Probability Theory I

Bocconi University, Milan, Italy

Sandra Fortini [1]

[1]These lecture notes are an edited version by Achilli B., Mascari F., Pandolfi A., Pesenti L., Romanò G., Straziota D. of the previous lecture notes, written by Peruzzi M.

Contents

Chapter 1

Probability Spaces

Probability deals with random experiments, which means observing something where we don't know the exact result, but we can say what are all the possible outcomes.

1.1 Definitions

Definition 1.1.1 (Sample space)**.** In a random experiment, the sample space is the space of all the possible outcomes of the experiment, and we denote it as Ω.

Example 1.1.2 (Tossing a coin)*.* The possible outcomes are "head" or "tail". Setting "head"$= 1$ and "tail"$= 0$ as a convention, we say that $\Omega = \{0, 1\}$.

Example 1.1.3 (Tossing a coin infinitely many times)*.* The sample space is a set of sequences $\Omega = \{\omega = (\omega_1, \omega_2, \dots) : \omega_i = 0, 1, i = 1, 2, \dots\}$.

Example 1.1.4 (Picking a random point from 0 to 1)*.* $\Omega = (0, 1]$.

Example 1.1.5 (Pricing an asset from time $t = 0$ to time $t = T$)*.* $\Omega = \{\omega : [0, 1] \to \mathbb{R} \text{ s.t. } \omega \text{ continuous and positive}\}$

Definition 1.1.6 (Event)**.** An event A is a subset of the sample space, $A \subseteq \Omega$.

Example 1.1.7 (Head at first toss for two tosses)*.* Toss a coin twice and consider the event A "heads at first toss". We have

$$\Omega = \{(0, 0), (0, 1), (1, 0), (1, 1)\} = \{0, 1\} \times \{0, 1\} = \{0, 1\}^2$$

and

$$A = \{(1, 0), (1, 1)\} = \{1\} \times \{0, 1\}.$$

Example 1.1.8 (Head at first toss for infinite tosses)*.* Toss a coin infinitely many times and consider the event A "heads at first toss". We have

$$\Omega = \{0, 1\}^\infty$$

and

$$A = \{\omega = (\omega_1, \omega_2, \dots) : \omega_1 = 1\} = \{1\} \times \{0, 1\}^\infty.$$

If A and B are two events, then

- A^c is the event "not A";

- $A \cap B$ is the event "A and B";
- $A \cup B$ is the event "A or B";
- $B \setminus A = B \cap A^c$ is the event "B, but not A";
- If $A \subseteq B$, then "A implies B" and "B is implied by A".

Moreover,

- the empty set $\emptyset$ corresponds to the impossible event;
- the whole set Ω corresponds to the certain event.

If $\{A_t\}_{t \in T}$ is a family of events,

- the union $\bigcup_{t \in T} A_t$ is the event that occurs if and only if at least one of the A_t's occur;
- the intersection $\bigcap_{t \in T} A_t$ is the event that occurs if and only if all the A_t's occur;

The following properties (De Morgan laws) hold

- $\left(\bigcup_{t \in T} A_t\right)^c = \bigcap_{t \in T} A_t^c$
- $\left(\bigcap_{t \in T} A_t\right)^c = \bigcup_{t \in T} A_t^c$

1.2 Algebra and σ-algebra

Among all the events, we want to select a class having some stability properties with respect to complementation and union; such a classed is called an algebra.

Definition 1.2.1 (Algebra)**.** A class $\mathcal{A}$ of subsets of Ω is an algebra if

(i) $\Omega \in \mathcal{A}$;

(ii) $A \in \mathcal{A} \implies A^c \in \mathcal{A}$;

(iii) $A, B \in \mathcal{A} \implies A \cup B \in \mathcal{A}$.

Given this definition, some properties follow

- $\emptyset \in \mathcal{A}$, as $\emptyset = \Omega^c \in \mathcal{A}$
- $\mathcal{A}$ closed under finite union, i.e.

$$\{A_1, A_2, \ldots, A_n\} \subseteq \mathcal{A} \implies \bigcup_{i=1}^{n} A_i = ((A_1 \cup A_2) \cup A_3) \cdots \cup A_n) \in \mathcal{A}.$$

- $\mathcal{A}$ closed under finite intersection, i.e.

$$\{A_1, A_2, \ldots, A_n\} \subseteq \mathcal{A} \implies \bigcap_{i=1}^{n} A_i = \left(\bigcup_{i=1}^{n} A_i^c\right)^c \in \mathcal{A}$$

So we can say that and algebra is a class stable by complementation and finite unions (or finite intersections).

If we require our class of subsets to be also closed by countable union, then we obtain a σ-algebra.

Definition 1.2.2 (σ-algebra)**.** A class $\mathcal{F}$ of subsets of Ω is a σ-algebra if

- $\Omega \in \mathcal{F}$;
- $A \in \mathcal{F} \implies A^c \in \mathcal{F}$;
- $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

Remark 1.2.3 (Finite union for a σ-algebra). A σ-algebra is an algebra because we can write a finite union as a countable union: $\bigcup_{i=1}^{n} A_i = A_1 \cup A_2 \cup \dots \cup A_n \cup \emptyset \dots$. However, an algebra is not a σ-algebra in general.

Example 1.2.4 (An algebra not stable by countable union). $\Omega = \mathbb{N}$, $A \subseteq \mathbb{N}$, A is finite if it contains a finite number of elements, A cofinite if A^c is finite. Consider $\mathcal{C} = \{A \subseteq \mathbb{N} : A \text{ finite or cofinite}\}$. $\mathcal{C}$ is an algebra but not a σ-algebra, indeed the set $\bigcup_{n=1}^{\infty}\{2n\}$ is the set of even numbers, which is the union of finite sets, but it is neither finite nor cofinite.

Example 1.2.5. $\Omega = \{0,1\}^{\infty}$, for fixed n consider the sequences of length n, $A \subseteq \{0,1\}^n$. A cylinder on A is defined as $C_n(A) = \{\omega = (\omega_1, \omega_2, \dots) : (\omega_1, \dots, \omega_n) \in A\}$. The class $\mathcal{C} = \{C_n(A) : A \subset \{0,1\}^n, n \geq 1\}$ is an algebra but is not a σ-algebra. For instance, $\omega = (0,0,\dots) = \bigcap_{n\in\mathbb{N}} C_n((0,0,\dots,0))$, but does not belong to $\mathcal{C}$.

There are two particular σ-algebras:

- $\mathcal{F}_0 = \{\emptyset, \Omega\}$, the trivial σ-algebra;
- $\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$, the power σ-algebra.

Those are respectively the smallest and largest σ-algebras on Ω, i.e. for any $\mathcal{F}$, σ-algebra on Ω, $\mathcal{F}_0 \subseteq \mathcal{F} \subseteq \mathcal{P}(\Omega)$.

Definition 1.2.6 (Generated σ-algebra)**.** Let $\mathcal{C}$ be a class of events of Ω. Then the σ-algebra generated by $\mathcal{C}$ is

$$\sigma(\mathcal{C}) = \bigcap_{\substack{\mathcal{G} \supseteq \mathcal{C} \\ \mathcal{G}\ \sigma\text{-algebra}}} \mathcal{G}.$$

Note that $\sigma(\mathcal{C})$ is well-defined as

- the intersection is not empty because it contains at least the power σ-algebra;
- it is a σ-algebra because it is the intersection of σ-algebras.

Moreover, it holds

- $\sigma(\mathcal{C}) \supseteq \mathcal{C}$;
- if $\mathcal{G} \supseteq \mathcal{C}$ and $\mathcal{G}$ is a σ-algebra, then $\mathcal{G} \supseteq \sigma(\mathcal{C})$.

In other words, $\sigma(\mathcal{C})$ is the smallest σ-algebra that includes $\mathcal{C}$.

Remark 1.2.7 (Expansion from the inside and generated σ-algebra). The expansion of a σ-algebra is an expansion from the outside. Hence, in general, we cannot just add events from the inside through complementation and countable unions/intersections. This is possible only in the discrete case.

Example 1.2.8. If we take $\Omega = (0,1]$ and $\mathcal{C} = \{$disjoint union of intervals $(a,b] : 0 \leq a \leq b \leq 1\}$ then $\mathcal{C}$ is an algebra and we define $\mathcal{B}((0,1]) = \sigma(\mathcal{C})$ the Borel σ-algebra on $(0,1]$.

- $\mathcal{B}((0,1]) \ni \{a\} = \bigcap_{n=1}^{\infty} \left(a - \frac{1}{n}, a\right];$
- $\mathcal{B}((0,1]) \supseteq \{$finite sets, countable sets$\}$;
- $\mathcal{B}((0,1]) \supseteq \{$open sets, closed sets$\}$.

Example 1.2.9. $\Omega = \{0,1\}^{\infty}$, $\mathcal{C} = \{C_n(A) : A \subseteq \{0,1\}^n, n \geq 1\}$, $\mathcal{F} = \sigma(\mathcal{C})$. Consider a point $x \in (0,1]$ and its binary representation $x = \sum_{i=1}^{\infty} \frac{x_i}{2^i}$, $x \sim (x_1, x_2, \dots)$. This suggests that $\Omega = (0,1]$ and $\Omega = \{0,1\}^{\infty}$ are very similar and also $\mathcal{F} = \sigma(\mathcal{C})$ and $\mathcal{B}((0,1])$ are similar.

1.3 Probability measures

Definition 1.3.1 (Finitely additive probability)**.** Let $\mathcal{A}$ be an algebra of events on a sample space Ω. Then a finitely additive probability (FAP) on $\mathcal{A}$ is a function $P : \mathcal{A} \to \mathbb{R}$ such that the following properties hold

(i) Non-negativeness: $P(A) \geq 0 \quad \forall A \in \mathcal{A}$;

(ii) Unitary total mass: $P(\Omega) = 1$;

(iii) Additivity: If $\{A_1, \dots, A_n\} \subseteq \mathcal{A}$ is a sequence of disjoint events (i.e. $A_i \cap A_j = \emptyset$ if $i \neq j$), then $P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$.

Remark 1.3.2 (Properties of a FAP).

- $P(\emptyset) = 0$, since $\Omega \cap \emptyset = \emptyset$, then $P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset)$, hence $P(\emptyset) = 0$.
- $P(A) = 1 - P(A^c)$, since $A \cap A^c = \emptyset$ and $A \cup A^c = \Omega$, then $1 = P(\Omega) = P(A) + P(A^c)$.
- $A \subseteq B \implies P(B \setminus A) = P(B) - P(A)$, thus $P(A) \leq P(B)$, since $A \cap (B \setminus A) = \emptyset$ and $A \cup (B \setminus A) = B$, then $P(B) = P(A) + P(B \setminus A) \geq P(A)$.
- $P(\bigcup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k+1} \sum_{i_1 < i_2 < \dots < i_k} P(A_{i_1} \cap \dots \cap A_{i_k})$, as it can be proved by induction. For $n = 2$, $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$. For $n = 3$, $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$.
- $P(\bigcup_{i=1}^{n} A_i) \leq \sum_{i=1}^{n} P(A_i)$, since we can consider $\tilde{A}_1 = A_1$ and $\tilde{A}_n = A_n \setminus \left(\bigcup_{k=1}^{n-1} A_k\right) \subseteq A_n$ for $n \geq 2$ and obtain

$$P\left(\bigcup_{i=1}^{n} A_i\right) = P\left(\bigcup_{i=1}^{n} \tilde{A}_i\right) = \sum_{i=1}^{n} P\left(\tilde{A}_i\right) = \sum_{i=1}^{n} P\left(A_i\right).$$

Definition 1.3.3 (Probability measure)**.** Let $\mathcal{A}$ be an algebra of events on a sample space Ω. A probability measure (PM) on $\mathcal{A}$ is a function $P : \mathcal{A} \to \mathbb{R}$ such that the following properties hold

(i) Non-negativeness: $P(A) \geq 0 \quad \forall A \in \mathcal{A}$;

(ii) Unitary total mass: $P(\Omega) = 1$;

(iii) Countable additivity (or σ-additivity): If $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ is a countable sequence of disjoint events such that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

Remark 1.3.4 (Finite additivity of a PM)*.* A probability measure is a finite additive probability. Indeed, we just need to prove that finite additivity holds. First, let us notice that for a PM P, $P(\emptyset) = 0$. In fact,

$$P(\emptyset) = P\left(\bigcup_{i=1}^{\infty} \emptyset\right) = \sum_{i=1}^{\infty} P(\emptyset) = \begin{cases} +\infty & \text{if } P(\emptyset) > 0; \\ 0 & \text{if } P(\emptyset) = 0. \end{cases}$$

Now, let $A_1, \dots, A_n \in \mathcal{F}$, then

$$\begin{aligned} P(A_1 \cup \dots \cup A_n) &= P(A_1 \cup \dots \cup A_n \cup \emptyset \cup \dots \cup \emptyset \cup \dots) \\ &= P(A_1) + \dots + P(A_n) + 0 + \dots + 0 + \dots \\ &= P(A_1) + \dots + P(A_n). \end{aligned}$$

1.3.1 Equivalent definitions to σ-additivity

In general, for a FAP countable additivity fails to hold, i.e.

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \neq \sum_{i=1}^{\infty} P(A_i),$$

Example 1.3.5 (A FAP not σ-additive). $\Omega = \mathbb{N}$, $\mathcal{A} = \{\text{finite and cofinite sets}\}$. We define

$$P(A) = \begin{cases} 0 & \text{if } A \text{ is finite} \\ 1 & \text{if } A \text{ is cofinite} \end{cases}$$

P is a FAP, since disjoint cofinite sets do not exists so we never have $1 + 1$. But if we consider $\mathbb{N} = \bigcup_{n=1}^{\infty} \{n\}$ we can see that

$$1 = P(\mathbb{N}) \neq \sum_{n=1}^{\infty} P(\{n\}) = 0$$

In general we are not allowed to go from the finite case to the countable case with the equality, but only one inequality is satisfied.

Proposition 1.3.6 (Superadditivity)**.** *If P is a finitely additive probability measure, $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ pairwise disjoint with $\bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$, then*

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) \geq \sum_{n=1}^{\infty} P(A_n).$$

Proof. $\bigcup_{i=1}^{\infty} A_i \supseteq \bigcup_{i=1}^{n} A_i$ and P is monotone. Hence,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) \geq P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i).$$

The statement holds by taking the limit as $n \to +\infty$. $\square$

This result is very useful because we only need to show the opposite inequality to prove that a FAP is a PM, i.e. σ-additivity holds.

Another way to prove if a FAP is σ-additive is to show other equivalent properties.

Definition 1.3.7 (Increasing/decreasing sequence of events)**.** We say that $\{A_n\}_{n\in\mathbb{N}}$ is an increasing (decreasing) sequence of events if $A_n \subseteq A_{n+1}$ ($A_n \supseteq A_{n+1}$), $\forall n \in \mathbb{N}$. For an increasing (decreasing) sequence of events, we define

$$\lim_n A_n := \bigcup_{n=1}^{\infty} A_n \quad \left(\lim_n A_n := \bigcap_{n=1}^{\infty} A_n\right).$$

For short, we write

$$A_n \nearrow A = \bigcup_{n=1}^{\infty} A_n \quad \left(A_n \searrow A = \bigcap_{n=1}^{\infty} A_n\right).$$

Definition 1.3.8 (Continuity on monotone sequences)**.** Let Ω be a sample space and $\mathcal{A}$ be an algebra on Ω. We say that a FAP P on $\mathcal{A}$ is continuous on monotone spaces if one of the following equivalent definition holds.

(1) For any $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ such that $A_n \nearrow A \in \mathcal{A}$, it holds

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n\to+\infty} P(A_n).$$

(2) For any $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ such that $A_n \searrow A \in \mathcal{A}$, it holds

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n\to+\infty} P(A_n).$$

Definition 1.3.9 (Continuity on $\emptyset$)**.** Let Ω be a sample space and $\mathcal{A}$ be an algebra on Ω. We say that a FAP P on $\mathcal{A}$ is continuous on $\emptyset$ if for any $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ such that $A_n \searrow \emptyset$, it holds

$$\lim_{n\to+\infty} P(A_n) = 0.$$

Remark 1.3.10 (Continuity on monotone sequences implies continuity on $\emptyset$)*.* It is trivial by the fact that $P(\emptyset) = 0$ to show that if P is continuous on monotone sequences, then P is continuous on $\emptyset$.

Actually the converse result holds as well as can be shown by the following theorem, which primarily gives the equivalence between continuity on monotone sequences and σ-additivity.

Theorem 1.3.11 (Equivalent properties to σ-additivity)**.** *Let Ω be a sample space, $\mathcal{A}$ an algebra on it and P be a FAP on $\mathcal{A}$.*

(i) If P is countably additive, then P is continuous on monotone sequences;

(ii) If P is continuous on $\emptyset$, then P is countably additive.

Proof. (i) Let $\{A_n\}_{n\in\mathbb{N}}$ be an increasing sequence. Let us define

$$\begin{aligned} B_1 &= A_1; \\ B_2 &= A_2 \setminus A_1; \\ &\vdots \\ B_n &= A_n \setminus A_{n-1} \quad \text{for } n \geq 2. \end{aligned}$$

Then, the B_i's are pairwise disjoint and $\bigcup_{i\in\mathbb{N}} A_i = \bigcup_{i\in\mathbb{N}} B_i$. Hence,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = P\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} P(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(A_n).$$

(ii) Let $\{B_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ pairwise disjoint and let $B := \bigcup_{n\in\mathbb{N}} B_n$.
Let us define

$$A_k := B \setminus \bigcup_{n=1}^{k} B_n,$$

then $A_k \searrow \emptyset$ and $P(A_k) \xrightarrow{k\to\infty} 0$. Consequently, since

$$P(A_k) = P\left(B \setminus \bigcup_{n=1}^{k} B_n\right) = P(B) - \sum_{n=1}^{k} P(B_n),$$

we get

$$P(A_k) = (B) - \sum_{n=1}^{k} P(B_n) \xrightarrow{k\to\infty} 0 \implies P(B) = \sum_{n=1}^{\infty} P(B_n).$$

□

1.4 Choice of the σ-algebra

If Ω is finite or countable, we can take $\mathcal{F} = \mathcal{P}(\Omega)$. For instance, if $\Omega = \{\omega_1, \omega_2, \dots\}$ then $p_i := P(\{\omega_i\})$ and $P(A) = \sum_{\omega_i \in A} p_i$. But what if Ω is not countable? We could still use $\mathcal{F} = \mathcal{P}(\Omega)$ but that is not a good choice in general as shown in Vitali's example.

Example 1.4.1 (Vitali's example). We want to define a probability measure, λ on $\Omega = (0, 1]$ that is *translation invariant*[1]. However, we will show that this is not possible if we take $\mathcal{F} = \mathcal{P}((0, 1])$ as our σ-algebra. Let us define

$$x \oplus y = \begin{cases} x + y & x + y \leq 1; \\ x + y - 1 & x + y > 1. \end{cases}$$

[1]We require the property of being translation invariant, since it very useful: it assures that the length of something does not change if we move it.

and the following equivalence relation:

$$x \sim y \quad \text{if } \exists r \in \mathbb{Q} \cap (0,1] \text{ such that } y = x \oplus r.$$

Hence, we can split $(0,1]$ into equivalence classes; take one point from each equivalence class for $\sim$ and put them together into a single set H. Let us define for $r \in \mathbb{R}$, $H \oplus r := \{h \oplus r_1 : h \in H\}$, then

- $r_1, r_2 \in \mathbb{Q} \cap (0,1] : r_1 \neq r_2 \implies (H \oplus r_1) \cap (H \oplus r_2) = \emptyset$;
- $\bigcup_{r \in \mathbb{Q} \cap (0,1]} H \oplus r = (0,1]$;

Now, let us suppose by contradiction that a translation invariant probability measure λ exists; then,

$$\begin{aligned} \underbrace{\lambda((0,1])}_{=1} = \lambda\left(\bigcup_{r \in \mathbb{Q} \cap (0,1]} H \oplus r\right) &= \sum_{r \in \mathbb{Q} \cap (0,1]} \lambda(H \oplus r) \\ &= \sum_{r \in \mathbb{Q} \cap (0,1]} \lambda(H) = \begin{cases} 0 & \lambda(H) = 0; \\ +\infty & \lambda(H) > 0. \end{cases} \end{aligned}$$

Here, the second equality is given by the countability of $\mathbb{Q} \cap (0,1]$ and the third by the translation invariance of λ. Thus, we have a contradiction. No translation invariant probability measure λ can be built on $((0,1], \mathcal{P}(0,1])$. In general, the power set is not appropriate when $\Omega = \mathbb{R}$ or $\{0,1\}^\infty$ because it is too large and imposes too many constraints on P.

A better way to define a σ-algebra and a measure consists in taking an algebra $\mathcal{A}$ on Ω and define P on $\mathcal{A}$. Then consider $\mathcal{F} = \sigma(\mathcal{A})$ and then extend P can be on $\mathcal{F}$. The possibility of this extension is guaranteed by Caratheodory extension theorem.

Theorem 1.4.2 (Carathèodory). *Let P be a probability measure on an algebra $\mathcal{A}$. Then there exists a unique probability measure P^* on $\sigma(\mathcal{A})$ that coincides with P on $\mathcal{A}$, i.e.*

$$P^*(A) = P(A) \quad \forall A \in \mathcal{A}.$$

Example 1.4.3 (Solution to Vitali's example). Let $\mathcal{A} = \{\text{finite union of disjoint intervals}\}$ and consider $\mathcal{B}((0,1]) = \sigma(\mathcal{A})$. If we define P on $\mathcal{A}$ as

$$P\left(\bigcup_{i=1}^n (a_i, b_i]\right) = \sum_{i=1}^n (b_i - a_i),$$

then P is a FAP on $\mathcal{A}$, since $P(A) \geq 0$ and $P((0,1]) = 1$. Let us now show that P is a PM by checking σ-additivity. Let $(A_i)_{i \in \mathbb{N}} \subseteq \mathcal{A}$ a sequence of disjoint sets such that $\bigcup_{i=1}^\infty A_i \in \mathcal{A}$, then

$$P\left(\bigcup_{i=1}^\infty A_i\right) \geq \sum_{i=1}^\infty P(A_i)$$

as in Proposition 1.3.6. For the other inequality, let us work with intervals, for the sake of simplicity, i.e. $A_i = (a_i, b_i]$ such that $\bigcup_{i=1}^\infty A_i = (a, b]$. (This can be easily extended to finite unions of disjoint intervals.) Now, we want to show that

$$P((a,b]) \leq \sum_{i=1}^\infty P((a_i, b_i]).$$

An intuitive way to prove this would be ordering the intervals; however, we may have some accumulation points, that could make our reasoning fail. Let us prove the inequality using compactness. Let $\varepsilon > 0$, then

$$[a+\varepsilon, b] \subseteq (a,b] = \bigcup_{i=1}^{\infty}(a_i, b_i] \subseteq \bigcup_{i=1}^{\infty}\left(a_i, b_i + \frac{\varepsilon}{2^i}\right).$$

Since $[a+\varepsilon, b]$ is a compact set and $\left(\left(a_i, b_i + \frac{\varepsilon}{2^i}\right)\right)_{i\in\mathbb{N}}$ is a countable family of open countable sets, then, by the topological definition of compactness, there exists N such that

$$(a+\varepsilon, b] \subset [a+\varepsilon, b] \subset \bigcup_{i=1}^{N}\left(a_i, b_i + \frac{\varepsilon}{2^i}\right) \subset \bigcup_{i=1}^{N}\left(a_i, b_i + \frac{\varepsilon}{2^i}\right]$$

By monotonicity,

$$P((a+\varepsilon, b]) \le \sum_{i=1}^{N} P\left(\left(a_i, b_i + \frac{\varepsilon}{2^i}\right]\right).$$

and using the definition of P on intervals and taking the limit

$$b - a - \varepsilon \le \sum_{i=1}^{\infty}(b_i - a_i) + \sum_{i=1}^{\infty}\frac{\varepsilon}{2^i} \le \sum_{i=1}^{\infty}(b_i - a_i) + \varepsilon,$$

which means

$$b - a \le \sum_{i=1}^{\infty}(b_i - a_i) + 2\varepsilon$$

and we can conclude since this holds for any $\varepsilon > 0$.

P is translation invariant on $\mathcal{A}$,and we can use Caratheodory extension theorem to state that it exists a unique PM P on $\mathcal{B}((0,1])$ such that $P((a,b]) = b - a$. However, are we sure that P is translation invariant on $\sigma(\mathcal{A}) = \mathcal{B}((0,1])$? We cannot address the problem directly because we cannot define P on $\mathcal{B}((0,1])$ but we can use the so-called π-λ theorem to show it.

Definition 1.4.4 (π-class)**.** A class $\mathcal{C} \subseteq \mathcal{P}(\Omega)$ is a π-class if it is closed under finite intersections.

Remark 1.4.5 (Algebra as π-class)*.* Trivially, an algebra is a π-class, a π-class is not an algebra in general.

Definition 1.4.6 (λ-class)**.** $\mathcal{L} \subseteq \mathcal{P}(\Omega)$ is a λ-class if

(i) $\Omega \in \mathcal{L}$

(ii) $A \in \mathcal{L} \implies A^c \in \mathcal{L}$

(iii) $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{L}$ disjoint $\implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{L}$

Theorem 1.4.7 (Dynkin or π-λ theorem)**.** *If $\mathcal{L}$ is a λ-class and $\mathcal{C}$ is a π-class such that $\mathcal{L} \supseteq \mathcal{C}$, then $\mathcal{L} \supseteq \sigma(\mathcal{C})$.*

Take the family of translation invariant sets

$$\mathcal{L} = \{A \in \mathcal{B}((0,1]) : P(A \oplus x) = P(A) \quad \forall x \in (0,1]\}$$

is a λ-class. In fact,

(i) $(0,1] \in \mathcal{L}$, trivial;

(ii) $A \in \mathcal{L} \implies A^c \in \mathcal{L}$, since

$$P(A^c \oplus x) = P((A \oplus x)^c) = 1 - P(A \oplus x) = 1 - P(A) = P(A^c);$$

(iii) $\{A_i\}_{i\in\mathbb{N}} \subseteq \mathcal{L}$ disjoint $\implies \bigcup_{i\in\mathbb{N}} A_i \in \mathcal{L}$, since

$$\begin{aligned} P\left(\left(\bigcup_{i\in\mathbb{N}} A_i\right) \oplus x\right) = P\left(\bigcup_{i\in\mathbb{N}} (A_i \oplus x)\right) &= \sum_{i\in\mathbb{N}} P(A_i \oplus x) \\ &= \sum_{i\in\mathbb{N}} P(A_i) = P\left(\bigcup_{i\in\mathbb{N}} A_i\right). \end{aligned}$$

On the other hand,

$$\mathcal{C} = \{(a,b], 0 \le a \le b \le 1\}$$

is a π-class. Indeed, the intersection of a left-open right-closed interval is either a left-open right-closed interval or the empty set. Lastly, $\mathcal{L} \supseteq \mathcal{C}$. Hence, $\mathcal{L} \supseteq \sigma(\mathcal{C}) = \mathcal{B}((0,1])$, by Dinkyn theorem.

Definition 1.4.8 (Determining class of a σ-algebra)**.** Let P_1, P_2 be probability measures on $(\Omega, \mathcal{F})$, a class $\mathcal{C} \subseteq \mathcal{F}$ is a determining class for $\mathcal{F}$ if

$$P_1(A) = P_2(A),\ \ \forall A \in \mathcal{C} \implies P_1 \equiv P_2 \quad \text{(i.e. } P_1(A) = P_2(A),\ \ \forall A \in \mathcal{F});$$

i.e. knowing that the probability measures coincides on $\mathcal{C}$ implies that they coincide everywhere in $\mathcal{F}$.

Theorem 1.4.9 (Sufficient condition for determining class)**.** *If $\mathcal{C}$ is a π-class and $\mathcal{F} = \sigma(\mathcal{C})$, then $\mathcal{C}$ is a determining class for $\mathcal{F}$.*

Proof. Let $\mathcal{L} = \{A \in \mathcal{F} : P_1(A) = P_2(A)\}$, then $\mathcal{L}$ is a λ-class. Indeed,

(i) $\Omega \in \mathcal{L}$, since $P_1(\Omega) = 1 = P_2(\Omega)$;

(ii) $A \in \mathcal{L} \implies A^c \in \mathcal{L}$, since $P_1(A^c) = 1 - P_1(A) = 1 - P_2(A) = P_2(A^c)$;

(iii) $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{L}$ disjoint $\implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{L}$, since

$$P_1\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P_1(A_i) = \sum_{i=1}^{\infty} P_2(A_i) = P_2\left(\bigcup_{i=1}^{\infty} A_i\right).$$

Since $\mathcal{C}$ is assumed to be a π-class and $\mathcal{C} \subseteq \mathcal{L}$ by definition of $\mathcal{C}$, by Dynkin theorem $\mathcal{L} \supseteq \sigma(\mathcal{C}) = \mathcal{F}$; i.e. $P_1(A) = P_2(A)$ for any $A \in \mathcal{F}$. $\square$

1.5 Product spaces

Let $(\Omega_1, \mathcal{F}_1)$, $(\Omega_2, \mathcal{F}_2)$ be two measurable spaces and define

$$\begin{aligned}\Omega_1 \times \Omega_2 &:= \{\omega = (\omega_1, \omega_2) : \omega_1 \in \Omega_1, \omega_2 \in \Omega_2\}\,;\\ \mathcal{F}_1 \otimes \mathcal{F}_2 &:= \sigma\left(\{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}\right).\end{aligned}$$

Proposition 1.5.1 (Rectangles as determining class)**.** *The class of rectangles*

$$\mathcal{C} = \{A_1 \times A_2 : A_1 \in \mathcal{F}_1, A_2 \in \mathcal{F}_2\}$$

is a determining class for $\mathcal{F}_1 \otimes \mathcal{F}_2$.

Proof. By definition, $\mathcal{F}_1 \otimes \mathcal{F}_2 = \sigma(\mathcal{C})$; hence according to Theorem 1.4.9, we just need to prove that $\mathcal{C}$ is a π-class.

$$\begin{aligned}(A_1 \times A_2) \cap (B_1 \times B_2) &= \{\omega = (\omega_1, \omega_2) : \omega_1 \in A_1, \omega_2 \in A_2, \omega_1 \in B_1, \omega_2 \in B_2\}\\ &= (A_1 \cap B_1) \times (A_2 \cap B_2).\end{aligned}$$

□

Example 1.5.2. For $\Omega = \mathbb{R}^k$, we can define the Borel σ-algebra, $\mathcal{B}(\mathbb{R}^k)$ in many equivalent ways,

- $\sigma\left(\{(a_1, b_1] \times \ldots (a_k, b_k]\}\right)$;
- $\sigma\left(\{\text{open sets}\}\right)$;
- $\mathcal{B}(\mathbb{R}) \otimes \cdots \otimes \mathcal{B}(\mathbb{R})$.

1.6 Regularity of probability measures

In this section, we consider a metric space (S, d) endowed with its Borel σ-algebra $\mathcal{B}(S)$, i.e. the σ-algebra generated by open sets. We will prove that for a probability measure[2] any Borel set can be approximated arbitrarily by closed sets from below, and by open sets from above.

Theorem 1.6.1 (Regularity of PMs)**.** *Every probability measure on $(S, \mathcal{B}(S))$ is regular, that is, for every $A \in \mathcal{B}(S)$ and $\varepsilon > 0$, there exist F closed and G open such that*

$$F \subseteq A \subseteq G \quad \textit{and} \quad P(G \setminus F) < \varepsilon$$

Proof. We will first prove the result when A is closed and subsequently extend it to $\mathcal{B}(S)$ using the "good set technique".

If A is closed, then

- $F = A$;

[2] Actually, this result can be immediately extended to finite measures

- $\forall \delta > 0$, let us define
$$G_\delta := \{x \in S : d(x, A) < \delta\}.$$
If $\delta_n \searrow 0$, then $G_{\delta_n} \searrow A$ since A is closed. Hence, by continuity on monotone sequences,
$$\lim_{n \to +\infty} P(G_{\delta_n}) = P(A)$$
Hence, for every $\varepsilon > 0$, there exists $\bar{n}$ large enough such that $P(G \setminus A) < \varepsilon$ for $G = G_{\delta_{\bar{n}}}$ open.

Consequently, for every closed set A,
$$F \subseteq A \subseteq G \quad \text{and} \quad P(G \setminus F) < \varepsilon.$$

Now let $\mathcal{C}$ be the family of closed sets
$$\mathcal{C} = \{A \subseteq S : A \text{ closed}\},$$
which is trivially a π-class. Let $\mathcal{L}$ be
$$\mathcal{L} = \{A \in \mathcal{B}(S) : \forall \varepsilon > 0,\ \exists\ F \text{ closed, } G \text{ open} : F \subseteq A \subseteq G \text{ and } P(G \setminus F) < \varepsilon\},$$
then $\mathcal{C} \subseteq \mathcal{L}$. If we now prove that $\mathcal{L}$ is a λ-class, then $\sigma(\mathcal{C}) = \mathcal{B}(S) \subseteq \mathcal{L}$ by Dynkin theorem.

(i) $S \in \mathcal{L}$. Trivial, as S is closed on itself;

(ii) $A \in \mathcal{L} \implies A^c \in \mathcal{L}$. If $A \in \mathcal{L}$, then for any $\varepsilon > 0$, there exist F closed, G open such that
$$F \subseteq A \subseteq G \text{ and } P(G \setminus F) < \varepsilon$$
Consequently, for A^c it holds that
$$G^c \subseteq A^c \subseteq F^c \text{ and } P(F^c \setminus G^c) = P(G \setminus F) < \varepsilon.$$
And everything is correct since G^c closed and F^c open.

(iii) $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{L}$ disjoint $\implies \bigcup_{n=1}^\infty A_n \in \mathcal{L}$. If $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{L}$, then for any $\varepsilon > 0$, for any $n \in \mathbb{N}$, there exist F_n closed and G_n open such that
$$F_n \subseteq A_n \subseteq G_n \text{ and } P(G_n \setminus F_n) < \frac{\varepsilon}{2^{n+1}}.$$
For $\bigcup_{n=1}^\infty A_n$, we can take
   - $G := \bigcup_{n=1}^\infty G_n$, which is open;
   - $F_0 := \bigcup_{n=1}^\infty F_n$, which is not said to be closed. However, we can approximate it up to an $N < +\infty$, i.e. there exist $N > 0$ such that
$$P(F_0 \setminus F) < \frac{\varepsilon}{2} \quad \text{where } F = \bigcup_{n=1}^N F_n \text{ closed.}$$

   Now it holds that
   - $F \subseteq F_0 \subseteq A \subseteq G$;

- $P(G \setminus F) = P(G \setminus F_0) + P(F_0 \setminus F) < \varepsilon$. In fact, the term $P(F_0 \setminus F) < \frac{\varepsilon}{2}$ for N large enough. While, the term

$$
\begin{aligned}
P(G \setminus F_0) = P\left(\bigcup_{i \in \mathbb{N}} G_i \setminus \bigcup_{j \in \mathbb{N}} F_j\right) &= P\left(\bigcup_{i \in \mathbb{N}} G_i \cap \bigcap_{j \in \mathbb{N}} F_j^c\right) \\
&\leq P\left(\bigcup_{i \in \mathbb{N}} (G_i \cap F_i^c)\right) \\
&\leq \sum_{i \in \mathbb{N}} P(G_i \setminus F_i) < \sum_{i \in \mathbb{N}} \frac{\varepsilon}{2^{i+1}} = \frac{\varepsilon}{2}
\end{aligned}
$$

□

Does the opposite hold? Is it possible to approximate the probability of a measurable set by an open inner set and a closed outer set? It is not possible as shown in the following example.

Example 1.6.2 (Impossible open inner and closed outer approximation of a PM). Let $S = [0,1]$, $A = \mathbb{Q} \cap [0,1]$ and P defined on intervals as $P([a,b]) = b-a$. Then, $P(A) = 0$, while the probability of the closure of A (i.e. the smallest closed set containing A) is $P(\bar{A}) = P([0,1]) = 1$.

If we consider $\mathbb{R}^k$ as our metric space, we have a stronger result, i.e. that the inner regularity is not given only by closed set, but also by compact set.

Proposition 1.6.3 (Regularity of PMs by compact sets on $\mathbb{R}$)**.** *When $S = \mathbb{R}^k$, then for every $A \in \mathcal{B}(\mathbb{R}^k)$ and any $\varepsilon > 0$, there exist G open and K compact such that*

$$K \subseteq A \subseteq G \quad \textit{and} \quad P(G \setminus K) < \varepsilon.$$

Proof. Let $\bar{B}_n = \left\{x \in \mathbb{R}^k : \|x\| \leq n\right\}$, then $\bar{B}_n \nearrow \mathbb{R}^k$. Hence,

$$P(\mathbb{R}^k \setminus \bar{B}_n) \xrightarrow{n \to \infty} 0;$$

i.e. for any $\varepsilon > 0$, there exists n_0 such that $P(\mathbb{R}^k \setminus \bar{B}_{n_0}) < \frac{\varepsilon}{2}$. By Theorem 1.6.1, there exist F closed and G open such that

$$F \subseteq A \subseteq G \text{ and } P(G \setminus F) < \frac{\varepsilon}{2}.$$

Now, let us set $K = \bar{B}_{n_0} \cap F$, which is compact as closed in a compact set ($\bar{B}_{n_0}$). Then

- $K \subseteq F \subseteq A \subseteq G$;
- $P(G \setminus K) = P(G \setminus F) + P(F \setminus K) < \varepsilon$. Indeed,

$$P(F \setminus K) = P(F \cap (F \cap \bar{B}_{n_0})^c) = P((F \cap F^c) \cup (F \cap \bar{B}_{n_0}^c)) \leq P(\bar{B}_{n_0}^c) < \frac{\varepsilon}{2}.$$

□

1.7 Kolmogorov extension theorem

1.7.1 Cylinder sets

Let us take as sample space the space of functions from a time space T, which can be either continuous or discrete, to $\mathbb{R}$; i.e.

$$\Omega = \mathbb{R}^T = \{\omega : T \to \mathbb{R}\} = \{\omega_t \in \mathbb{R} : t \in T\}$$

We want to construct a σ-algebra on Ω. In order to do so we start with the algebra of cylinders. Fix $n \in \mathbb{N}$ and $t_1, \dots, t_n \in T$ and define the cylinder set

$$C_{t_1,\dots,t_n}(A) = \Big\{\omega \in \mathbb{R}^T : (\omega_{t_1}, \dots, \omega_{t_n}) \in A\Big\} \quad \text{for } A \in \mathcal{B}(\mathbb{R}^n). \tag{1.1}$$

For instance, if $A = A_1 \times \cdots \times A_n$ is a rectangle in $\mathcal{B}(\mathbb{R}^n)$, then we are fixing some times $t_1, \dots, t_n$ and check that at time t_i the path is in A_i.

The family of cylinder sets

$$\mathcal{A} = \{C_{t_1,\dots,t_n}(A) : A \in \mathcal{B}(\mathbb{R}^n),\ t_1, \dots, t_n \in T,\ n \geq 1\}$$

is an algebra but not a σ-algebra. Consequently, we consider the generated σ-algebra of cylinders, i.e. $\mathcal{F} = \sigma(\mathcal{A})$.

However, we have that the representation of a cylinder is not unique. Hence, to have consistency, we need that some necessary properties hold.

Proposition 1.7.1 (Consistency properties of cylinder sets)**.** *The family of cylinder sets defined as in* (1.1) *satisfies the following consistency properties.*

(I) For any $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, then

$$C_{t_1,\dots,t_n,t_{n+1}}(A_1 \times \cdots \times A_n \times \mathbb{R}) = C_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n).$$

(II) For any $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ and π a permutation of $\{1, \dots, n\}$, then

$$C_{t_{\pi(1)},\dots,t_{\pi(n)}}(A_{\pi(1)} \times \cdots \times A_{\pi(n)}) = C_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n).$$

Proof. (I) Let $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, then

$$\begin{aligned} C_{t_1,\dots,t_n,t_{n+1}}(A_1 \times \cdots \times A_n \times \mathbb{R}) &= \Big\{\omega \in \mathbb{R}^T : (\omega_{t_1}, \dots, \omega_{t_{n+1}}) \in A_1 \times \cdots \times A_n \times \mathbb{R}\Big\} \\ &= \Big\{\omega \in \mathbb{R}^T : (\omega_{t_1}, \dots, \omega_{t_n}) \in A_1 \times \cdots \times A_n\Big\} \\ &= C_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n). \end{aligned}$$

(II) Let $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ and π a permutation of $\{1, \dots, n\}$, then

$$\begin{aligned} C_{t_{\pi(1)},\dots,t_{\pi(n)}}(A_{\pi(1)} \times \cdots \times A_{\pi(n)}) &= \Big\{\omega \in \mathbb{R}^T : (\omega_{t_{\pi(1)}}, \dots, \omega_{t_{\pi(n)}}) \in A_{\pi(1)} \times \cdots \times A_{\pi(n)}\Big\} \\ &= \Big\{\omega \in \mathbb{R}^T : (\omega_{t_1}, \dots, \omega_{t_n}) \in A_1 \times \cdots \times A_n\Big\} \\ &= C_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n). \end{aligned}$$

$\square$

1.7.2 Kolmogorov consistency conditions

Let us assume that there exists a PM P on $\mathcal{A}$, and denote

$$P_{t_1,\dots,t_n}(A) := P(C_{t_1,\dots,t_n}(A)) \quad \text{for any } A \in \mathcal{A} \tag{1.2}$$

the probability measure on $\mathcal{B}(\mathbb{R}^n)$ for any $n \in \mathbb{N}$ and any $t_1, \dots, t_n \in T$. Then, given the consistency conditions on cylinders as in Proposition 1.7.1, the following conditions must hold for P.

(I) Let $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, then

$$\begin{aligned} P_{t_1,\dots,t_n,t_{n+1}}(A_1 \times \cdots \times A_n \times \mathbb{R}) &= P(C_{t_1,\dots,t_n,t_{n+1}}(A_1 \times \cdots \times A_n \times \mathbb{R})) \\ &= P(C_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n)) \\ &= P_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n). \end{aligned}$$

(II) Let $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ and π a permutation of $\{1, \dots, n\}$, then

$$\begin{aligned} P_{t_{\pi(1)},\dots,t_{\pi(n)}}(A_{\pi(1)} \times \cdots \times A_{\pi(n)}) &= P(C_{t_{\pi(1)},\dots,t_{\pi(n)}}(A_{\pi(1)} \times \cdots \times A_{\pi(n)})) \\ &= P(C_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n)) \\ &= P_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n). \end{aligned}$$

In particular, we say that our family of probability measures in (1.2) satisfies Kolmogorov consistency conditions.

Definition 1.7.2 (Kolmogorov consistency conditions)**.** A family of probability measures

$$\{P_{t_1,\dots,t_n} \text{ PM on } \mathcal{B}(\mathbb{R}^n) : t_1, \dots, t_n \in T, n \in \mathbb{N}\}$$

satisfies Kolmogorov consistency conditions if

(I) for any $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$, then

$$P_{t_1,\dots,t_n,t_{n+1}}(A_1 \times \cdots \times A_n \times \mathbb{R}) = P_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n); \tag{KCC1}$$

(II) for any $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R})$ and π a permutation of $\{1, \dots, n\}$, then

$$P_{t_{\pi(1)},\dots,t_{\pi(n)}}(A_{\pi(1)} \times \cdots \times A_{\pi(n)}) = P_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n). \tag{KCC2}$$

Remark 1.7.3 (KCC for PM on Borel sets)*.* We can actually extend Kolmogorov consistency conditions from rectangles to Borel sets using a compact notation.

(I) If we iterate the reasoning in (KCC1), then

$$P_{t_1,\dots,t_n,\dots,t_{n+k}}(A_1 \times \cdots \times A_n \times \mathbb{R}^k) = P_{t_1,\dots,t_n}(A_1 \times \cdots \times A_n).$$

Let us now set

$$\begin{gathered} \mathbf{t} \text{ as the time vector i.e. } \mathbf{t} = (t_1, \dots, t_n, \dots, t_{n+k}); \\ \varphi_k \text{ as the projection on } \mathbb{R}^n \text{ i.e. } \varphi_k(x_1, \dots, x_n, \dots, x_{n+k}) = (x_1, \dots, x_n). \end{gathered}$$

Then,

$$A_1 \times \cdots \times A_n \times \mathbb{R}^k = \varphi_k^{-1}(A_1 \times \cdots \times A_n).$$

Hence,

$$P_{\mathbf{t}}(\varphi_k^{-1}(A_1 \times \cdots \times A_n)) = P_{\varphi_k(\mathbf{t})}(A_1 \times \cdots \times A_n).$$

for any $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}^n)$. Applying Proposition 1.5.1, then

$$P_{\mathbf{t}}(\varphi_k^{-1}(A)) = P_{\varphi_k(\mathbf{t})}(A)$$

for any $A \in \mathcal{B}(\mathbb{R}^n)$, which is equivalent to (KCC1).

(II) Let us set

$$\mathbf{t} \text{ as the time vector i.e. } \mathbf{t} = (t_1, \ldots, t_n);$$
$$\pi \text{ as a permutation on } \{1,\ldots,n\} \text{ i.e. } \pi(x_1, \ldots, x_n) = (x_{\pi(1)}, \ldots, x_{\pi(n)}).$$

Then, by (KCC2),

$$A_{\pi^{-1}(1)} \times \cdots \times A_{\pi^{-1}(n)} = \pi^{-1}(A_1 \times \cdots \times A_n).$$

Hence,

$$P_{\mathbf{t}}(\pi^{-1}(A_1 \times \cdots \times A_n)) = P_{\pi(\mathbf{t})}(A_1 \times \cdots \times A_n)$$

for any $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R}^n)$. Applying Proposition 1.5.1, then

$$P_{\mathbf{t}}(\pi^{-1}(A)) = P_{\pi(\mathbf{t})}(A)$$

for any $A \in \mathcal{B}(\mathbb{R}^n)$, which is equivalent to (KCC2).

Hence, summing everything up, if $\psi = \varphi_k \circ \pi$, where φ_k is a projection and π is a permutation, then

$$P_{\psi(\mathbf{t})}(A) = P_{\varphi_k(\pi(\mathbf{t}))}(A) = P_{\pi(\mathbf{t})}(\varphi_k^{-1}(A)) = P_{\mathbf{t}}(\pi^{-1}(\varphi_k^{-1}(A))) = P_{\mathbf{t}}(\psi^{-1}(A))$$

for any $A \in \mathcal{B}(\mathbb{R}^n)$.

Remark 1.7.4 (KCC for cylinders of Borel sets). For cylinders, we similarly have that

$$C_{\psi(\mathbf{t})}(A) = C_{\mathbf{t}}(\psi^{-1}(A)) \quad \text{for any } A \in \mathcal{B}(\mathbb{R}^n),$$

where $\psi = \varphi_k \circ \pi$, with φ_k projection and π permutation.

Example 1.7.5 (Tossing a coin infinitely many times). Let $\Omega = \{0,1\}^\infty$, then the cylinders are defined as

$$C_{t_1,\ldots,t_n}(A) = \{\omega \in \{0,1\}^\infty : (\omega_{t_1}, \ldots, \omega_{t_n}) \in A\} \quad \text{for } A \in \{0,1\}^N.$$

Let us define the family of probability measures of

$$P_{t_1,\ldots,t_n}(A) := \frac{|A|}{2^n}$$

define the family of probability measures on $(\{0,1\}^n, \mathcal{P}(\{0,1\}^n))$. Then this family satisfies Kolmogorov consistency conditions. Indeed,

(I)

$$P_{t_1,\dots,t_n,t_{n+1}}(A\times\{0,1\}) = \frac{|A\times\{0,1\}|}{2^{n+1}} = \frac{2|A|}{2^{n+1}} = \frac{|A|}{2^n};$$
$$P_{t_1,\dots,t_n}(A) = \frac{|A|}{2^n}.$$

(II)

$$P_{t_{\pi(1)},\dots,t_{\pi(n)}}(A_{\pi(1)}\times\dots\times A_{\pi(n)}) = \frac{|A_{\pi(1)}\times\dots\times A_{\pi(n)}|}{2^n} = \frac{|A_1|\dots|A_n|}{2^n};$$
$$P_{t_1,\dots,t_n}(A_1\times\dots\times A_n) = \frac{|A_1|\dots|A_n|}{2^n}.$$

1.7.3 Kolmogorov extension theorem

If we have a probability measure defined on $\mathcal{A}$, can we extend it to $\mathcal{F}$? The answer is that it is actually possible provided that Kolmogorov consistency conditions hold.

Theorem 1.7.6 (Kolmogorov extension theorem)**.** *Let us consider a family of probability measures $\{P_{t_1,\dots,t_n}\}$ on $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$, defined for all $n\ge 1$ and $t_1,\dots,t_n\in T$ time set, such that* (KCC1) *and* (KCC2) *are satisfied. Then, there exists a unique probability measure P on $(\mathbb{R}^T,\mathcal{F})$ where $\mathcal{F}$ is the cylinder σ-algebra of $\mathbb{R}^T$, such that*

$$P(C_{t_1,\dots,t_n}(A)) = P_{t_1,\dots,t_n}(A)$$

for all $t_1,\dots,t_n\in T, A\in\mathcal{B}(\mathbb{R}^n)$.

Proof. We will proceed by steps:

(1) Define P on $\mathcal{A}$;

(2) Show P is "well defined";

(3) Show P is a FAP on $\mathcal{A}$;

(4) Show P is continuous on $\emptyset$ (i.e. P is a PM on $\mathcal{A}$);

(5) Apply Carathéodory.

(1) We define
$$P(C_{\mathbf{t}}(A)) := P_{\mathbf{t}}(A).$$

(2) We need to show that if $C_{\mathbf{t}}(A)=C_{\mathbf{s}}(B)$, then $P(C_{\mathbf{t}}(A))=P(C_{\mathbf{s}}(B))$. We can find a vector $\mathbf{u}$ containing all the elements of both $\mathbf{t}$ and $\mathbf{s}$. Hence, there exist
$$\psi_1:\mathbf{t}=\psi_1(\mathbf{u}) \quad\text{and}\quad \psi_2:\mathbf{s}=\psi_2(\mathbf{u}).$$
Then, we obtain $\psi_1^{-1}(A)=\psi_2^{-1}(B)$ since
$$C_{\mathbf{u}}(\psi_1^{-1}(A)) = C_{\psi_1(\mathbf{u})}(A) = C_{\mathbf{t}}(A) = C_{\mathbf{s}}(B) = C_{\psi_2(\mathbf{u})}(B) = C_{\mathbf{u}}(\psi_2^{-1}(B)).$$
Consequently, for P holds
$$\begin{aligned}P(C_{\mathbf{t}}(A)) = P_{\mathbf{t}}(A) = P_{\psi_1(\mathbf{u})}(A) = P_{\mathbf{u}}(\psi_1^{-1}(A))\\ = P_{\mathbf{u}}(\psi_2^{-1}(B)) = P_{\psi_2(\mathbf{u})}(B) = P_{\mathbf{s}}(B) = P(C_{\mathbf{s}}(B)).\end{aligned}$$

(3) Let us prove the properties of a FAP

- $P(C_{\mathbf{t}})(A) = P_{\mathbf{t}}(A) \geq 0$;
- $P(\mathbb{R}^T) = P(C_t(\mathbb{R})) = P_t(\mathbb{R}) = 1$;
- If $C_{\mathbf{t}}(A) \cap C_{\mathbf{s}}(B) = \emptyset$, then, taking $\mathbf{t} = \psi_1(\mathbf{u})$, $\mathbf{s} = \psi_2(\mathbf{u})$,

$$C_{\mathbf{u}}(\psi_1^{-1}(A)) \cap C_{\mathbf{u}}(\psi_2^{-1}(B)) = \emptyset \implies \psi_1^{-1}(A) \cap \psi_2^{-1}(B) = \emptyset.$$

Hence,

$$\begin{aligned} P(C_{\mathbf{t}}(A) \cup C_{\mathbf{s}}(B)) &= P(C_{\mathbf{u}}(\psi_1^{-1}(A)) \cup C_{\mathbf{u}}(\psi_2^{-1}(B))) \\ &= P(C_{\mathbf{u}}(\psi_1^{-1}(A) \cup \psi_2^{-1}(B))) \\ &= P_{\mathbf{u}}(\psi_1^{-1}(A) \cup \psi_2^{-1}(B)) \\ &= P_{\mathbf{u}}(\psi_1^{-1}(A)) + P_{\mathbf{u}}(\psi_2^{-1}(B)) \\ &= P_{\mathbf{t}}(A) + P_{\mathbf{s}}(B) \\ &= P(C_{\mathbf{t}}(A)) + P(C_{\mathbf{s}}(B)) \end{aligned}$$

(4) We now want to show that if $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ such that $A_n \searrow \emptyset$, then $P(A_n) \to 0$. Since the A_n's are decreasing ($A_1 \supseteq A_2 \supseteq \dots$), we have that the time steps are increasing (i.e. we are adding more constraints) as n increases. Indeed, let $A_1 = C_{t_1,\dots,t_k}(D_1)$;

$$\omega \in A_2 \implies \omega \in A_1 \implies (\omega_{t_1}, \dots, \omega_{t_k}) \in D_1$$

For instance,

$$\begin{aligned} A_1 &= C_{t_1,\dots,t_{k_1}}(D_1); \\ A_2 &= C_{t_1,\dots,t_{k_1},\dots,t_{k_1+k_2}}(D_2); \\ A_3 &= C_{t_1,\dots,t_{k_1},\dots,t_{k_1+k_2},\dots,t_{k_1+k_2+k_3}}(D_3). \end{aligned}$$

If there is no unitary increase in the indices (i.e. $k_1 = k_2 = k_3 = 1$), then we can create one as

$$\begin{aligned} A'_1 &= \mathbb{R}^T = C_{t_1}(\mathbb{R}); \\ A'_2 &= \mathbb{R}^T = C_{t_1,t_2}(\mathbb{R} \times \mathbb{R}); \\ &\vdots \\ A'_{k_1} &= A_1 = C_{t_1,\dots,t_{k_1}}(D_1); \\ A'_{k_1+1} &= A_1 = C_{t_1,\dots,t_{k_1},t_{k_1+1}}(D_1 \times \mathbb{R}); \\ &\vdots \\ A'_{k_1+k_2} &= A_2 = C_{t_1,\dots,t_{k_1},\dots,t_{k_1+k_2}}(D_2); \\ A'_{k_1+1} &= A_2 = C_{t_1,\dots,t_{k_1},\dots,t_{k_1+k_2},t_{k_1+k_2+1}}(D_2 \times \mathbb{R}); \\ &\vdots \end{aligned}$$

In general, we get that for every $n \in \mathbb{N}$,

$$A'_n = C_{t_1,\dots,t_n}(H_n) \quad \text{with } H_n \in \mathcal{B}(\mathbb{R}^n).$$

Hence, we obtain that $A'_n \searrow \emptyset$ and $P(A_n) \to 0$ if $P(A'_n) \to 0$. Hence, we now want to show that $P(A'_n) \to 0$. Let us drop the $'$ for the sake of simplicity. By contradiction, there exists $\varepsilon > 0$ such that

$$P(A_n) > \varepsilon \quad \forall n \in \mathbb{N}.$$

Since $P_{t_1,\dots,t_n}$ is a PM on $\mathcal{B}(\mathbb{R}^n)$, by Proposition 1.6.3,

$$\exists\, K_n \subseteq H_n \text{ compact} : P_{t_1,\dots,t_n}(H_n \setminus K_n) < \frac{\varepsilon}{2^{n+1}}$$

Let $B_n = C_{t_1,\dots,t_n}(K_n)$. We cannot take the sequence $\{B_n\}_{n\in\mathbb{N}}$ directly as we do not know whether it is increasing or not. Consequently, let

$$C_n = \bigcap_{j=1}^{n} B_j \quad \text{for } n \in \mathbb{N}. \tag{1.3}$$

which is a decreasing sequence. Then,

$$\begin{aligned}
P(A_n \setminus C_n) &= P(A_n \cap C_n^c) = P\left(A_n \cap \bigcup_{j=1}^{n} B_j^c\right) = P\left(\bigcup_{j=1}^{n} (A_n \setminus B_j)\right) \\
&\le P\left(\bigcup_{j=1}^{n} (A_j \setminus B_j)\right) \le \sum_{j=1}^{n} P(A_j \setminus B_j) \\
&= \sum_{j=1}^{n} P_{t_1,\dots,t_n}(H_j \setminus K_j) = \sum_{j=1}^{n} \frac{\varepsilon}{2^{j+1}} \le \sum_{j=1}^{\infty} \frac{\varepsilon}{2^{j+1}} = \frac{\varepsilon}{2}
\end{aligned}$$

Consequently, for any $n \in \mathbb{N}$

$$P(C_n) = P(A_n) - P(A_n \setminus C_n) > \frac{\varepsilon}{2}.$$

This implies that $C_n \neq \emptyset$ for all $n \in \mathbb{N}$, i.e. there exists $\omega^{(n)} \in C_n$ for any $n \in \mathbb{N}$. Hence, by (1.3),

$$\omega^{(n)} \in B_j = C_{t_1,\dots,t_j}(K_j) \quad \text{for } j \le n,$$

i.e.

$$(\omega_{t_1}^{(n)}, \dots, \omega_{t_j}^{(n)}) \in K_j \quad \text{for } n \ge j.$$

For $j = 1$, $\omega_{t_1}^{(n)} \in K_1$ for any $n \ge 1$, then

$$\exists (n_1) \subseteq (n) : \omega_{t_1}^{(n_1)} \longrightarrow u_1 \in K_1.$$

This hols since K_1 is compact, so for every sequence it exists a subsequence that converges to a point in K_1. Now, For $j = 2$, $\left(\omega_{t_1}^{(n_1)}, \omega_{t_2}^{(n_1)}\right) \in K_2$ for any $n_1 \ge 2$, then

$$\exists (n_2) \subseteq (n_1) : \left(\omega_{t_1}^{(n_2)}, \omega_{t_2}^{(n_2)}\right) \longrightarrow (u_1, u_2) \in K_2.$$

Continuing this way, for a generic j, we have

$$\left(\omega_{t_1}^{(n_j)}, \dots, \omega_{t_j}^{(n_j)}\right) \longrightarrow (u_1, \dots, u_j) \in K_j$$

Let us now take $\bar{\omega} \in \mathbb{R}^T$ such that $\bar{\omega}_{t_j} = u_j$ for any j, then

$$\bar{\omega} \in C_{t_1,\ldots,t_j}(K_j) = B_j \quad \text{for any } j.$$

Hence,

$$\bar{\omega} \in \bigcap_{j=1}^{\infty} B_j \subseteq \bigcap_{j=1}^{\infty} A_j \qquad \left(\Longrightarrow \bigcap_{j=1}^{\infty} A_j \neq \emptyset \right).$$

And we have got to a contradiction.

□

1.8 Sequences of events

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $\{A_n\}_{n\in\mathbb{N}}$ a general sequence of events in $\mathcal{F}$

Definition 1.8.1 (Almost-sure/null events)**.** An event A is called almost sure if $P(A) = 1$. An event A is called null if $P(A) = 0$.

Remark 1.8.2 (Intersection/union of almost-sure/null events)*.* For a sequence of almost sure events $\{A_n\}_{n\in\mathbb{N}}$, it holds $P\left(\bigcap_{n=1}^{\infty} A_n\right) = 1$. For a sequence of null events $\{A_n\}_{n\in\mathbb{N}}$, it holds $P\left(\bigcup_{n=1}^{\infty} A_n\right) = 0$.

Definition 1.8.3 (liminf/limsup of a sequence of events)**.** For a sequence of events $\{A_n\}_{n\in\mathbb{N}} \subset \mathcal{F}$, we define

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k = \lim_{n\to\infty} \bigcap_{k=n}^{\infty} A_k;$$
$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \lim_{n\to\infty} \bigcup_{k=n}^{\infty} A_k.$$

Remark 1.8.4 (Relation between liminf and limsup)*.* The relationship between liminf and limsup is established by the complement operator as

$$(\liminf A_n)^c = \left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \right)^c = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k^c = \limsup A_n^c;$$
$$(\limsup A_n)^c = \left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k \right)^c = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c = \liminf A_n^c.$$

Moreover, $\liminf A_n \subseteq \limsup A_n$.

What is the meaning of these two events? Let us call ω the outcome of an experiment. We say that an event A occurs if the result of the experiment is $\omega \in A$. Hence,

$$\omega \in \limsup A_n \iff \omega \in \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$
$$\iff \forall n \geq 1 \quad \omega \in \bigcup_{k=n}^{\infty} A_k \iff \forall n \geq 1 \quad \exists k \geq n : \omega \in A_k.$$

In other words, regardless of n, we can always find a subsequent occurrence of the event, which means that infinitely many A_n's occur, i.e.

$$\limsup A_n = \{A_n \text{ infinitely often}\} = \{A_n \text{ i.o.}\}$$

On the other hand,

$$\begin{aligned} \omega \in \liminf A_n &\iff \omega \in \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k \\ &\iff \exists n \geq 1 \quad : \omega \in \bigcap_{k=n}^{\infty} A_k \iff \exists n \geq 1 : \forall k \geq n : \omega \in A_k. \end{aligned}$$

In other words, starting from some event A_n, every subsequent event occurs, which means that the A_n's occur ultimately.

$$\liminf A_n = \{A_n \text{ ultimately}\} = \{A_n \text{ ult.}\}$$

Example 1.8.5. Tossing a coin infinitely many times can be represented with $\Omega = \{0,1\}^\infty$ where 0 = tails and 1 = heads. The event $A_n = \{\omega : \omega_n = 0\}$ leads to $\limsup A_n = \{\text{infinitely many tails}\}$ and $\liminf A_n = \{\text{tails ultimately}\}$.

Definition 1.8.6 (Limit of a sequence of events)**.** We say that a sequence of events $\{A_n\}$ has limit if

$$\liminf_n A_n = \limsup_n A_n$$

In which case the limit is

$$\lim_n A_n = \liminf_n A_n = \limsup_n A_n$$

Proposition 1.8.7 (Properties of the limit of a sequence of events)**.** *1.* $P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n)$*;*

2. if $\exists \lim A_n$ *then* $P(\lim A_n) = \lim P(A_n)$*.*

Proof. 1. Considering that $\liminf P(A_n) \leq \limsup P(A_n)$ because of the properties of sequences of numbers, let us prove that $\limsup P(A_n) \leq P(\limsup A_n)$.

$$P(\limsup A_n) = P\left(\bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k\right) = \lim P\left(\bigcup_{k=n}^{\infty} A_k\right) = \limsup P\left(\bigcup_{k=n}^{\infty} A_k\right) \geq \limsup P(A_n).$$

where we have used the fact that $\{\bigcup_{k=n}^{\infty} A_k\}_{n\geq 1}$ is a decreasing sequence and the monotonicity of P on monotone sequences.

2. Setting $P(\lim A_n) = L$, then

$$L = P(\liminf A_n) \leq \liminf P(A_n) \leq \limsup P(A_n) \leq P(\limsup A_n) = L.$$

□

Example 1.8.8. $\Omega = \{0,1\}^\infty$, $A_n = \{\omega : \omega_n = 0\}$

$$\begin{aligned}
P(\liminf A_n) = P(\{\text{tails ult.}\}) &= P(\lim_n \bigcap_{k=n}^{\infty} A_k) = \lim_n P(\bigcap_{k=n}^{\infty} A_k) \\
&= \lim_n \lim_{N\to\infty} P(\bigcap_{k=n}^{N} A_k) = \lim_n \lim_{N\to\infty} \frac{1}{2^{N-n+1}} = 0 \\
P(\limsup A_n) = P(\{\text{tails i.o.}\}) &= 1 - P(\{\limsup A_n\}^c) \\
&= 1 - P(\liminf A_n^c) = 1 - 0 = 1.
\end{aligned}$$

In general, finding null or almost certain events for the limsup and liminf of a sequence of events is important because they are related to the behaviour at infinity of the sequence.

From Proposition 1.8.7 we can see that a necessary condition for $P(\limsup A_n) = 0$ is that $\limsup P(A_n) = 0$, i.e. $P(A_n) \to 0$. The next lemma provides a sufficient condition.

Lemma 1.8.9 (Borel-Cantelli first lemma)**.** *If* $\sum_{n=1}^\infty P(A_n) < \infty$ *then*

$$P(\limsup A_n) = 0.$$

Proof.

$$P(\limsup A_n) = \lim P\left(\bigcup_{k=n}^{\infty} A_k\right) \le \lim \sum_{k=n}^{\infty} P(A_k) = 0,$$

where the last equality follows because the remainder of a converging series converges to 0. □

Example 1.8.10. Take $\Omega = \{0,1\}^\infty$ and define B_n as a block of "tails" of length r_n starting from toss $n+1$. How can we define r_n such that we have a finite number of blocks, i.e. $P(\limsup B_n) = 0$? In the previous example, we showed that $P(B_n \text{ i.o.}) = 1$ when $r_n = 1$. Now $P(B_n) = 2^{-r_n}$, then

$$\text{if } \sum_{n=1}^{\infty} P(B_n) = \sum_{n=1}^{\infty} \frac{1}{2^{r_n}} < \infty \quad \text{then} \quad P(\limsup B_n) = P(B_n \text{ i.o.}) = 0$$

by Lemma 1.8.9. If $r_n = n$ then $P(B_n \text{ i.o.}) = 0$. If $r_n = \lfloor (1+\varepsilon)\log_2 n \rfloor$ with $\varepsilon > 0$, then

$$\sum_{n=1}^{\infty} P(B_n) = \sum_{n=1}^{\infty} \frac{1}{2^{\lfloor (1+\varepsilon)\log_2 n \rfloor}} \le \sum_{n=1}^{\infty} \frac{1}{2^{(1+\varepsilon)\log_2 n - 1}} \le 2 \sum_{n=1}^{\infty} \frac{1}{n^{1+\varepsilon}} < \infty.$$

Hence, if $r_n \to \infty$ at a rate $(1+\varepsilon)\log_2 n$ or faster, then $P(B_n \text{ i.o.}) = 0$.

Definition 1.8.11 (Independent sequence of events)**.** A sequence of events $\{A_n\}_{n\in\mathbb{N}}$ is said to be stochastically independent (and we write $\perp\!\!\!\perp$) if

$$P\left(\bigcap_{i=1}^{n} A_{t_i}\right) = \prod_{i=1}^{n} P(A_{t_i}) \quad \forall n, \quad \forall t_1, \dots, t_n,$$

i.e. information on some of these events does not change the probability of the other ones.

Remark 1.8.12 (Independence for complement events). If the A_n's are independent, so are the A_n^c's. In fact, take $A_1^c, A_2, \dots$. We have

$$\begin{aligned}
P(A_1^c \cap A_2 \cap \dots \cap A_n) &= P((A_2 \cap \dots \cap A_n) \setminus (A_1 \cap \dots \cap A_n)) \\
&= \prod_{j=2}^{n} P(A_j) - \prod_{i=1}^{n} P(A_i) \\
&= \prod_{j=2}^{n} P(A_j)(1 - P(A_1)) = \left(\prod_{j=2}^{n} P(A_j) \right) P(A_1^c)
\end{aligned}$$

and we can repeat the same steps for all the A_i's.

Lemma 1.8.13 (Borel-Cantelli second lemma). *Let $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of independent events. If $\sum_{n=1}^{\infty} P(A_n) = +\infty$, then*

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

Proof. We want to show $P(\limsup A_n) = 1$ or equivalently $P(\liminf A_n^c) = 0$.

$$\begin{aligned}
P(\liminf A_n^c) = \lim_{n\to\infty} P(\bigcap_{k=n}^{\infty} A_k^c) &= \lim_{n\to\infty} \lim_{N\to\infty} P(\bigcap_{k=n}^{N} A_k^c) \\
&= \lim_{n\to\infty} \lim_{N\to\infty} \prod_{k=n}^{N} P(A_k^c) \\
&= \lim_{n\to\infty} \lim_{N\to\infty} \prod_{k=n}^{N} (1 - P(A_k)) \\
&\le \lim_{n\to\infty} \lim_{N\to\infty} \prod_{k=n}^{N} e^{-P(A_k)} \\
&= \lim_{n\to\infty} \lim_{N\to\infty} e^{-\sum_{k=n}^{N} P(A_k)} = 0
\end{aligned}$$

□

Example 1.8.14 (Importance of independence for BC second lemma). Is it possible to get rid of the $\perp\!\!\!\perp$ assumption in Lemma 1.8.13? No. For example, take $\Omega = \{0,1\}^\infty$, and $A_n = \{\omega_1 = 0, \omega_2 = 0, \dots, \omega_{\lfloor \log_2 n \rfloor} = 0\}$. Then

$$\sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} \frac{1}{2^{\lfloor \log_2 n \rfloor}} \ge \sum_{n=1}^{\infty} \frac{1}{2^{\log_2 n}} = \sum_{n=1}^{\infty} \frac{1}{n} = +\infty.$$

However, $P(A_n \text{ i.o.}) = P(\{\text{all tails}\}) = 0$

Example 1.8.15. With $\Omega = \{0,1\}^\infty$ and $B_n = \{\omega_{n+1} = 0, \dots, \omega_{n+r_n} = 0\}$, take $r_n = \lfloor \log_2 n \rfloor$. In this situation we cannot apply Lemma 1.8.13 because the tosses are overlapping. So let us consider disjoint blocks of tosses by defining a sequence of integers:

$$n_1 = 2, \quad n_{k+1} = n_k + \lfloor \log_2 n_k \rfloor.$$

Take C_k a block of tails of length $\lfloor \log_2 n_k \rfloor$ starting from toss $n_k + 1$, $C_k = \{\omega_{n_k+1} = 0, \dots, \omega_{n_{k+1}} = 0\}$, $r_{n_k} = \lfloor \log_2 n_k \rfloor = n_{k+1} - n_k$. Now $C_k \perp\!\!\!\perp$.

$$\begin{aligned}
\sum_{k=1}^{\infty} P(C_k) &= \sum_{k=1}^{\infty} \frac{1}{2^{\lfloor \log_2 n_k \rfloor}} \geq \sum_{k=1}^{\infty} \frac{1}{2^{\log_2 n_k}} = \sum_{k=1}^{\infty} \frac{1}{n_k} \\
&= \sum_{k=1}^{\infty} \frac{n_{k+1} - n_k}{\lfloor \log_2 n_k \rfloor n_k} \geq \sum_{k=1}^{\infty} \frac{n_{k+1} - n_k}{n_k \log_2 n_k} \\
&= \sum_{k=1}^{\infty} \sum_{j=n_k+1}^{n_{k+1}} \frac{1}{n_k \log_2 n_k} \\
&\geq \sum_{k=1}^{\infty} \sum_{j=n_k+1}^{n_{k+1}} \frac{1}{j \log_2 j} = \sum_{j=3}^{\infty} \frac{1}{j \log_2 j} = +\infty
\end{aligned}$$

By Lemma 1.8.13, $P(C_k \text{ i.o.}) = 1$, and since $\{C_k \text{ i.o.}\} \subseteq \{B_n \text{ i.o.}\}$ we have that $P(\{B_n \text{ i.o.}\}) \geq P(\{C_k \text{ i.o.}\})$. Therefore $P(B_n \text{ i.o.}) = 1$ as well.

1.9 Independence

An important concept in probability is the one of conditional probability, i.e. the probability of an event knowing that another event occurs.

Definition 1.9.1 (Conditional probability)**.** In a probability space $(\Omega, \mathcal{F}, P)$, for $A, B \in \Omega$ such that $P(A), P(B) > 0$, the probability of A given B is defined as

$$P(A|B) := \frac{P(A \cap B)}{P(B)}.$$

Notice that we can write

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c) = P(A|B)P(B) + P(A|B^c)(1 - P(B)).$$

Hence, $P(A)$ is a convex combination of $P(A|B)$ and $P(A|B^c)$. In general, in fact, $P(A)$ depends on B occurring or not.

Definition 1.9.2 (Independent events)**.** Two events A and B are independent (we use the symbol $\perp\!\!\!\perp$) if $P(A \cap B) = P(A)P(B)$.

And this can be written in terms of conditional probability as $P(A|B) = P(A|B^c) = P(A)$ meaning the events don't influence each other.

Remark 1.9.3 (Independence for almost-sure/null events). If $P(A) = 0 \implies P(A \cap B) = 0$ so $A \perp\!\!\!\perp B \;\; \forall B$. The null event is always independent on every other event and the same holds for almost sure events.

Definition 1.9.4 (Independence of a sequence of events)**.** Let $\{A_n\}_{n \in \mathbb{N}}$ sequence of events. They are mutually independent if $\forall N, \forall k_1, \dots, k_N$ distinct

$$P(A_{k_1} \cap A_{k_2} \cap \dots \cap A_{k_N}) = P(A_{k_1})P(A_{k_2}) \dots P(A_{k_N})$$

and this is equivalent to $\forall j, \forall k_1, \dots, k_N$ distinct, $P(A_j | A_{k_1} \cap A_{k_2} \cap \dots \cap A_{k_N}) = P(A_j)$, knowing that some of these occur doesn't change the probability of A_j.

Remark 1.9.5 (Independence for three event). If we have three events A,B and C we have to prove four constraints for independence

1. $P(A \cap B) = P(A)P(B)$;

2. $P(A \cap C) = P(A)P(C)$;

3. $P(C \cap B) = P(C)P(B)$;

4. $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Example 1.9.6 (Independence and pairwise independence). Roll a die, $\Omega = \{1, 2, 3, 4, 5, 6\}$, and consider $A = \{1, 2, 3, 4\}$, $B = \{2, 3, 4\}$, $C = \{4, 5, 6\}$, so that $P(A) = \frac{2}{3}$, $P(B) = \frac{1}{2}$, $P(C) = \frac{1}{2}$. We have that $P(A \cap B \cap C) = \frac{1}{6} = P(A)P(B)P(C)$ but these events are not mutually independent, as $B \subset A$, so the last condition alone is not sufficient.

Example 1.9.7. Toss a coin twice and consider $A =$"H" at first, $B =$"H" at second, $C =$ equal outcomes. We can see that these events are two by two mutually independent but the three of them are not as $P(A \cap B \cap C) = P(A \cap B) = P(A)P(B) \neq P(A)P(B)P(C)$.

Definition 1.9.8 (Independence of a family of events). Let $\{A_t\}_{t \in T}$ be a family of events. They are independent if $\forall n, \ \forall t_1, \ldots, t_n \in T \ A_{t_1}, \ldots, A_{t_n} \perp\!\!\!\perp$

Example 1.9.9. Toss a coin infinitely many times, $A_n =$"H" at toss n, $\{A_n\} \perp\!\!\!\perp$.

Independence is not really a definition that concerns events. If we have $A \perp\!\!\!\perp B$ we have seen that

- $P(A|B) = P(A) = P(A|B^c)$

- $P(B|A) = P(B) = P(B|A^c)$

the point is that information about A occurring or not does not change the probability of B, so it is really about information. In $(\Omega, \mathcal{F}, P)$ consider a sub-σ-algebra $\mathcal{G} \subset \mathcal{F}$. Information is about events in $\mathcal{G}$.

Example 1.9.10. Roll a die, $\mathcal{G} = \{\Omega, \ \emptyset, \ \text{even}, \ \text{odd}\}$. We still don't know the outcome, but if we know which event of $\mathcal{G}$ it is we have some information.

Definition 1.9.11 (Independence of classes of events). Let $\mathcal{C}_1, \ldots, \mathcal{C}_n$ be classes of events. They are said independent if $\forall A_1 \in \mathcal{C}_1, \ldots, A_n \in \mathcal{C}_n$ we have $A_1, \ldots, A_n \perp\!\!\!\perp$.

Example 1.9.12. Roll a die. Consider $\mathcal{C}_1 = \{\text{even}, \text{odd}\}$, $\mathcal{C}_2 = \{\leq 2, > 2\}$. Then $\mathcal{C}_1 \perp\!\!\!\perp \mathcal{C}_2$ because probability of the outcome being even or odd has nothing to do with the outcome being smaller or grater than two.

Suppose $\mathcal{C}_1, \ldots, \mathcal{C}_n \perp\!\!\!\perp$ and consider subclasses $\mathcal{C}'_1 \subset \mathcal{C}_1, \ldots, \mathcal{C}'_2 \subset \mathcal{C}_n$. Then if we take events $\forall A_1 \in \mathcal{C}'_1, \ldots, A_n \in \mathcal{C}'_n$ they are also in $\mathcal{C}_1, \ldots, \mathcal{C}_n$, so they are independent. We never loose independence while restricting a class, but can we keep independence extending?

Theorem 1.9.13 (Independence of generated σ-algebras). *Let $\mathcal{C}_1, \ldots, \mathcal{C}_n$ be independent π-classes. Then $\sigma(\mathcal{C}_1), \ldots, \sigma(\mathcal{C}_n)$ are independent σ-algebras.*

Proof. Without loss of generality we can assume that $\forall i$, $\mathcal{C}_i \ni \Omega$. It is enough to show that $\sigma(\mathcal{C}_1), \mathcal{C}_2, \dots, \mathcal{C}_n$ are $\perp\!\!\!\perp$ because then we have $\sigma(\mathcal{C}_1)$ as another π-class and we can apply the same reasoning again. Let $B_2 \in \mathcal{C}_2, \dots, B_n \in \mathcal{C}_n$. Take the set $\mathcal{L}$ defined as follows:

$$\mathcal{L} = \{B \in \mathcal{F} : P(B \cap B_2 \cap \cdots \cap B_n) = P(B)P(B_2) \cdots P(B_n)\}$$

$\mathcal{L}$ is a $\lambda - class$, and $\mathcal{L} \supseteq \mathcal{C}_1$ because $\mathcal{C}_1$ is a class that is independent of the others, hence it satisfies the condition in $\mathcal{L}$. By Dynkin's lemma, $\mathcal{L} \supseteq \sigma(\mathcal{C}_1)$. We also have that $\forall B_1 \in \sigma(\mathcal{C}_1)$,

$$P(B_1 \cap B_2 \cap \cdots \cap B_n) = P(B_1) \cdots P(B_n)$$

We can repeat this $\forall B_2, \dots, B_n$ and by taking some of the $B_i = \Omega$ we have the independence condition for a subset of events. □

Example 1.9.14. Toss a coin twice, $\Omega = \{0,1\}^2$ and take $A_1 = \{\omega_1 = 1\}$, $A_2 = \{\omega_2 = 1\}$, $A_3 = \{\omega_1 = \omega_2\}$. Take two classes: $\mathcal{C}_1 = \{A_1, A_2\}$ and $\mathcal{C}_2 = \{A_3\}$. The two classes are independent but $\sigma(\mathcal{C}_1)$ and $\sigma(\mathcal{C}_2)$ are not independent, because $A_1 \cap A_2 \in \mathcal{C}_1$ and $A_1 \cap A_2 \subseteq A_3 \in \mathcal{C}_2$.

Theorem 1.9.15 (Disjoint block independence)**.** *Take this array (with finite or infinite number of rows each with finite or infinite length) of independent events:*

$$\begin{matrix} A_{11} & A_{12} & \dots \\ A_{21} & A_{22} & \\ \vdots & \vdots & \ddots \end{matrix}$$

Let $\mathcal{G}_i = \sigma(\{A_{ij}, j \in \mathbb{N}\})$. Then $\{\mathcal{G}_n\}_{n\geq 1}$ are independent.

It is convenient to put the events in an array, but in general it means that σ-algebras generated by blocks of independent events are independent.

Proof. Take the π-classes $\mathcal{C}_i = \{\bigcap_{j\in J} A_{ij} : J \text{ finite}\}$. When $J = \emptyset$, $\bigcap_{j\in\emptyset} A_{ij} = \Omega$. These generate $\mathcal{G}_i$ and are also independent. Indeed, let us take $B_{i_1} \in \mathcal{C}_{i_1}, \dots, B_{i_n} \in \mathcal{C}_{i_n}$. Then

$$B_{i_k} = \bigcap_{j_k \in J_k} A_{i_k j_k}$$

and

$$\begin{aligned} P\left(\bigcap_{k=1}^{n} B_{i_k}\right) &= P\left(\bigcap_{k=1}^{n} \bigcap_{j_k\in J_k} A_{i_k j_k}\right) = \prod_{k=1}^{n} \prod_{j_k\in J_k} P(A_{i_k j_k}) = \prod_{k=1}^{n} P\left(\bigcap_{j_k\in J_k} A_{i_k j_k}\right) \\ &= \prod_{k=1}^{n} P(B_{i_k}) \end{aligned}$$

So $\sigma(\mathcal{C}_1), \dots, \sigma(\mathcal{C}_{i_n}) \;\; \perp\!\!\!\perp \forall i_1, \dots i_n$ and $\;\forall n$. □

Definition 1.9.16 (Tail σ-algebra)**.** Let $\{A_n\}$ be a sequence of events. The tail σ-algebra is defined as $\mathcal{T}(\{A_n\}) = \bigcap_{n=1}^{\infty} \sigma(\{A_n, A_{n+1}, \dots\})$.

$\mathcal{T}$ is the σ-algebra of the events such that it is possible to establish whether A occurs by looking at the tail of the sequence. So for example $\liminf A_n$ is a tail event and belongs to $\mathcal{T}$.

Example 1.9.17. $\limsup A_n$ is a tail event, indeed

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k = \bigcap_{n=N}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

because we don't care when we start, so for every N $\limsup A_n \in \sigma(\{A_N, A_{N+1}, \dots\})$. The same holds for $\liminf A_n$.

Theorem 1.9.18 (Kolmogorov 0-1 Law)**.** *If $\{A_n\}_{n\in\mathbb{N}}$ are independent and if $A \in \mathcal{T}(\{A_n\})$, then*

$$P(A) = \begin{cases} 0 \\ 1 \end{cases}$$

Proof. Using the sequence, fix n and construct this array:

$$\begin{array}{ll} A_1 & \\ A_2 & \\ \vdots & \\ A_n & \\ A_{n+1} & A_{n+2} \dots \end{array}$$

Then by the disjoint blocks theorem we have

$$\sigma(A_1), \sigma(A_2), \dots, \sigma(A_n), \sigma(\{A_{n+1}, A_{n+2}, \dots\}) \;\; \perp\!\!\!\perp$$

And $A \in \sigma(\{A_{n+1}, A_{n+2}, \dots\})$. Therefore $A, A_1, A_2, \dots, A_n$ are $\perp\!\!\!\perp$, $\forall n$. This means $A, A_1, A_2, \dots$ are $\perp\!\!\!\perp$. Now take another array:

$$\begin{array}{lll} A & & \\ A_1 & A_2 & A_3 \dots \end{array}$$

By the disjoint blocks theorem, $\sigma(A) \perp\!\!\!\perp \sigma(\{A_1, A_2, \dots\}) \ni A$. This means that $A \perp\!\!\!\perp A$, therefore $P(A) = P(A \cap A) = P(A)P(A) = P(A)^2$ and $P(A)$ is either 0 or 1. □

Example 1.9.19. Toss a coin infinitely many times. We want to know if it exists a limit frequency of heads, so we consider the event

$$A = \{\omega : \liminf \frac{\sum_{i=1}^n \omega_i}{n} = \limsup \frac{\sum_{i=1}^n \omega_i}{n}\}$$

and the sequence $A_n = \{\omega : \omega_n = 1\}$.

$$\liminf \frac{\sum_{i=1}^n \omega_i}{n} = \liminf \left(\frac{\omega_1}{n} + \frac{\sum_{i=2}^n \omega_i}{n} \right) = \liminf \frac{\sum_{i=2}^n \omega_i}{n} = \liminf \frac{\sum_{i=N}^n \omega_i}{n}$$

so $A \in \mathcal{T}(A_n)$ and A_n are independent, so $P(A) = 0$ or $P(A) = 1$.

Chapter 2

Random Variables and Random Vectors

Recall that a function $g : (\Omega, \mathcal{F}) \mapsto (\Omega', \mathcal{F}')$ is $\mathcal{F}/\mathcal{F}'$-measurable if $g^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{F}'$. Here are some properties of functions on measurable spaces:

- Let $g : (\Omega, \mathcal{F}) \mapsto (\Omega', \mathcal{F}')$. Then $\{g^{-1}(B) : B \in \mathcal{F}'\}$ is a sigma-algebra;
- Let $g : (\Omega, \mathcal{F}) \mapsto (\Omega', \mathcal{F}')$. Then $\{B \in \mathcal{F}' : g^{-1}(B) \in \mathcal{F}\}$ is a sigma-algebra;
- Let $\mathcal{C}' \subset \mathcal{F}'$ be a class of subsets of Ω' such that $\mathcal{F}' = \sigma(\mathcal{C}')$ and let $g : (\Omega, \mathcal{F}) \mapsto (\Omega', \mathcal{F}')$. If $g^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{C}'$, then g is $\mathcal{F}/\mathcal{F}'$-measurable;
- If $g : (\Omega, \mathcal{F}) \mapsto (\Omega' \mathcal{F}')$ is measurable and $h : (\Omega', \mathcal{F}') \mapsto (\Omega''.\mathcal{F}'')$ is measurable, then $h(g)$ is $\mathcal{F}/\mathcal{F}''$-measurable;
- If S and S' are metric spaces and $g : (S, \mathcal{B}(S)) \mapsto (S', \mathcal{B}(S'))$ is continuous, then g is measurable.

2.1 Definitions of random variable and random vector

Definition 2.1.1 (Random Variable)**.** A random variable on $(\Omega, \mathcal{F}, P)$ is a measurable function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, so by definition $\forall B \in \mathcal{B}(\mathbb{R})$, $X^{-1}(B) \in \mathcal{F}$.

The meaning of the inverse image is $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$. The measurable condition is needed to talk about the probability $P(X \in B)$, indeed we want P defined on events like $(X \in B) = \{\omega : X(\omega) \in B\}$.

Definition 2.1.2 (Indicators)**.** For a set $A \in \mathcal{F}$, we define the indicator function on A

$$\mathbb{1}_A = \begin{cases} 1 \text{ if } \omega \in A \\ 0 \text{ otherwise} \end{cases}$$

which is always measurable.

Definition 2.1.3 (Simple Random Variable)**.** A random variable which takes a finite number of values is called a simple random variable and can be written as a combination of indicators

$$X = \sum_{i=1}^{\infty} a_i \mathbb{1}_{A_i}$$

where $a_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$.

Proposition 2.1.4 (Simple random variable approximation)**.** *We can approximate a random variable with a sequence of simple random variable. Namely,* $\forall X\ \exists (X_n)$ *simple random variables such that* $\forall \omega\ X_n(\omega) \to X(\omega)$*.*

Proof. Assume $X \geq 0$, then:

$$X_n = \sum_{k=0}^{n2^n-1} \frac{k}{2^n} \mathbb{1}_{\left(\frac{k}{2^n} \leq X < \frac{k+1}{2^n}\right)}$$

this is a simple measurable function by measurability of X. For any fixed $\omega \in \Omega$, $\exists n_0 : \forall n \geq n_0, n > X(\omega)$; by construction $X(\omega) - X_n(\omega) < \frac{1}{2^n} \to 0$ for $n \to \infty$. For a general X we can define the positive and negative part as

$$X^+ = max(X, 0) \quad X^- = -min(X, 0)$$

which are both positive and satisfy $X = X^+ - X^-$. Once we have the converging sequences for X^+ and X^-, we sum them to get a sequence converging to X. $\square$

Proposition 2.1.5 (Equivalent definitions to measurability for a r.v.)**.** $X : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$*,* X *is measurable if and only if one of the following conditions hold*

- $(X \leq a) \in \mathcal{F}\ \ \forall a \in \mathbb{R}$
- $(X < a) \in \mathcal{F}\ \ \forall a \in \mathbb{R}$
- $(X > a) \in \mathcal{F}\ \ \forall a \in \mathbb{R}$
- $(X \geq a) \in \mathcal{F}\ \ \forall a \in \mathbb{R}$

Definition 2.1.6 (Random vectors)**.** A random vector is a vector of random variables. We can also see it as a measurable function $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ such that $X(\omega) = (X_1(\omega), \ldots, X_k(\omega))$.

Proposition 2.1.7 (Measurability as a component-wise property)**.** $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ *is measurable if and only if* X_i *is measurable for any* i*.*

Proof. Suppose X_i measurable and let us take $\mathcal{C} = \{B_1 \times \cdots \times B_k : B_i \in \mathcal{B}(\mathbb{R}), i = 1, \ldots, k\}$, then

$$X^{-1}(B_1 \times \cdots \times B_k) = ((X_1, \ldots X_k) \in B_1 \times \cdots \times B_k) = \bigcap_{i=1}^{\infty} (X_i \in B_i) \in \mathcal{F}$$

is measurable and it can be trivially extended on $\mathcal{B}(\mathbb{R}^n)$ using Dynkin's theorem. Conversely, suppose $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is measurable, then the coordinates $X = (X_1, \ldots, X_k)$ are measurable:

$$X_i^{-1}(B) = (X_i \in B_i) = ((X_1, \ldots, X_k) \in \mathbb{R} \times \cdots \times B_i \times \mathbb{R} \times \cdots \times \mathbb{R}) \in \mathcal{F}.$$

$\square$

Proposition 2.1.8 (Measurability of functions of r.v.'s)**.** *Since random vectors are measurable functions the following properties hold:*

- *If X is a random vector and $g : \mathbb{R}^k \to \mathbb{R}^j$ is a measurable function, then $Y = g(X)$ is a random vector*
- *If X is a random vector, then* $\max(X_1, \dots, X_k)$ *and* $\min(X_1, \dots, X_k)$ *are random variables*
- *If X_n is a random variable $\forall n$ (a sequence of random variables) then the following are random variables:* $\sup X_n, \inf X_n, \limsup X_n, \liminf X_n$ *and* $\lim X_n$ *(if it exists).*

Remark 2.1.9 (Sub σ-algebra)*.* We usually call $\mathcal{F}$ the universe σ-algebra and we interpret it as the set of all possible events that one might be interested in. Consider now a sub-sigma-algebra $\mathcal{G} \subseteq \mathcal{F}$ and assume that for any event $A \in \mathcal{G}$, we know whether A occurs or not. This means that, even if we don't know the precise outcome of the experiment, we still have some information about it (look at the next examples). We interpret sub-σ-algebras as container of (partial) information on the outcome of the experiment.

Example 2.1.10.

- $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$ then knowing whether any element of $\mathcal{G} = \{\emptyset, \Omega, \{2, 4, 6\}, \{1, 3, 5\}\}$ occurs or not, means that we have information on whether the outcome is even or odd.
- $\Omega = \{0, 1\}^\infty$ and $\mathcal{F}$ = cylinder sigma-algebra, then information on the first n tosses is represented by the set of cylinders $\mathcal{G}_n = \{C_{1,\dots,n}(A) : \textit{ where } A \subseteq \{0, 1\}^n\}$.

Definition 2.1.11 ($\mathcal{G}$-measurability)**.** Let $(\Omega, \mathcal{F}, P)$ probability space and $\mathcal{G} \subseteq \mathcal{F}$ a sub-σ-algebra of $\mathcal{F}$. We say that X is $\mathcal{G}$-measurable if $X : (\Omega, \mathcal{G}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ is measurable, that is

$$(X \in B) \in \mathcal{G} \ \ \forall B \in \mathcal{B}(\mathbb{R}^k)$$

and if this is true, then $\forall B \in \mathcal{B}(\mathbb{R}^k)$ we can say whether $(X \in B)$ or not based on information in $\mathcal{G}$.

If X is $\mathcal{G}$-measurable then the information contained in $\mathcal{G}$ allows to determine the value of X.

Example 2.1.12. Toss a coin infinitely many times. $\Omega = \{0, 1\}^\infty$ and consider for each n the σ-algebra of cylinders of dimension n $\mathcal{G}_n$ (defined above). Let $X = \#$ heads until toss m. Here, X is $\mathcal{G}_n$-measurable if and only if $n \geq m$ i.e. we know the number of "H" in the first m tosses only once we have observed them.

Definition 2.1.13 (Sigma-algebra generated by a random vector)**.** In $(\Omega, \mathcal{F}, P)$, with X random vector, the sigma-algebra generated by X is:

$$\sigma(X) = \bigcap_{\substack{\mathcal{G} \subseteq \mathcal{F} \\ X \text{ is } \mathcal{G}\text{-meas.}}} \mathcal{G}$$

$\sigma(X)$ is the smallest sigma-algebra with respect to which X is measurable. Indeed, if X is $\mathcal{G}$-measurable, then $\sigma(X) \subseteq \mathcal{G}$.

Theorem 2.1.14 (σ-algebra generated by a r.v.)**.** *Let $(\Omega, \mathcal{F}, P)$ be a probability space, X a random vector. Then*

$$\sigma(X) = \{X^{-1}(B) : B \in \mathcal{B}(\mathbb{R}^k)\}.$$

Proof. Let $\mathcal{G}$ be the right hand side, one can easily see that it is a sigma-algebra.

- ($\subset$) By construction X is $\mathcal{G}$-measurable and $\mathcal{G}$ is a sigma algebra. Then $\sigma(X) \subset \mathcal{G}$, by def. of $\sigma(X)$.
- ($\supset$) Since X is $\sigma(X)$-measurable, $X^{-1}(B) \in \sigma(X)$ for any $B \in \mathcal{B}(\mathbb{R}^n)$

$\square$

Theorem 2.1.15 (Doob-Dynkin). *Let $(\Omega, \mathcal{F}, P)$ be a probability space, $X : (\Omega, \mathcal{F}) \to (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ a random vector and Y a random variable. Then:*

$$Y \textit{ is } \sigma(X)\textit{-measurable} \iff \exists g : (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k) \to (\mathbb{R}, \mathcal{B}(\mathbb{R})) \textit{ measurable} : Y = g(X)$$

Idea: if Y is $\sigma(X)$-measurable then the information about X allows to determine Y, so there must be a way of recovering X from Y.

Proof. ($\Longleftarrow$) If $Y = g(X)$, then $(Y \in B) = (g(X) \in B) = (X \in g^{-1}(B)) \in \sigma(X)$ and so Y is measurable with respect to $\sigma(X)$.
($\Longrightarrow$) Suppose Y is $\sigma(X)-$measurable.

1. Let's start by assuming that Y is a simple random variable, i.e. it takes on a finite number of different values. This means that for $a_1, \dots, a_N$ we have $A_i = (Y = a_i) = (X \in B_i)$ for some B_i Borel set, so the representation is not unique. But we can find $B_1, \dots, B_N$ such that $B_i \subset \{X(\omega) : \omega \in \Omega\}$ and $B_i \cap B_j = \emptyset \ \forall i \neq j$, and we can restrict to the range of X and discard the rest. We define g as
$$g(x) = \begin{cases} a_1 & x \in B_1 \\ a_2 & x \in B_2 \\ \dots \\ a_N & x \in B_N \\ 0 & x \in (B_1 \cup \dots \cup B_N)^c \end{cases}$$
if we take a point $\omega \in \Omega$, $g(X(\omega)) = a_i$ if $X(\omega) \in B_i$, and so $Y = \sum_{i=1}^N a_i \mathbb{1}_{A_i} = \sum_{i=1}^N a_i \mathbb{1}_{(X \in B_i)} = \sum_{i=1}^N a_i \mathbb{1}_{B_i}(X) = g(X)$.

2. Consider $Y \geq 0$. Using the construction in 2.1.4, we can approximate Y from below with a sequence of simple random variables $Y_n \uparrow Y$. Since Y_n is a function of Y, then it is measurable with respect to $\sigma(X)$. Moreover, since Y_n is simple, $Y_n = g_n(X), \ \forall n$. Now define the set $M = \{x \in \mathbb{R} : g_n(x) \text{ converges}\}$ and notice that $\forall \omega \in \Omega, \ X(\omega) \in M$, since $g_n(X(\omega)) = Y_n(\omega)$ and $Y_n(\omega)$ converges. So define
$$g(x) = \mathbb{1}_M(x) \lim g_n(x)$$
For any $\omega \in \Omega$, we have:
$$g(X(\omega)) = \lim_{n\to\infty} g_n(X(\omega)) \mathbb{1}_M(X(\omega)) = \lim_{n\to\infty} Y_n(\omega) = Y(\omega).$$

3. Finally, taking Y as a general random variable, we have $Y = Y^+ - Y^-$ and both Y^+ and Y^- are $\sigma(X)$-measurable since they are non-negative, so $Y = g_1(X) - g_2(X) = g(X)$.

$\square$

2.2 Probability distribution of a Random Vector

Definition 2.2.1 (Probability distribution)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space and $X = (X_1, \dots, X_k)$ a random vector. We define the probability distribution of X as a probability measure μ on $\mathcal{B}(\mathbb{R}^k)$ such that

$$\mu(B) = P_X(B) = P(X^{-1}(B)) = P(X \in B), \quad \forall B \in \mathcal{B}(\mathbb{R}^k).$$

We can show that also the converse is true

Theorem 2.2.2 (From distribution to r.v.)**.** *Let μ be a probability measure on $\mathcal{B}(\mathbb{R}^k)$. Then $\exists$ $(\Omega, \mathcal{F}, P)$ probability space, and X random vector on it such that $\mu = PX^{-1}$.*

Proof. Take $\Omega = \mathbb{R}^k$ and take $\mathcal{F} = \mathcal{B}(\mathbb{R}^k)$ and $P = \mu$. With $X(\omega) = \omega$ we are done. □

2.2.1 Cumulative distribution function

$\forall x \in \mathbb{R}^k$ let S_x be the set of "south-west of x", defined as

$$S_x = \{(S_1, \dots, S_k) \in \mathbb{R}^k : S_i \le x_i \forall i\}.$$

Definition 2.2.3 ((Cumulative) distribution function)**.** We define the (cumulative) distribution function (CDF) of X as:

$$F(x) = \mu(S_x) = P(X_1 \le x_1, \dots, X_k \le x_k)$$

Knowing F, one can compute the probability that X belongs to any rectangle $R = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_k, b_k]$ as follows.

Proposition 2.2.4 (Measure of a rectangle through a CDF)**.** *Let $V = \{v = (v_1, \dots, v_k) : v_i \in \{a_i, b_i\}, i = 1, \dots, k\}$ be the set of vertices of the rectangle $R = (a_1, b_1] \times (a_2, b_2] \times \cdots \times (a_k, b_k]$. For any $v \in V$, we define its sign as:*

$$sign(v) = \begin{cases} +1 & \text{if the number of } a_i \text{ is even or 0} \\ -1 & \text{if the number of } a_i \text{ is odd} \end{cases}$$

so that:

$$\mu(R) = \Delta_R(F) := \sum_{v \in V} sign(v) F(v) \tag{2.1}$$

Proof.

$$\begin{aligned}
\mu(R) &= P\left(\bigcap_{i=1}^{k} (a_i < X_i \le b_i)\right) \\
&= P(X_1 \le b_1, \dots, X_k \le b_k) - P\left(\bigcup_{i=1}^{k} \underbrace{(X_1 \le b_1, \dots, X_i \le a_i, \dots X_k \le b_k)}_{A_i}\right) \\
&= F(b_1, \dots, b_k) - \sum_{j=1}^{k} (-1)^{j+1} \sum_{i_1 < \cdots < i_j} \underbrace{P(A_{i_1} \cap \cdots \cap A_{i_j})}_{F(b_1, \dots, a_{i_1}, \dots, a_{i_j}, \dots, b_k)} \\
&= F(b_1, \dots, b_k) - \sum_{j=1}^{k} \underbrace{(-1)^{j+1}}_{n.\ of\ a_i} \sum_{i_1 < \cdots < i_j} F(b_1, \dots, a_{i_1}, \dots, a_{i_j}, \dots, b_k) \\
&= \Delta_R(F)
\end{aligned}$$

□

Theorem 2.2.5 (Properties of a CDF). *A function $F : \mathbb{R}^k \to \mathbb{R}$ is the distribution function of some random variable if and only if the following conditions are satisfied:*

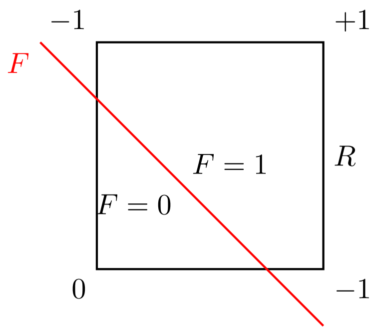
1. $\lim_{x_i \to -\infty} F(x_1, \dots, x_k) = 0 \ \forall i$

2. $\lim_{\substack{x_1 \to +\infty \\ \dots \\ x_k \to +\infty}} F(x_1, \dots, x_k) = 1$

3. *F is continuous from above:* $\lim_{h_i \to 0^+ \forall i} F(x_1 + h_1, \dots, x_k + h_k) = F(x_1, \dots, x_k)$

4. $\forall R \ \ \Delta_R(F) \geq 0$

Remark 2.2.6 (Monotonicity of a CDF in each argument). Note that property (4) implies that F is monotone in each argument, however the reverse is not true. Indeed, take F which is zero on the left of a line in the plane and one on the right. This function satisfies the first three conditions and it is monotone in each component, but it is not a distribution function: if we compute $\Delta_R(F)$ for a rectangle that has only one vertex on the left of the line we get $\Delta_R(F) = 1 - 1 - 1 + 0 = -1$.



Recap on measure theory

Definition 2.2.7 (Measure). $\mu : \Omega \to [0, +\infty]$ is a measure if

- it is a set function: $\mu(\emptyset) = 0$
- it is σ-additive: $\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ where A_n's are disjoint

Definition 2.2.8 (σ-finite measure). A measure μ is σ-finite if $\exists A_1, A_2, \dots \in \mathcal{F}$ such that $\Omega = \bigcup_{i=1}^{\infty} A_i$ and $\mu(A_i) < +\infty$

Definition 2.2.9 (Measurable function). $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is a measurable function if and only if $f^{-1}(B) \in \mathcal{F}$ where $B \in \mathcal{B}(\mathbb{R})$.

Definition 2.2.10 (Integration w.r.t. a measure). f measurable function, we can define $\int_\Omega f d\mu$

- $f = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i} \geq 0 \implies \int_\Omega f d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$
- $f \geq 0 \implies \int_\Omega f d\mu = \sup\{\int_\Omega g d\mu : g \in B_0^+, g \leq f\}$
- f general, then $f = f^+ - f^-$ and $\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu$ provided at least one is finite

Proposition 2.2.11 (Properties of the integral)**.**

- *$f = g$ a.e. means $\mu(\{\omega : f(\omega) \neq g(\omega)\}) = 0$ and implies that $\int f d\mu = \int g d\mu$*
- *$\int_A f d\mu$ where $A \in \mathcal{F}$, $\int_A f d\mu = \int f \mathbb{1}_A d\mu$*
- *$f = g$ a.e. $\iff \forall A \in \mathcal{F}, \int_A f d\mu = \int_A g d\mu$*

2.3 Radon-Nikodym theorem

Definition 2.3.1 (Absolutely continuous measures)**.** Given two measures μ, ν on the space $(\Omega, \mathcal{F})$. We say that ν is absolutely continuous with respect to μ if $\mu(A) = 0 \implies \nu(A) = 0$, and we write $\nu \ll \mu$.

Example 2.3.2. $f \geq 0$ and $f : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}^k))$, with σ-finite measure μ, and $\int_A f d\mu = 0$ if $\mu(A) = 0$ then ν defined as $\nu(A) = \int_A f d\mu$ is a σ-finite measure on $\mathcal{F}$ and $\nu \ll \mu$.

Definition 2.3.3 (Singular measures)**.** Given two measures μ, ν on the space $(\Omega, \mathcal{F})$. We say that μ and ν are (mutually) singular if $\exists S_\mu, S_\nu \in \mathcal{F}$ such that $S_\mu \cap S_\nu = \emptyset$, $\mu(S_\mu^c) = 0$, $\nu(S_\nu^c) = 0$, and we write $\nu \perp \mu$.

Example 2.3.4. On $\Omega = \mathbb{R}$, consider the Lebesgue measure $\lambda((a, b]) = b - a$ and μ the counting measure on a countable set S, $\mu(A) = |A \cap S|$. Then $\lambda \perp \mu$, take $S_\mu = S$ and $S_\lambda = \mathbb{R} \setminus S$, then $\lambda(S) = 0$ and $\mu(S^c) = 0$

Theorem 2.3.5 (Radon-Nikodym)**.** *Let μ, ν be σ-finite measures on $(\Omega, \mathcal{F})$, then:*

$$\nu \ll \mu \iff \exists f \text{ measurable and non-negative such that } \nu(A) = \int_A f d\mu$$

Moreover, f is essentially unique, that is if $\nu(A) = \int_A g d\mu \ \forall A \in \mathcal{F}$, then $f = g$ μ-almost everywhere. The function f is called the Radon-Nikodym derivative (or density) of ν with respect to μ and denoted by $f = \frac{d\mu}{d\nu}$.

Theorem 2.3.6 (Lebesgue decomposition)**.** *$(\Omega, \mathcal{F})$ measurable space, μ, ν measures on $\mathcal{F}$, both σ-finite. Then*

$$\nu = \nu_{ac} + \nu_s$$

with $\nu_{ac} \ll \mu$ and $\nu_s \perp \mu$. This decomposition is unique, thus

$$\nu(A) = \int_A f d\mu + \nu_s(A)$$

2.3.1 Singular continuous distributions

Take X random vector with probability distribution μ, and take the Lebesgue measure $\lambda((a_1, b_1] \times \cdots \times (a_k, b_k]) = \prod_{i=1}^k (b_i - a_i)$ as the reference measure. Then $\mu = \mu_{ac} + \mu_s$ and we know that $\mu_{ac}(A) = \int_A f d\lambda$, f density function. We now want to analyze the singular part. Define $D = \{x \in \mathbb{R}^k\} : \mu(\{x\}) > 0\}$. Then for $x \in D$, $\mu(\{x\}) = \mu_s(\{x\})$ because the Lebesgue measure on a single point is zero.

Lemma 2.3.7 (Cardinality of the set of disjoint non-null events)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measured space, μ finite measure, and let $\mathcal{C} = \{B_t\}_{t \in T}$ be a class of disjoint events such that $\mu(B_t) > 0 \ \forall t$. Then $\mathcal{C}$ is countable at most.*

Proof. Fix k and let $t_1, \dots, t_n \in T$ be such that $P(B_{t_i}) > \frac{1}{k}$ $\forall i$. Then since

$$\mu\left(\bigcup_{i=1}^{n} B_{t_i}\right) = \sum_{i=1}^{n} \mu(B_{t_i}) \geq \frac{n}{k}$$

$n \leq k\mu(\Omega)$ is bounded. Then:

$$\mathcal{C} = \bigcup_{k=1}^{\infty} \{B_t : \mu(B_t) > \frac{1}{k}\}$$

i.e. it is a countable union of finite sets (for a fixed k), that is countable at most. □

Thanks to this lemma we can state that D is countable at most. Now define the probability mass function as $m(x) = \mu(\{x\})$, $x \in D$. Then we have the discrete component of μ as $\mu_D(A) = \sum_{x \in A \cap D} m(x) = \mu_s(A \cap D) \leq \mu_s(A)$, μ_D is a finite measure. If we now consider $\mu_s(A) - \mu_D(A) \geq 0$ $\forall A$ is a measure, let's call it singular continuous component $\mu_{sc} = \mu_s - \mu_D$ (which means $\mu_s = \mu_{sc} + \mu_D$). Then every probability distribution μ can be decomposed as

$$\mu = \mu_{ac} + \mu_d + \mu_{sc}$$

$$\mu(A) = \int_A f d\lambda + \sum_{x \in A \cap D} m(x) + \mu_{sc}(A)$$

. There are situations in which the singular continuous component is not trivial

Example 2.3.8. X, Y random variable such that $X + Y = 1$, $X \sim U[0,1]$. The joint distribution of X, Y is $P_{X,Y} \perp \lambda$ (the support of (X, Y) is a straight line, which has λ-measure 0). Moreover $P_{X,Y}$ has no discrete component, hence it is singular continuous.

In general, this happens when the random vector $X = (X_1, \dots, X_k) \in \mathbb{R}^k$ and the support of μ is a variety with lower dimension.

Example 2.3.9. $X \sim \mathcal{N}(0, \Sigma)$ with Σ singular, which means that $\exists a : a^T \Sigma a = 0$. This condition defines a plane: $a^T \Sigma a = V(a^T X) \implies a^T X = c$. So, since for $x \in S$, where $S = \{x : a^T x = c\}$, $\lambda(S) = 0$, $\mu \perp \lambda$ has no absolutely continuous component, and since $\mu(\{x\}) = 0$ it has no discrete component $\implies \mu = \mu_{sc}$.

Example 2.3.10. $\Omega = \{0,1\}^\infty$ i.e. we toss a coin infinitely many times. If toss n is heads, then the player wins $\frac{1}{2^n}$. We denote by X the total winnings. Thus we can write

$$X_n = \begin{cases} 1 & \omega_n = 1 \\ 0 & \omega_n = 0 \end{cases} \qquad X = \sum_{n=1}^{\infty} \frac{X_n}{2^n}, \qquad 0 \leq X \leq 1$$

- If $p_0 = p_1 = \frac{1}{2}$ then $X \sim U[0,1]$ and so X is absolutely continuous
- if $p_0 \neq p_1$ then P_X is singular continuous, $P(X = x) = 0$ so it has also no discrete component, and it has no density

We are going to show that the probability distribution of X is singular continuous. Consider a real number $x \in [0,1]$ and its binary representation $x = \sum_{i=1}^{\infty} \frac{u_i}{2^i}$, $x = (u_1, u_2, \dots)$. On principle the representation is not unique, but if we discard terminating sequences it is. We have that $P(X = x) = P(X_i = u_i, i = 1, 2, \dots) = \prod_{i=1}^{\infty} P(u_i) = 0$, and so $\mu_D = 0$, and we want to show that also the absolutely continuous component is zero. We first need two results

- Every monotone function is almost everywhere differentiable and $F(b) - F(a) \geq \int_a^b F'(x)\lambda(dx)$
- If $F(x) = \int_{-\infty}^x f(u)\lambda(du)$, with $f(u) \geq 0$ and $\int_{-\infty}^x f(u)\lambda(du) < \infty$ then $F'(x) = f(x)$ λ-a.e.

Let us denote by $F = \mu(-\infty, x]$ and $F_{ac}(x) = \mu_{ac}(-\infty, x]$. Since $\mu = \mu_{ac} + \mu_{sc}$, $F_{ac}(b) - F_{ac}(a) \leq F(b) - F(a)$ $\forall a, b$. If $F'(x) = 0$ λ-a.e. $\implies$ $F'_{ac}(x) = 0$ λ-a.e., and since $F_{ac}(x) = \int_{-\infty}^x f d\lambda \implies f = 0$ λ-a.e., so $\mu_{ac} = 0$. We now need to show that $F'(x) = 0$. Consider x such that $\exists F'(x)$ and let k_n such that $\frac{k_n}{2^n} \leq x < \frac{k_n+1}{2^n}$. Then

$$\begin{aligned} F'(x) &= \lim_{n\to\infty} \frac{F\left(\frac{k_n+1}{2^n}\right) - F\left(\frac{k_n}{2^n}\right)}{\frac{1}{2^n}} = \lim_{n\to\infty} 2^n P(\frac{k_n}{2^n} \leq X < \frac{k_n+1}{2^n}) = \\ &= \lim_{n\to\infty} 2^n P(u_1)P(u_2)\dots P(u_n) \end{aligned}$$

if $F'(x) \neq 0$, $\lim_{n\to\infty} \frac{2^n P(u_1)P(u_2)\dots P(u_n)}{2^{n+1}P(u_1)P(u_2)\dots P(u_{n+1})} = \lim_{n\to\infty} \frac{1}{2P(u_{n+1})} = 1$, which implies $\lim_{n\to\infty} P(u_{n+1}) = \frac{1}{2}$, but this is not possible.

2.4 Independent Random Vectors

Theorem 2.4.1 (Fubini-Tonelli, Recap on measure theory)**.** *Let* $(\Omega, \mathcal{F}, \mu), (\Omega', \mathcal{F}', \mu')$ *measured spaces and consider the product space* $(\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}', \mu \times \mu')$ *where* $(\mu \times \mu')(B \times B') = \mu(B)\mu'(B)$*. Let* $g : (\Omega \times \Omega', \mathcal{F} \otimes \mathcal{F}') \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ *measurable. It can be proved that*

$$\int_{\Omega\times\Omega'} g(\omega, \omega')d(\mu \times \mu')(\omega.\omega') = \int_\Omega \int_{\Omega'} g(\omega, \omega')d\mu'(\omega')d\mu(\omega)$$

We are going to consider integrals with respect to the Lebesgue measure, we denote with λ_n the Lebesgue measure on $\mathcal{B}(\mathbb{R}^n)$, $\lambda_n = \lambda_1 \times \lambda_1 \times \cdots \times \lambda_1$. We'll be dealing with random vectors with different dimensions, so we drop the index and simply write λ.

Definition 2.4.2 (Independent random vectors)**.** The random vectors $X_1, \dots, X_n$ are said to be stochastically independent if $\sigma(X_1), \dots, \sigma(X_n)$ are independent

This means that the information we have on one random vector does not include information on the probability of other random vectors being in a Borel set, i.e.:

$$P(X_i \in B_i | X_1 \in B_1, \dots X_{i-1} \in B_{i-1}, X_{i+1} \in B_{i+1}, \dots, X_n \in B_n) = P(X_i \in B_i)$$

In fact, since $\sigma(X_i) = \{(X_i \in B_i) : B_i \text{ Borel set}\}$ and $\sigma(X_1)\dots\sigma(X_n) \perp\!\!\!\perp$, then $P(X_1 \in B_1, \dots, X_n \in B_n) = P(\bigcap_{i=1}^n (X_i \in B_i)) = \prod_{i=1}^n P(X_i \in B_i)$ $\forall B_1, \dots, B_n$.

If we denote by $P_{X_1,\dots,X_n}$ the probability distribution of $(X_1, \dots, X_n)$ and by P_{X_i} the probability distribution of X_i the above definition reads in terms of probabilities distribution as

$$P_{X_1,\dots,X_n}(B_1 \times \cdots \times B_n) = P_{X_1}(B_1)P_{X_2}(B_2)\dots P_{X_n}(B_n) = P_{X_1} \times P_{X_2} \times \cdots \times P_{X_n}(B_1 \times \cdots \times B_n).$$

This holds for every rectangle and rectangles are a determining class, so we can say that

$$X_1, \dots, X_n \perp\!\!\!\perp \iff P_{X_1,\dots,X_n} = P_{X_1} \times P_{X_2} \times \cdots \times P_{X_n}.$$

2.4.1 Criteria for independence

1. We can give a characterization in terms of distribution functions

$$\begin{aligned} F_{X_1,\dots,X_n}(x_1,\dots,x_n) &= P_{X_1,\dots,X_n}(S_{x_1,\dots,x_n}) = P_{X_1,\dots,X_n}(S_{x_1}\times\cdots\times S_{x_n}) \\ &= P_{X_1}\times P_{X_2}\times\cdots\times P_{X_n}(S_{x_1}\times\cdots\times S_{x_n}) \\ &= P_{X_1}(S_{x_1})P_{X_2}(S_{x_2})\dots P_{X_n}(S_{x_n}) \\ &= F_{X_1}(x_1)F_{X_2}(x_2)\dots F_{X_n}(x_n). \end{aligned}$$

So we can say that

$$X_1,\dots,X_n \perp\!\!\!\perp \iff F_{X_1,\dots,X_n}(x_1,\dots,x_n) = F_{X_1}(x_1)F_{X_2}(x_2)\dots F_{X_n}(x_n)$$

2. In the case where the probability distribution is absolutely continuous with respect to the Lebesgue measure, $P_{X_1,\dots,X_n} \ll \lambda$ we want to show that $P_{X_i} \ll \lambda$. We know that there exists a density function $f(x_1,\dots,x_n)$ such that $P_{X_1,\dots,X_n}(B) = \int_B(x_1,\dots,x_n)d\lambda(x_1,\dots,x_n)$. If we want to compute

$$\begin{aligned} &P_{X_i}(B_i) = P_{X_1,\dots,X_n}(\mathbb{R}^{d_1}\times\cdots\times B_i\times\cdots\times\mathbb{R}^{d_n}) \\ &= \int_{\mathbb{R}^{d_1}\times\cdots\times B_i\times\cdots\times\mathbb{R}^{d_n}} f(x_1,\dots,x_n)d\lambda(x_1,\dots,x_n) \\ &= \int_{B_i}\left[\int_{\mathbb{R}^{d_1}\times\cdots\times\mathbb{R}^{d_{i-1}}\times\mathbb{R}^{d_{i+1}}\times\cdots\times\mathbb{R}^{d_n}} f(x_1,\dots,x_n)d\lambda(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n)\right]d\lambda(x_i) \end{aligned}$$

and the term in square brackets is a function such that integrated in $d\lambda(x_i)$ we obtain P_{X_i}, so it is a density function of X_i. Note that in general the converse is not true, $P_{X_i} \ll \lambda \not\Rightarrow P_{X_1,\dots,X_n} \ll \lambda$

Example 2.4.3. X, Y random variable such that $X + Y = 1, \quad X, Y \sim U[0,1]$. $U[0,1]$ is absolutely continuous with respect to the Lebesgue measure but the joint distribution of X, Y is $P_{X,Y}$ is singular continuous.

Now suppose $P_{X_1,\dots,X_n} \ll \lambda$ and $X_1,\dots,X_n \perp\!\!\!\perp$, which means $P(X_1 \in B_1,\dots,X_n \in B_n) = P(X_1 \in B_1)\dots P(X_n \in B_n)$. Writing each side in terms of density functions we obtain

$$\int_{B_1\times\cdots\times B_n} f(x_1,\dots,x_n)d\lambda(x_1,\dots,x_n) = \int_{B_1\times\cdots\times B_n} f_{X_1}(x_1)\dots f_{X_n}(x_n)d\lambda(x_1)\dots d\lambda(x_n)$$

from which we can deduce that $f(x_1,\dots,x_n) = f_{X_1}(x_1)\dots f_{X_n}(x_n)$ λ-almost everywhere.
In this case ($X_1,\dots,X_n \perp\!\!\!\perp$) the converse is also true: if $P_{X_i} \ll \lambda\ \forall i = 1,\dots,n$ and $X_1,\dots,X_n \perp\!\!\!\perp$ then $P_{X_1,\dots,X_n} \ll \lambda$

$$\begin{aligned} &P_{X_1,\dots,X_n}(B_1\times\cdots\times B_n) = P_{X_1}(B_1)P_{X_2}(B_2)\dots P_{X_n}(B_n) = \\ &= \int_{B_1} f_{X_1}(x_1)d\lambda(x_1)\cdots\int_{B_n} f_{X_n}(x_n)d\lambda(x_n) = \\ &= \int_{B_1\times\cdots\times B_n} f_{X_1}(x_1)\dots f_{X_n}(x_n)d\lambda(x_1,\dots,x_n) \end{aligned}$$

so it exists the joint density of $X_1,\dots,X_n$ and it is $f_{X_1}(x_1)\dots f_{X_n}(x_n)$.

$$X_1,\dots,X_n \perp\!\!\!\perp \iff f(x_1,\dots,x_n) = \prod_{i=1}^n f_{X_i}(x_i)\ \lambda-\text{a.e.}$$

3. Suppose we have $P_{X_1,\dots,X_n}$ discrete, then there will be a probability mass function $m(x_1,\dots,x_n)$ and a set D countable at most such that $P_{X_1,\dots,X_n}(D) = 1$ and $P_{X_1,\dots,X_n}(\{(x_1,\dots,x_n)\}) > 0$ $(x_1,\dots,x_n) \in D$. Then, P_{X_i} is also discrete and $D_i = \pi_i^{-}(D)$ where π_i projections and

$$m_{X_i}(x_i) = \sum_{(x_1,\dots,x_{i-1},x_{i+1},\dots,x_n):(x_1,\dots,x_n)\in D} m(x_1,\dots,x_n)$$

$P_{X_1,\dots,X_n}$ is discrete if and only if P_{X_i} are discrete and $D \subset D_1 \times \cdots \times D_n$. It is an if and only if because projections cannot be singular continuous, while in the absolutely continuous case this can happen.

$$X_1,\dots,X_n \perp\!\!\!\perp \iff m(x_1,\dots,x_n) = m(x_1)m(x_2)\dots m(x_n)$$

Definition 2.4.4 (Independence for a family of r.v.'s)**.** If you have a family of random vectors $\{X_t\}_{t\in T}$ they are independent if and only if $\forall n\ t_1,\dots,t_n \in T\ X_{t_1}\dots X_{t_n} \perp\!\!\!\perp$

Example 2.4.5 (Gaussian white noise). $\{X_t\}_{t\geq 0} \perp\!\!\!\perp\ \forall t,\ X_t \sim \mathcal{N}(0,\sigma^2)$

If $\{X_t\}_{t\in T} \perp\!\!\!\perp$ and we take a sub family $T' \subset T$ then $\{X_t\}_{t\in T'} \perp\!\!\!\perp$, you never loose independence reducing the family.

Theorem 2.4.6 (Disjoint block independence for r.v.'s)**.** *Let the following be an array of independent random vectors*

$$\begin{matrix} X_{11} & X_{12} & \dots \\ X_{21} & X_{22} & \\ \vdots & \vdots & \ddots \end{matrix}$$

Let $\mathcal{G}_i = \sigma(\{X_{ij}\}_{j\in\mathbb{N}})$*. Then* $\{\mathcal{G}_n\}_{n\geq 1}$ *are independent.*

Remark 2.4.7. If we take $T_i = t_i(X_{i1}\dots X_{ik_1})$ function of some random vectors in row i, we have that this is measurable with respect to $\mathcal{G}_i$, which implies $\sigma(T_i) \subset \mathcal{G}_i \implies \{\sigma(T_i)\}_{i\geq 1} \perp\!\!\!\perp \implies \{T_i\}_{i\geq 1} \perp\!\!\!\perp$

Theorem 2.4.8 (Kolmogorow 0-1 law)**.** *If* $X_{k_1}, X_{k_2},\dots$ *are* $\perp\!\!\!\perp$ *and if* $A \in \mathcal{T}_k(\{X_{k_1}, X_{k_2},\dots\})$ *then* $P(A) = 0$ *or* $P(A) = 1$*.*

Corollary 2.4.9 (Constancy of a r.v. in the tail σ-algebra)**.** *If a random variable* Y *is measurable with respect to* $\mathcal{T}(\{X_n\})$ *and* $\{X_n\} \perp\!\!\!\perp$ *then there exists* $c \in \mathbb{R}$ *such that* $Y = c$ *almost surely.*

Proof. $P(Y \leq y) = 0$ or $P(Y \leq y) = 1\ \forall y$ □

Example 2.4.10. If (X_n) i.i.d. sequence and $|X_n| \leq C$, then:

$$\limsup_n X_n \overset{a.s.}{=} const. = \textit{ess-sup}\ P_{X_1}$$

Where $s = \textit{ess-sup}\ P_x$ if for any $\epsilon > 0$, $P_X(s+\epsilon,+\infty) = 0$ and $P_X(s-\epsilon,s) > 0$

2.5 Functions and transformations

Definition 2.5.1 (Riemann integral in $\mathbb{R}$)**.** Let $g : [a, b] \to \mathbb{R}$ be a measurable function. The Riemann integral, if it exists is a number r satisfying $\forall \epsilon > 0\ \exists \delta > 0$ such that for every finite partition $\{I_j : j \in J\}$ of $[a, b]$ with I_j intervals and $\lambda(I_j) < \delta$ and for every $x_j \in I_j\ (j \in J)$.

$$\left| r - \sum_j g(x_j)\lambda(I_j) \right| < \epsilon$$

Not all measurable functions are Riemann integrable. Note that Riemann integrability requires the regularity of the function while Lebesgue not. However, Riemann integrability implies Lebesgue one, and the integrals coincide.

Remark 2.5.2 (Riemann integrability of continuous functions). Continuous functions are Riemann integrable by Fundamental Theorem of calculus, and we have that, if G is continuously differentiable on $[a, b]$ with derivative g, then

$$\int_a^b g\ dx = G(b) - G(a)$$

Proposition 2.5.3 (Radon Nikodym derivative of λT^{-1} with respect to λ)**.** *Let U, V be two open subsets of $\mathbb{R}^k$, and let $T : U \to V$ be one to one, continuously differentiable with det $J_{T^{-1}} \neq 0$ on V, where $J_{T^{-1}}$ is the Jacobian matrix of T^{-1} then*

$$\frac{d(\lambda T^{-1})}{d\lambda} = |det\ J_{T^{-1}}|$$

Definition 2.5.4 (Integrals with respect the counting measure on $\mathbb{R}^k$)**.** Let $D = \{x_1, x_2, \dots\}$ be a countable subsets of $\mathbb{R}^k$ and let μ_c be the counting measure on D:

$$\mu_c(A) = \underbrace{|A \cap D|}_{Number of Points}$$

and in particular $\mu_c(\{x_i\}) = 1 \quad \forall x_i \in D$. Let $g : \mathbb{R}^k \to \mathbb{R}^+$ be a measurable function

$$\int g(x)\mu_c(dx) = \int_D g(x)\mu_c(dx) = \lim_{n\to\infty} \sum_{i=1}^{n} g(x_i) = \sum_{i=1}^{\infty} g(x_i)$$

Note that for a general function we can work with g_+, g_-.

Proposition 2.5.5 (Function of independent random variables)**.** *Let X, Y be random variables with $X \perp\!\!\!\perp Y$ and $X \sim \mu$, $Y \sim \nu$. The probability distribution of (X, Y) is $P((X, Y) \in B) = (\mu \times \nu)(B)$. For fixed x let $B_x = \{y : (x, y) \in B\}$ is the section of B in x. Then*

$$\begin{aligned} P((X, Y) \in B) &= \int 1_B(x, y)d(\mu \times \nu)(x, y) \\ &= \int \left(\int 1_B(x, y)d\nu(y) \right) d\mu(x) \\ &= \int \nu(B_x)d\mu(x). \end{aligned}$$

Example 2.5.6. X, Y random variables, i.i.d. and with density $f(x) = \alpha e^{-\alpha x} \mathbb{1}_(0, \infty)(x)$, with $\alpha > 0$. We want to find the probability distribution of $Z = \frac{Y}{X}$. $Z > 0$. Fix $z > 0$, then

$$\begin{aligned} P(Z > z) = P(\frac{Y}{X} > z) &= \int_0^\infty P(\frac{Y}{X} > z) \alpha e^{-\alpha x} dx \\ &= \int_0^\infty P(Y > zx) \alpha e^{-\alpha x} dx = \int_0^\infty e^{-\alpha z x} \alpha e^{-\alpha x} dx \\ &= \alpha \int_0^\infty e^{-\alpha x (z+1)} dx = \frac{1}{z+1} \end{aligned}$$

while for $z < 0$ $P(\frac{Y}{X} > z) = 1$. So

$$f_Z(z) = \frac{d}{dz}(1 - P(Z > z)) = \begin{cases} 0 & z < 0 \\ \frac{1}{(z+1)^2} & z > 0 \end{cases}$$

and

$$F_Z(z) = \begin{cases} 0 & Z \leq 0 \\ 1 - \frac{1}{z+1} & Z > 0 \end{cases}$$

Definition 2.5.7 (Convolution)**.** Let μ, ν be finite measures on $\mathcal{B}(\mathbb{R}^k)$. The convolution between μ and ν is

$$(\mu * \nu)(A) = \int \nu(A - x) \mu(dx)$$

where $A - x = \{y - x : y \in A\}$.

$\mu * \nu$ is a finite measure and $(\mu * \nu)(\mathbb{R}^k) = \mu(\mathbb{R}^k)\nu(\mathbb{R}^k)$. If μ, ν are probability measures then $\mu * \nu$ is a probability measure. There are some properties:

1. $\mu * \nu = \nu * \mu$
2. $(\nu * \mu) * \eta = \nu * (\mu * \eta)$
3. If $F(x) = \mu(-\infty, x]$ and $G(x) = \nu(-\infty, x]$ then $(\mu * \nu)(-\infty, x] = (F * G)(x)$ where
$$(F * G)(x) = \int F(x - y) G(dy)$$
4. If $\nu \ll \eta$ and $\mu \ll \eta$ with $f = \frac{d\nu}{d\eta}$ and $g = \frac{d\mu}{d\eta}$, then $\mu * \nu \ll \eta$ and
$$\frac{d(\mu * \nu)}{d\eta}(x) = \int f(x - y) g(y) \eta(dy)$$

Proposition 2.5.8 (Sum of independent r.v.'s)**.** *Let $X \perp\!\!\!\perp Y$ be random vectors such that $X \sim \mu$ and $Y \sim \nu$. Then $X + Y \sim \mu * \nu$.*

Proof.

$$P((X + Y) \in B) = \int P(Y \in B - x) dP_X(x) = \int \nu(B - x) d\mu(x)$$

□

Example 2.5.9. X, Y i.i.d. $\sim$ Poisson(λ). Then

$$m(x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, \ldots$$

$$\begin{aligned} m_{X+Y} &= \sum_x m(z-x)m(x) = \sum_{x=0}^{z} \frac{e^{-\lambda}\lambda^{z-x}}{(z-x)!}\frac{e^{-\lambda}\lambda^x}{x!} = e^{-2\lambda}\lambda^x \sum_{x=0}^{z} \frac{1}{(z-x)!x!} \\ &= \frac{e^{-2\lambda}\lambda^z}{z!}\sum_{x=0}^{z}\frac{z!}{(z-x)!x!}1^x 1^{z-x} = \frac{e^{-2\lambda}(2\lambda)^z}{z!} \end{aligned}$$

That is to say $X + Y \sim$ Poisson(2λ)

Theorem 2.5.10 (Transformations of random vectors)**.** *Let X be a random vector with absolutely continuous distribution and density f_X. Let U be an open set such that $f_X(x) > 0$ for $x \in U$. Let $g : U \to V$ be a one to one, continuously differentiable function such that*

$$\det(J_{g^{-1}}) \neq 0$$

Then $Y = g(X)$ is absolutely continuous with density

$$f_Y(y) = f_X(g^{-1}(y)) \left|\det(J_{g^{-1}}(y))\right| \mathbb{1}_V(y)$$

Proof. Let $B \subset V$

$$\begin{aligned} P(Y \in B) = P(g(X) \in B) = P(X \in g^{-1}(B)) &= \int_U \mathbb{1}_{g^{-1}(B)}(x) f_X(x) d\lambda(x) \\ &= \int_U \mathbb{1}_B(g(X)) f_X(g^{-1}(g(x))) d\lambda(x) = \int_V \mathbb{1}_B(y) f_X(g^{-1}(y)) d(\lambda g^{-1})(y) \\ &= \int_V \mathbb{1}_B(y) f_X(g^{-1}(y)) \left|\det(J_{g^{-1}}(y))\right| d\lambda(y) \end{aligned}$$

□

Example 2.5.11. $X, Y \perp\!\!\!\perp\!\sim \varepsilon(\alpha)$, $Z = X + Y$, $W = \frac{X}{X+Y}$. Find $P_{(Z,W)}$

$$f_{(Z,W)}(z, w) = f_{(X,Y)}(x(z,w), y(z,w))|\det(J_{g^{-1}}(z,w))|\mathbb{1}_V(z,w)$$

We start by looking for U, an open set such that

$$P((X,Y) \in U) = 1 \quad V = g(U) \quad (z,w) = g(x,y)$$

$$f_X(x) = \alpha e^{-\alpha x}\mathbb{1}_{(0,\infty)}(x) \quad X \in (0,\infty) \text{ and the same holds for } Y$$

So we take $U = (0,\infty) \times (0,\infty)$. Now we look for $g^{-1}(z,w)$.

$$\begin{cases} z = x + y & y = z(1-w) \\ w = \frac{x}{x+y} & x = wz \end{cases}$$

And this is a 1-1 function. Also, $z > 0$ and $0 < x < 1$ as a consequence of $x > 0$ and $y > 0$.

$$\det(J_{g^{-1}}) = -wz - (1-w)z = -z$$

Therefore $z > 0 \implies |\det(J_{g^{-1}}(z,w))| = z$. We obtain the joint density

$$f_{XY}(x,y) = \alpha e^{-\alpha x}\alpha e^{-\alpha y}\mathbb{1}_{(0,\infty)}(x)\mathbb{1}_{(0,\infty)}(y) = \alpha^2 e^{-\alpha(x+y)}\mathbb{1}_U(x,y)$$

$$f_{ZW}(z,w) = \alpha^2 z e^{-\alpha z}\mathbb{1}_{(0,\infty)}(z)\mathbb{1}_{(0,1)}(w)$$

We see that $Z \perp\!\!\!\perp W$, $W \sim U(0,1)$ and $Z \sim \Gamma(2,\alpha)$.

2.6 Convergence of sequence of random variables

In this section we have $\{X_n\}_{n\in\mathbb{N}}$ sequence of random variables, i.e. functions $X_n(\omega), \omega \in \Omega$ defined on an underlying probability space $(\Omega, \mathcal{F}, P)$.

Definition 2.6.1 (Almost sure convergence)**.** We say that X_n converges to X almost surely, and write $X_n \xrightarrow{as} X$, if $P(\{\omega \in \Omega : X_n(\omega) \to X(\omega)\}) = 1$.

Definition 2.6.2 (Convergence in probability)**.** We say that X_n converges to X in probability, and write $X_n \xrightarrow{p} X$, if for all $\varepsilon > 0$ $P(|X_n - X| > \varepsilon) \to 0$ as $n \to \infty$.

Starting from the definition of almost sure convergence we can say that

$$\begin{aligned} P(X_n(\omega) \to X(\omega)) = 1 &\iff P(\{\omega \in \Omega : \forall\varepsilon > 0 \ \exists n_0 : \forall n \geq n_0 \ |X_n - X| < \varepsilon\}) = 1 \\ &\iff P\left(\left\{\omega \in \Omega : \forall k \ \exists n_0 : \forall n \geq n_0 \ |X_n(\omega) - X(\omega)| < \frac{1}{k}\right\}\right) = 1 \\ &\iff P\left(\bigcap_{k=1}^{\infty}\bigcup_{n_0=1}^{\infty}\bigcap_{n=n_0}^{\infty} |X_n - X| < \frac{1}{k}\right) = 1 \end{aligned}$$

Since in general the probability of a countable intersection can be one if and only if the probability of each event is one

$$P\left(\bigcap_{k=1}^{\infty} A_k\right) = 1 \iff P(A_k) = 1 \ \forall k$$

in our case we take

$$A_k = \bigcup_{n_0=1}^{\infty}\bigcap_{n=n_0}^{\infty}\left\{|X_n - X| < \frac{1}{k}\right\}$$

therefore

$$X_n \xrightarrow{as} X \iff \forall k \ P\left(\bigcup_{n_0=1}^{\infty}\bigcap_{n=n_0}^{\infty}\left\{|X_n - X| < \frac{1}{k}\right\}\right) = 1$$

is the same as

$$\forall\varepsilon > 0 \quad P(\liminf |X_n - X| < \varepsilon) = 1$$

or alternatively

$$\forall\varepsilon > 0 \quad P(\limsup |X_n - X| > \varepsilon) = 0$$

Therefore we can say

$$X_n \xrightarrow{as} X \iff \forall\varepsilon > 0 \ P(\limsup |X_n - X| > \varepsilon) = 0$$

Finally note that since $\limsup(P(A_n)) \leq P(\limsup(A_n))$, we have

$$X_n \xrightarrow{as} X \implies P(|X_n - X| > \varepsilon) \to 0$$

In other words, almost sure convergence implies convergence in probability, so convergence in probability is a necessary condition to almost sure convergence. There is also a sufficient condition. We collect these observation in the following proposition:

Proposition 2.6.3 (Properties of a.s. convergence and convergence in probability)**.** *Given a sequence (X_n) and X of r.v. on $(\Omega, \mathcal{F}, P)$. Then:*

- *(Characterization of a.s.-convergence)*

$$X_n \xrightarrow{as} X \iff P(\limsup |X_n - X| > \varepsilon) = 0 \ \forall \varepsilon$$

- *(Necessary condition for a.s.-convergence)*

$$X_n \xrightarrow{as} X \implies X_n \xrightarrow{p} X$$

- *(Sufficient condition, corollary of Borel-Cantelli)*

$$\sum_{n=1}^{\infty} P(|X_n - X| > \varepsilon) < \infty \implies P(\limsup |X_n - X| > \varepsilon) = 0 \ \forall \varepsilon \iff X_n \xrightarrow{as} X$$

Example 2.6.4. Let $\{X_n\} \perp\!\!\!\perp$ be such that

$$X_n = \begin{cases} 0 & 1 - \frac{1}{n^2} \\ 1 & \frac{1}{n^2} \end{cases}$$

Then notice that $X_n \xrightarrow{p} 0$ because $P(|X_n - 0| > \varepsilon) = P(X_n > \varepsilon) = \frac{1}{n^2} \to 0$ and $\sum P(X_n > \varepsilon) = \sum \frac{1}{n^2} < \infty$, so X_n converges almost surely to X.

Example 2.6.5. Let $\{X_n\} \perp\!\!\!\perp$ be such that

$$X_n = \begin{cases} 0 & 1 - \frac{1}{n} \\ 1 & \frac{1}{n} \end{cases}$$

Then notice that $X_n \xrightarrow{p} 0$ because $P(|X_n - 0| > \varepsilon) = P(X_n > \varepsilon) = \frac{1}{n} \to 0$. Does $X_n \xrightarrow{as} 0$? We need $P(\limsup |X_n - X| > \varepsilon) = 0$. But here, we would like that $P(\limsup\{X_n > \varepsilon\}) = 0$. However, note that by BC2, $\sum P(X_n > \varepsilon) = \sum \frac{1}{n}$ diverges, hence $P(\limsup\{X_n > \varepsilon\}) = 1$. So X_n does not converge almost surely to X.

Theorem 2.6.6 (Equivalent definition to convergence in probability)**.** $X_n \xrightarrow{p} X \iff \forall (n') \subseteq (n)$ $\exists$ *a subsequence* $(n'') \subseteq (n')$ *such that* $X_{n''} \xrightarrow{as} X$

Proof. ($\implies$) Let $X_n \xrightarrow{p} X$ i.e. $\forall \varepsilon > 0$ $P(|X_n - X| > \varepsilon) \to 0$, which means:

$$\forall \varepsilon > 0, \ \forall \delta > 0, \ \exists n_0 : \forall n \geq n_0 \ \ P(|X_n - X| > \varepsilon) < \delta$$

For any sequence (n'), for any $k \in \mathbb{N}$ take $\varepsilon = \frac{1}{k}$, $\delta = \frac{1}{2^k}$, then we can find $n_k \in (n'), n_k > n_{k-1}$: $P(|X_{n_k} - X| > \frac{1}{k}) < \frac{1}{2^k}$, and we call $(n'') = (n_k)$. Therefore

$\sum_{k=1}^{\infty} P(|X_{n_k} - X| > \frac{1}{k}) = \sum_{k=1}^{\infty} \frac{1}{2^k} < \infty$ and by Borel Cantelli 1 $P(\{|X_{n_k} - X| > \frac{1}{k} \text{ i.o.}\}) = 0$. Now we should have ε fixed instead of $\frac{1}{k}$, but notice that for k large enough $\frac{1}{k} < \varepsilon$, so $(\{|X_{n_k} - X| > \varepsilon \text{ i.o.}\}) \subset (\{|X_{n_k} - X| > \frac{1}{k} \text{ i.o.}\})$ therefore $P(\{|X_{n_k} - X| > \varepsilon \text{ i.o.}\}) \leq P(\{|X_{n_k} - X| > \frac{1}{k} \text{ i.o.}\}) = 0$.

($\Longleftarrow$) Let's suppose that $X_n \not\to_p X$. Then (look at equation above):

$$\exists \varepsilon > 0,\ \exists \delta > 0,\ \exists (n') : P(|X_{n'} - X| > \varepsilon) > \delta\ \ \forall n' \in (n')$$

In particular this will be true for any subsequence of (n'):

$$\forall (n'') \subseteq (n')\ \ P(|X_{n''} - X| > \varepsilon) > \delta\ \ \forall n'' \in (n'')$$

This means that $X_{n''} \not\to_p X \implies X_{n''} \not\to_{as} X$. This contradicts the assumption of the existence of an a.s.-converging subsequence. □

There are two useful applications of this theorem:

Proposition 2.6.7 (Uniqueness of the limit in probability). *The limit in probability is essentially unique: if $X_n \xrightarrow{p} X$ and $X_n \xrightarrow{p} Y$ then $X = Y$ a.s.*

Proof. $\exists (n')$ and (n'') such that $X_{n'} \xrightarrow{as} X$ and $X_{n''} \xrightarrow{as} Y$, but along the subsequence the limit is the same so $X_{n''} \xrightarrow{as} Y$. This means $X = Y$ a.s. □

Proposition 2.6.8 (Convergence in probability through continuous functions). *If $X_n \xrightarrow{p} X$ and f is a continuous function then $f(X_n) \xrightarrow{p} f(X)$.*

Proof. Fix $(n'') \subseteq (n')$ such that $X_{n''} \xrightarrow{as} X$. Then $f(X_{n''}) \xrightarrow{as} f(X)$. Using the above result, we have found a sub-sequence converging a.s. to $f(X)$ which means that $f(X_n) \xrightarrow{p} f(X)$. □

Recap on measure theory

Let $(\Omega, \mathcal{F}, \mu)$ a measured space, μ sigma-finite, $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We define the integral of g on Ω with respect to the measure μ, $\int_\Omega g d\mu$ as

- if $g = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ then $\int_\Omega g d\mu = \sum_{i=1}^{n} a_i \mu(A_i)$
- if $g \geq 0$ it exists a non-decreasing succession g_n of simple, non-negative such that $g_n \nearrow g$, then $\int_\Omega g d\mu = \lim_{n\to\infty} \int_\Omega g_n d\mu$
- if g is integrable, $g = g_+ - g_-$, then $\int_\Omega g d\mu = \int_\Omega g_+ d\mu - \int_\Omega g_- d\mu$

Definition 2.6.9 (Integrable function). g is said to be integrable if $\int_\Omega g_+ d\mu < +\infty$ and $\int_\Omega g_- d\mu < +\infty$, that is $\iff \int_\Omega |g| d\mu < +\infty$.

Proposition 2.6.10 (Properties of the integral).

- *Monotonicity:* $g \leq h \implies \int_\Omega g d\mu \leq \int_\Omega h d\mu$
- *Linearity:* $\int_\Omega (ag + bh) d\mu = a \int_\Omega g d\mu + b \int_\Omega h d\mu$
- *Linearity with respect to the measure:* $\int_\Omega g d(\alpha\mu + \beta\nu) = \alpha \int_\Omega g d\mu + \beta \int_\Omega g d\nu$

Definition 2.6.11 (push-forward measure)**.** Let $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$ be measurable spaces, $g : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ measurable function. Let μ be a sigma finite measure on $\mathcal{F}$. The measure induced by g on $\mathcal{F}'$ is defined as

$$\mu'(B) = \mu(g^{-1}(B)).$$

It can be proved that μ' is indeed a sigma-finite measure. We write $\mu' = \mu g^{-}1$.

Proposition 2.6.12 (Change of variables)**.** *Let $(\Omega, \mathcal{F}), (\Omega', \mathcal{F}')$ be measurable spaces, $g : (\Omega, \mathcal{F}) \to (\Omega', \mathcal{F}')$ measurable function, $h : (\Omega', \mathcal{F}') \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ measurable function. Let μ be a sigma finite measure on $\mathcal{F}$. It can be proved that*

$$\int_{\Omega'} h(\omega') d(\mu g^{-1}(\omega') = \int_{\Omega} h(g(\omega)) d\mu(\omega).$$

Proposition 2.6.13 (Change of measure)**.** *Let $(\Omega, \mathcal{F}, \mu)$ be a measured space, ν a sigma-finite measure $\nu \ll \mu$, $g : (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ measurable. It can be proved that*

$$\int g d\nu = \int g \frac{d\nu}{d\mu} d\mu$$

and denoting by $f = \frac{d\nu}{d\mu}$ the density (Radon-Nikodym derivative) of ν with respect to μ

$$\int g d\nu = \int g \cdot f d\mu.$$

Let $(\Omega, \mathcal{F}, \mu)$ be a measured space, (g_n) be a sequence of measurable functions $g_n \to g$ μ-a.e.

Proposition 2.6.14 (Fatou lemma)**.** *If $g_n \leq 0$ then*

$$\int \liminf g_n d\mu \leq \liminf \int g_n d\mu.$$

Theorem 2.6.15 (Monotone convergence)**.** *If $g_n \leq 0$ and $g_n \leq g_{n+1}$ μ-a.e., then*

$$\lim_{n\to\infty} \int g_n d\mu = \int g d\mu.$$

Theorem 2.6.16 (Dominated convergence)**.** *If there exists h measurable such that $\forall n$ $|g_n| \leq h$ μ-a.e. and h is integrable then*

$$\lim_{n\to\infty} \int g_n d\mu = \int g d\mu.$$

Remark 2.6.17. In monotone convergence theorem the limit can also be $+\infty$ while in dominated convergence theorem the limit is always finite.

2.7 Expectation of Random Variable

Definition 2.7.1 (Expectation of a random variable)**.** Let $(\Omega, \mathcal{F}, P)$ probability space, X random variable, we define the expectation of X as $E(X) = \int_{\Omega} X dP$.

The expectation has the same properties of the Lebesgue integral, thus it is monotone and linear and also linear with respect to the measure.

- if $X = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}$ then $E(X) = \sum_{i=1}^{n} a_i P(A_i)$
- if $X \geq 0$ then $\exists X_n$ sequence of simple, positive, $X_n \to X$ then $E(X) = \lim_{n\to\infty} E(X_n)$. This can converge or be ∞.
- $X = X^+ - X^-$ then $E(X) = E(X^+) - E(X^-)$ provided at least one is finite.

and we say that X is integrable if $E(X^+) < \infty$ and $E(X^-) < \infty \iff E(|X|) < \infty$. Denoting by $L^1(\Omega, \mathcal{F}, P) = \{X : E(|X|) < \infty\}$ we say X integrable $X \in L^1$.

We would like to be able to compute the expectation of a function of a random vector, which we can do applying the change of variable formula. For instance, let $(\Omega, \mathcal{F}, P)$ be a probability space, X a random vector of dimension k with probability distribution μ which is a PM on $\mathcal{B}(\mathbb{R}^k)$. Take $g : (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k)) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then we have

$$E(g(X)) = \int_\Omega g(X(\omega))P(d\omega) = \int_{\mathbb{R}^k} g(x)\mu(dx)$$

since the measure induced by X on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is $PX^{-1} = \mu$. How can we compute $\int_{\mathbb{R}^k} g(x)\mu(dx)$? We know that μ can be decomposed as $\mu = \mu_{ac} + \mu_D + \mu_{sc}$, so

$$\int g d\mu = \int g d\mu_{ac} + \int g d\mu_D + \int g d\mu_{sc}$$

- $\mu_{ac} \ll \lambda$, call $f = \frac{d\mu_{ac}}{d\lambda}$, so that $\int g d\mu_{ac} = \int_{\mathbb{R}^k} g f d\lambda$
- $D = \{x : \mu(\{x\}) > 0\}$, $\mu_D \ll \lambda_c$ where λ_c is the counting measure on D, $\lambda_c(A) = |A \cap D|$ and $\frac{d\mu_D}{d\lambda_c}(x) = m(x)$. Then $\int g d\mu_D = \int g(x)m(x)d\lambda_c(x) = \sum_{x\in D} g(x)m(x)$ because the only points that matter are the ones where λ_c puts mass so λ_c is almost everywhere equal to a simple function

At the end we can say

$$E(g(X)) = \int_{\mathbb{R}^k} g(x)f(x)d\lambda(x) + \sum_{x\in D} g(x)m(x) + \int g(x)d\mu_{sc}(x)$$

Suppose we have two random vectors $X_1, X_2 \perp\!\!\!\perp$, $\mu_{X_1X_2} = \mu_{X_1} \times \mu_{X_2}$. Then

$$\begin{aligned}
E(g_1(X_1)g_2(X_2)) &= \int g_1(x_1)g_2(x_2)d(\mu_{X_1} \times \mu_{X_2})(x_1, x_2) \\
&= \int \left[\int g_1(x_1)g_2(x_2)d\mu_{X_2}(x_2)\right] d\mu_{X_1}(x_1) \\
&= \int g_1(x_1) \left[\int g_2(x_2)d\mu_{X_2}(x_2)\right] d\mu_{X_1}(x_1) \\
&= \int g_2(x_2)d\mu_{X_2}(x_2) \int g_1(x_1)d\mu_{X_1}(x_1) \\
&= E(g_2(X_2))E(g_1(X_1))
\end{aligned}$$

2.7.1 Expectation and limits

In general $X_n \xrightarrow{as} X$ does not imply $E(X_n) \to E(X)$, but in the following we are going to see some theorems to handle limits.

Proposition 2.7.2 (Fatou's Lemma). *If $X_n \geq 0$ then $E(\liminf X_n) \leq \liminf E(X_n)$.*

Application 2.7.3. $X_n \geq 0$, $X_n \in L^1\ \forall n$, $X_n \xrightarrow{as} X$, under which conditions can we say that $X \in L^1$? If $X = \liminf X_n$ then $E(X) \leq \liminf E(X_n)$, so if $\liminf E(X_n) < +\infty$ then $X \in L^1$. More in general, $X_n \in L^1\ \forall n$, $X_n \xrightarrow{as} X$, $|X_n| \xrightarrow{as} |X|$ then by Fatou lemma we can say $E(|X|) \leq \liminf E(|X_n|)$, so if the latter is $< +\infty$ then $X \in L^1$.

Proposition 2.7.4 (Monotone convergence)**.** *If $X_n \geq 0$ and $X_{n+1} \geq X_n$ a.s. $\forall n$ and $X_n \xrightarrow{as} X$ then $E(X_n) \to E(X)$.*

Application 2.7.5. To see an application of monotone convergence we will prove the following formula: if $X \geq 0$ then

$$E(X) = \int_0^\infty (1 - F(t))dt$$

where F is the distribution function of X.

Proof.

- First suppose X is simple and takes values $x_1, \ldots, x_n$ with $0 \leq x_1 < \cdots < x_n$. Then X is a discrete random variable and

$$\begin{aligned}
E(X) &= \sum_{i=1}^n x_i P(X = x_i) = \sum_{i=1}^n \int_0^{x_i} 1 dt P(X = x_i) \\
&= \sum_{i=1}^n \int_0^{x_i} P(X = x_i)dt = \sum_{i=1}^n \int_0^\infty \mathbb{1}_{[0,x_i]}(t)P(X = x_i)dt \\
&= \int_0^\infty \sum_{i=1}^n \mathbb{1}_{[0,x_i]}(t)P(X = x_i)dt \\
&= \int_0^\infty \sum_{i=1}^\infty \mathbb{1}_{[t,\infty)}(x_i)P(X = x_i)dt \\
&= \int_0^\infty P(X \geq t)dt
\end{aligned}$$

  Recall $P(X \geq t) = P(X > t)$ a.e. with respect to the Lebesgue measure because $P(X \geq t) \neq P(X > t) \iff P(X = t) > 0$ and there can be an at most countable number of points where this is true – otherwise $P(\Omega) > 1$. Therefore the integrals are the same.

$$E(X) = \int_0^\infty P(X \geq t)dt = \int_0^\infty P(X > t)dt = \int_0^\infty (1 - F(t))dt$$

- Now with a more general X such that $X \geq 0$, then $\exists \{X_n\}_{n \geq 1}$ a sequence of increasing and simple random variables such that $X_n \to X\ \forall \omega$. From the above point,

$$E(X_n) = \int_0^\infty (1 - F_n(t))dt = \int_0^\infty P(X_n > t)dt$$

  On the left hand side, we apply the monotone convergence theorem: $E(X_n) \to E(X)$. On the right hand side, $P(X_n > t) = E(\mathbb{1}_{(t,\infty)}(X_n))$ where $n \to \infty$ and t is fixed. Let's take t such that $P(X = t) = 0$. Then $\mathbb{1}_{(t,\infty)}(\cdot)$ is continuous on a set O such that $P(X \in O) = 0$. Also, $\mathbb{1}_{(t,\infty)}(X_n) \leq \mathbb{I}_{(t,\infty)}(X_{n+1})$ because

$\{X_n\}_{n\geq 1}$ is increasing so $X_n > t \implies X_{n+1} > t$. By the continuity of $\mathbb{1}_{(t,\infty)}(\cdot)$, $\mathbb{1}_{(t,\infty)}(X_n) \to \mathbb{1}_{(t,\infty)}(X)$. By the monotone convergence theorem we have:

$$E(\mathbb{1}_{(t,\infty)}(X_n)) \to E(\mathbb{1}_{(t,\infty)}(X))$$

because $E(\cdot)$ preserves the monotonicity. Therefore $1 - F_n(t) \to 1 - F(t)$ $\forall t$: $P(X = t) = 0$ and this set is at most countable. On the left hand side, since $P(X_n > t) \to P(X > t)$, we obtain

$$\int_0^\infty P(X_n > t)dt \to \int_0^\infty P(X > t)dt = \int_0^\infty (1 - F(t))dt$$

The limits on the right hand side and the left hand side must be the same, thus

$$E(X) = \int_0^\infty (1 - F(t))dt$$

□

Alternative proof. By Fubini's theorem (everything is non-negative),

$$\int_0^\infty P(X > x)dx = \int_0^\infty \int_\Omega \mathbb{1}_{(X(\omega)>x)}dP(\omega)dx = \int_\Omega \int_0^\infty \mathbb{1}_{(X(\omega)>x)}dx\, dP(\omega) =$$
$$= \int_\Omega X(\omega)dP(\omega) = E(X).$$

□

Proposition 2.7.6 (Dominated convergence)**.** *If* $X_n \xrightarrow{as} X$ *and* $\exists Y$ *integrable such that* $|X_n| \leq Y$ $\forall n$ *a.s., then* $E(X_n) \to E(X)$.

2.8 Moment generating function

Definition 2.8.1 (Moment generating function of X)**.** Let X be a random variable and $I = \{s \in \mathbb{R} : E(e^{sX}) < \infty\}$ a set of real number. For $s \in I$ we define $M(s) = E(e^{sX})$ the moment generating function of X.

Note that $0 \in I$, and I is convex, so it is basically an interval.

Theorem 2.8.2 (Taylor expansion at 0 of a MGF)**.** *If* $\exists s_0 > 0$ *such that* $[-s_0, s_0] \subseteq I$ *then all the moments of* X *are finite* $\forall k$ $E(|X|^k) < +\infty$ *and*

$$M(s) = \sum_{k=0}^\infty \frac{s^k E(X^k)}{k!}$$

for $s \in [-s_0, s_0]$.

If the function can be written in this way it is analytic and the expression coincides with the Taylor expansion in zero $E(X^k) = M^{(k)}(0)$.

Proof. Fix k and consider $\frac{|s|^k|X|^k}{k!} \leq \sum_{j=0}^{\infty} \frac{|s|^j|X|^j}{j!} = e^{|sX|} \leq e^{|s_0X|}$ and $E(e^{|s_0X|}) \leq E(e^{s_0X} + e^{-s_0X}) < +\infty$ so $E(|X|^k) < +\infty$. The goal is to show that we can exchange the limit and the integral is to find a dominating random variable Y that is integrable to apply the dominated convergence theorem. We have:

$$\left|\sum_{k=0}^{n} \frac{s^k X^k}{k!}\right| \leq \sum_{k=0}^{n} \frac{|s|^k|X|^k}{k!} \leq \sum_{k=0}^{\infty} \frac{|s|^k|X|^k}{k!} = e^{|sX|} \leq e^{|s_0X|}$$

Therefore $e^{|s_0X|}$ is dominating, and it is integrable. Therefore $Y = e^{|sX|}$ is integrable, and by the dominated convergence theorem:

$$E\left(\sum_{k=0}^{n} \frac{s^k X^k}{k!}\right) = \sum_{k=0}^{n} \frac{s^k E\left(X^k\right)}{k!} \to E\left(\sum_{k=0}^{\infty} \frac{s^k X^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{s^k E\left(X^k\right)}{k!}$$

Therefore

$$E\left(\sum_{k=0}^{\infty} \frac{s^k X^k}{k!}\right) = \sum_{k=0}^{\infty} \frac{s^k E\left(X^k\right)}{k!}$$

□

Example 2.8.3. If $X \sim \exp(\lambda)$, and we want to find the moments we can compute

$$M(s) = E(e^{sX}) = \int_0^{\infty} e^{sX} \lambda e^{-\lambda x} dx = \lambda \int_0^{\infty} e^{-x(\lambda - s)} dx < \infty$$

for $s < \lambda$ so our domain for the moment generating function is $I = (-\infty, \lambda)$

$$M(s) = \frac{1}{1 - \frac{s}{\lambda}} = \sum_{k=0}^{\infty} \left(\frac{s}{\lambda}\right)^k = \sum_{k=0}^{\infty} \frac{s^k}{\lambda^k} = \sum_{k=0}^{\infty} \frac{s^k E(X^k)}{k!}$$

i.e. if we can expand

$$\frac{E(X^k)}{k!} = \frac{1}{\lambda^k} \ \forall k$$

Therefore

$$E(X^k) = \frac{k!}{\lambda^k}.$$

If you cannot find $[-s_0, s_0] \subset I$ the moment generating function is useless

Example 2.8.4. $X \sim$ Cauchy, $f(x) = \frac{1}{\pi(1+x^2)}$ then $I = \{0\}$ and $M_X(0) = 1$. $X \sim$ lognormal, $I = (-\infty, 0]$.

2.9 Uniform integrability

Let X be an integrable random variable on $(\Omega, \mathcal{F}, P)$. Then

$$\lim_{\alpha \to \infty} E(|X| \mathbb{1}_{(|X| > \alpha)}) = 0$$

since $|X|\mathbb{1}_{(|X|>\alpha)} \to 0$ as $\alpha \to \infty$ and $|X|\mathbb{1}_{(|X|>\alpha)} \leq |X| < \infty$. Now consider a finite number of integrable random variables $X_1, \ldots, X_n$. We can find $\forall \varepsilon > 0 \ \exists \alpha_i : E(|X_i|\mathbb{1}_{(|X_i|>\alpha_i)}) < \varepsilon$ and if we take $\alpha = \max(\alpha_1, \ldots, \alpha_n)$ we get

$$E(|X_i|\mathbb{1}_{(|X_i|>\alpha)}) \leq E(|X_i|\mathbb{1}_{(|X_i|>\alpha_i)}) < \varepsilon$$

so $\sup_i E(|X_i|\mathbb{1}_{(|X_i|>\alpha_i)}) < \varepsilon$, that is $\sup_i E(|X_i|\mathbb{1}_{(|X_i|>\alpha_i)}) \to 0$ as $\alpha \to \infty$.

If instead we take a sequence $X_1, X_2, \dots$ (so an infinite number) of integrable random variables, we have again that $\forall \varepsilon > 0\ \exists \alpha_i : E(|X_i|\mathbb{1}_{(|X_i|>\alpha_i)}) < \varepsilon$, but in this case $\sup_i \alpha_i$ can be $+\infty$ so in general we don't have $\sup_i E(|X_i|\mathbb{1}_{(|X_i|>\alpha)}) \to 0$ as $\alpha \to \infty$.

Definition 2.9.1 (Uniform integrability (UI))**.** A sequence of random variable $\{X_n\}$ is uniformly integrable if

$$\forall \varepsilon > 0\ \exists \alpha : \sup_n E(|X_n|\mathbb{1}_{(|X_n|>\alpha)}) < \varepsilon$$

which means $\sup_n E(|X_n|\mathbb{1}_{(|X_n|>\alpha)}) \to 0$ as $\alpha \to \infty$.

Idea: uniformly in n, the tails of the sequence don't matter.

Example 2.9.2. $X_n \sim \mathcal{N}(0, n)$, then $\forall n\ E(|X_n|) = \int |x| \frac{1}{\sqrt{2\pi n}} e^{-\frac{1}{2}\frac{x^2}{n}} dx < +\infty$, so X_n integrable. But $X_n = \sqrt{n}Z$ in distribution, where $Z = \mathcal{N}(0, 1)$, thus

$$\sup_n E(|X_n|\mathbb{1}_{(|X_n|>\alpha)}) = \sup_n E(\sqrt{n}|Z|\mathbb{1}_{(\sqrt{n}|Z|>\alpha)}) = \sup_n \sqrt{n} E(|Z|\mathbb{1}_{(|Z|>\frac{\alpha}{\sqrt{n}})}) = +\infty$$

so its not uniformly integrable.

Proposition 2.9.3 (Necessary condition for UI)**.** *X_n uniformly integrable* $\implies \sup_n E(|X_n|) < +\infty$

Proof. We can write $\sup_n E(|X_n|) = \sup_n \left[E(|X_n|\mathbb{1}_{(|X_n|>\alpha)}) + E(|X_n|\mathbb{1}_{(|X_n|\le\alpha)}) \right]$. We can fix α in order to have $E(|X_n|\mathbb{1}_{(|X_n|>\alpha)}) < \varepsilon$ and of course $E(|X_n|\mathbb{1}_{(|X_n|\le\alpha)}) \le \alpha$, so $\sup_n E(|X_n|) \le \alpha + \varepsilon < +\infty$. ☐

Example 2.9.4. The condition is not sufficient. Take

$$X_n = \begin{cases} 0 & 1 - \frac{1}{n} \\ 1 & \frac{1}{n} \end{cases}$$

$X_n \ge 0$ and $\sup_n E(|X_n|) = 1 < +\infty$. However this sequence is not uniformly integrable: $\sup_n E(|X_n|\mathbb{1}_{(|X_n|>\alpha)}) = 1 \not\to 0$.

Proposition 2.9.5 (Sufficient condition for UI)**.** *If* $\exists p > 1$ *such that* $\sup_n E(|X_n|^p) < +\infty$ *then X_n is uniformly integrable.*

Proof.

$$\begin{aligned}
\sup_n E(|X_n|\mathbb{1}_{(|X_n|>\alpha)}) &= \sup_n E\left(\frac{|X_n|^p}{|X_n|^{p-1}} \mathbb{1}_{(|X_n|>\alpha)} \right) \\
&\le \sup_n E\left(\frac{|X_n|^p}{\alpha^{p-1}} \mathbb{1}_{(|X_n|>\alpha)} \right) \\
&\le \frac{1}{\alpha^{p-1}} \underbrace{\sup_n E(|X_n|^p)}_{<\infty} \to 0
\end{aligned}$$

as $\alpha \to \infty$. ☐

Example 2.9.6. $X_n \sim \mathcal{N}(0, \sigma_n^2)$, $\sigma_n^2 = E(X_n^2)$, $\sup_n \sigma_n^2 < +\infty \implies X_n$ uniformly integrable.

Theorem 2.9.7 (Uniform integrability criterion)**.** *Suppose we have a sequence of random variables such that $X_n \to X$ almost surely and $X_n \in L^1$, $\forall n$. Then the following conditions are equivalent*

1. *X_n is uniformly integrable*
2. *$X \in L^1$ and $E(|X_n - X|) \to 0$ as $n \to \infty$*
3. *$E(|X_n|) \to E(|X|) < +\infty$ (usually the easiest to verify)*

Proof. We will write $X_{n\alpha} = X_n \mathbb{1}_{|X_n| \leq \alpha}$ and $X_n^{(\alpha)} = X_n \mathbb{1}_{|X_n| > \alpha}$. Analogously $X_\alpha = X \mathbb{1}_{|X| \leq \alpha}$ and $X^{(\alpha)} = X \mathbb{1}_{|X| > \alpha}$. We can always choose α such that $P(X = \alpha) = 0$ and this implies $X_{n\alpha} \to X_\alpha$ almost surely and $X_n^{(\alpha)} \to X^{(\alpha)}$ almost surely.

- (1) $\implies$ (2) Assume X_n uniformly integrable. Then $E(|X|) = E(\liminf_n |X_n|) \leq \liminf_n E(|X_n|) \leq \sup_n E(|X_n|) < +\infty$ so X is integrable. Then in $E(|X_n - X|) \leq E(|X_{n\alpha} - X_\alpha| + |X_n^{(\alpha)}| + |X^{(\alpha)}|)$ take α such that $\sup_n E(|X_n^{(\alpha)}|) < \varepsilon$ and $E(|X^{(\alpha)}|) < \varepsilon$. Then $|X_{n\alpha} - X_\alpha| \to 0$ as $n \to \infty$ and is bounded by 2α (since they are both smaller then α in absolute value) so we can use dominated convergence and get that $|X_{n\alpha} - X_\alpha| \to 0$. Then, $\forall \varepsilon > 0$ $\limsup_n E(|X_n - X|) \leq 2\varepsilon \implies \limsup_n E(|X_n - X|) = 0$.
- (2) $\implies$ (3) $X \in L^1$, $E(|X_n - X|) \to 0 \implies E(|X_n|) \to E(|X|) < +\infty$. Indeed, $|E(|X_n|) - E(|X|)| = |E(|X_n| - |X|)| \leq E(|X_n - X|) \to 0$.
- (3) $\implies$ (1) $E(|X_n|) \to E(|X|) < +\infty \implies \forall \varepsilon\ \exists \alpha : \limsup E(|X_n| \mathbb{1}_{(|X_n| > \alpha)}) < \varepsilon$. We get rid of the absolute value assuming $X_n \geq 0, X \geq 0$. If we can prove $\limsup_n E(X_n^{(\alpha)}) \leq E(\limsup_n X_n^{(\alpha)})$ then this is equal to $E(X^{(\alpha)})$ and for α large enough this is less than ε. The inequality we want to prove is similar to Fatou lemma but with lim sup. We get

$$
\begin{aligned}
\limsup_n E(X_n^{(\alpha)}) &= \limsup_n E(X_n - X_{n\alpha}) \leq \limsup_n E(X_n) + \limsup_n E(-X_{n\alpha}) \\
&\leq E(X) - \liminf_n E(X_{n\alpha}) \leq E(X) - E(\liminf_n X_{n\alpha}) \\
&\leq E(\limsup_n (X_n - X_{n\alpha})) \leq E(\limsup X_n^{(\alpha)})
\end{aligned}
$$

□

Proposition 2.9.8 (Convergence of the expectation and UI)**.** *If X_n is uniformly integrable and $X_n \to X$ almost surely then $E(X_n) \to E(X)$.*

Proof. $|E(X_n) - E(X)| = |E(X_n - X)| \leq E(|X_n - X|) \to 0$ as $n \to \infty$. □

Suppose $X_n \to X$, $E(X_n) \to E(X)$, $X_n \geq 0$. Then Fatou lemma $E(\liminf X_n) \leq \liminf E(X_n) \implies E(\limsup X_n) \leq \limsup E(X_n)$, in this case the inequation is the opposite.

Example 2.9.9. Toss a coin infinitely many times, $Z_i = 2\mathbb{1}_{A_i}$, A_i ="H" at toss i, $P(A_i) = \frac{1}{2}$. Consider $X_n = \prod_{i=1}^n Z_i$, then $X_n \to 0$ almost surely because if you observe a tail the product is zero. Then $E(X_n) = \prod_{i=1}^n E(Z_i) = 1$, so $0 = E(\limsup X_n) \not\geq \limsup E(X_n) = 1$.

2.9.1 Expectation and series

Proposition 2.9.10 (Expectation and series for non negative r.v.'s). *$X_n \geq 0$ $\forall n \implies$ $E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n)$.*

Proof. $E\left(\lim_{n\to\infty} \sum_{i=1}^{n} X_i\right) = \lim_{n\to\infty} E\left(\sum_{i=1}^{n} X_i\right) = \lim_{n\to\infty} \sum_{i=1}^{n} E(X_i) = \sum_{n=1}^{\infty} E(X_i)$ using monotone convergence. □

Proposition 2.9.11 (Expectation and series for general r.v.'s). *If $\sum_{n=1}^{\infty} E(|X_n|) < +\infty \implies E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E(X_n)$.*

Proof. $|\sum_{n=1}^{\infty} X_n| \leq \sum_{n=1}^{\infty} |X_n| \in L^1$ and by dominated convergence $E\left(\sum_{n=1}^{\infty} |X_n|\right) = \sum_{n=1}^{\infty} E(|X_n|) < +\infty$. □

2.10 Inequalities

Proposition 2.10.1 (Jensen inequality). *Let X be a random variable with finite expectation $E(X)$. Let U be an interval such that $P(X \in U) = 1$ and let φ a convex function on U. Then*

$$E(\varphi(X)) \geq \varphi(E(X))$$

Moreover, if φ is strictly convex on an interval V such that $P(X \in V) > 0$, then $E(\varphi(X)) > \varphi(E(X))$.

Proof. $\forall x_0 \in U$ $\exists a, b \in \mathbb{R}$ such that $\varphi(x_0) = ax_0 + b$ and $\varphi(x) \geq ax_0 + b$.
Now choose $x_0 = E(X)$, so that

$$\varphi(E(X)) = aE(X) + b.$$

Furthermore

$$\begin{gathered}\varphi(X(w)) \geq aX(w) + b \text{ with probability 1} \\ \Downarrow \\ E(\varphi(X)) \geq aE(X) + b = \varphi(E(X)) \\ \Downarrow \\ E(\varphi(X)) \geq \varphi(E(X)).\end{gathered}$$

□

Example 2.10.2. If $\varphi(X) = |X|$, then $E(|X|) \geq |E(X)|$. If $\varphi(X) = X^2$, then $E(X^2) \geq E^2(X)$.

Proposition 2.10.3 (Markov inequality). *Let X be a random variable and $p > 0$. Then*

$$\forall \alpha > 0 \;\; P(|X| \geq \alpha) \leq \frac{E(|X|^p)}{\alpha^p}.$$

Proof.

$$\begin{aligned}E(|X|^p) = E(|X|^p \mathbb{1}_{(|X|<\alpha)}) + E(|X|^p \mathbb{1}_{(|X|\geq\alpha)}) \geq E(|X|^p \mathbb{1}_{(|X|\geq\alpha)}) \geq \\ \geq \alpha^p E(\mathbb{1}_{(X>\alpha)}) = \alpha^p P(|X| \geq \alpha).\end{aligned}$$

□

If we take $p = 2$ and we replace X with the centered variable $X - E(X)$, we get the Chebyshev inequality.

Proposition 2.10.4 (Chebyshev's inequality)**.** *Let X be a random variable. Then*

$$P(|X - E(X)| > \alpha) \leq \frac{Var(X)}{\alpha^2},$$

where $Var(X)$ is the variance of X.

Proof. Direct consequence of Markov inequality. □

Definition 2.10.5 (Conjugate numbers)**.** Two numbers $p, q > 1$ are conjugate if $\frac{1}{p} + \frac{1}{q} = 1$.

Example 2.10.6. Some couples of conjugate numbers are: (2, 2), (3, 3/2), (4, 4/3).

Proposition 2.10.7 (Hölder's inequality)**.** *Let p, q be conjugate numbers and X, Y random variables. Then*

$$E(|XY|) \leq (E(|X|^p))^{1/p}(E(|X|^q))^{1/q}.$$

We introduce the notation

$$\|X\|_p = (E(|X|^p))^{1/p}.$$

Note that here it is just a notation, but later we will define the concept of norm in L^p space, that uses this notation, in the same way.

$$E(|XY|) = \|X\|_p \|Y\|_q$$

Proof. We want to show that $\frac{E|XY|}{\|X\|_p\|Y\|_q} \leq 1$. First of all let us recall a property of numbers:

$$\frac{1}{p}a^p + \frac{1}{q}b^q \geq ab, \quad \forall a, b \geq 0$$

To prove it, let's assume that $a, b > 0$ (otherwise it is obvious). Let's write $a = e^s$ and $b = e^t$. Then, by convexity of the exponential function and keeping in mind that $\frac{1}{q} = 1 - \frac{1}{p}$,

$$\frac{1}{p}a^p + \frac{1}{q}b^q = \frac{1}{p}e^{sp} + \frac{1}{q}e^{sq} \geq e^{\frac{1}{p}ps + \frac{1}{q}qs} = e^s e^t = ab.$$

Hence we can take $a = \frac{|X|}{\|X\|_p}$ and $b = \frac{|Y|}{\|Y\|_q}$ so that, assuming $\|X\|_p, \|Y\|_q > 0$,

$$\frac{1}{p}\frac{|X|^p}{\|X\|_p^p} + \frac{1}{q}\frac{|Y|^q}{\|Y\|_q^q} \geq \frac{|XY|}{\|X\|_p\|Y\|_q}$$

We take expectations on both sides and we recall that $\|X\|_p^p = E(|X|^p)$ and $\|Y\|_q^q = E(|Y|^q)$:

$$\frac{1}{p} \cdot 1 + \frac{1}{q} \cdot 1 \geq \frac{E(|XY|)}{\|X\|_p\|Y\|_q}.$$

Hence

$$E(|XY|) \leq \|X\|_p\|Y\|_q.$$

If $\|X\|_p = 0$ then $(E(|X|^p))^{1/p} = 0$, that is $E(|X|^p) = 0$. This implies that $X = 0$ a.s. and so the inequality would be $0 \leq 0$, that is true. □

A particular case of the last proposition is the Cauchy-Schwarz inequality, that is obtained taking $p = q = 2$.

Proposition 2.10.8 (Cauchy-Schwarz inequality). *Let X, Y be two random variables. Then*

$$E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}.$$

Proof. Hölder inequality with $p = q = 2$. $\square$

Application 2.10.9. By Jensen and Cauchy-Schwarz inequalities we get

$$|E(XY)| = E(|XY|) \leq \sqrt{E(X^2)E(Y^2)}$$

Then taking $X - E(X)$ and $Y - E(Y)$ in place of X and Y

$$|\mathrm{Cov}(X,Y)| \leq \sqrt{Var(X)Var(Y)}.$$

Furthermore it can be proved that

$$|\mathrm{Cov}(X,Y)| = \sqrt{Var(X)Var(Y)} \iff \frac{Y - E(Y)}{Var(Y)} = \frac{X - E(X)}{Var(X)}.$$

Proposition 2.10.10 (Lyapounov's inequality). *If $\beta \geq \alpha \geq 1$ then $\|X\|_\beta \geq \|X\|_\alpha$*

Proof. Consider Holder's inequality

$$E(|XY|) \leq (E|X|^p)^{\frac{1}{p}}(E|Y|^q)^{\frac{1}{q}}$$

and apply it taking $Y = 1$, $X = |X|^\alpha$ and $p = \frac{\beta}{\alpha}$

$$E(|X|^\alpha) \leq \left(E(|X|^{\alpha\frac{\beta}{\alpha}})\right)^{\frac{\alpha}{\beta}} = (E(|X|^\beta))^{\frac{\alpha}{\beta}}$$
$$\Updownarrow$$
$$E(|X|^\alpha)^{\frac{1}{\alpha}} \leq (E(|X|^\beta))^{\frac{1}{\beta}}$$
$$\Updownarrow$$
$$\|X\|_\alpha \leq \|X\|_\beta.$$

$\square$

Proposition 2.10.11 (Minkowsky's inequality). *Let X, Y be two random variables and $p \geq 1$. Then*

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p$$

Proof. If $p = 1$, by the triangular inequality $E(|X + Y|) \leq E(|X| + |Y|) = E(|X|) + E(|Y|)$. If $p > 1$,

$$\begin{aligned} E(|X+Y|^p) = E(|X+Y|^{p-1}|X+Y|) &\overset{(1)}{\leq} E(|X+Y|^{p-1}|X|) + E(|X+Y|^{p-1}|Y|) \\ &\overset{(2)}{\leq} \left[E\left(|X+Y|^{(p-1)q}\right)\right]^{\frac{1}{q}}(E|X|^p)^{\frac{1}{p}} + \left[E\left(|X+Y|^{(p-1)q}\right)\right]^{\frac{1}{q}}(E|Y|^p)^{\frac{1}{p}} \\ &\leq [E(|X+Y|^p)]^{\frac{1}{q}}\|X\|_p + (E|X+Y|^p)^{\frac{1}{q}}\|Y\|_p. \end{aligned}$$

Where

(1) Triangle inequality;

(2) Holder inequality on both addends using $|X+Y|^{p-1}$ as first random variable (of the two of the Holder inequality) in both cases and X, Y as second random variables respectively in the first and second addend.

Therefore we have

$$E(|X+Y|^p) \leq [E(|X+Y|^p)]^{\frac{1}{q}} (\|X\|_p + \|Y\|_p)$$
$$\Updownarrow$$
$$(E|X+Y|^p|)^{1-\frac{1}{q}} \leq \|X\|_p + \|Y\|_p$$
$$\Updownarrow$$
$$(E|X+Y|^p|)^{\frac{1}{p}} \leq \|X\|_p + \|Y\|_p$$
$$\Updownarrow$$
$$\|X+Y\|_p \leq \|X\|_p + \|Y\|_p$$

$\square$

Chapter 3

L^p spaces

3.1 Random variables and L^p spaces

Definition 3.1.1 (L^p space)**.** Let $p \geq 1$ and consider $(\Omega, \mathcal{F}, P)$.

$$L^p(\Omega, \mathcal{F}, P) = L^p = \{X \text{ on } (\Omega, \mathcal{F}, P) : E(|X|^p) < \infty\}.$$

Proposition 3.1.2 (L^p as a vector space)**.** *L^p is a linear/vector space:*

1. *if $X \in L^p$ and $a \in \mathbb{R}$, then $aX \in L^p$;*

2. *if $X, Y \in L^p$, then $X + Y \in L^p$.*

Proof. 1. $X \in L^p$ and $E(|X|^p) < \infty$. Let $(Z_n)_{n\in\mathbb{N}}$ be a sequence of simple, ≥ 0 random variables such that $Z_n \uparrow |X|^p$, then $\lim_n E(Z_n) < \infty$ (by Monotone convergence). Then $|a|^p Z_n \uparrow |a|^p |X|^p = |aX|^p$ and $\lim_n E(|a|^p Z_n) = |a|^p E(|X|^p) < \infty$.

2. $X, Y \in L^p$.

$$\begin{aligned} E(|X+Y|^p) &\leq E(2^p \max(|X|^p, |Y|^p)) \leq 2^p E(|X|^p + |Y|^p) = \\ &= 2^p[\underbrace{E(|X|^p)}_{<\infty} + \underbrace{E(|X|^q)}_{<\infty}] < \infty. \end{aligned}$$

□

Introduce the norm $\|X\|_p = (E(|X|^p))^{\frac{1}{p}}$.

1. $\|X\|_p \geq 0$ and $\|X\|_p = 0$ if and only if $X = 0$ a.s.

2. $\|aX\|_p = |a|\|X\|_p$

3. $\|X+Y\|_p \leq \|X\|_p + \|Y\|_q$ by Minkowski

Definition 3.1.3 (Convergence in L^p)**.** $X_n \xrightarrow{L^p} X$ if $\|X_n - X\|_p \to 0$ as $n \to \infty$.

$$\begin{aligned} X_n \xrightarrow{L^p} X &\iff (E(|X_n - X|^p))^{\frac{1}{p}} \to 0 \\ &\iff E(|X_n - X|^p) \to 0. \end{aligned}$$

Proposition 3.1.4 (Inclusion of L^p spaces)**.** *Take L^p, L^q with $1 \le p < q$. Then $L^q \subset L^p$, that is*

$$E(|X|^q) < \infty \implies E(|X|^p) < \infty.$$

Furthermore by Lyapounov $\|X\|_p \le \|Y\|_q$.

Proposition 3.1.5 (Relationship between $\xrightarrow{L^p}$ and $\xrightarrow{p}$)**.** *Convergence in L^p implies convergence in probability:*

$$\|X_n - X\|_p \to 0 \implies X_n \xrightarrow{p} X.$$

Proof.

$$P(|X_n - X| > \varepsilon) \le \frac{E(|X_n - X|^p)}{\varepsilon^p} \to 0 \text{ as } n \to \infty$$

Where the inequality is given by Markov inequality. □

Definition 3.1.6 (Cauchy sequence)**.** $(x_n)_{n\in\mathbb{N}}$ is a Cauchy sequence if $\|x_n - x_m\| \to 0$ as $n, m \to \infty$.

Definition 3.1.7 (Complete space)**.** A space is complete if every Cauchy sequence converges.

Example 3.1.8. $\mathbb{R}$ is complete.

Theorem 3.1.9 (Completeness of L^p)**.** *L^p is complete.*

Proof. $(X_n)_{n\in\mathbb{N}}$ Cauchy means that

$$\forall \varepsilon > 0 \;\; \exists n_0 : \|X_n - X_m\|_p < \varepsilon \;\; \forall n, m \ge n_0.$$

This can be written as

$$\forall k \in \mathbb{N} \;\; \exists n_k : \|X_n - X_m\|_p^p < \frac{1}{2^{kp}} \frac{1}{2^k} \;\; \forall n, m \ge n_k.$$

Then

$$\sum_{k=1}^{\infty} P(|X_{n_k} - X_{n_{k+1}}| > \frac{1}{2^k}) \le \sum_{k=1}^{\infty} \frac{E(|X_{n_k} - X_{n_{k+1}}|^p)}{1/2^{kp}} \le \sum_{k=1}^{\infty} \frac{1/2^{kp} 1/2^k}{1/2^{kp}} = 1 < +\infty,$$

where the first inequality is by Markov inequality. By Borel-Cantelli first lemma

$$P\left(|X_{n_k} - X_{n_{k+1}}| > \frac{1}{2^k} \text{ i.o.}\right) = 0.$$

Hence, if we take

$$H = \left\{|X_{n_k} - X_{n_{k+1}}| > \frac{1}{2^k} \text{ i.o.}\right\}^c = \left\{|X_{n_k} - X_{n_{k+1}}| \le \frac{1}{2^k} \text{ ult.}\right\},$$

then $P(H) = 1$ and $\forall \omega \in H$,

$$|X_{n_k}(\omega) - X_{n_{k+1}}(\omega)| \le \frac{1}{2^k}$$

for n_k large enough.

Hence, for $h > k$,

$$|X_{n_k}(\omega) - X_{n_h}(\omega)| \leq |X_{n_k}(\omega) - X_{n_{k+1}}(\omega)| + \cdots + |X_{n_{h-1}}(\omega) - X_{n_h}(\omega)|$$
$$\leq \sum_{j=k}^{h} \frac{1}{2^j} \leq \sum_{j=k}^{\infty} \frac{1}{2^j} \to 0 \text{ as } k \to +\infty,$$

that means that $(X_{n_k}(\omega))_{n\in\mathbb{N}}$ is a Cauchy sequence in $\mathbb{R}$. This implies that

$$\forall \omega \in H \ \exists \lim_{k\to\infty} X_{n_k}(\omega) = X(\omega) \text{ with } X \text{ r.v.}$$

and so $X_{n_k} \xrightarrow{as} X$. Now we want to show also convergence in L^p.

$$E(|X_{n_k} - X|^p) \leq E(\liminf_{j\to\infty} |X_{n_k} - X_{n_j}|^p) \leq \liminf_{j\to\infty} E(|X_{n_k} - X_{n_j}|^p)$$
$$\leq \liminf_{j\to\infty} \frac{1}{2^{kp}2^k} \to 0 \text{ as } k \to +\infty,$$

where the second inequality is due to Fatou Lemma. Hence, $X_n \xrightarrow{L^p} X$, indeed

$$\|X_n - X\|_p \leq \|X_n - X_{n_k}\|_p + \|X_{n_k} - X\|_p \to 0.$$

because both addends are smaller than every $\varepsilon > 0$, respectively because $(X_n)_{n\in\mathbb{N}}$ is Cauchy and because of almost sure convergence. □

So L^p is a complete, normed, linear space, that is L^p is a Banach space. However L^p is not separable. In general, you can always approximate X with simple random variables

$$X_n = \sum_{i=1}^{k_n} a_i^{(n)} \mathbb{1}_{A_i^{(n)}}$$

with $a_i^{(n)} \in \mathbb{Q}$, the problem is with $A_i^{(n)} \in \mathcal{F}$. If $\mathcal{F}$ is not too large we can approximate $\{A_i^{(n)}\}$ with a countable set, so that the following theorem holds

Theorem 3.1.10 (Separability of L^p). *If $\mathcal{F}$ is countably-generated (i.e. there exists a sequence $\{A_n\}$ of subsets of X, such that $\mathcal{F} = \sigma(A_1, A_2, \dots)$), then $L^p(\Omega, \mathcal{F}, P)$ is separable.*

3.1.1 L^2 as an Hilbert space

An interesting case is L^2: on this space we can define an inner product

$$E(XY) = \langle X, Y \rangle$$

which is linear, symmetric and $\langle X, X\rangle \geq 0$ and $\langle X, X\rangle = 0 \iff X = 0$. So the L^2 norm

$$\|X\|_2 = \sqrt{\langle X, X\rangle} = \sqrt{E(X^2)}$$

comes from an inner product, which means that L^2 is an Hilbert space.

Chapter 4

Laws of large numbers

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables on a probability space $(\Omega, \mathcal{F}, P)$. For every n, let

$$S_n = \sum_{k=1}^{n} X_k$$

then $\frac{S_n}{n}$ is the average of $X_1, \ldots, X_n$.

4.1 Weak and strong laws

Definition 4.1.1 (Weak law of large numbers (WLLN))**.** We say that $\{X_n\}$ obeys a weak law of large numbers if $\frac{S_n}{n}$ converges in probability.

Definition 4.1.2 (Strong law of large numbers (SLLN))**.** We say that $\{X_n\}$ obeys a strong law of large numbers if $\frac{S_n}{n}$ converges in almost surely.

There are many theorems ensuring the convergence that differ in the assumptions and in the limit.

Theorem 4.1.3 (WLLN for uncorrelated random variables)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of square integrable random variables, i.e. $X_n \in L^2\ \ \forall n$, such that $\forall n\ E(X_n) = \mu$, $Var(X_n) = \sigma^2$ and $Cov(X_n, X_{n+k}) = 0\ \ \forall k \geq 1$. Then:*

$$\frac{S_n}{n} \xrightarrow{p} \mu$$

Proof. We will prove $\frac{S_n}{n} \to \mu$ in L^2 because, as shown above, this implies convergence in probability. Notice that

$$E\left(\frac{S_n}{n}\right) = \sum_{k=1}^{n} E\left(\frac{X_k}{n}\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \mu$$

Therefore, since the X_k are uncorrelated and so $Var(\sum_{k=1}^{n} X_k) = \sum_{k=1}^{n} Var(X_k)$,

$$E\left(\frac{S_n}{n} - \mu\right)^2 = Var\left(\frac{S_n}{n}\right) = \frac{1}{n^2} Var(S_n) = \frac{\sum_{k=1}^{n} Var(X_k)}{n^2} = \frac{\sigma^2}{n} \to 0$$

and the argument follows by Chebychev's inequality. $\square$

If the correlation is not zero but is weak for random variables far apart then the WLLN still holds.

Theorem 4.1.4 (WLLN for asintotically uncorrelated random variables)**.** *Let* $\{X_n\}_{n\in\mathbb{N}}$ *be a sequence of square integrable random variables, i.e.* $X_n \in L^2$ $\forall n$*, such that* $\forall n$ $E(X_n) = \mu$*,* $Var(X_n) = \sigma^2$*,* $Cov(X_n, X_{n+k}) = \gamma(k)$ $\forall k \geq 1$*. If* $\gamma(k) \to 0$ *as* $k \to \infty$*, then:*

$$\frac{S_n}{n} \xrightarrow{p} \mu.$$

Proof. We compute $Var(\frac{S_n}{n})$

$$\begin{aligned}
Var\left(\frac{S_n}{n}\right) &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} \sum_{i,j=1}^{n} Cov(X_i, X_j) \\
&= \frac{1}{n^2}(n\sigma^2 + 2(n-1)\gamma(1) + 2(n-2)\gamma(2) + \cdots + 2\gamma(n-1)) \\
&\leq \frac{1}{n^2}(2n|\gamma(0)| + 2n|\gamma(1)| + 2n|\gamma(2)| + \cdots + 2n|\gamma(n-1)| \\
&= 2\sum_{i=0}^{n-1} \frac{|\gamma(k)|}{n} \to 0
\end{aligned}$$

□

Theorem 4.1.5 (Kolmogorov SLLN)**.** *Let* $\{X_n\}_{n\in\mathbb{N}}$ *be a sequence of independent, identically distributed random variables with finite expectation, i.e.* $X_n \in L^1$ $\forall n$*,* $E(X_n) = \mu < \infty$*. Then:*

$$\frac{S_n}{n} \xrightarrow{as} \mu.$$

Proof. (1) If the thesis is true for non-negative random variables, then it is true in general:

$$\frac{\sum_{k=1}^{n} X_k}{n} = \frac{\sum_{k=1}^{n} X_k^+}{n} - \frac{\sum_{k=1}^{n} X_k^-}{n} \xrightarrow{as} E(X_1^+) - E(X_1^-) = E(X_1) = \mu,$$

where the positive and the negative part of the X_k's converge to their mean by assumption of this point.

(2) We have not assumed a finite variance (L^2) for X_k's. In order to get a finite variance, we truncate the X_k by defining $X_k^* = X_k \mathbb{1}_{(X_k \leq k)}$. Notice how X_k^* is closer and closer to X_k as k increases. Also define $S_n^* = \sum_{k=1}^{n} X_k^*$.

We want to show that $\frac{S_n^*}{n} \xrightarrow{as} \mu$ implies that $\frac{S_n}{n} \xrightarrow{as} \mu$. In order to do this, we prove that $P(X_k^* \neq X_k \text{ i.o.}) = 0$, indeed if the two sums differ only for a finite number of elements, then the limit is the same. Now observe that:

$$\begin{aligned}
\sum_{k=1}^{\infty} P(X_k^* \neq X_k) &\overset{(1)}{=} \sum_{k=1}^{\infty} P(X_k > k) = \sum_{k=1}^{\infty} \int_{k-1}^{k} P(X_k > k)dt \\
&\overset{(2)}{\leq} \sum_{k=1}^{\infty} \int_{k-1}^{k} P(X_k > t)dt \overset{(3)}{=} \int_{0}^{\infty} P(X_1 > t)dt = E(X_1) \\
&< \infty,
\end{aligned}$$

where (1) is due to definition of K_k^*, (2) is because $t < k$ inside the integral, since the domain of integration is $(k-1, k)$, and (3) is due the identical distribution of the X_k's.

Therefore by BC1 we obtain $P(X_k^* \neq X_k \text{ i.o.}) = 0$ and so we now have that

$$\frac{S_n^*}{n} \xrightarrow{as} \mu \implies \frac{S_n}{n} \xrightarrow{as} \mu$$

which would mean that we can continue with X_k^*.

(3) Now it remains to prove that $\frac{S_n^*}{n} \xrightarrow{as} \mu$ (it will be done in the following three points).

Let $\alpha > 1$ and $u_n = \lfloor \alpha^n \rfloor$. u_n is a subsequence of n and we want to show that

$$\frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n} \xrightarrow{as} 0$$

that has as a sufficient condition

$$\sum_{n=1}^{\infty} P\left(\left|\frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n}\right| > \varepsilon\right) < +\infty.$$

So we compute the summands of this series

$$\begin{aligned}
P\left(\left|\frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n}\right| > \varepsilon\right) &\overset{Cheb.}{\leq} \frac{Var(S_{u_n}^*)}{u_n^2\varepsilon^2} = \frac{\sum_{k=1}^{u_n} Var(X_k^*)}{u_n^2\varepsilon^2} \leq \frac{\sum_{k=1}^{u_n} E(X_k^{*2})}{u_n^2\varepsilon^2} \\
&= \frac{\sum_{k=1}^{u_n} E(X_k^2 \mathbb{1}_{(X_k \leq k)})}{u_n^2\varepsilon^2} \leq \frac{\sum_{k=1}^{u_n} E(X_k^2 \mathbb{1}_{(X_k \leq u_n)})}{u_n^2\varepsilon^2} \\
&\overset{iid}{=} \frac{u_n E(X_1^2 \mathbb{1}_{(X_1 \leq u_n)})}{u_n^2\varepsilon^2} = \frac{E(X_1^2 \mathbb{1}_{(X_1 \leq u_n)})}{u_n\varepsilon^2}.
\end{aligned}$$

Now summing these terms we obtain (we remove ε^2, but this does not change anything):

$$\sum_{n=1}^{\infty} \frac{E(X_1^2 \mathbb{1}_{(X_1 \leq u_n)})}{u_n} \overset{MCT}{=} E\left(X_1^2 \sum_{n=1}^{\infty} \frac{1}{u_n} \mathbb{1}_{(X_1 \leq u_n)}\right) = (\star)$$

Let us define $N = N(\omega) = \inf\{n \in \mathbb{N} : u_n > X_1(\omega)\}$. Then

$$\begin{aligned}
(\star) &= E\left(X_1^2 \sum_{n=N}^{\infty} \frac{1}{u_n}\right) \leq E\left(X_1^2 \sum_{n=N}^{\infty} \frac{2}{\alpha^n}\right) = E\left(X_1^2 \sum_{n=0}^{\infty} \frac{2}{\alpha^{n+N}}\right) \\
&= E\left(X_1^2 \frac{2}{\alpha^N} \sum_{n=0}^{\infty} \frac{1}{\alpha^n}\right) \leq \frac{2}{1-\frac{1}{\alpha}} E\left(\frac{X_1^2}{\alpha^N}\right) \leq 2E\left(\frac{X_1^2}{u_N}\right) \leq 2E\left(\frac{X_1^2}{X_1}\right) \\
&= 2E(X_1)\frac{1}{1-\frac{1}{\alpha}} < \infty.
\end{aligned}$$

Therefore the series converges and by BC1 the sufficient condition is verified and so we obtain the result we were aiming at, that is $\frac{S_{u_n}^* - E(S_{u_n}^*)}{u_n} \xrightarrow{as} 0$.

(4) From Step 3 we know that $\frac{(S^*_{u_n} - E(S^*_{u_n}))}{u_n} \xrightarrow{as} 0$, now we want to show that $\frac{S^*_{u_n}}{u_n} \xrightarrow{as} \mu$ (and the last point will be $\frac{S^*_n}{n} \xrightarrow{as} \mu$).

We also have that

$$E(X^*_k) = E(X_k \mathbb{1}_{(X_k \le k)}) = E(X_1 \mathbb{1}_{(X_1 \le k)}) \to \mu,$$

by MCT, since $X_1 \mathbb{1}_{(X_1 \le k)} \uparrow X_1$. On the other hand, we have

$$\frac{E(S^*_{u_n})}{u_n} = \frac{E\left(\sum_{k=1}^{u_n} X^*_k\right)}{u_n} = \frac{\sum_{k=1}^{u_n} E\left(X^*_k\right)}{u_n}$$

We now use Cesàro sums: if $a_n \to a$, then $\frac{1}{n}\sum_{k=1}^{n} a_k \to a$. In our case, $a_n = E(X^*_n) \to a = \mu$. Therefore

$$\frac{E(S^*_{u_n})}{u_n} = \frac{\sum_{k=1}^{u_n} E(X^*_k)}{u_n} \to \mu.$$

Now, applying the triangle inequality, we obtain:

$$\left|\frac{S^*_{u_n}}{u_n} - \mu\right| \le \left|\frac{S^*_{u_n} - E(S^*_{u_n})}{u_n}\right| + \left|\frac{E(S^*_{u_n})}{u_n} - \mu\right| \xrightarrow{as} 0$$

Where the first term in the sum converges a.s. to 0 for the third point and the second is a sequence of real numbers that converges to 0 for what we have just shown.

(5) Up until now we have a.s. convergence for a subsequence represented by u_n. We now want to show that

$$\frac{S^*_k}{k} \xrightarrow{as} \mu.$$

With α fixed, $\forall k \ \exists n = n(k)$ such that $u_{n-1} \le k \le u_n$ and so

$$(\star_1) = \frac{S^*_{u_{n-1}}}{k} \le \frac{S^*_{u_n}}{k} \le \frac{S^*_{u_n}}{k} = (\star_2).$$

But then

$$(\star_1) \ge \frac{S^*_{u_{n-1}}}{u_{n-1}} = \frac{S^*_{u_{n-1}}}{u_{n-1}} \frac{u_{n-1}}{u_n} \xrightarrow{as} \mu\frac{1}{\alpha},$$

since $\frac{S^*_{u_{n-1}}}{u_{n-1}} \xrightarrow{as} \mu$ and $\frac{u_{n-1}}{u_n} \sim \frac{\alpha^{n-1}}{\alpha^n} \to \frac{1}{\alpha}$, and also

$$(\star_2) \le \frac{S^*_{u_n}}{u_{n-1}} = \frac{S^*_{u_n}}{u_n}\frac{u_n}{u_{n-1}} \to \mu\alpha,$$

since $\frac{S^*_{u_n}}{u_n} \xrightarrow{as} \mu$ and $\frac{u_n}{u_{n-1}} \sim \frac{\alpha^n}{\alpha^{n-1}} \to \alpha$.

To conclude, $\forall \alpha > 1$, with probability 1

$$\mu\frac{1}{\alpha} \le \liminf \frac{S^*_k}{k} \le \limsup \frac{S^*_k}{k} \le \mu\alpha.$$

Now we take α of the kind $1+\frac{1}{j}$, $j \in \mathbb{N}$ (indeed if we took $\forall\alpha$ we would have had some problems with the following step, however it is sufficionet to take a countable number of α).

$$P\left\{\bigcap_{j=1}^{\infty}\left(\frac{\mu}{1+1/j} \le \liminf \frac{S_k^*}{k} \le \limsup \frac{S_k^*}{k} = \mu\right)\right\} = 1.$$

Hence

$$P\left\{\liminf \frac{S_k^*}{k} = \limsup \frac{S_k^*}{k} = \mu\right\} = 1.$$

□

Example 4.1.6. Let $\{X_n\}$ be i.i.d. random variables with distribution function F. Let

$$F_n(x) = \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{(-\infty,x]}(X_k)$$

the empirical distribution function. Then

$$E(\mathbb{1}_{(-\infty,x]}(X_k)) = P(X_k \le x) = F(x)$$

so we can say that $F_n(x) \to F(x)$ almost surely as $n \to \infty$.

What can we say if $E(X_n) = +\infty$?

Theorem 4.1.7 (SLLN: infinite case)**.** *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables with expectation $E(X_n) = +\infty$. Then:*

$$\frac{S_n}{n} \xrightarrow{as} +\infty$$

Proof. Without loss of generality, we can assume that $X_n \ge 0$. Fix $M > 0$ and let $X_n^* = X_n \mathbb{1}_{(X_n<M)}$. Then

$$\frac{S_n}{n} = \frac{\sum_{k=1}^n X_k}{n} \ge \frac{\sum_{k=1}^n X_k^*}{n} \xrightarrow{as} E(X_1^*),$$

where the convergence is guaranteed by the Kolmogorov SLLN. Then as $M \to \infty$ we have that $E(X_1^*) \to \infty$ by MCT and so

$$\liminf_n \frac{S_n}{n} \ge E(X_1^*) \to \infty.$$

□

Theorem 4.1.8. *Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables such that $\exists E(X_n)$. Then $\frac{S_n}{n} \xrightarrow{as} E(X_1)$, both if the expectation is finite or infinite.*

Proof. Just join last theorem and Kolmogorov SLLN. □

Remark 4.1.9 (Behaviour of an i.i.d. sequence with no expectation defined)*.* If $\{X_n\}_{n\in\mathbb{N}}$ are i.i.d. and $\nexists E(X_n)$ then there are three possibilities

1. $\frac{S_n}{n} \to +\infty$ almost surely

2. $\frac{S_n}{n} \to -\infty$ almost surely
3. $P((\liminf \frac{S_n}{n} = -\infty) \cap (\limsup \frac{S_n}{n} = +\infty)) = 1$

Example 4.1.10. $X_n \sim$ Cauchy, independent, with density $f(x) = \frac{1}{\pi(1+x^2)}$.
$E(X_n^+) = +\infty$ and $E(X_n^-) = +\infty$, so $\nexists E(X_n)\ \forall n$.
Then $P((\liminf \frac{S_n}{n} = -\infty) \cap (\limsup \frac{S_n}{n} = +\infty)) = 1$.

List of Theorems