to show that \mathbb{P}_0 can be uniquely extended to a σ -additive probability measure \mathbb{P} on $\mathcal{A}_1 \otimes \mathcal{A}_2$. So we can define $(\Omega_1, \mathcal{A}_1, \mathbb{P}_1) \times (\Omega_2, \mathcal{A}_2, \mathbb{P}_2) := (\Omega, \mathcal{A}, \mathbb{P})$.

2.1.2 Countably many spaces

Let $(\Omega_n, \mathcal{A}_n, \mathbb{P}_n)$ be probability spaces for all $n \in \mathbb{N}$. Define

- $\Omega := \times_{n \in \mathbb{N}} \Omega_n$
- $\mathcal{A} := \sigma(\{A_1 \times \dots \times A_n \times \Omega_{n+1} \times \dots \mid A_i \in \mathcal{A}_i, n \in \mathbb{N}\})$
- $\mathbb{P}_0[A_1 \times \dots \times A_n \times \Omega_{n+1} \times \dots] := \mathbb{P}_1[A_1] \mathbb{P}_2[A_2] \dots \mathbb{P}_n[A_n] \cdot 1 \cdot 1 \cdot \dots$

Construction.

For all $n \in \mathbb{N}$, define

$$(\Omega^{(n)}, \mathcal{A}^{(n)}, \mathbb{P}^{(n)}) := \bigotimes_{i=1}^n (\Omega_i, \mathcal{A}_i, \mathbb{P}_i).$$

By abuse of notation, we identify

$$(\Omega^{(n)}, \mathcal{A}^{(n)}, \mathbb{P}^{(n)}) \quad \text{and} \quad (\Omega, \{A^{(n)} \times \Omega_{n+1} \times \dots \mid A^{(n)} \in \mathcal{A}^{(n)}\}, \mathbb{P}^{(n)}).$$

Observe that $\mathcal{A}^{(n-1)} \subseteq \mathcal{A}^{(n)}$, since

$$\underbrace{(A^{(n-1)} \times \Omega_n)}_{\in \mathcal{A}^{(n)}} \times \Omega_{n+1} \times \dots$$

Now define

$$\mathcal{R} := \bigcup_{n \in \mathbb{N}} \mathcal{A}^{(n)}.$$

\mathcal{R} is not a σ -algebra, but it is an algebra over Ω . Now, if $A \in \mathcal{R}$, then $A \in \mathcal{A}^{(j)}$ for some $j \in \mathbb{N}$. Thus we can define the measure \mathbb{P} on \mathcal{R} given by

$$\mathbb{P}[A] := \mathbb{P}^{(j)}(A), \quad (*)$$

which is well-defined since $\mathbb{P}^{(n)} \upharpoonright \mathcal{A}^{(n-1)} = \mathbb{P}^{(n-1)}$ for all $n \in \mathbb{N}$. Of course \mathbb{P} is finitely additive.

Theorem 2.5. \mathbb{P} is σ -additive on \mathcal{R} . This also implies that \mathbb{P} can be uniquely extended to $\sigma(\mathcal{R}) = \mathcal{A}$.

Proof. We need more notation. Define

- $\Omega^{(k,n)} := \Omega_{k+1} \times \dots \times \Omega_1$, so $\Omega^{(n)} = \Omega^{(0,n)}$.
- $\Omega^{[k]} := \Omega^{(k,\infty)}$ so $\Omega = \Omega^{[0]}$.
- $\mathcal{A}^{(k,n)} := \mathcal{A}_{k+1} \otimes \dots \otimes \mathcal{A}_n$, so $\mathcal{A}^{(n)} = \mathcal{A}^{(0,n)}$.
- $\mathbb{P}^{(k,n)}$ is defined in the obvious way, and we have $\mathbb{P}^{(n)} = \mathbb{P}^{(0,n)}$.
- By abuse of notation, define $\mathcal{R}^{[k]} := \bigcup_{n > k} \mathcal{A}^{(k,n)}$, which is an algebra.
- $\mathbb{P}^{[k]}$ on $\mathcal{R}^{[k]}$ is defined in the obvious way, similar to (*). Of course it is finitely additive. So $\mathbb{P}^{[0]} = \mathbb{P}$.

Now consider the following lemma, whose proof is left as an easy exercise:

A finitely additive measure on a ring (algebra) of set is σ -additive on \mathcal{R} if and only if it's continuous at \emptyset .

Let (B_n) be a sequence in $\mathcal{R} = \mathcal{R}^{[0]}$ such that $B_n \searrow 0$. Thanks to the lemma, if we show that $\mathbb{P}^{[0]}[B_n] \searrow 0$ we are done. Of course $B_n \in \mathcal{A}^{(k_n)}$ for some $k_n \in \mathbb{N}$. Observe that we can suppose w.l.o.g. that $(k_n)_{n \in \mathbb{N}}$ is strictly ascending (because otherwise we can simply define $\overline{k}_1 := k_1$ and $\overline{k}_n := \max\{k_n, \overline{k}_{n-1} + 1\}$), and therefore, again w.l.o.g. $B_n \in \mathcal{A}^{(n)}$ (because if there is a “jump”, between some k_n and k_{n+1} , we can just fill it “extending” the sequence ??? giusto?).

Now assume towards a contradiction $\mathbb{P}^{[0]}[B_n] \searrow \varepsilon > 0$. We will show that this implies $\bigcup_{n \in \mathbb{N}} B_n \neq \emptyset$, which is a contradiction.

For $E \in \mathcal{R}$ and $(\omega_1, \dots, \omega_n) \in \Omega_1 \times \dots \times \Omega_n$, define

$$E(\omega_1, \dots, \omega_n) := \{\boldsymbol{\omega} = (\omega_{n+1}, \omega_{n+2}, \dots) \in \Omega^{[n]} \mid (\omega_1, \dots, \omega_n, \boldsymbol{\omega}) \in E\} \in \mathcal{R}^{[n]}.$$

Let $C_n := \{\omega_1 \in \Omega_1 \mid \mathbb{P}^{[0]}[B_n(\omega_1)] > \frac{\varepsilon}{2}\}$. So we have

$$\varepsilon \leq \mathbb{P}[B_n] = \int_{\Omega_1} \mathbb{P}^{(1,n)}[B_n(\omega_1)] d\mathbb{P}_1(\omega_1) = \int_{C_n} \dots + \int_{\Omega_1 \setminus C_n} \dots \leq \mathbb{P}_1[C_n] + \frac{\varepsilon}{2}.$$

Thus $\mathbb{P}_1[C_n] \geq \frac{\varepsilon}{2}$ for all $n \in \mathbb{N}$, and since $(C_n)_{n \in \mathbb{N}}$ is clearly decreasing, this means $\bigcap_{n \in \mathbb{N}} C_n \neq \emptyset$. So there exists $a_1 \in \bigcap_{n \in \mathbb{N}} C_n$.

Now continue with $B_n(a_1) \in \mathcal{A}^{(1,n)} \subseteq \mathcal{R}^{[1]}$. We have that $B_n(a_1) \searrow$ and $\mathbb{P}^{[1]}[B_n(a_1)] \geq \frac{\varepsilon}{2}$. Following the same argument we can find $a_2 \in \Omega_2$ such that $B_n(a_1, a_2) \searrow$ and $\mathbb{P}^{[2]}[B_n(a_1, a_2)] \geq \frac{\varepsilon}{4}$. By induction, for all k we can find a_1, \dots, a_k such that

$$B_n(a_1, \dots, a_k) \searrow \quad \text{and} \quad \mathbb{P}^{[k]}[B_n(a_1, \dots, a_k)] \geq \frac{\varepsilon}{2^k}.$$

We claim $(a_1, a_2, \dots) \in \bigcap_{n \in \mathbb{N}} B_n$. Observe that

$$B_n(a_1, \dots, a_n) = \begin{cases} \emptyset, & \text{if no element of } B_n \text{ begins with } (a_1, \dots, a_n) \\ \Omega_{n+1} \times \Omega_{n+2} \times \dots, & \text{otherwise (??? pourquoi?).} \end{cases}$$

Since $\mathbb{P}^{[n]}[B_n(a_1, \dots, a_n)] \geq \frac{\varepsilon}{2^n}$, the first case can never happen. Therefore

$$(a_1, \dots, a_n, a_{n+1}, a_{n+2}, \dots) \in B_n$$

for all $n \in \mathbb{N}$. □

2.2 Birkhoff's ergodic theorem

Definition 2.6. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A measurable function $T: \Omega \rightarrow \Omega$ is called *measure preserving* if for all $A \in \mathcal{A}$

$$\mathbb{P}[T^{-1}[A]] = \mathbb{P}[A].$$

Remark. T is measure preserving if and only if $\mathbb{E}[f \circ T] = \mathbb{E}[f]$ for all $f \in L^1(\Omega, \mathcal{A}, \mathbb{P})$.

Proof. Exercise. □

Examples.

1. Let $\Omega := [0, 1)$, $\mathcal{A} := \mathcal{B}_{[0,1)}$, $\mathbb{P} := \lambda$, where λ is the Lebesgue measure on $[0, 1)$. It's easy to check that $T(x) := \langle 2x \rangle := 2x - \lfloor 2x \rfloor$ is measure preserving.

2. Let X be a real-valued random variable with distribution $P := P^X$. More in general: let \mathfrak{X} be a separable complete metrix space, \mathcal{B} the Borel σ -algebra, P a probability measure on \mathcal{B} . We can construct a sequence (X_n) of \mathfrak{X} -valued iid random variables with distribution P as follows: we define

$$(\Omega, \mathcal{A}, \mathbb{P}) := \bigotimes_{n=1}^{\infty} (\mathfrak{X}, \mathcal{B}, P) = (\mathfrak{X}^{\mathbb{N}}, \mathcal{B} \otimes \mathcal{B} \otimes \dots, P^{\mathbb{N}}).$$

So we can write every $\omega \in \mathfrak{X}^{\mathbb{N}}$ as $\omega = (x_1, x_2, \dots)$. Defining $X_n(\omega) := x_n$ for all $n \in \mathbb{N}$, we obtain what we wanted.

Furthermore, defining $T(\omega) := (x_2, x_3, \dots)$ we get a measure preserving function, since $A \in \mathcal{A} \Rightarrow T^{-1}[A] = \mathfrak{X} \times A$. T is called *shift*.

Let $f \in L^1(\Omega, \mathcal{A}, \mathbb{P})$, T m.p. (measure preserving). We define for all $n \geq 0$

$$S_n f(\omega) := \sum_{k=0}^{n-1} f(T^k(\omega)).^1$$

We call $\frac{1}{n} S_n f$ *time average* and $\mathbb{E}(f) = \int_{\Omega} f d\mathbb{P}$ *space average*.

We will soon prove that $S_n f \rightarrow \mathbb{E}(f)$ a.s.

Theorem 2.7 (Maximal Ergodic Lemma). Let $F_n(\omega) := \max\{S_j f(\omega) \mid j = 0, \dots, n\}$. Then

$$\int_{[F_n > 0]} f d\mathbb{P} \geq 0.$$

Proof. First observe that $F_n \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ (because f is integrable, so also the sum is integrable, and thus also the maximum). We have

$$\begin{aligned} F_n &\geq S_j f \\ \Rightarrow F_n \circ T &\geq S_j f \circ T \\ \Rightarrow f + F_n \circ T &\geq f + S_j f \circ T = S_{j+1} f \\ \Rightarrow f + F_n \circ T &\geq \max\{S_j f \mid j = 1, \dots, n\}. \end{aligned}$$

If $F_n(\omega) > 0$, then $\max\{S_j f(\omega) \mid j = 1, \dots, n\} = F_n(\omega)$. This means that on $[F_n > 0]$ we have $f \geq F_n - F_n \circ T$. Thus

$$\begin{aligned} \int_{[F_n > 0]} f d\mathbb{P} &\geq \int_{[F_n > 0]} F_n d\mathbb{P} - \int_{[F_n > 0]} F_n \circ T d\mathbb{P} \\ &\geq \int_{\Omega} F_n d\mathbb{P} - \int_{\Omega} F_n \circ T d\mathbb{P} \\ &= \mathbb{E}[F_n] - \mathbb{E}[F_n \circ T] = 0, \end{aligned}$$

where the last equality follows by the last remark. □

¹Recall that the empty sum is defined as 0.

Corollary 2.8. Suppose $f \in L^1$ and $\mathbb{E}[f] = \int_{\Omega} f \, d\mathbb{P} > 0$. Then

$$\exists \omega \in \Omega \, \forall n \geq 1 \, (S_n f(\omega) > 0).$$

Proof. Assume to the contrary that

$$\forall \omega \in \Omega \, \exists n \geq 1 \, (S_n f(\omega) \leq 0).$$

Let $0 < \varepsilon < \mathbb{E}[f]$. Define $g := \varepsilon - f$. We trivially have $S_n g = n\varepsilon - S_n f$. Thus

$$\forall \omega \in \Omega \, \exists n \geq 1 \, (S_n g(\omega) \geq n\varepsilon > 0),$$

Which implies

$$\forall \omega \in \Omega \, \exists n \geq 1 \, (F_n g(\omega) > 0).$$

Define $A_n := [F_n g > 0]$. We have that $A_n \nearrow \Omega$. Applying maximal ergodic lemma we obtain

$$\int_{\Omega} \mathbb{1}_{A_n} g \, d\mathbb{P} = \int_{A_n} g \, d\mathbb{P} \geq 0.$$

Observe that $|\mathbb{1}_{A_n} g| \leq |g| \in L^1$. By dominated convergence we obtain

$$\int_{\Omega} g \, d\mathbb{P} \geq 0.$$

But since $f = \varepsilon - g$ this implies

$$\mathbb{E}[f] = \int_{\Omega} f \, d\mathbb{P} = \int_{\Omega} \varepsilon \, d\mathbb{P} - \int_{\Omega} g \, d\mathbb{P} \leq \int_{\Omega} \varepsilon \, d\mathbb{P} = \varepsilon,$$

contradiction. □

Definition 2.9. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $T: \Omega \rightarrow \Omega$ measure preserving. The T -invariant σ -algebra is

$$\mathcal{J} := \{A \in \mathcal{A} \mid T^{-1}[A] = A \text{ a.s.}\}^2$$

T is called *ergodic* if \mathcal{J} is *trivial w.r.t.* \mathbb{P} , i.e. if

$$A \in \mathcal{J} \Rightarrow \mathbb{P}[A] \in \{0, 1\}.$$

Lemma 2.10.

1. $f: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B})$ is \mathcal{J} -measurable iff $f = f \circ T$ a.s.

²We say that $A = B$ a.s. if $\mathbb{P}[(A \setminus B) \cup (B \setminus A)] = 0$. Observe that $\mathbb{P}[A] = 0 \Leftrightarrow A = \emptyset$ a.s. and $\mathbb{P}[A] = 1 \Leftrightarrow A = \Omega$ a.s.

2. \mathcal{J} is trivial iff every \mathcal{J} -measurable function is a.s. constant.

Proof. Exercise. □

Remark. Suppose that $T: \Omega \rightarrow \Omega$ is such that $T^{-1}[A] = A$ a.s. Define

$$\tilde{A} := \limsup T^{-n}[A] = \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} T^{-n}[A] \stackrel{\text{a.s.}}{=} \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A = A.$$

Observe also that

$$T^{-1}[\tilde{A}] = T^{-1} \left[\bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} T^{-n}[A] \right] = \bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} T^{-(n+1)}[A] = \bigcap_{k \geq 1} \bigcup_{n \geq k} T^{-n}[A] = \tilde{A}.$$

So $T^{-1}[\tilde{A}] = \tilde{A}$ and $\tilde{A} \stackrel{\text{a.s.}}{=} A$.

Theorem 2.11 (Birkhoff's pointwise ergodic theorem). Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, $T: \Omega \rightarrow \Omega$ measure preserving, and \mathcal{J} the T -invariant σ -algebra. Suppose $f \in L^1(\Omega, \mathcal{A}, \mathbb{P})$ and $\mathbb{E}[f] = \int_{\Omega} f \, d\mathbb{P} > 0$. Then there exists $Mf \in L^1(\Omega, \mathcal{J}, \mathbb{P})$ such that the following hold:

1. $\frac{1}{n} S_n f \rightarrow Mf$ a.s.
2. $\int_A f \, d\mathbb{P} = \int_A Mf \, d\mathbb{P}$, for all $A \in \mathcal{J}$.

In particular, if \mathcal{J} is trivial, by last lemma this implies

$$Mf = \int_{\Omega} f \, d\mathbb{P}.$$

Proof. Define

$$M^+ f(\omega) := \limsup \frac{1}{n} S_n f(\omega) \quad \text{and} \quad M^- f(\omega) := \liminf \frac{1}{n} S_n f(\omega).$$

Claim 1: $M^+ f$ and $M^- f$ are \mathcal{J} -measurable.

Proof: Observe that $S_{n+1} f = f + S_n f \circ T$. So $S_n f \circ T = S_{n+1} f - f$, and thus

$$\begin{aligned} M^+ f \circ T(\omega) &= \limsup \frac{1}{n} S_n f \circ T(\omega) = \limsup \frac{1}{n} f + \frac{1}{n} S_{n+1} f(\omega) \\ &= \limsup \frac{1}{n} S_{n+1} f(\omega) = \limsup \frac{1}{n} S_n f(\omega) = M^+ f(\omega). \end{aligned}$$

Thanks to last lemma we are done. ■

Claim 2: $M^+ f < \infty$ a.s. (and $M^- f > -\infty$ a.s.).

Proof: Define $A := [M^+ f = \infty]$. Observe that $A \in \mathcal{J}$ by Claim 1, since it's the preimage of a singleton under a J -measurable function. This implies $A = T^{-1}[A]$, whereby it immediately follows that $T[A] \subseteq A$.

Now suppose towards a contradiction that $\mathbb{P}[A] > 0$. Then there exists $c \in (0, \infty)$ s.t. $c\mathbb{P}[A] > \int_A f d\mathbb{P}$. Now define $g := c - f$. Of course $\int_A g d\mathbb{P} > 0$. Now consider the restriction

$$T|_A : (A, \mathcal{A}_A, \mathbb{P}|_A) \rightarrow (A, \mathcal{A}_A, \mathbb{P}|_A).$$

By Corollary 2.8 we have

$$\exists \omega_0 \in A \forall n \geq 1 (S_n g(\omega_0) > 0).$$

Now observe that trivially $S_n g = nc - S_n f$. So

$$\frac{1}{n} S_n f(\omega_0) = c - \frac{1}{n} S_n g(\omega_0) < c$$

for all $n \geq 1$. This implies $\infty = M^+ f(\omega_0) \leq c < \infty$, contradiction. ■

For all $N \in \mathbb{N}$ define $\Omega_N := [-N \leq M^- f \leq M^+ f \leq N]$. Clearly we have $\Omega_N \nearrow [-\infty < M^- f \leq M^+ f < \infty] \stackrel{\text{a.s.}}{=} \Omega$. Furthermore $T[\Omega_N] \subseteq \Omega_N$ by the same argument as above, since $\Omega_N \in \mathcal{J}$.

Claim 3: $\int_{\Omega_N} M^+ f d\mathbb{P} \leq \int_{\Omega_N} f d\mathbb{P}$ and $\int_{\Omega_N} f d\mathbb{P} \leq \int_{\Omega_N} M^- f d\mathbb{P}$.

Proof: Suppose to the contrary that $\int_{\Omega_N} M^+ f d\mathbb{P} > \int_{\Omega_N} f d\mathbb{P}$. Then there exists $\varepsilon > 0$ such that

$$\int_{\Omega_N} M^+ f d\mathbb{P} > \int_{\Omega_N} f + \varepsilon d\mathbb{P}.$$

We have

$$g := (M^+ f - f - \varepsilon)|_{\Omega_N} \in L^1(\Omega_N, \mathcal{A}_{\Omega_N}, \mathbb{P})$$

and obviously $\int_{\Omega_N} g d\mathbb{P} > 0$. By Corollary 2.8 we get

$$\exists \omega_0 \in \Omega_N \forall n \geq 1 (S_n g(\omega_0) > 0).$$

Now consider

$$\frac{1}{n} S_n g = \frac{1}{n} S_n (M^+ f) - \frac{1}{n} S_n f - \frac{n}{n} \varepsilon,$$

Observe that $\frac{1}{n} S_n (M^+ f) = M^+ f$, because $M^+ f \circ T = M^+$ (cfr. proof of Claim 1). Thus

$$\frac{1}{n} S_n f(\omega_0) < M^+ f(\omega_0) - \varepsilon.$$

Taking the latter inequality to the limit $n \rightarrow \infty$ we obtain

$$M^+ f(\omega_0) \leq M^+ f(\omega_0) - \varepsilon,$$

and since by Claim 2 $M^+ f(\omega_0)$ is finite almost surely (??? eh ok, ma se per caso non lo è?), we get $\varepsilon \leq 0$, contradiction.

The other inequality follows by the very same argument. ■

So

$$\int_{\Omega_N} f \, d\mathbb{P} \leq \int_{\Omega_N} M^- f \, d\mathbb{P} \leq \int_{\Omega_N} M^+ f \, d\mathbb{P} \leq \int_{\Omega_N} f \, d\mathbb{P}$$

that is, the equality holds. Hence

$$\int_{\Omega_N} M^+ f - M^- f \, d\mathbb{P} = 0.$$

But $M^+ f - M^- f \geq 0$ immediately by definition, and so we trivially get $M^+ f - M^- f = 0$ a.s. on Ω_N , i.e. $M^+ f = M^- f =: M f = \lim_{n \rightarrow \infty} \frac{1}{n} S_n f$ a.s. on Ω_N . Since $\Omega_N \nearrow \Omega$ a.s., this means $\lim_{n \rightarrow \infty} \frac{1}{n} S_n f = M f$ a.s. on Ω , so (1) is proved.

Now write $f = f^+ - f^-$ and $M f = M f^+ - M f^-$. Since $\int_{\Omega_N} M f \, d\mathbb{P} = \int_{\Omega_N} f \, d\mathbb{P}$, we have

$$\int_{\Omega} M f^{\pm} \, \mathbb{1}_{\Omega_N} \, d\mathbb{P} = \int_{\Omega} f^{\pm} \, \mathbb{1}_{\Omega_N} \, d\mathbb{P},$$

and applying monotone convergence we get

$$\int_{\Omega} M f \, d\mathbb{P} = \int_{\Omega} f \, d\mathbb{P}.$$

Therefore $M f \in L^1(\Omega, \mathcal{J}, \mathbb{P})$.

Now let $A \in \mathcal{J}$. We must show that

$$\int_A M f \, d\mathbb{P} = \int_A f \, d\mathbb{P}. \tag{*}$$

- If $\mathbb{P}[A] = 0$, (*) is trivially true.
- If $\mathbb{P}[A] > 0$ and $T^{-1}[A] = A$, repeating the the proof all over again on $(A, \mathcal{A}_A, \mathbb{P}|_A)$ we find that (*) holds.
- If $\mathbb{P}[A] > 0$ and $T^{-1}[A] = A$ a.s., then consider $\tilde{A} := \limsup T^{-n}[A] \stackrel{\text{a.s.}}{=} A$. By last remark, $T^{-1}[\tilde{A}] = \tilde{A}$, and so by last point (*) holds for \tilde{A} . Since $\tilde{A} = A$ a.s., (*) also holds for A .

So (2) is proved, and the proof is complete. □

2.3 The Strong Law of Large Numbers

By Birkhoff's pointwise ergodic theorem we can deduce the Strong Law of Large Numbers. Let's repeat the statement:

Theorem 2.12 (Strong law of large numbers). Let (X_n) be iid and suppose $\mathbb{E}[X_n] = \mu < \infty$ for all $n \in \mathbb{N}$. Then $\overline{X}_n \rightarrow \mu$ a.s.

Proof. Let P be the distribution of the random variables. We can assume w.l.o.g. that the random variables live in the probability space $(\mathbb{R}, \mathcal{B}, P)^{\mathbb{N}}$.³ That is, $\Omega = \mathbb{R}^{\mathbb{N}}$, $\mathbb{P} := P^{\mathbb{N}}$ is the product measure on the product σ -algebra \mathcal{A} (cfr. 2.1.2), and for all $\Omega \ni \omega = (x_n)$ with $x_n \in \mathbb{R}$ we define $X_n(\omega) := x_n$. This way, the X_n are iid and have distribution P (cfr. second example at page 15).

Now define $T(\omega) := (x_{n+1})$, which is measure preserving (cfr. same example). Finally, define $f(\omega) := x_1$, i.e. $f = X_1$. So

$$\int_{\Omega} f \, d\mathbb{P} = \mathbb{E}[X_1] = \mu. \quad (*)$$

Of course $f \circ T^n(\omega) = x_n$. So

$$S_n f = \sum_{k=0}^{n-1} X_k.$$

By Birkhoff's pointwise ergodic theorem, there is $M f$ \mathcal{J} -measurable such that $\frac{1}{n} S_n f \rightarrow M f$ a.s.

Claim: T is ergodic.

Proof: Later. ■

Putting together (*) and Lemma 2.10(2), we obtain $M f = \mu$ a.s., and so

$$\frac{1}{n} S_n f \rightarrow \mu$$

a.s., as wanted. □

Definition 2.13. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let $(X_n)_{n \geq 1}$ be a sequence of (real) random variables. We call this *stochastic process with discrete time n* . Let

$$\mathcal{A}(X_1, \dots, X_n) := \sigma(X_1, \dots, X_n) = \{[(X_1, \dots, X_n) \in B] \mid B \in \mathcal{B}_{\mathbb{R}^n}\} \subseteq \mathcal{A}.^4$$

³From the third sheet of exercises: *The distribution of a given function of a sequence of i.i.d. random variables (X_n) does not depend on the concrete realisation of those random variables. That is, you may choose your favorite model for the underlying probability space and construction of the random variables.* So, if we start with a random variable X on $(\Omega, \mathcal{A}, \mathbb{P})$, and its distribution is P^X , we can actually look at it as the identity function on $(\mathbb{R}, \mathcal{B}, P^X)$ (??? giusto?).

Let

$$\mathcal{A}(X_1, X_2, \dots) := \bigvee_{n=1}^{\infty} \mathcal{A}(X_1, \dots, X_n) := \sigma \left(\bigcup_{n=1}^{\infty} \mathcal{A}(X_1, \dots, X_n) \right).^5$$

The *tail σ -algebra* of $(X_n)_{n \geq 1}$ is

$$\mathcal{T} := \bigcap_{k=1}^{\infty} \mathcal{A}(X_k, X_{k+1}, \dots).^6$$

Theorem 2.14 (Kolmogorov's 0–1 Law). If (X_n) are independent, then \mathcal{T} is trivial, so $\mathbb{P}[\mathcal{T}] = \{0, 1\}$.

Proof. Take $A \in \mathcal{T}$. Of course $A \in \mathcal{A}(X_{k+1}, X_{k+2}, \dots)$ for all $k \in \mathbb{N}$. But it's easy to check (nota: aggiungere cfr a esercizi) that $\mathcal{A}(X_{k+1}, X_{k+2}, \dots)$ is independent of $\mathcal{A}(X_1, \dots, X_k)$ for all $k \in \mathbb{N}$. This means that A is independent of every set in $\bigcup_{k=1}^{\infty} \mathcal{A}(X_1, \dots, X_k)$, and thus (nota: aggiungere perché, o almeno cfr al libro) of every set in $\sigma(\bigcup_{k=1}^{\infty} \mathcal{A}(X_1, \dots, X_k)) = \mathcal{A}(X_1, X_2, \dots)$. But of course $A \in \mathcal{A}(X_1, X_2, \dots)$. So

$$\mathbb{P}[A \cap A] = \mathbb{P}[A] \cdot \mathbb{P}[A] = \mathbb{P}[A]^2,$$

i.e. $\mathbb{P}[A] = \{0, 1\}$. □

We finally prove the claim in the proof of the Strong law of large numbers:

Lemma 2.15. T is ergodic.

Proof. Take $A \in \mathcal{J}$, i.e. $T^{-1}[A] = A$ a.s. We can assume w.l.o.g. that $T^{-1}[A] = A$, because otherwise we consider $\tilde{A} := \limsup T^{-1}[A]$ and we have already seen that $\tilde{A} = A$ a.s. and $T^{-1}[\tilde{A}] = \tilde{A}$.

So $T^{-1}[A] = A$. Recall that A is a set of sequences and T is the shift. So

$$T^{-1}[A] = \mathbb{R} \times A,$$

⁴We can see at this σ -algebra as the family of events it is possible to observe after the first n “experiments”.

⁵We can see at this σ -algebra as the family of events it is possible to observe after a finite amount of “experiments”. (??? giusto? non sono mica troppo sicuro...)

⁶“The tail sigma-algebra is the sigma-algebra of sets B such that, for every integer N one can build B from the sets A_n with $n \geq N$ only. For example the \limsup/\liminf of $(A_n)_n$ is also the \limsup/\liminf of $(A_{n+N})_n$ hence the \limsup/\liminf is in the tail sigma-algebra. There is no measure involved here.

In the independent case with respect to a probability P , the tail sigma-algebra is trivial in the sense that it contains only sets of P -probability zero or one. For example the \limsup/\liminf of any sequence which is independent for P has P -probability zero or one. This is a property of P in relation with the sigma-algebras considered but definitely not a property of the sigma-algebras alone.” (source)

and thus

$$T^{-k}[A] = \mathbb{R}^k \times A = A.$$

But $\mathbb{R}^k \times A$ is clearly in $\mathcal{A}(X_{k+1}, X_{k+2}, \dots)$. This means

$$A \in \bigcap_{k \in \mathbb{N}} \mathcal{A}(X_{k+1}, X_{k+2}, \dots) = \mathcal{T},$$

and so $\mathbb{P}[A] \in \{0, 1\}$ by Kolmogorov's 0–1 law. □

Remark. The shift is defined on a product space. If you don't have one: see exercise sheet.

Chapter 3

Weak convergence and Central Limit Theorem

Notation. $C_\infty(\mathbb{R})$ is the family of all the functions $\mathbb{R} \rightarrow \mathbb{R}$ which are continuous and bounded. $C_0(\mathbb{R})$ is the family of all the functions $\mathbb{R} \rightarrow \mathbb{R}$ which are continuous and such that $\lim_{x \rightarrow \pm\infty} = 0$.

Definition 3.1. Let P_n be probability measures on $(\mathbb{R}, \mathcal{B})$ and P another (probability¹) measure. We say that (P_n) *converges weakly to P* if

$$\int_{\mathbb{R}} f \, dP_n \rightarrow \int_{\mathbb{R}} f \, dP$$

for all $f \in C_\infty(\mathbb{R})$.

Examples. Let $P_n \sim N(0, \frac{1}{n^2})$. We want to show that $P_n \rightarrow \delta_0$ weakly, i.e. that

$$\int_{\mathbb{R}} f(x) \frac{n}{\sqrt{2\pi}} e^{-n^2 \frac{x^2}{2}} \, dx \rightarrow f(0)$$

for all $f \in C_\infty(\mathbb{R})$. This is true because

$$\int_{\mathbb{R}} |f(x) - f(0)| \frac{n}{\sqrt{2\pi}} e^{-n^2 \frac{x^2}{2}} \, dx \rightarrow 0.$$

We know by the Law of Large Numbers that $\frac{1}{n} \overline{X}_n - \mu \rightarrow 0$. Now we want to study the “speed” of convergence.

Definition 3.2. A sequence (X_n) of random variables *converges weakly to* a probability measure P (a random variable X) if $P^{X_n} \rightarrow P$ weakly (resp. $P^{X_n} \rightarrow P^X$ weakly).

¹Actually P doesn't have to be a *probability* measure, because if P_n converges weakly to P then taking $f := 1 \in C_\infty(\mathbb{R})$ we have $\int_{\mathbb{R}} f \, dP = 1$.

Recall that a probability measure on \mathbb{R} defines a distribution function $F_P(x) = \mathbb{P}((-\infty, x])$. By Exercise 15, also the converse holds, that is: there exists a unique probability measure P on $(\mathbb{R}, \mathcal{B})$ s.t. $P[(a, b]] = F(b) - F(a)$ for all $a < b$.

Lemma 3.3. If $F : \mathbb{R} \rightarrow \mathbb{R}$ is monotone increasing, then its set of discontinuities is at most countable.

Proof. Let $F(x^-), F(x^+)$ denote the left and right hand limits of F in x respectively. Let A be the set of points where F is not continuous. Then for any $x \in A$, there is a rational, say, $f(x)$ such that $F(x^-) < f(x) < F(x^+)$. For $x_1 < x_2$, we have that $F(x_1^+) \leq F(x_2^-)$. Thus $f(x_1) \neq f(x_2)$ if $x_1 \neq x_2$. Thus f is an injection $A \rightarrow \mathbb{Q}$. \square

Definition 3.4. Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a function. The *continuities* of F are the elements of

$$\text{Cont}(F) := \{x \mid F \text{ is continuous at } x\}.$$

Remark. $\text{Cont}(F)$ is dense for every function (trivial by last lemma).

Lemma 3.5 (Helly-Bray). Let F_n, F be distribution functions in the extended sense. If $F_n(x) \rightarrow F(x)$ for all $x \in \text{Cont}(F)$, then

$$\int_{(a,b]} g(x) dF_n(x) \rightarrow \int_{(a,b]} g(x) dF(x)$$

for all $g \in C(\mathbb{R})$ and $(a, b] \in \text{Cont}(F)$.

Remark. The presence of $(a, b]$ instead of $[a, b]$ in the statement is important. For example

$$\int_{\{a\}} g(x) dF(x) = g(a)(F(a^+) - F(a^-)) \underbrace{\neq 0}_{\text{if } g(a) \neq 0 \text{ and } F(a^+) \neq F(a^-)},$$

i.e. a singleton might not have measure zero. Of course this can't happen if $a \in \text{Cont}(F)$.

Proof. Every continuous function over a closed bounded interval is uniformly continuous². Thus g is uniformly continuous on every interval $[a, b]$, that is

$$\forall \varepsilon > 0 \exists \delta \forall x, y \in [a, b] \left(|x - y| < \delta \Rightarrow |g(x) - g(y)| < \frac{\varepsilon}{4} \right). \quad (*)$$

Since $\text{Cont}(F)$ is dense, we can find

$$a = x_0 < x_1 < \dots < x_k = b$$

²cfr. Heine-Cantor Theorem.

such that $x_i \in \text{Cont}(F)$ and $x_i - x_{i-1} < \delta$. Now define

$$R_n := \sum_{i=1}^k g(x_i)(F_n(x_i) - F_n(x_{i-1})) \quad \text{and} \quad R := \sum_{i=1}^k g(x_i)(F(x_i) - F(x_{i-1})).$$

We have

$$\begin{aligned} \left| \int_{(a,b]} g \, dF - R \right| &= \left| \sum_{i=1}^k \int_{(x_{i-1}, x_i]} (g(x) - g(x_i)) \, dF(x) \right| \\ &\leq \left| \sum_{i=1}^k \int_{(x_{i-1}, x_i]} (g(x) - g(x_i)) \, dF(x) \right| \quad (\text{by } (*)) \\ &= \sum_{i=1}^k \frac{\varepsilon}{4} (F(x_i) - F(x_{i-1})) \\ &= \frac{\varepsilon}{4} (F(b) - F(a)) \\ &\leq 1 \cdot \frac{\varepsilon}{4}. \quad (\text{because } F \text{ distr. fun. in the ext. sense}) \end{aligned}$$

In the same way, we can show that $|\int_{(a,b]} g \, dF_n - R_n| \leq \varepsilon/4$. Now observe that, since $x_i \in \text{Cont}(F)$,

$$|R_n - R| = \left| \sum_{i=1}^k g(x_i) \left(\underbrace{F_n(x_i) - F(x_{i-1})}_{\rightarrow 0} - \underbrace{(F_n(x_{i-1}) - F(x_{i-1}))}_{\rightarrow 0} \right) \right| \leq \underbrace{\varepsilon/2}_{\text{for } n \text{ big enough}}$$

So

$$\begin{aligned} \left| \int_{(a,b]} g \, dF_n - \int_{(a,b]} g \, dF \right| - |R_n - R| &\leq \left| \int_{(a,b]} g \, dF_n - R_n - \int_{(a,b]} g \, dF + R \right| \\ &\leq \left| \int_{(a,b]} g \, dF_n - R_n \right| + \left| \int_{(a,b]} g \, dF - R \right| \\ &\leq \frac{\varepsilon}{2}. \end{aligned}$$

Thus $\left| \int_{(a,b]} g \, dF_n - \int_{(a,b]} g \, dF \right| \leq \varepsilon/2 + |R_n - R| < \varepsilon$ for n big enough, and therefore

$$\int_{(a,b]} g \, dF_n \rightarrow \int_{(a,b]} g \, dF.$$

The proof is complete.

Observe also that if $a \in \text{Cont}(F)$

$$\int_{\{a\}} g(x) \, dF_n(x) = g(a)(F_n(a^+) - F_n(a^-)) \rightarrow 0.$$

□

Theorem 3.6. Let P_n, P be probability measures on \mathbb{R} and F_n, F their distribution functions. Then

$$P_n \rightarrow P \text{ weakly} \iff \forall x \in \text{Cont}(F) [F_n(x) \rightarrow F(x)].$$

Proof. “ \Rightarrow ”: Let $x \in \text{Cont}(F)$. Of course $F_n(x) = \int_{\mathbb{R}} \mathbb{1}_{(-\infty, x]} dP_n$. We replace $\mathbb{1}_{(-\infty, x]}$ by

$$g_\delta(y) := \begin{cases} 1, & y \leq x \\ \dots & x < y < x + \delta \\ 0, & y \geq x + \delta, \end{cases}$$

so that

$$\text{INSERIRE DISEGNO}$$

We define also $h_\delta(y)$ in this way:

$$\text{INSERIRE DISEGNO}$$

Now take $\varepsilon > 0$. Since $P_n \rightarrow P$ weakly, we have that

$$\left| \int_{\mathbb{R}} g_\delta dP_n - \int_{\mathbb{R}} g_\delta dP \right| < \frac{\varepsilon}{6}$$

for n big enough, and the same holds for h_δ . Observe also that, since $x \in \text{Cont}(F)$, there exists δ such that

$$|F(x \pm \delta) - F(x)| < \frac{\varepsilon}{6}. \quad (*)$$

Putting everything together we obtain

$$\begin{aligned} \int_x^{x+\delta} g_\delta dP_n &\leq \int_{x-\delta}^{x+\delta} h_\delta dP_n \\ &< \int_{x-\delta}^{x+\delta} h_\delta dP + \frac{\varepsilon}{2} \\ &\leq F(x + \delta) - F(x - \delta) + \frac{\varepsilon}{6} && \text{(because } h_\delta \leq 1) \\ &< \frac{\varepsilon}{2} && \text{(by } (*)) \end{aligned}$$

for n big enough. So

$$\begin{aligned} |F(x) - F_n(x)| &= \left| \int_{\mathbb{R}} g_\delta dP - \int_{\mathbb{R}} g_\delta dP_n - \int_x^{x+\delta} g_\delta dP + \int_x^{x+\delta} g_\delta dP_n \right| \\ &\leq \left| \int_{\mathbb{R}} g_\delta dP - \int_{\mathbb{R}} g_\delta dP_n \right| + \left| \int_x^{x+\delta} g_\delta dP \right| + \underbrace{\left| \int_x^{x+\delta} g_\delta dP_n \right|}_{\leq |F(x+\delta) - F(x)|} \\ &< \frac{\varepsilon}{6} + \frac{\varepsilon}{2} + \frac{\varepsilon}{6} < \varepsilon \end{aligned}$$

for n big enough.

“ \Leftarrow ”: Suppose $F_n(x) \rightarrow F(x)$ for all $x \in \text{Cont}(F)$. Let $g \in C_\infty(\mathbb{R})$, and suppose $|g(x)| \leq M$ for all $x \in \mathbb{R}$. Choose $c \in \text{Cont}(F)$ big enough that

$$F(-c) < \frac{\varepsilon}{8M} \quad \text{and} \quad 1 - F(c) < \frac{\varepsilon}{8M}.$$

By hypothesis it's possible to find n big enough that

$$|F_n(\pm c) - F(\mp c)| \leq \frac{\varepsilon}{8M}.$$

Now observe that

$$\begin{aligned} \left| \int_{\mathbb{R}} g \, dP_n - \int_{\mathbb{R}} g \, dP \right| &\leq \left| \int_{-c}^c g \, dP_n - \int_{-c}^c g \, dP \right| + \int_{|x| \geq c} |g| \, dP + \int_{|x| \geq c} |g| \, dP_n \\ &\leq \frac{\varepsilon}{8} + M(F_n(-c) + 1 - F_n(c^-)) + M(F(-c) + 1 - \underbrace{F(c^-)}_{=F(c)}) \\ &< \frac{\varepsilon}{8} + M(F_n(-c) + 1 - F_n(c^-)) + \frac{\varepsilon}{4} \\ &< \frac{\varepsilon}{8} + \frac{\varepsilon}{4} + \frac{\varepsilon}{8} + \frac{\varepsilon}{4} < \varepsilon. \end{aligned}$$

The proof is complete.

The following observation will be useful: if F is a distribution function in the extended sense, everything works for $g \in C_0(\mathbb{R})$ (the limit measure is a sub-probability-measure). \square

Definition 3.7. Let P_n be finite measures on $(\mathbb{R}, \mathcal{B})$ and P another measure. We say that (P_n) *converges vaguely* to P if

$$\int_{\mathbb{R}} f \, dP_n \rightarrow \int_{\mathbb{R}} f \, dP$$

for all $f \in C_0(\mathbb{R})$ such that $\lim_{|x| \rightarrow \infty} f(x) = 0$.

Remark. Of course if $P_n \rightarrow P$ weakly then $P_n \rightarrow P$ vaguely.

Theorem 3.8. Let P_n, P be probability measures on $(\mathbb{R}, \mathcal{B})$.

1. If $P_n(B) \rightarrow P(B)$ for all $B \in \mathcal{B}$, then $P_n \rightarrow P$ weakly.
2. Let θ be a σ -finite measure on $(\mathbb{R}, \mathcal{B})$. Suppose that

$$P_n(dx) = f_n(x)d\theta(x) \quad \text{and} \quad P(dx) = f(x)d\theta(x),$$

that is

$$P_n(B) = \int_{\mathbb{R}} f \, d\theta \quad \text{and} \quad P(B) = \int_{\mathbb{R}} f \, d\theta.$$

If $f_n \rightarrow f$ σ -almost everywhere, then $P_n \rightarrow P$ weakly.

Proof. No proof. □

Theorem 3.9 (Helly's theorem). Let F_n be a sequence of distribution functions on \mathbb{R} in the extended sense. Then there exists a distribution function (in the extended sense) F and a subsequence F_{n_k} such that

$$F_{n_k}(x) \rightarrow F(x)$$

for all $x \in \text{Cont}(F)$.

Proof. We split the proof in two parts.

1. Let D be a dense countable set (e.g. \mathbb{Q}). Of course $(F_n(x))_{n \in \mathbb{N}}$ is a bounded sequence for all $x \in D$. (??? non capisco come procede...chiedere)
So there exist $(n_k)_{k \in \mathbb{N}}$ and $F_0: D \rightarrow \mathbb{R}$ such that

$$F_{n_k}(x) \rightarrow F_0(x)$$

for all $x \in D$.

2. Suppose $F_n(x) \rightarrow F_0(x)$ for all $x \in D$. We now want to show that we can find an F from F_0 such that the property is preserved. First observe that clearly on D we have $0 \leq F_0 \leq 1$ and F_0 is monotone increasing.

- (a) Since D is dense, for all $x \in \mathbb{R}$ there is a sequence (x_k) in D such that $x_k \searrow x$ strictly. We define

$$F(x) := \lim_{k \rightarrow \infty} F_0(x_k).$$

(??? perché esiste? credo sia perché F_0 è chiaramente anche continua a destra su D . In caso aggiungere.)

- (b) F is well define. Indeed, take two sequences in D such that $x_k \searrow x$ and $y_l \searrow x$ strictly. Now fix l . Of course there exists $k(l)$ such that $x_{k(l)} < y_l$. By monotonicity we get

$$\begin{aligned} F_0(x_k) &\leq F_0(y_l) && \text{for all } k \geq k(l) \\ \Rightarrow \lim_{k \rightarrow \infty} F_0(x_k) &\leq F_0(y_l) && \text{for all } l \in \mathbb{N} \\ \Rightarrow \lim_{k \rightarrow \infty} F_0(x_k) &\leq \lim_{l \rightarrow \infty} F_0(y_l). \end{aligned}$$

The other inequality follows by the very same argument.

- (c) We have that $0 \leq F \leq 1$, and F is monotone increasing. Furthermore, $F(-\infty) = 0$ (exercise).

- (d) If $x < y < z$ with $x, z \in D$, then $F_0(x) \leq F(y) \leq F_0(z)$. Indeed, take (y_k) in D such that $y_k \searrow y$. Then $x < y_k < z$ for all k big enough. Since F_0 is monotone, we get $F_0(x) \leq F_0(y_k) \leq F_0(z)$ for all k big enough, and because $F_0(y_k) \rightarrow F(y)$ the assertion follows.
- (e) F is right-continuous. Indeed, let $x \in \mathbb{R}$ and let (x_n) be a sequence in \mathbb{R} such that $x_n \searrow x$ strictly. Then, since D is dense, there exists a sequence (y_n) in D such that $y_n \searrow x$ strictly and $y_n > x_n$ for all $n \in \mathbb{N}$. By monotonicity and by point (d) we obtain

$$F(x) \leq F(x_n) \leq F_0(y_n) \rightarrow F(x).$$

- (f) If $x \in \text{Cont}(F)$, (x_n) is a sequence in D and $x_n \nearrow x$ strictly, then $F_0(x_n) \nearrow F(x)$. Indeed, since $x_n - 2^{-n} < x_n < x$, (almost) by point (d) we get

$$\underbrace{F(x_n - 2^{-n})}_{\rightarrow F(x) \text{ by continuity}} \leq F_0(x_n) \leq F(x).$$

- (g) Let $x \in \text{Cont}(F)$. Then we can find sequences (x_k) and (y_k) in D such that $x_k < x_{k+1} < x < y_{k+1} < y_k$ and $x_k \rightarrow x$ and $y_k \rightarrow x$. Then

$$F_n(x_k) \leq F_n(x) \leq F_n(y_k),$$

and so, taking the limit $n \rightarrow \infty$,

$$F_0(x_k) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F_0(y_k),$$

and so, taking the limit $k \rightarrow \infty$,

$$F(x) \leq \liminf_n F_n(x) \leq \limsup_n F_n(x) \leq F(x),$$

which finally means $\lim_{n \rightarrow \infty} F_n(x) = F(x)$.

□

3.1 Characteristic functions

Definition 3.10. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space and let X be an a.s. finite random variable, P its distribution on \mathbb{R} and F the associated distribution function (in the extended sense). The associated *characteristic function* $\varphi_X: \mathbb{R} \rightarrow \mathbb{C}$ (also called φ_P or φ_F) is defined by

$$\varphi_X(t) := \mathbb{E} [e^{itX}] = \int_{\Omega} e^{itX} d\mathbb{P} = \int_{\mathbb{R}} e^{itx} dP(x) = \int_{\mathbb{R}} e^{itx} dF(x).$$

Remark. Clearly, if $X_n \rightarrow P$ weakly, we have that

$$\mathbb{E}[e^{itX_n}] = \int_{\mathbb{R}} e^{itx} dP_{X_n}(x) \xrightarrow{3} \int_{\mathbb{R}} e^{itx} dP(x) = \mathbb{E}[e^{itX}],$$

i.e. $\varphi_{P_{X_n}}(t) \rightarrow \varphi_P(t)$ for all $t \in \mathbb{R}$.

3.1.1 Properties of characteristic functions

Definition 3.11. The first derivative of $\varphi(t)$ is defined as

$$\varphi'(t) := \lim_{h \rightarrow 0} \frac{\varphi(t+h) - \varphi(t)}{h}$$

if the limit exists. We also defined for all $k \in \mathbb{N}$ the k -th moment of X as

$$m_k := m_k(X) := \mathbb{E}[X^k]$$

if X^k is integrable. Sometimes we write $m_k(P)$ or $m_k(F)$ instead of $m_k(X)$. We also defined for all $k \in \mathbb{N}$ the k -th absolute moment of X as

$$|m|_k := |m|_k(X) := \mathbb{E}[|X|^k]$$

if X^k is integrable.

Remark. Of course $m_k(F) = \int_{\mathbb{R}} x^k dF_k(x)$. Furthermore, if $m_k < \infty$ then $m_l < \infty$ for all $l \leq k$ (??? perché? non c'è qualche problema con indici pari/dispari?).

Moreover, if F is a distribution function in a generalized sense, then trivially $m_0(F) = F(\infty) - F(-\infty) = F(\infty) - 0 = F(\infty) \leq 1$.

Theorem 3.12. Let X be a random variable. Call $\varphi := \varphi_X$. We have:

1. $|\varphi(t)| \leq 1$, $\varphi(0) = 1$.
2. φ is uniformly continuous.
3. If m_k exists, then φ is k times differentiable and

$$\varphi^{(l)}(0) = i^l m_l$$

for all $l = 0, \dots, k$.

4. If $\varphi^{(2n)}(0)$ exists, then m_{2n} is finite.

³Recall that $\int_A f(x) + ig(x) d\mu := \int_A f(x) d\mu + i \int_A g(x) d\mu$. Since $e^{itx} = \cos(tx) + i \sin(tx)$ and of course $\cos, \sin \in C_{\infty}(\mathbb{R})$, everything works.

Proof.

1. Of course $|e^{itx} = 1|$, so by complex analysis we know that $|\mathbb{E}[e^{itx}]| \leq 1$.
2. Observe that

$$\varphi(t+h) - \varphi(t) = \int_{\mathbb{R}} (e^{i(t+h)x} - e^{itx}) dF(x) = \int_{\mathbb{R}} e^{itx} (e^{ihx} - 1) dF(x).$$

So

$$|\varphi(t+h) - \varphi(t)| = \int_{\mathbb{R}} \underbrace{|e^{itx}(e^{ihx} - 1)|}_{\leq 2 \text{ (thus integrable)}} dF(x) \xrightarrow{h \rightarrow 0} 0$$

by Dominated convergence. Since the t has disappeared, φ is uniformly continuous (??? pensarci e poi in caso chiedere meglio).

3. If we show the following we are done:

$$\varphi^{(l)}(t) = \int_{\mathbb{R}} (ix)^l e^{itx} dF(x).$$

The proof proceeds by induction on l . Suppose $l = 1$. Then

$$\frac{\varphi(t+h) - \varphi(t)}{h} = \int_{\mathbb{R}} e^{itx} \frac{e^{ihx} - 1}{h} dF(x).$$

Observe that

$$\begin{aligned} \left| \frac{e^{ihx} - 1}{h} \right| &= \left| \frac{\cos(hx) - 1}{h} + i \frac{\sin(hx) - \sin(0h)}{h} \right| \\ &= | -x \sin(S_1 x) + i \cos(S_2 x) | \quad (\text{by Mean value theorem}) \\ &\leq 2|x|, \end{aligned}$$

which is integrable since $|m|_1$ exists. So by Dominated convergence we get

$$\lim_{h \rightarrow 0} \frac{\varphi(t+h) - \varphi(t)}{h} = \int_{\mathbb{R}} e^{itx} ix dF(x).$$

The inductive case is very similar.

4. For any $f: \mathbb{R} \rightarrow \mathbb{R}$, we define the *difference operator*

$$\Delta_h f(t) := f(t+h) - f(t-h),$$

and so we define recursively

$$\Delta_h^k f(t) := \Delta_h(\Delta_h^{k-1} f)(t).$$

So

$$f^{(2n)}(0) = \lim_{h \rightarrow 0} \frac{\Delta_h^{2n} f(0)}{(2h)^{2n}}.$$

Now take $f(t) := e^{ixt}$ (with x fixed). It's easy to check that

$$\Delta_h f(t) = e^{ixt}(e^{ixh} - e^{-ixh}) \quad \text{and} \quad \Delta_h^k f(t) = e^{ixt}(e^{ixh} - e^{-ixh})^k.$$

Observe that the last factor doesn't depend on t . So

$$\varphi^{(2n)}(0) = \lim_{h \rightarrow 0} \int_{\mathbb{R}} \frac{1}{(2h)^{2n}} e^{ix0} (2i \sin(hx))^{2n} dF(x) = \lim_{h \rightarrow 0} \int_{\mathbb{R}} \left(\frac{i \sin(hx)}{h} \right)^{2n} dF(x).$$

Now define

$$f_h(x) := \begin{cases} \left(\frac{\sin(hx)}{h} \right)^{2n}, & \text{if } h \neq 0 \\ x^{2n}, & \text{if } h = 0. \end{cases}$$

We have

$$\begin{aligned} |\varphi^{(2n)}(0)| &= \liminf_{h \rightarrow 0} \int_{\mathbb{R}} f_h(x) dF(x) \\ &\geq \int_{\mathbb{R}} \liminf_{h \rightarrow 0} f_h(x) dF(x) && \text{(by Fatou lemma)} \\ &= \int_{\mathbb{R}} \lim_{h \rightarrow 0} f_h(x) dF(x) && \text{(since the limit exists)} \\ &= \int_{\mathbb{R}} x^{2n} dF(x) = m_{2n}. \end{aligned}$$

This implies $m_{2n} \leq \infty$. Now apply point (c) and we are done. \square

So, if $|m|_n < \infty$, then m_k is finite for all $k \leq n$ and

•

$$\begin{aligned} \varphi(t) &= \sum_{k=0}^{n-1} m_k \frac{(it)^k}{k!} + \underbrace{t^n \int_0^1 \frac{(1-u)^{n-1}}{(n-1)!} \varphi^{(n)}(tu) du}_{\text{error term}} && \text{(Taylor's series)} \\ &= \sum_{k=0}^n m_k \frac{(it)^k}{k!} + \underbrace{t^n \int_0^1 \frac{(1-u)^{n-1}}{(n-1)!} (\varphi^{(n)}(tu) - \varphi^{(n)}(0)) du}_{=: E_n}. \end{aligned}$$

- $|\varphi^{(n)}(t)| \leq |m|_n$, because $\varphi^{(l)}(t) = \int_{\mathbb{R}} (ix)^l e^{itx} dF(x)$.

- $|E_n| \leq 2 \frac{|m|_n}{n!}$.

Suppose now that all moments exist. If

$$|t| <^4 g = \frac{1}{\limsup_{k \rightarrow \infty} \left(\frac{|m|_k}{k!} \right)^{1/k}}, \quad (\text{radius of convergence})$$

then $t^n E_n \rightarrow 0$ and $\varphi(t) = \sum_{k=0}^{\infty} m_k \frac{(it)^k}{k!}$.

Corollary 3.13. If $\limsup_{k \rightarrow \infty} \left(\frac{|m|_k}{k!} \right)^{1/k} = 0$, then

$$\varphi(t) = \sum_{k=0}^{\infty} m_k \frac{(it)^k}{k!}$$

for all $t \in \mathbb{R}$.

Examples. One important application of the last corollary is the computation of the moments of the normal distribution (see exercises sheet).

Theorem 3.14 (Inversion Theorem). Let $F(x)$ be a distribution function in the extended sense and $\varphi(t)$ its characteristic function. Then, for all $a, b \in \text{Cont}(F)$ such that $a < b$ we have

$$F(b) - F(a) = \lim_{c \rightarrow \infty} \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt.$$

Proof. First of all, observe that the integrand function has no pole, since it is equal to $b - a$ in $t = 0$ (just apply l'Hopital's rule). Recall that

$$\lim_{c \rightarrow \infty} \int_{-c}^c \frac{\sin t}{t} dt = \pi \quad \text{and} \quad \lim_{c \rightarrow \infty} \int_0^c \frac{\sin t}{t} dt = \frac{\pi}{2}.$$

Now define

$$I_c(a, b) := \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \varphi(t) dt = \frac{1}{2\pi} \int_{-c}^c \frac{e^{-ita} - e^{-itb}}{it} \int_{\mathbb{R}} e^{itx} dF(x) dt$$

Observe that

$$g(t, x) := \frac{1}{2\pi} \frac{e^{-ita} - e^{-itb}}{it} e^{itx}$$

⁴replace $|m_k|$ by $|m|_k$.

is bounded and continuous. Applying Fubini's Theorem on the product measure $\lambda_{[-c,c]} \times P_F$ we get

$$\begin{aligned}
I_c(a, b) &= \frac{1}{2\pi} \int_{\mathbb{R}} \int_{-c}^c \frac{e^{-it(x-a)} - e^{-it(x-b)}}{it} dt dF(x) \\
&= \frac{1}{2\pi} \int_{\mathbb{R}} \left[\underbrace{\int_{-c}^c \frac{\cos t(x-a) - \cos t(x-b)}{it} dt}_{\text{even function}} + \underbrace{\int_{-c}^c \frac{\sin t(x-a) - \sin t(x-b)}{t} dt}_{\text{odd function}} \right] dF(x) \\
&= \frac{1}{\pi} \int_{\mathbb{R}} \left[\int_0^c \frac{\sin t(x-a)}{t} dt \right] dF(x) - \frac{1}{\pi} \int_{\mathbb{R}} \left[\int_0^c \frac{\sin t(x-b)}{t} dt \right] dF(x) \\
&\stackrel{5}{=} \frac{1}{\pi} \int_{\mathbb{R}} \underbrace{\int_{c(x-b)}^{c(x-a)} \frac{\sin u}{u} du}_{=: f_c(x)} dF(x).
\end{aligned}$$

Observe that

$$f_c(x) \xrightarrow{c \rightarrow \infty} \begin{cases} 0, & x < a \vee x > b; \\ \pi, & a < x < b; \\ \frac{\pi}{2}, & x = a \vee x = b. \end{cases}$$

Therefore

$$\begin{aligned}
\lim_{c \rightarrow \infty} I_c(a, b) &= \lim_{c \rightarrow \infty} \frac{1}{\pi} \int_{\mathbb{R}} f_c(x) dF(x) \\
&\stackrel{6}{=} \frac{1}{\pi} \int_{\mathbb{R}} \lim_{c \rightarrow \infty} f_c(x) dF(x) \\
&= \frac{1}{\pi} \left[\pi(F(b^-) - F(a)) + \frac{\pi}{2}(F(a) - F(a^-)) + \frac{\pi}{2}(F(b) - F(b^-)) \right] \\
&= F(b) - F(a). \quad (\text{since } a, b \in \text{Cont}(F))
\end{aligned}$$

□

Examples (important). Let (X_n) be a sequence of iid random variables with $\mathbb{E}[X_n] = \mu$, $\text{Var}[X_n] = \sigma^2$, and so $\text{Var}[\bar{X}_n] = \frac{1}{n^2} \text{Var}[X_1 + \dots + X_n]$. We want to consider

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}.$$

⁵We use the substitution $u := t(x-a)$, and so $dt = \frac{du}{x-a}$. We do the same with $t(x-b)$.

⁶ $h(\xi) := \int_0^\xi \frac{\sin t}{t} dt$ is continuous and $\lim_{\xi \rightarrow \infty} h(\xi) = \frac{\pi}{2}$. So $h(\xi)$ is bounded, let's say $|h(\xi)| \leq M$. Thus $|f_c(x)| = |h(\xi_1) + h(\xi_2)| \leq 2M$ is bounded and hence integrable. Hence we can apply Dominated convergence.

W.l.o.g. $\mu = 0$ (otherwise replace X_n by $X_n - \mu$ in the definition of \bar{X}_n). We have

$$\varphi(t) := \varphi_{X_n}(t) \stackrel{\text{Taylor}}{=} m_0 + m_1 \frac{it}{1!} + m_2 \frac{(it)^2}{2!} + t^2 E_2(t) = 1 - \sigma^2 \frac{t^2}{2} + t^2 E_2(t), \quad (*)$$

where $E_2(t)$ is as above, and so

$$|E_2| = \left| \int_0^1 \frac{1-u}{1!} [\varphi''(tu) - \varphi''(0)] du \right| \leq \frac{1}{2} \sup_{\substack{0 \leq u \leq t \\ (\text{w.l.o.g. } t \geq 0)}} |\varphi''(u) - \varphi''(0)|.$$

So

$$\begin{aligned} \varphi_{\bar{X}_n/(\sigma/\sqrt{n})}(t) &= \varphi_{\sqrt{n} \frac{\bar{X}_n}{\sigma}}(t) \\ &= \varphi_{\frac{1}{\sqrt{n}\sigma}(X_1 + \dots + X_n)}(t) \\ &= \left(\varphi_{\frac{X_1}{\sqrt{n}\sigma}}(t) \right)^n \\ &= \varphi \left(\frac{t}{\sqrt{n}\sigma} \right)^n \quad (\varphi_{aX}(t) = \mathbb{E}[e^{itax}] = \varphi_X(at)) \\ &= \left(1 - \frac{\sigma^2 t^2}{2n\sigma^2} + \frac{t^2}{n\sigma^2} E_2 \left(\frac{t}{\sqrt{n}\sigma} \right) \right)^n \quad (\text{by } (*)) \\ &\rightarrow e^{-\frac{t^2}{2}} \text{ for } n \rightarrow \infty, \end{aligned}$$

where the last asymptotic evaluation holds because $\frac{t}{\sqrt{n}\sigma} \rightarrow 0$ and thus $\frac{t^2}{n\sigma^2} E_2 \left(\frac{t}{\sqrt{n}\sigma} \right) = O(\frac{t}{\sqrt{n}\sigma})$ for $n \rightarrow \infty$. Finally, recall that $(1 + \frac{x}{n})^n \rightarrow e^x$. Therefore

$$\varphi_{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}(t) \xrightarrow{n \rightarrow \infty} e^{-\frac{t^2}{2}} = \varphi_{N(0,1)}(t),$$

where the last equality follows by the exercises.

Theorem 3.15 (Continuity Theorem). Let $F_n(x)$ be a sequence of distribution functions in the strict sense. Then

1. If $P_{F_n} \rightarrow P$ (probability on \mathbb{R}) weakly (and thus $F_n(x) \rightarrow F(x)$ for all $x \in \text{Cont}(F)$, where F is distribution function in the strict sense), then

$$\varphi_{F_n}(t) \rightarrow \varphi_F(t) = \varphi_P(t),$$

where P is the distribution associated to F .

2. If $\varphi_n := \varphi_{F_n}$ converges pointwise to a function φ which is continuous in $t = 0$, then φ is the characteristic function of a (strict) distribution function F , and

$$F_n(x) \rightarrow F(x)$$

for all $x \in \text{Cont}(F)$.

Proof. The statement (a) is immediate. Let's prove (b). Observe that $\varphi_n(t)$ is integrable, since the real and the imaginary parts are both bounded and continuous. So, by Dominated convergence,

$$\int_0^u \varphi_n(t) dt \rightarrow \int_0^u \varphi(t) dt \quad ((*))$$

for all $u \in \mathbb{R}$. Now

$$\int_0^u \varphi_n(t) dt = \int_0^u \int_{\mathbb{R}} e^{itx} dF_n(x) dt = \int_{\mathbb{R}} \frac{e^{iux} - 1}{ix} dF_n(x),$$

where the last equality holds by Fubini's theorem, since everything is bounded and continuous. Now observe that the integrand function goes to u for $x \rightarrow 0$, and so it's in $C_0(\mathbb{R})$. Now apply Helly's Theorem: let $(n_k)_{k \in \mathbb{N}}$ be a sequence and F a distribution function in the *extended* sense such that

$$F_{n_k} \rightarrow F(x)$$

for all $x \in \text{Cont}(F)$. Then, by weak convergence we get

$$\int_0^u \varphi_{n_k}(t) dt = \int_{\mathbb{R}} \frac{e^{iux} - 1}{ix} dF_{n_k}(x) \rightarrow \int_{\mathbb{R}} \frac{e^{iux} - 1}{ix} dF(x) = \int_0^u \varphi_F(t) dt =: h_2(u)$$

and also, by (*),

$$\int_0^u \varphi_{n_k}(t) dt \int_0^u \varphi(t) dt =: h_1(u).$$

So $h_1(u) = h_2(u)$ for all $u \in \mathbb{R}$. Now observe that φ_F is continuous everywhere on \mathbb{R} because it's a characteristic function. So $h_2'(t) = \varphi_F(t)$. Furthermore, φ is continuous at 0, so $h_1'(0) = \varphi(0)$. Hence

$$\varphi_F(0) = \int_{\mathbb{R}} e^{i0x} dF(x) = \int_{\mathbb{R}} 1 dF(x) = F(+\infty) - F(-\infty) = F(+\infty)$$

is equal to

$$\lim_{n \rightarrow \infty} \varphi_n(0) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} e^{i0x} dP(x) = \lim_{n \rightarrow \infty} 1 = 1,$$

so $F(+\infty) = 1$, i.e. F is a distribution function in the strict sense. Moreover, $\varphi_F(t) = h'_1(t)$, which depends only on φ .

Summing up: for every subsequence (F_{n_k}) which converges to a function F on $\text{Cont}(F)$, F is the distribution function on a probability measure and $\varphi_F(t) = h'_1(t)$. By the inversion theorem, we have that F is determined by $h'_1(\bullet)$, so F is determined by $\varphi(\bullet)$.

Claim: $F_n(x) \rightarrow F(x)$ for all $x \in \text{Cont}(F)$.

Proof: Assume to the contrary that $F(x_0) \not\rightarrow F(x_0)$ for some $x_0 \in \text{Cont}(F)$. This means that there exists a subsequence (n_k) such that

$$F_{n_k}(x_0) \rightarrow y_0 \neq F(x_0).$$

Thus, by Helly's theorem, there exist a (sub-)subsequence (n_{k_l}) and a distribution function (in the extended sense) G such that

$$F_{n_{k_l}}(x) \rightarrow G(x)$$

for all $x \in \text{Cont}(G)$. But since $\varphi_G = h'_1 = \varphi_F$, we get $G = F$, i.e. $F_{n_{k_l}}(x_0) \rightarrow F(x_0)$, contradiction. ■

Conclusion: F is strict and $F_n(x) \rightarrow F(x)$ for all $x \in \text{Cont}(F)$. Observe that this also means (a posteriori) that $\varphi = \varphi_F$. □

Theorem 3.16 (Central Limit Theorem). Let (X_n) be a sequence of iid random variables with expected value μ and finite variance σ^2 . Define $\bar{X}_n := \frac{1}{n}(X_1 + \dots + X_n)$. Then

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \rightarrow N(0, 1) \text{ weakly.}$$

Proof. Call $Y_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$. We know that

$$\varphi_{Y_n}(t) \rightarrow e^{-\frac{t^2}{2}} = \varphi_{N(0,1)}(t).$$

Apply point (b) of Continuity Theorem and we are done. □

What follows is historically the oldest version of the Central Limit Theorem.

Remark (De Moivre - Laplace, 1718). Let (X_n) be iid, $X_n \sim B(1, \theta)$, $S_n := X_1 + \dots + X_n \sim B(n, \theta)$. By the exercise sheets we know that

$$\varphi_{S_n}(t) = (\varphi_{X_1}(t))^n = (1 - \theta + \theta e^{it})^n$$

and $\mathbb{E}[S_n] = n\theta$, $\text{Var}[S_n] = n\theta(1 - \theta)$. Now:

$$Y_n := \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{S_n - n\theta}{\sqrt{n\theta(1 - \theta)}} = \frac{\bar{X}_n - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}.$$

Recall that $\varphi_{aX+b}(t) = e^{itb}\varphi_X(at)$. We obtain

$$\begin{aligned}
\varphi_{Y_n}(t) &= e^{-it\frac{n\theta}{\sqrt{n\theta(1-\theta)}}} \left(1 - \theta + \theta e^{i\frac{t}{\sqrt{n\theta(1-\theta)}}} \right)^n \\
&= \left((1 - \theta) e^{-it\frac{1}{\sqrt{n}}\sqrt{\frac{\theta}{1-\theta}}} + \theta e^{it\frac{1}{\sqrt{n}}\sqrt{\frac{\theta}{1-\theta}}} \right)^n \\
&= \left(\sum_{k=0}^{\infty} \frac{1}{k!} \left(\frac{it}{\sqrt{n}} \right)^k \left[(1 - \theta) \left(\sqrt{\frac{\theta}{1-\theta}} \right)^k (-1)^k + \theta \left(\sqrt{\frac{1-\theta}{\theta}} \right)^k \right] \right)^n \\
&= \left(1 - 0t - \frac{t^2}{2n} \cdot 1 + O\left(\frac{t^3}{n^{3/2}}\right) \right)^n \rightarrow e^{-\frac{t^2}{2}}.
\end{aligned}$$

Chapter 4

Conditional expectation and Martingales

4.1 Conditional expectation on atomic algebras

Definition 4.1. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space. A sub- σ -algebra $\mathcal{F} \subseteq \mathcal{A}$ is called *atomic* if it has at most countably many smallest nonempty sets. More precisely: $\Omega = \bigsqcup_{j \in J} C_j$, where $C_j \in \mathcal{A}$, J is countable and $\mathcal{F} = \sigma(\{C_j : j \in J\})$, i.e.

$$\mathcal{F} = \left\{ \bigsqcup_{j \in I} C_j : I \subseteq J \right\}.$$

Remark.

- Let Y be a random variable, and suppose that the value set $\{y_j : j \in J\} \subseteq \mathbb{R}$ is at most countable. Define $C_j := Y^{-1}[\{y_j\}] = [Y = y_j]$ and $\mathcal{F} := \sigma(Y)$. Of course \mathcal{F} is atomic.
- If \mathcal{F} is atomic, then $f : (\Omega, \mathcal{F}) \rightarrow (\Omega', \mathcal{A}')$ is \mathcal{F} -measurable iff it is constant on each C_j (??? secondo me “ \Rightarrow ” non vale in generale...dipende da come è fatta \mathcal{A}').

Let $X : (\Omega, \mathcal{A}, \mathbb{R}) \rightarrow (\mathbb{R}, \mathcal{B})$ an integrable random variable (not necessarily \mathcal{F} -measurable) and let $\mathcal{F} \subseteq \mathcal{A}$ be atomic. We look for a “simplified version” of X which is \mathcal{F} -measurable, i.e. we want Y \mathcal{F} -measurable such that

$$\forall F \in \mathcal{F} : \int_F X \, d\mathbb{P} = \int_F Y \, d\mathbb{P},$$

which is equivalent to

$$\forall F \in \mathcal{F} : \mathbb{E}[X \mathbb{1}_F] = \mathbb{E}[Y \mathbb{1}_F],$$

which, since \mathcal{F} is atomic, is equivalent to

$$\text{for all atoms } C_j \in \mathcal{F} : \mathbb{E}[X \mathbb{1}_{C_j}] = \mathbb{E}[Y \mathbb{1}_{C_j}],$$

and also $Y|_{C_j} = y_j$ is constant on C_j . How should we take y_j ? Observe that we are asking for the following to hold:

$$\mathbb{E}[X \mathbb{1}_{C_j}] = \mathbb{E}[y_j \mathbb{1}_{C_j}] = y_j \cdot \mathbb{P}[C_j],$$

so

$$y_j = \begin{cases} \frac{\mathbb{E}[X \mathbb{1}_{C_j}]}{\mathbb{P}[C_j]}, & \mathbb{P}[C_j] > 0 \\ k, & \mathbb{P}[C_j] = 0, \end{cases}$$

where k is arbitrary, but the preferred choice is $k = 0$. Observe that intuitively y_j turns out to be the local average of X over the block C_j .

Hence our conditions determine Y uniquely a.s.

Examples. If $X = \mathbb{1}_A$, where $A \in \mathcal{A}$, then

$$y_j = \mathbb{P}[A|C_j] = \frac{\mathbb{P}[A \cap C_j]}{\mathbb{P}[C_j]}.$$

Definition 4.2. Let Y be as above. Y is called *conditional expectation of X w.r.t. \mathcal{F}* . We denote it by $\mathbb{E}[X|\mathcal{F}]$. Keep in mind that it is determined only a.s.

Proposition 4.3. The following hold:

1. X \mathcal{F} -measurable $\Rightarrow \mathbb{E}[X|\mathcal{F}] = X$. In particular, $\mathbb{E}[k|\mathcal{F}] = k$ for all k constant.
2. $X \geq 0$ a.s. $\Rightarrow \mathbb{E}[X|\mathcal{F}] \geq 0$ a.s.
3. $\mathbb{E}[\lambda_1 X_1 + \lambda_2 X_2 | \mathcal{F}] = \lambda_1 \mathbb{E}[X_1|\mathcal{F}] + \lambda_2 \mathbb{E}[X_2|\mathcal{F}]$ a.s.
4. If W is an \mathcal{F} -measurable random variable such that $W \cdot X$ is integrable, then $\mathbb{E}[W \cdot X | \mathcal{F}] = W \cdot \mathbb{E}[X|\mathcal{F}]$.
5. Let $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \mathcal{A}$ be two atomic sub- σ -algebras. Then $\mathbb{E}[X|\mathcal{F}_1] = \mathbb{E}[\mathbb{E}[X|\mathcal{F}_2]|\mathcal{F}_1]$.
6. If X is independent of \mathcal{F} (i.e. $\sigma(X)$ is independent of \mathcal{F}), then $\mathbb{E}[X|\mathcal{F}] = \mathbb{E}[X]$.
7. If $0 \leq X_n \nearrow X$, then $\mathbb{E}[X_n|\mathcal{F}] \nearrow \mathbb{E}[X|\mathcal{F}]$.
8. **Jensen's inequality.** If $g: \mathbb{R} \rightarrow \mathbb{R}$ is convex and $\mathbb{E}[g(X)]$ is finite, then $g(\mathbb{E}[X|\mathcal{F}]) \leq \mathbb{E}[g(X)|\mathcal{F}]$. (observe that a trivial example is given by $\mathcal{F} = \{\emptyset, \Omega\}$ and $g(x) = x^2$)

Proof. (1),(2) and (3) are clear.

4. Call $Y := \mathbb{E}[WX|\mathcal{F}]$. First observe that WY is \mathcal{F} -measurable (pensarci), so the equality actually makes sense. To prove that it holds, we follow the steps of the construction of the integral:

- (a) Suppose $W = \mathbb{1}_E$ for some $E \in \mathcal{F}$. Then, for all $F \in \mathcal{F}$, we have

$$\begin{aligned} \int_F WX \, d\mathbb{P} &= \int_F \mathbb{1}_E X \, d\mathbb{P} = \int \mathbb{1}_{E \cap F} X \, d\mathbb{P} = \int_{E \cap F} X \, d\mathbb{P} = \\ &= \int_{E \cap F} Y \, d\mathbb{P} = \int_F \mathbb{1}_E Y \, d\mathbb{P} = \int_F WY \, d\mathbb{P}, \end{aligned}$$

$$\text{i.e. } \mathbb{E}[WX|\mathcal{F}] = W \mathbb{E}[X|\mathcal{F}].$$

- (b) Suppose $W = \sum_{j \in J_0} w_j \mathbb{1}_{E_j}$ with J_0 finite. Then

$$\begin{aligned} \mathbb{E}[WX|\mathcal{F}] &= \mathbb{E} \left[\sum_{j \in J_0} w_j \mathbb{1}_{E_j} X | \mathcal{F} \right] \\ &= \sum_{j \in J_0} w_j \mathbb{E}[\mathbb{1}_{E_j} X | \mathcal{F}] && \text{(by (3))} \\ &= \sum_{j \in J_0} w_j \mathbb{1}_{E_j} \mathbb{E}[X | \mathcal{F}] && \text{(by (a))} \\ &= W \mathbb{E}[X | \mathcal{F}]. \end{aligned}$$

- (c) As usual, suppose $W \geq 0$ and find a sequence $0 \leq W_n \nearrow W$. Then use monotone convergence.

- (d) Write $W = W^+ - W^-$ and proceed as always.

5. Later.

6. Write $\Omega = \bigsqcup_{j \in J} C_j$ and call $Y := \mathbb{E}[X|\mathcal{F}]$. By definition of Y , $Y = y_i$ on C_i . Observe that since X and \mathcal{F} are independent, X and $\mathbb{1}_F$ are independent random variables for all $F \in \mathcal{F}$. So

$$y_j = \frac{\mathbb{E}[X \mathbb{1}_{C_j}]}{\mathbb{P}[C_j]} = \frac{\mathbb{E}[X] \mathbb{E}[\mathbb{1}_{C_j}]}{\mathbb{P}[C_j]} = \mathbb{E}[X] \frac{\mathbb{P}[C_j]}{\mathbb{P}[C_j]} = \mathbb{E}[X]$$

7. Easy.

8. A convex function is not necessarily differentiable, but $g'(a^-)$ exists for all $a \in \mathbb{R}$.

□

4.2 Conditional expectation on general algebras

So far, we defined the conditional expectation w.r.t. atomic σ -algebras only. We now want to extend the definition to any σ -algebra. Furthermore, $g(x) \geq g(a) + g'(a^-)(x - a)$ (??? perché?). Now take $a := \mathbb{E}[X|\mathcal{F}]$. We get

$$g(X) \geq g(\mathbb{E}[X|\mathcal{F}]) + g'(\mathbb{E}[X|\mathcal{F}]^-(X - \mathbb{E}[X|\mathcal{F}])).$$

Observe that $g(\mathbb{E}[X|\mathcal{F}])$ and $g'(\mathbb{E}[X|\mathcal{F}]^-)$ are \mathcal{F} -measurable, while X might not be. So, by (b) and (d) we obtain

$$\mathbb{E}[g(X)|\mathcal{F}] \geq g(\mathbb{E}[X|\mathcal{F}]) + g'(\mathbb{E}[X|\mathcal{F}]^-) \underbrace{\mathbb{E}[X - \mathbb{E}[X|\mathcal{F}]|\mathcal{F}]}_{\mathbb{E}[X|\mathcal{F}] - \mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{F}] = 0} = g(\mathbb{E}[X|\mathcal{F}]).$$

Definition 4.4. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space, X an integrable random variable and $\mathcal{F} \subseteq \mathcal{A}$ a sub- σ -algebra. An \mathcal{F} -measurable random variable Y is called *conditional expectation of X with respect to \mathcal{F}* if

$$\int_F X \, d\mathbb{P} = \int_F Y \, d\mathbb{P}$$

for all $F \in \mathcal{F}$.

The problems now are: does there always exist such an Y ? And if it exists, is it unique?

Lemma 4.5. If it exists, the conditional expectation is unique.

Proof. Suppose $\int_F Y \, d\mathbb{P} = \int_F Y' \, d\mathbb{P}$ (finite) for all $F \in \mathcal{F}$, where Y and Y' are measurable. Then $\int_F Y - Y' \, d\mathbb{P} = 0$ for all $F \in \mathcal{F}$. Now consider $E := [Y \geq Y']$. Of course $E \in \mathcal{F}$ (bah...). So $\int_E Y - Y' \, d\mathbb{P} = 0$ means¹ that $Y = Y'$ a.s. on E . Repeating the same argument exchanging Y and Y' shows that $Y = Y'$ a.s. on E^c , and we are done. \square

So we have proved the uniqueness. The proof of existence is less trivial, and it requires the Theorem of Radon-Nikodym. So we skip it in this course.

4.3 Martingales

Definition 4.6. Let $(\mathcal{F}_n)_{n \geq 1}$ be an increasing sequence of sub- σ -algebras of \mathcal{A} (i.e. a *filtration* on \mathcal{A}).² A *martingale* (resp. *sub-martingale*, *super-martingale*) w.r.t. $(\mathcal{F}_n)_{n \geq 1}$ is a sequence of random variables $(X_n)_{n \geq 1}$ with the following properties (everything a.s.):

¹Define $E_n := [Y - Y' \geq \frac{1}{n}] \in \mathcal{F}$. Then $0 = \int_{E_n} Y - Y' \, d\mathbb{P} \geq \frac{1}{n} \mathbb{P}[E_n]$, so $\mathbb{P}[E_n] = 0$. Of course $E_n \nearrow E$, so passing to the limit: $\mathbb{P}[Y - Y' \geq 0] = \lim_n \mathbb{P}[E_n] = 0$.

²Intuitively, as n increases, you can distinguish more and more things.

1. X_n is \mathcal{F}_n -measurable for all $n \in \mathbb{N}$ (adaptedness).
2. $\mathbb{E}[|X_n|] < \infty$ for all $n \in \mathbb{N}$.
3. $\mathbb{E}[X_n | \mathcal{F}_{n-1}] = X_{n-1}$ (resp. $\leq X_{n-1}$, $\geq X_{n-1}$).

Remark. The typical case of a martingale is a stochastic process $(Y_n)_{n \geq 1}$ defined on $(\Omega, \mathcal{A}, \mathbb{P})$. $Y_n: (\Omega, \mathcal{A}) \rightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, or more generally $\rightarrow (\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, or even more generally $\rightarrow (\Omega', \mathcal{A}')$. Then we define $\mathcal{F}_n := \sigma(Y_1, \dots, Y_n)$ and we consider the random vector

$$(Y_1, \dots, Y_n): (\Omega, \mathcal{A}) \rightarrow (\Omega', \mathcal{A}')^n.$$

Finally, usually $X_n := g_n(Y_1, \dots, Y_n)$, where $g_n: (\Omega', \mathcal{A}')^n \rightarrow \mathbb{R}$ is a measurable function w.r.t. \mathcal{F}_n . We call $\mathbb{E}[X | Y_1, \dots, Y_n] := \mathbb{E}[X | \mathcal{F}_n]$. So $(X_n)_{n \in \mathbb{N}}$ is a martingale if $\mathbb{E}[|X_n|] < \infty$ and $\mathbb{E}[X_n | Y_1, \dots, Y_{n-1}] = X_{n-1}$ for all $n \in \mathbb{N}$.

Examples.

1. Let (Y_n) be a sequence of independent, integrable real random variables. Define $S_n := Y_1 + \dots + Y_n$ and $S_0 := 0$ (which is a markov chain). Then

$$\mathbb{E}[S_n | \mathcal{F}_{n-1}] = \mathbb{E}[Y_n + S_{n-1} | \mathcal{F}_{n-1}] = \mathbb{E}[Y_n | \mathcal{F}_{n-1}] + \mathbb{E}[S_{n-1} | \mathcal{F}_{n-1}] = \mathbb{E}[Y_n] + S_{n-1}.$$

So:

- If $\mathbb{E}[Y_n] = 0$ for all $n \in \mathbb{N}$, then (S_n) is a martingale.
- If $\mathbb{E}[Y_n] \leq 0$ for all $n \in \mathbb{N}$, then (S_n) is a supermartingale.
- If $\mathbb{E}[Y_n] \geq 0$ for all $n \in \mathbb{N}$, then (S_n) is a submartingale.

An interesting question is: for which measurable functions $g: \mathbb{R} \rightarrow \mathbb{R}$ is the sequence of $g(S_n)$'s a martingale?

2. Let (Y_n) be a sequence of independent, integrable positive random variables. Let \mathcal{F}_n be as above. Define $A_n := Y_1 \cdot \dots \cdot Y_n$. Then

$$\begin{aligned} \mathbb{E}[A_n | \mathcal{F}_{n-1}] &= \mathbb{E}[A_n | Y_1, \dots, Y_{n-1}] \\ &= \mathbb{E}[A_{n-1} \cdot Y_n | Y_1, \dots, Y_{n-1}] \\ &= A_{n-1} \mathbb{E}[Y_n | Y_1, \dots, Y_{n-1}] && \text{(by (d))} \\ &= \mathbb{E}[Y_n] A_{n-1}. \end{aligned}$$

So:

- If $\mathbb{E}[Y_n] = 1$ for all $n \in \mathbb{N}$, then (S_n) is a martingale.
- If $\mathbb{E}[Y_n] \leq 1$ for all $n \in \mathbb{N}$, then (S_n) is a supermartingale.

- If $\mathbb{E}[Y_n] \geq 1$ for all $n \in \mathbb{N}$, then (S_n) is a submartingale.
3. Given a filtration (\mathcal{F}_n) and X an integrable random variable, $X_n := \mathbb{E}[X|\mathcal{F}_n]$ is a martingale by property (e).

Remark. Observe that Birkhoff's ergodic theorem states that $Mf = \mathbb{E}[f|\mathcal{I}]$.

Theorem 4.7 (Martingale Convergence Theorem). Let $(X_n)_{n \geq 1}$ be a supermartingale w.r.t. $(\mathcal{F}_n)_{n \geq 1}$ and suppose $\mathbb{E}[|X_n|] \leq M < \infty$. Then $X_n \rightarrow X$ a.s. for some random variable X .

COPIARE

Definition 4.8. Given a filtration (\mathcal{F}_n) , a *stopping time* is a random variable $t: \Omega \rightarrow \mathbb{N}_0 \cup \{+\infty\}$ such that $[t \geq n] \in \mathcal{F}_n$ for all $n \in \mathbb{N}$.

Remark. An important example of stopping time is the following: let (Y_n) be a stochastic process and let $\mathcal{F}_n := \sigma(Y_0, \dots, Y_n)$. Intuitively, we can see (Y_0, \dots, Y_n) as the information up to time n . So, if t is a stopping time for (\mathcal{F}_n) , it means that we can decide whether $t(\omega) \leq n$ or not by looking only at $Y_0(\omega), \dots, Y_n(\omega)$.

Lemma 4.9 (Stopping Lemma). Let $\mathcal{F}_0 \subseteq \dots \subseteq \mathcal{F}_N \subseteq \mathcal{A}$, and let X_0, \dots, X_N be a finite supermartingale w.r.t. $(\mathcal{F}_n)_{0 \leq n \leq N}$. Let s and t be two stopping times with $s \leq t \leq N$ a.s. Then

$$\mathbb{E}[X_s] \geq \mathbb{E}[X_t],$$

where $X_s(\omega) := X_{s(\omega)}(\omega)$.

Proof. Observe that

$$\begin{aligned} X_s &= \sum_{n=0}^N X_n \mathbb{1}_{[s=n]} \\ &= \sum_{n=0}^N X_n (\mathbb{1}_{[s \leq n]} - \mathbb{1}_{[s \leq n-1]}) \\ &= \sum_{n=0}^N X_n \mathbb{1}_{[s \leq n]} - \sum_{n=0}^{N-1} X_{n+1} \mathbb{1}_{[s \leq n]} \\ &= \sum_{n=0}^{N-1} (X_n - X_{n+1}) \mathbb{1}_{[s \leq n]} + X_N \mathbb{1}_{[s \leq N]} \\ &= \sum_{n=0}^{N-1} (X_n - X_{n+1}) \mathbb{1}_{[s \leq n]} + X_N. \end{aligned}$$

Similarly,

$$X_t = \sum_{n=0}^{N-1} (X_n - X_{n+1}) \mathbb{1}_{[t \leq n]} + X_N.$$

So

$$X_s - X_t = \sum_{n=0}^{N-1} (X_n - X_{n+1}) (\mathbb{1}_{[s \leq n]} - \mathbb{1}_{[t \leq n]}).$$

Define $A_n := [s \leq n] \setminus [t \leq n] \in \mathcal{F}_n$. We get

$$\mathbb{E}[X_s - X_t] = \sum_{n=0}^{N-1} \mathbb{E}[X_n - X_{n+1}] \mathbb{1}_{A_n}$$

and we are done because $\mathbb{E}[X_n \mathbb{1}_{A_n}] - \mathbb{E}[X_{n+1} \mathbb{1}_{A_n}] \geq 0$, since

$$\mathbb{E}[X_{n+1} \mathbb{1}_{A_n}] = \mathbb{E}[\mathbb{E}[X_{n+1} \mathbb{1}_{A_n} | \mathcal{F}_n]] = \mathbb{E}[\mathbb{1}_{A_n} \mathbb{E}[X_{n+1} | \mathcal{F}_n]] = \text{COPIARE}$$

□

Definition 4.10. Let $(r_n)_{n \geq 0}$ be a sequence in \mathbb{R} . We define $D((r_n) | [a, b]) \in \mathbb{N} \cup \{+\infty\}$ as

$$\begin{aligned} D((r_n) | [a, b]) &:= \text{number of down-crossings of } [a, b] \text{ by } (r_n) \\ &= \lim_{N \rightarrow \infty} D_N(r_0, \dots, r_n | [a, b]) \\ &= \text{maximum number of pairs } (r_{n_i}, r_{m_i}) \text{ with} \\ &\quad \dots < n_i < m_i < n_{i+1} < m_{i+1} < \dots \text{ and } r_{n_i} > b, r_{m_i} < a. \end{aligned}$$

Theorem 4.11 (Downcrossing Theorem). Let X_0, \dots, X_N be a (finite) martingale w.r.t. $\mathcal{F}_0 \subseteq \dots \subseteq \mathcal{F}_N$. Define $D_N := D_N(X_0, \dots, X_N | [a, b])$. Then

$$\mathbb{E}[D_N] \leq \frac{\mathbb{E}[X_0] + \mathbb{E}[|X_N|] + |b|}{b - a}.$$

We will prove the Downcrossing theorem later. We can now provide a proof of the super-Martingale Convergence Theorem.

Proof. Observe that

$$\mathbb{E}[D_N((X_0, \dots, X_N) | [a, b])] \leq \frac{2M + |b|}{b - a}$$

for all $n \in \mathbb{N}$. By monotone convergence we can deduce that

$$\mathbb{E}[D((X_n)_{n \geq 0} | [a, b])] \leq \frac{2M + |b|}{b - a} < \infty,$$

and so $D((X_n) \mid [a, b]) < \infty$ a.s. for all $[a, b]$. Now define

$$\Omega_{[a,b]} := \{\omega \in \Omega : D((X_n(\omega)) \mid [a, b]) < \infty\},$$

which is a measurable set (boring exercise). By what we just showed, $\mathbb{P}[\Omega_{[a,b]}] = 1$ for all $[a, b]$. Now, defining

$$\Omega_0 := \bigcap_{\substack{[a,b] \\ a,b \in \mathbb{Q}}} \Omega_{[a,b]}$$

we easily get $\mathbb{P}[\Omega_0] = 1$ (just pass to the complement). Pick $\omega \in \Omega_0$ arbitrarily and define $r_n := X_n(\omega)$. We claim that (r_n) converges. First observe that we just proved $D((r_n) \mid [a, b]) < \infty$ for all $[a, b]$ with $a, b \in \mathbb{Q}$. Suppose towards a contradiction that $\limsup r_n > \liminf r_n$. So there exists $[a, b]$ with $a, b \in \mathbb{Q}$ which is strictly in the middle. By definition of \limsup and \liminf , it's easy to see that this implies $D((r_n) \mid [a, b]) = \infty$, contradiction. Thus, for all $\omega \in \Omega_0$ there exists $X(\omega) := \lim_n X_n(\omega) \in [-\infty, +\infty]$.

It is left to show that X is a.s. finite (??? perché? mica è nell'enunciato...). By hypothesis $\mathbb{E}[|X_n|] \leq M$ for all $n \in \mathbb{N}$, and hence by Fatou's Lemma $\mathbb{E}[|X|] \leq M$, whereby X is a.s. finite. \square

Now we prove the Downcrossing theorem:

Proof. Define $t_0 := 0$ and t_n as follows: if n is odd, then

$$t_n := \begin{cases} \min\{i \leq t_{n-1} : i \leq N, X_i \geq b\}, & \text{if such } i \text{ exists} \\ N, & \text{otherwise.} \end{cases}$$

If n is even:

$$t_n := \begin{cases} \min\{i \leq t_{n-1} : i \leq N, X_i \leq a\}, & \text{if such } i \text{ exists} \\ N, & \text{otherwise.} \end{cases}$$

These are stopping times (exercise). If $t_n = N$, it means that $n \geq N$ or $n \geq 2D_N + 2$. Let $m \geq \frac{N+1}{2}$. Of course $D_N \leq N$ by definition. So $2m + 1 \geq 2D_N + 2$. Define

$$\tilde{X} := X_{t_1} + \sum_{j=1}^m (X_{t_{2j+1}} - X_{t_{2j}}) \tag{*}$$

$$\begin{aligned} &= X_{t_1} - X_{t_2} + X_{t_3} - \dots + X_{t_{2m-1}} - X_{t_{2m}} + X_{t_{2m+1}} \\ &= \sum_{i=1}^{D_N} (X_{t_{2i-1}} - X_{t_{2i}}) + X_{t_{2D_N+1}}, \end{aligned} \tag{**}$$

where the last equality holds because $t_{2D_N+2} = t_{2D_N+3} = \dots = N$. Now look at (*):

$$\mathbb{E}[\tilde{X}] = \mathbb{E}[X_{t_1}] + \sum_{j=1}^m \underbrace{(\mathbb{E}[X_{t_{2j+1}}] - \mathbb{E}[X_{t_{2j}}])}_{\leq 0 \text{ by stopping lemma}} \leq \mathbb{E}[X_{t_1}] \leq \mathbb{E}[X_0].$$

Look at (**): of course $\tilde{X} \leq (b-a)D_N + X_{t_{2D_N+1}}$. Now:

CASE 1. If $t_{2D_N+1} < N$, then $X_{t_{2D_N+1}} \geq b$.

CASE 2. If $t_{2D_N+1} = N$, then $X_{t_{2D_N+1}} = X_N$.

Thus $X_{t_{2D_N+1}} \geq \min\{b, X_N\}$. So

$$(b-a)D_N \leq \tilde{X} + \max\{-b, -X_N\} \leq \begin{cases} \tilde{X} + |b| + |X_N| \\ \tilde{X} - X_N + (X_N - b)^+ \end{cases}.$$

Therefore, looking at (*) again,

$$(b-a)\mathbb{E}[D_N] \leq \begin{cases} \mathbb{E}[X_0] + |b| + \mathbb{E}[|X_N|], & \text{which is our bound} \\ \mathbb{E}[X_0] - \mathbb{E}[X_N] + \mathbb{E}[(X_N - b)^+] = \mathbb{E}[(X_N - b)^+], & \text{if martingale.} \end{cases}$$

□

Lemma 4.12 (Stopping Lemma for Martingales). Let $\mathcal{F}_0 \subseteq \dots \subseteq \mathcal{F}_N \subseteq \mathcal{A}$, and let X_0, \dots, X_N be a finite w.r.t. $(\mathcal{F}_n)_{0 \leq n \leq N}$. Let s and t be two stopping times with $s \leq t \leq N$ a.s. Then

$$\mathbb{E}[X_s] = \mathbb{E}[X_t],$$

where $X_s(\omega) := X_{s(\omega)}(\omega)$.

Question. What if $t < \infty$ a.s., but is not bounded. For example, let (Y_n) be a sequence of iid random variables with $\mathbb{P}[Y_n = \pm 1] = \frac{1}{2}$. Define $S_0 := 0$ and $S_n := \sum_{k=1}^n Y_k$, which is a martingale. It's easy to check that $\mathbb{P}[\exists n : S_n = 1] = 1$. Now define $t := \inf\{n : S_n = 1\}$ (although inf and min are the same on a set with probability 1), which is a stopping time. So we have that $\mathbb{E}[S_t] = 1$, while obviously $\mathbb{E}[S_n] = 0$. Thus the equality claimed in the Stopping Lemma does not hold in this case.

Remark. In general, let t be a stopping time and suppose $\mathbb{P}[t < \infty] = 1$. Define $t_n := \min\{t, n\}$, which is a stopping time. We have

$$X_t - X_{t_n} := X_t \mathbb{1}_{[t > n]} - X_n \mathbb{1}_{[t > n]},$$

so

$$\mathbb{E}[X_t] = \underbrace{\mathbb{E}[X_{t_n}]}_{=\mathbb{E}[X_0] \text{ by stopping lemma}} + \mathbb{E}[X_t \mathbb{1}_{[t > n]}] - \mathbb{E}[X_n \mathbb{1}_{[t > n]}].$$

If $\mathbb{E}[|X_t|] < \infty$, then $|X_t|\mathbb{1}_{[t>n]} \searrow 0$ a.s., and so by dominated convergence we get

$$\mathbb{E}[|X_t|\mathbb{1}_{[t>n]}] \xrightarrow{n \rightarrow \infty} 0.$$

Theorem 4.13 (Optional Sampling Theorem). Let (X_n) be a martingale w.r.t. (\mathcal{F}_n) , t a stopping time. If

1. $\mathbb{E}[|X_t|] < \infty$;
2. $\mathbb{E}[|X_t|\mathbb{1}_{[t \leq n]}] \rightarrow 0$,

then $\mathbb{E}[X_t] = \mathbb{E}[X_0]$.

Proof. See last remark. □

How can we verify in practice whether condition (2) of last theorem holds?

Definition 4.14. Let $(X_i)_{i \in I}$ be a family of random variables. $(X_i)_{i \in I}$ is called *uniformly integrable* if

$$\forall \varepsilon > 0 \exists \delta_\varepsilon > 0 \left[A \in \mathcal{A} \wedge \mathbb{P}[A] < \delta \Rightarrow \forall i \in I \mathbb{E}[|X_i|\mathbb{1}_A] < \varepsilon \right].$$

The following is a variant of the Optional Sampling Theorem:

If (X_n) is a uniformly integrable martingale and t is an a.s. finite stopping time with $\mathbb{E}[|X_t|] < \infty$, then $\mathbb{E}[X_t] = \mathbb{E}[X_0]$.

Lemma 4.15. If $\mathbb{E}[|X_i|^2] \leq c < \infty$ for all $i \in I$, then $(X_i)_{i \in I}$ is uniformly integrable.

Proof. Let $\varepsilon > 0$, $\delta := \frac{\varepsilon^2}{nc}$ and $A \in \mathcal{A}$ with $\mathbb{P}[A] < \delta$. Then

$$\begin{aligned} \mathbb{E}[|X_i|\mathbb{1}_A] &= \mathbb{E}\left[|X_i|\mathbb{1}_{A \cap [|X_i| \geq \frac{2c}{\varepsilon}]}\right] + \mathbb{E}\left[|X_i|\mathbb{1}_{A \cap [|X_i| < \frac{2c}{\varepsilon}]}\right] \\ &\leq \sqrt{\mathbb{E}[X_i^2]} \cdot \sqrt{\mathbb{P}[A \cap [|X_i| \geq \frac{2c}{\varepsilon}]]} + \frac{2c}{\varepsilon} \mathbb{P}[A \cap [|X_i| < \frac{2c}{\varepsilon}]] \\ &\hspace{15em} \text{(by Cauchy-Schwarz and } \mathbb{1}^2 = \mathbb{1}) \\ &\leq \sqrt{c} \cdot \sqrt{\underbrace{\text{Var}[X_i] \cdot \frac{\varepsilon^2}{4c^2}}_{\leq c} + \frac{2c}{\varepsilon} \mathbb{P}[A]} \hspace{5em} \text{(by Tchebychef)} \\ &\leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \end{aligned}$$

□